BESIII Electromagnetic Calorimeter Fast Simulation

Tong Liu[1], Qiaojia Ge[2], Wenxing Fang[1], Weidong Li[1], Rui Zhang[2], Ke Li[1]

[1] IHEP

[2] NJU

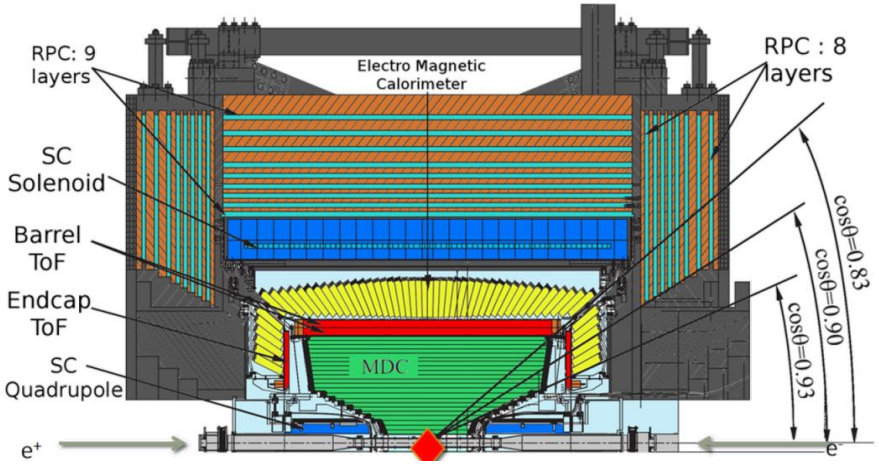
Outline

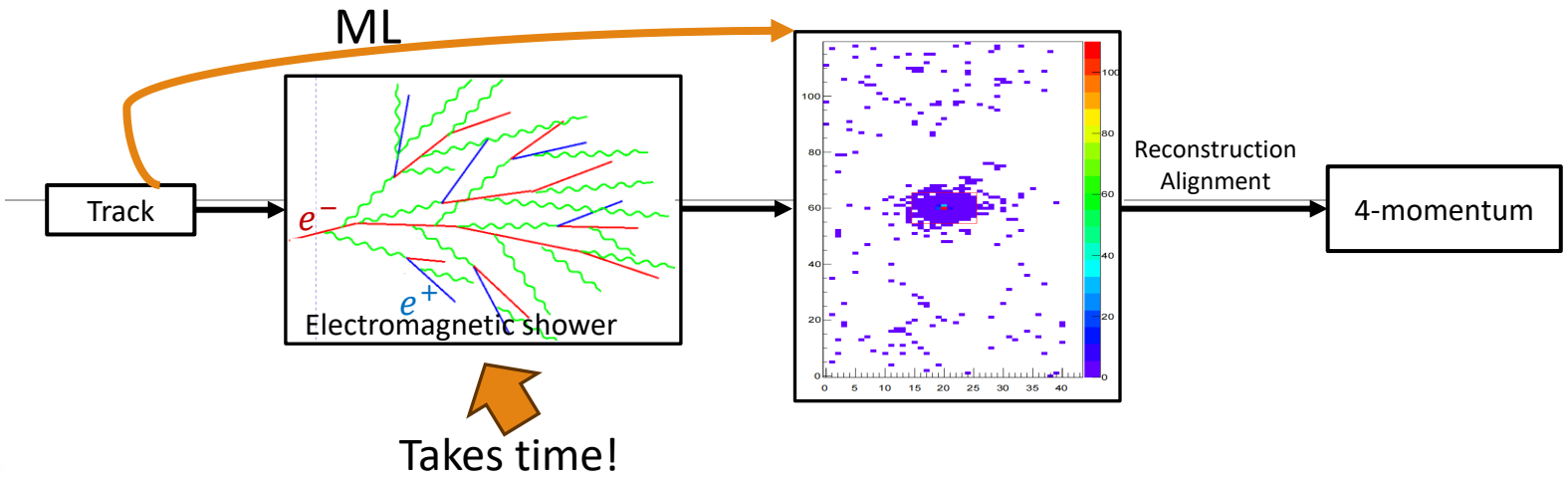
- Introduction & Training samples
- EMC simulation with GAN
- EMC simulation with MoE-GAN
- Summary

Introduction & Training samples

Introduction

- Improved data luminosity requires bigger MC
- MC simulation, especially for electromagnetic calorimeter (EMC), takes large CPU resources
  - Traditional: Geant4, gradually calculate the next state, complex due to secondary particles
  - ML: without Geant4, calculate hit map from input conditions
- EMC: 44 layer*120 crystals in barrel
  - Focus on barrel region firstly to avoid energy leakage caused by the gap

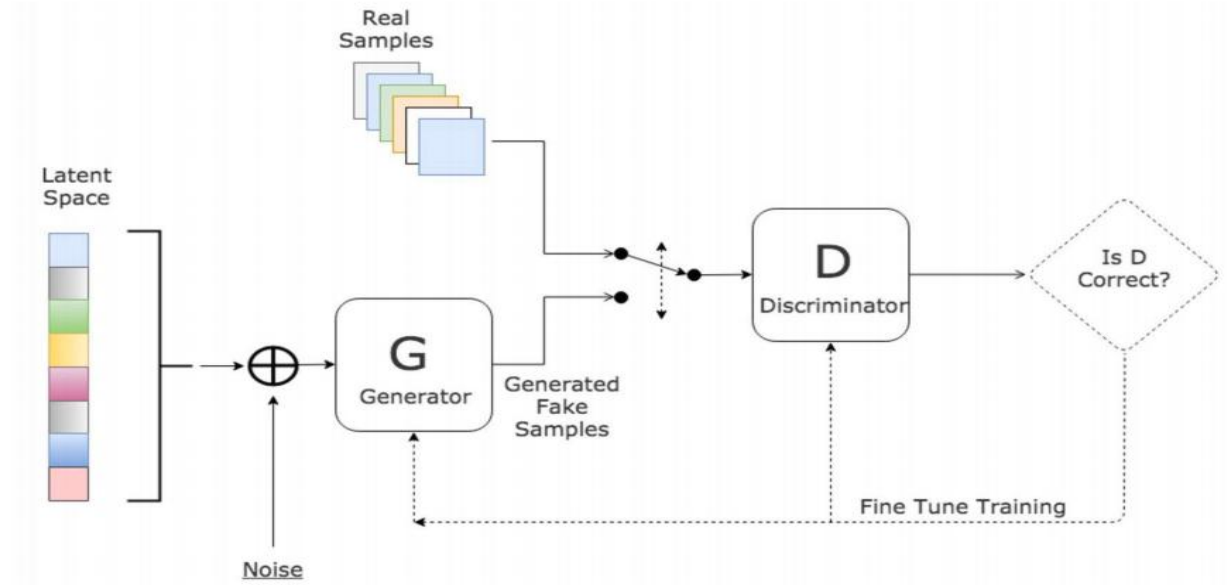




Introduction

arXiv:1406.2661

- Generative Adversarial Networks (GAN)
  - A discriminator (D) tries to discriminate the real and fake data; a generator (G) to produce fake data, tries to confuse discriminator
  - Train D and G alternately to improve performance

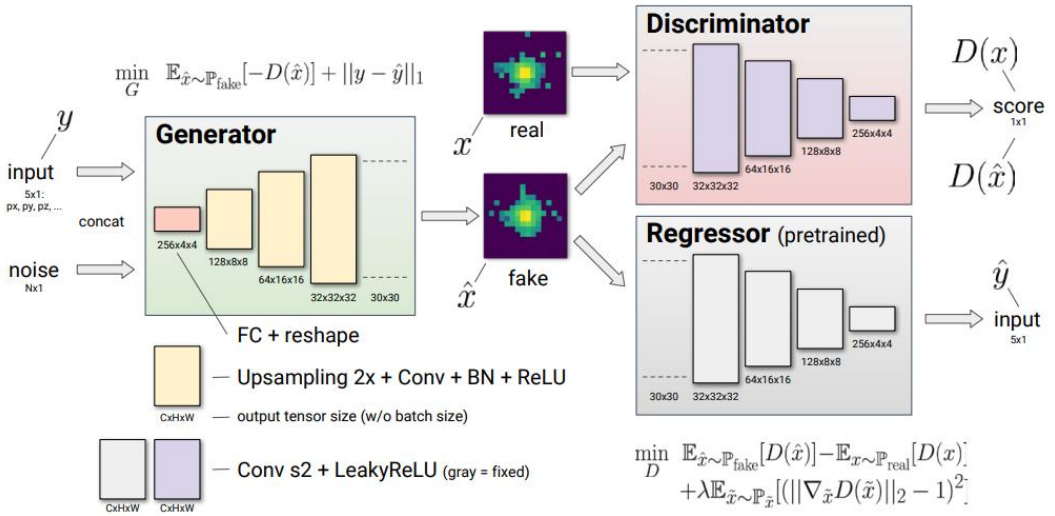


$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))]$$

Introduction

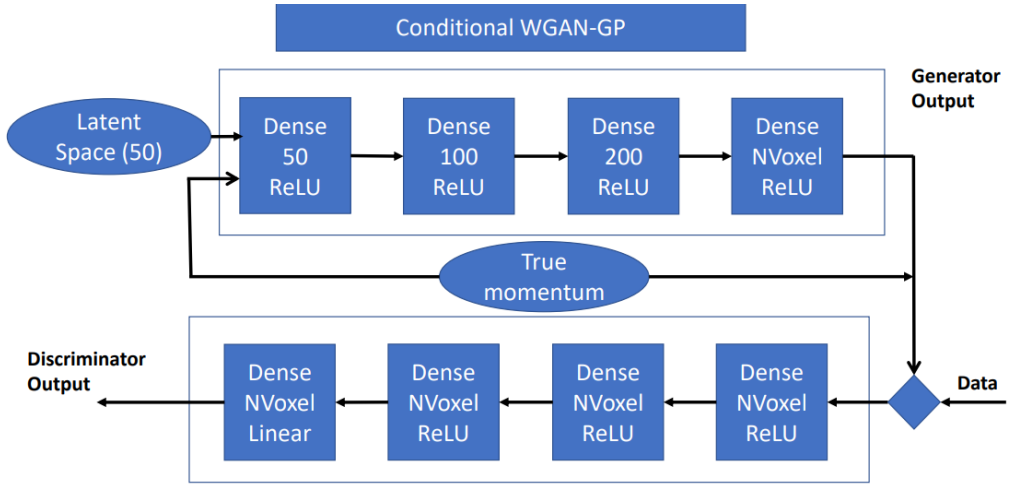
- ML, especially GAN, is used for simulation at LHCb and ATLAS
  - A pretrained regressor, for further constrain G result
  - Train ~ 100 GANs for different ATLAS detector regions
- BESIII is an ideal place to perform ML-simulation: simpler detector, smaller condition space

EPJ Web Conf. **214** 02034 (2019)



caloGAN @ LHCb

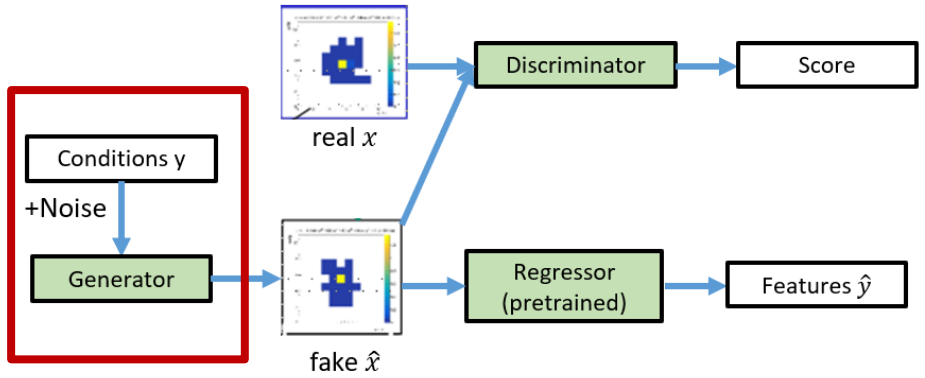
ACAT 2024

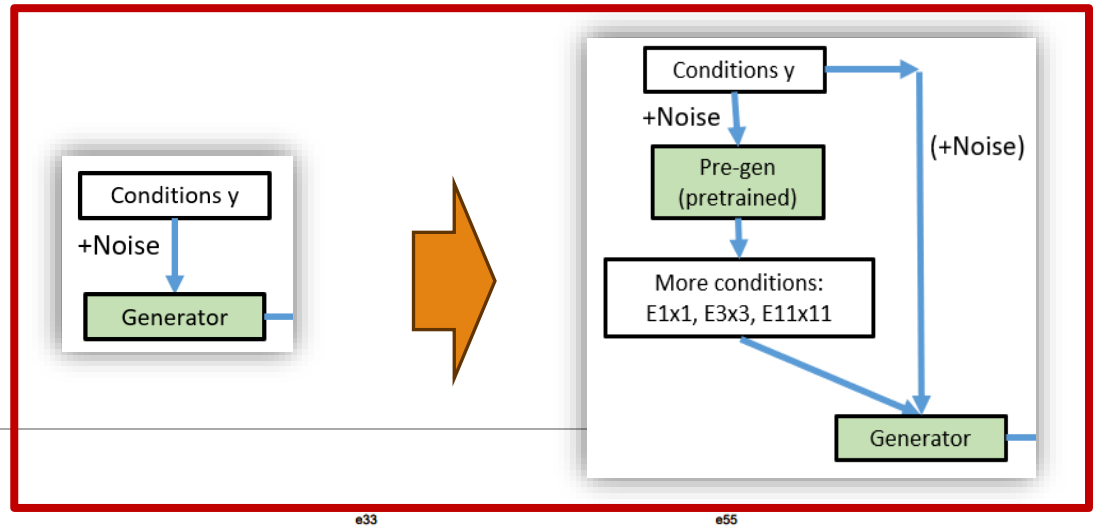


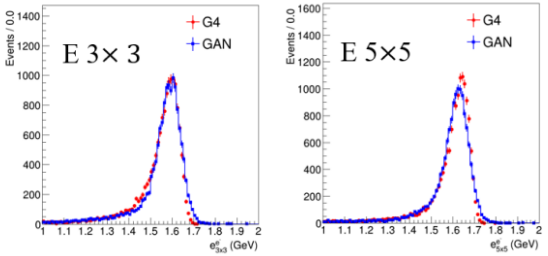
FastCaloGAN @ ATLAS

An Advanced Condition Generator

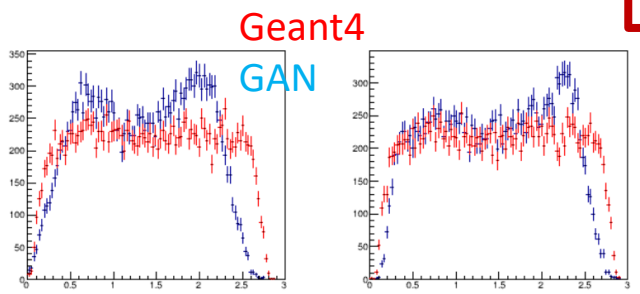
- BESIII tests the GAN simulation based on Bhabha ($e^+e^- \to e^+e^-$) events
- The model does not work well in a larger condition space
- To cover the full condition space, a pretrained pre-gen model is integrated



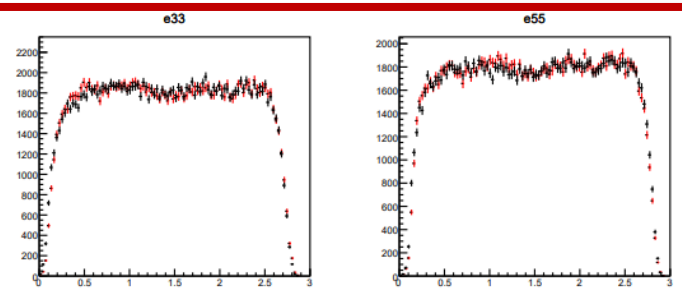




GAN for Bhabha events



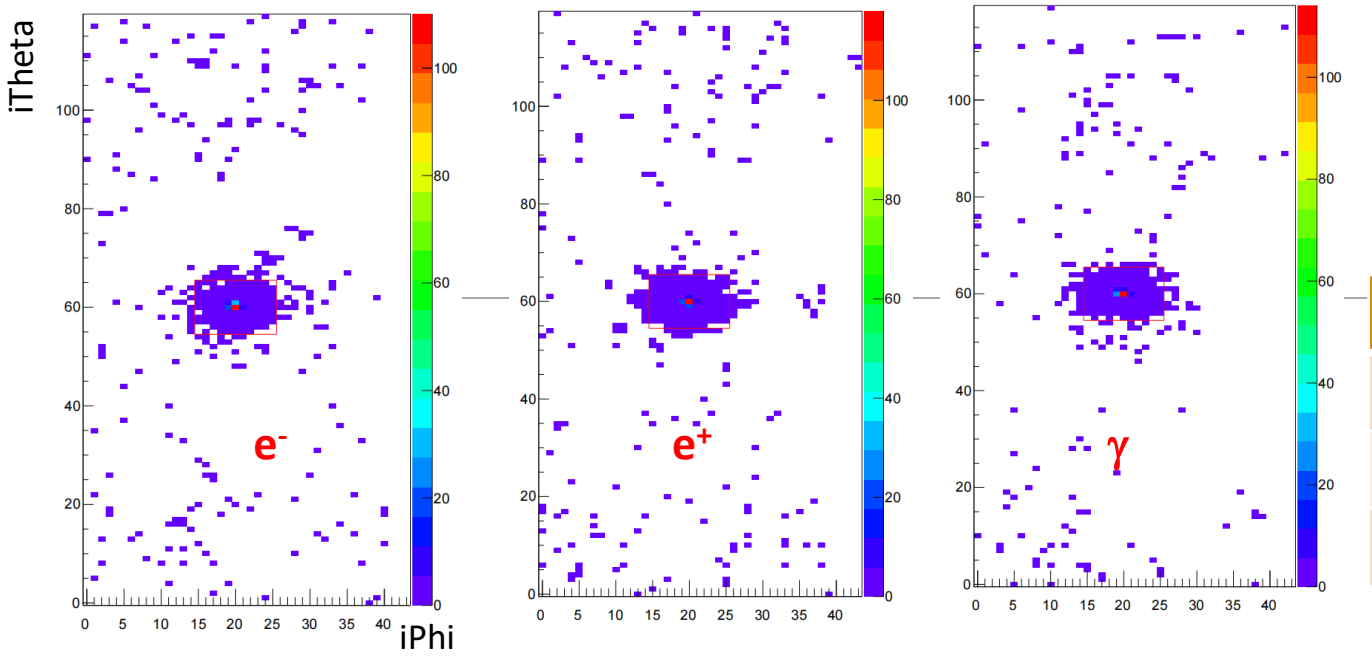
GAN in large condition space



Pre-gen + GAN

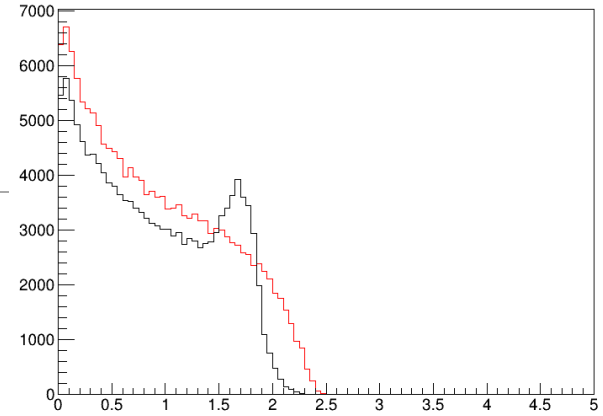
Samples

- We simulate ~1M single-track events for $e^+/e^-/\gamma$ as training set
  - Full condition space sample: $0 < P < 3\ \mathrm{GeV}/c, 0 < \theta < 2\pi, 0 < \phi < 2\pi$
  - Single-momentum samples: $P = 0.5, 0.8, 1.2, \ldots 2.5\ \mathrm{GeV}/c$
  - Remove random trigger, randomness of IP, readout cutoff
- The 11x11 region contains nearly the entire detector response



EMC hit map in barrel region from 100 events

	E_{11x11} / E_{tot} (%)
e^+	99.7
e^-	99.8
γ	99.8

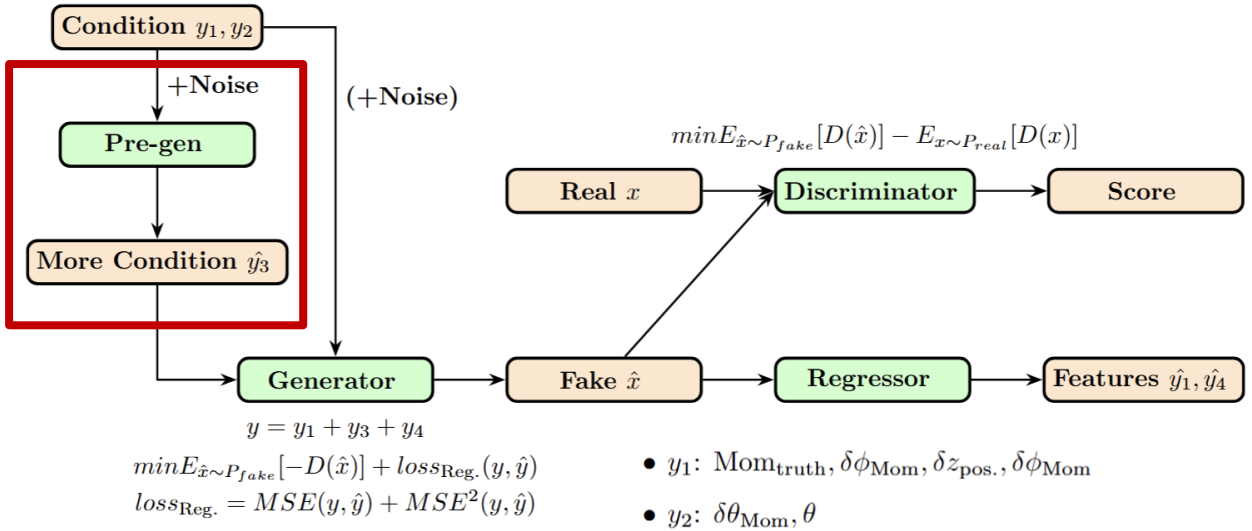


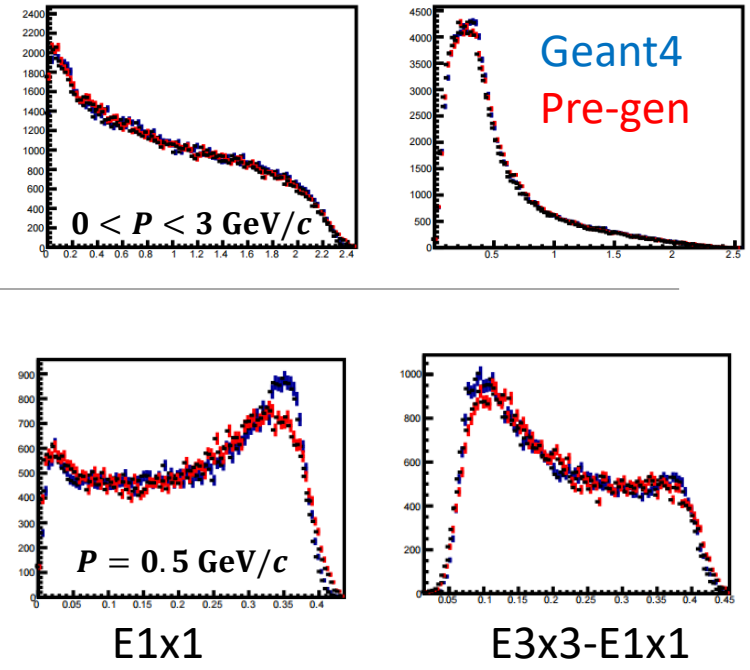
E1x1 w/ and w/o readout cutoff

EMC simulation with GAN

EMC simulation with ML - GAN

- Similar strategy as LHCb case: G + D + regressor (pretrained) + pre-gen (pretrained)
- The pre-gen model: an advanced condition generator
  - The full space sample: the basic training
  - Single-momentum samples: compare the resolution after each 5 steps and apply an additional optimization step
- Use $\hat{y}_{3\,\mathrm{truth}}$ in training, $\hat{y}_{3\,\mathrm{pre-gen}}$ in simulating

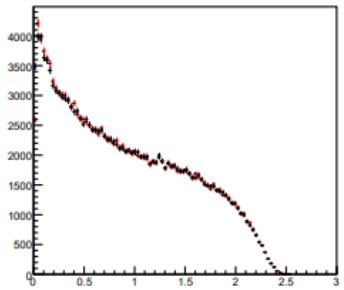




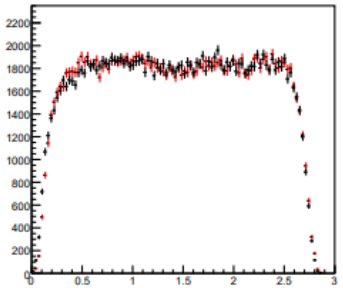
EMC simulation with ML – GAN for e^+

- ML reachs comparable accuracy to Geant4
  - In the full condition space, and for the single-momentum sample
- ~ 2 s for 50k tracks

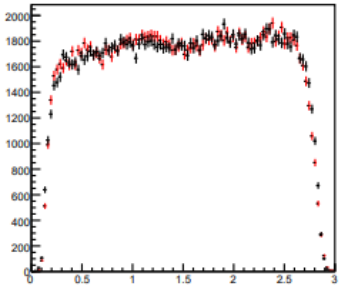
$\boldsymbol{0 < P < 3\ \mathrm{GeV}/c}, 0 < \theta < 2\pi, 0 < \phi < 2\pi$



E1x1



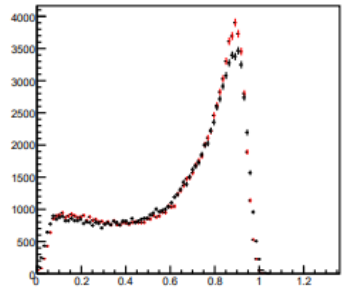
E3x3



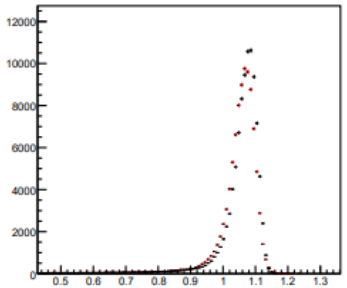
E11x11

Geant4
GAN

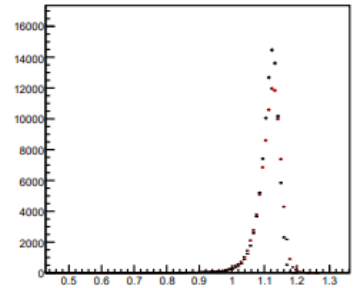
$\boldsymbol{P = 1.2\ \mathrm{GeV}/c}, 0 < \theta < 2\pi, 0 < \phi < 2\pi$



E1x1

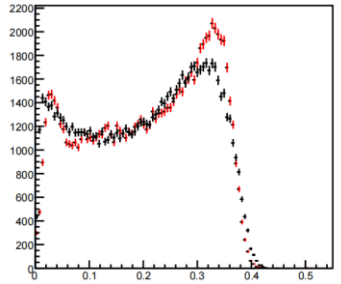


E3x3

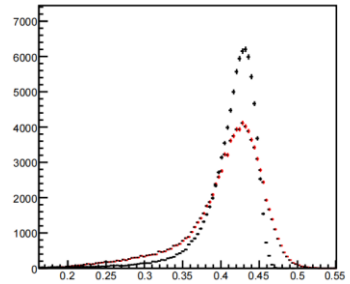


E11x11

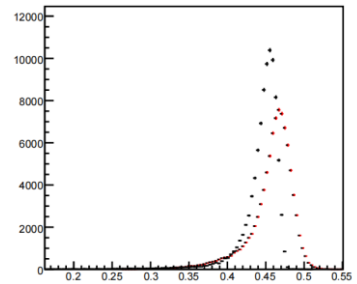
$\boldsymbol{P = 0.5\ \mathrm{GeV}/c}, 0 < \theta < 2\pi, 0 < \phi < 2\pi$



E1x1



E3x3



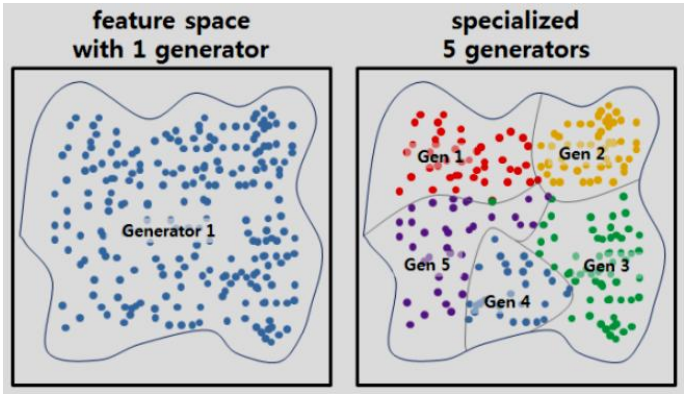
E11x11

EMC simulation with MoE-GAN

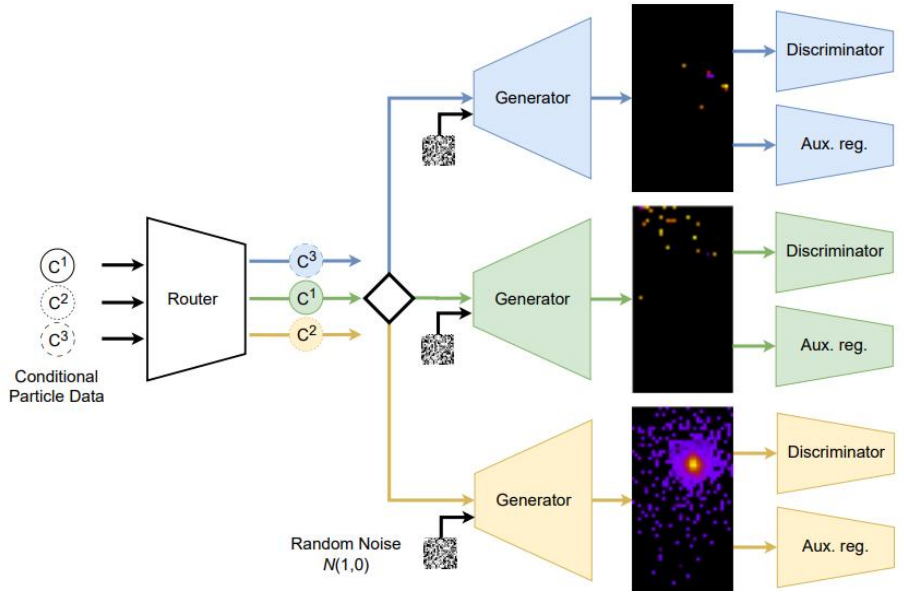
EMC simulation with MoE-GAN

- Mixture of Experts (MoE): a neural network architecture that scales model capacity by using multiple specialized “expert” sub-networks, with a gating mechanism that activates only a subset of experts per input token.

1805.02481

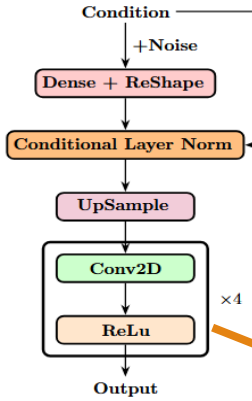


2508.20991 (ZDC @ ALICE)

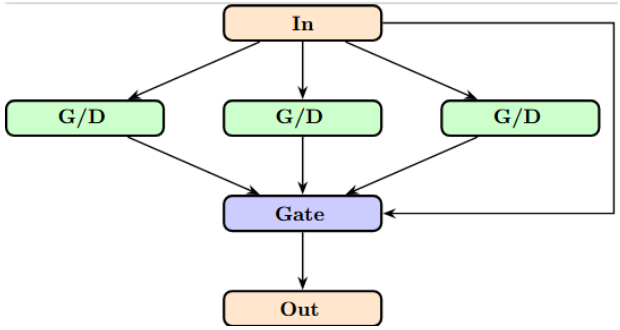


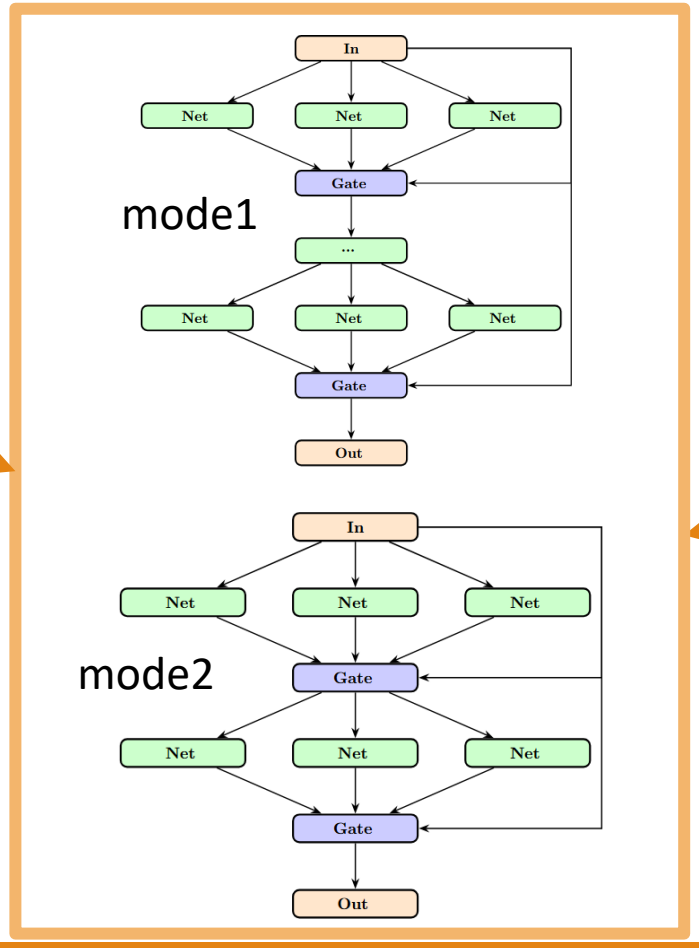
EMC simulation with MoE-GAN

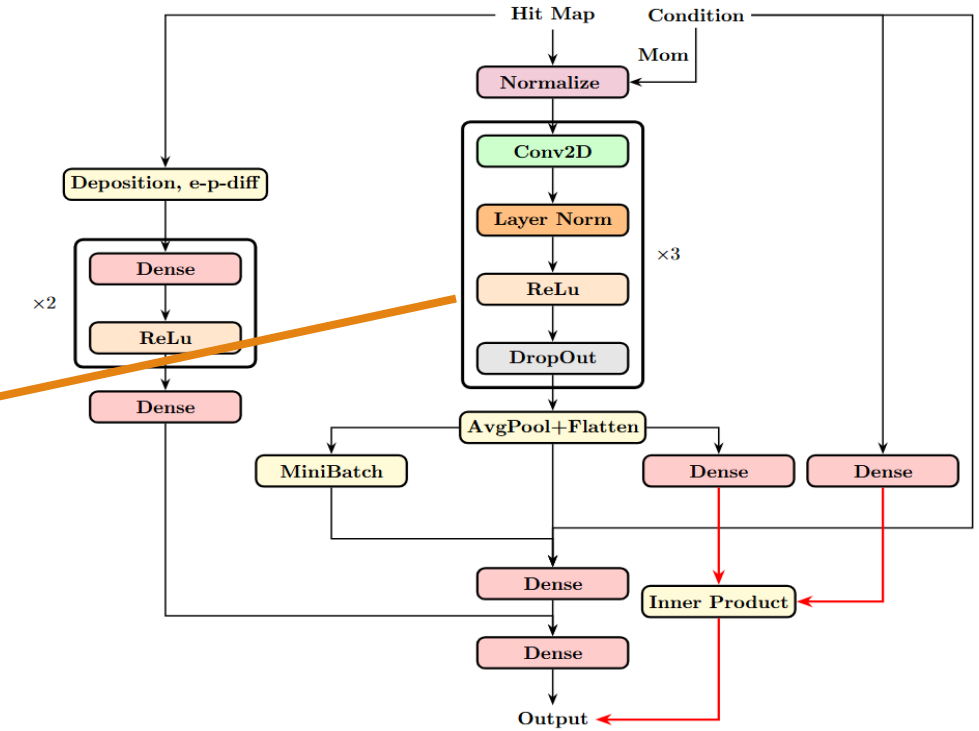
- Possible design: among convolutional layers, or at the whole-model level



Generator





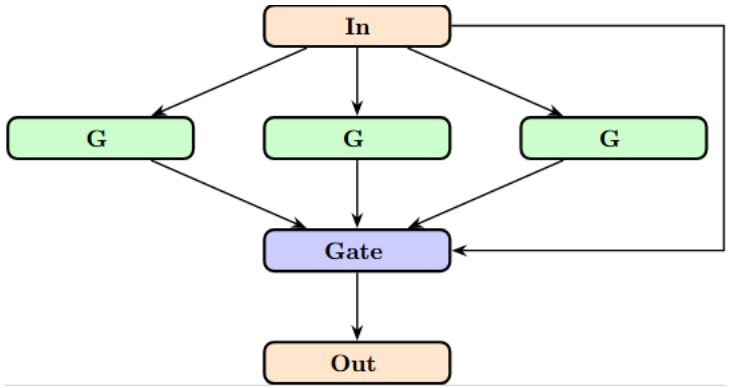


Discriminator

EMC simulation with MoE-GAN

- Speed is acceptable
- Improved quality for low-momentum tracks

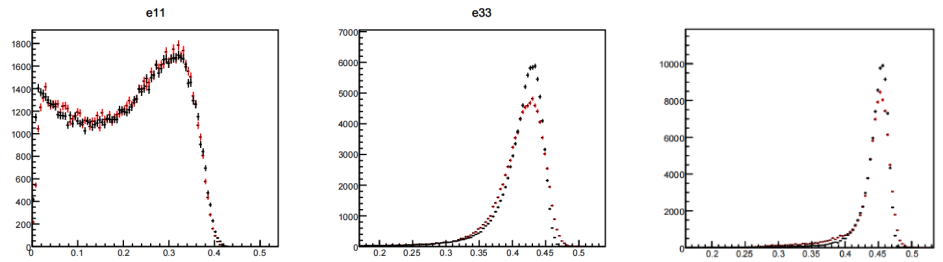
	Time * 95k evts
Nominal	~3~4 s
2 Generator	~5 s
3 Generator	~7 s
10 Generator	~11 s



$P = 0.5\ \mathrm{GeV}/c$

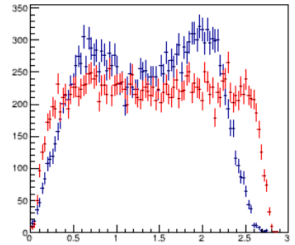
nominal

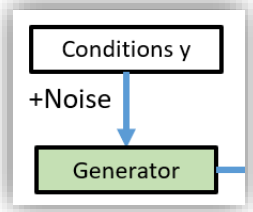
3Gen



Summary & Next

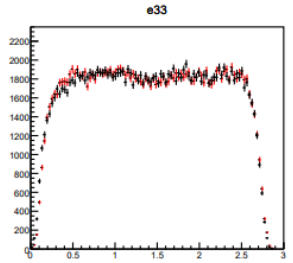
- Simulation for electromagnetic calorimeter with ML reachs comparable accuracy to Geant4
  - For the full EMC detector
  - Within full condition space
- A pretrained pre-gen model improves performance in the full condition space
  - More condition for Generator
- MoE improve the accuracy with limited affect on speed

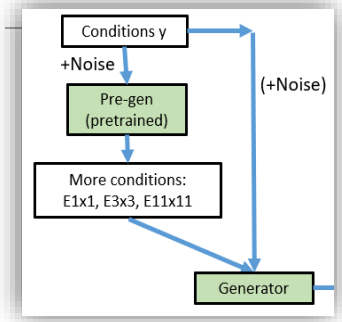




GAN w/o pre-gen

- Next:
  - Test other MoE design

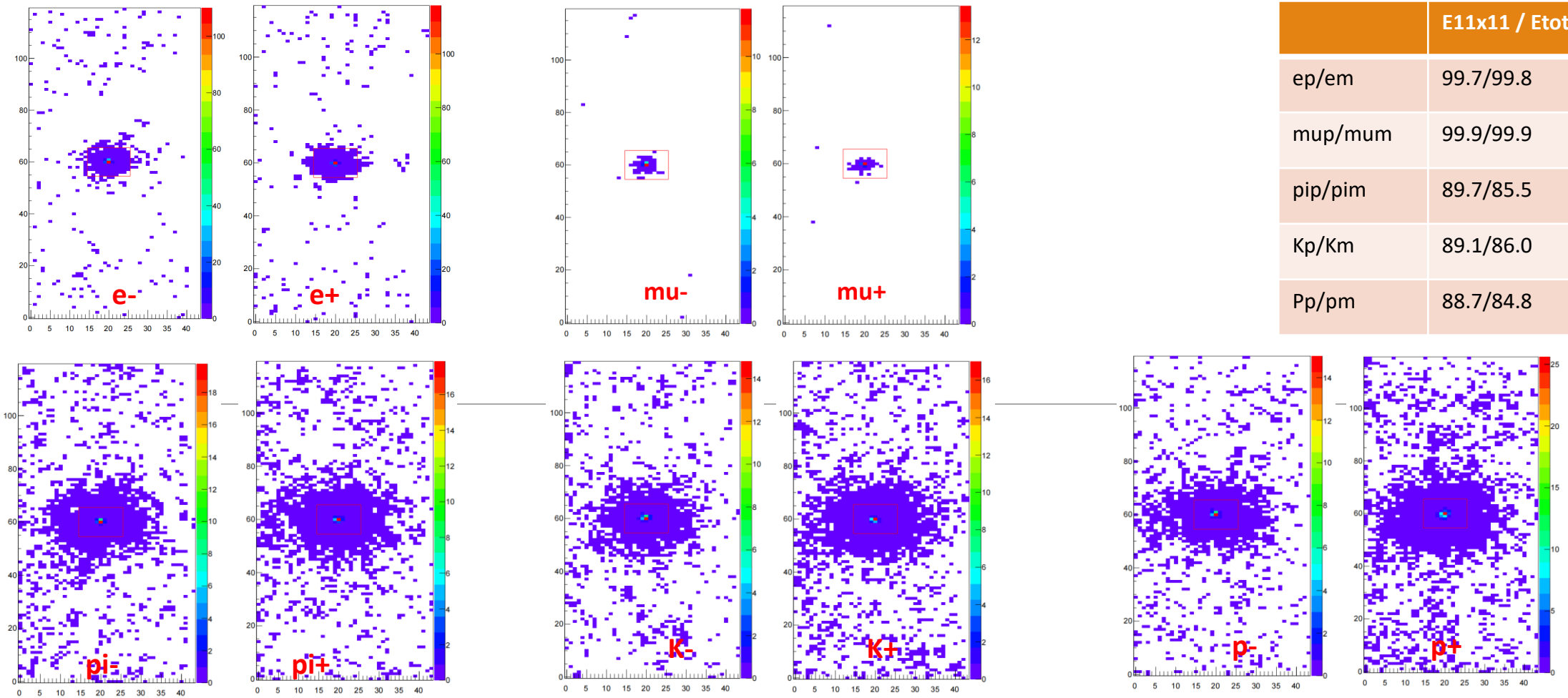




pre-gen + GAN

Readout region

- EMC hit map in barrel region from 100 events with P>1 GeV/c



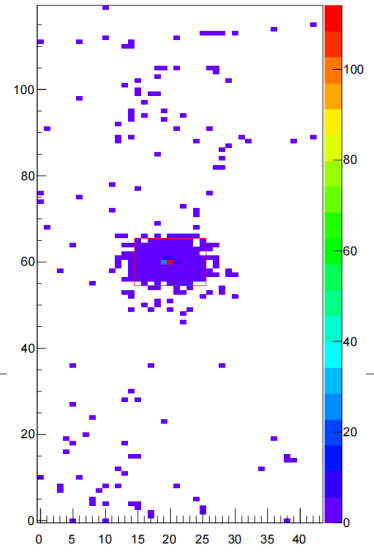
	E11x11 / Etot (%)
ep/em	99.7/99.8
mup/mum	99.9/99.9
pip/pim	89.7/85.5
Kp/Km	89.1/86.0
Pp/pm	88.7/84.8

Readout region

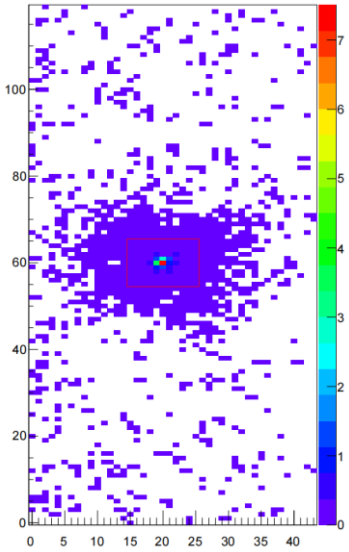
- EMC hit map in barrel region from 100 events with P>1 GeV/c

	E11x11 / Etot (%)
Gamma	99.8
n	87.7
nbar	91.6

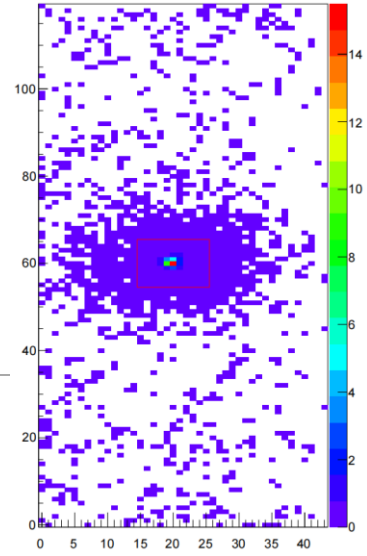
gamma



n



nbar



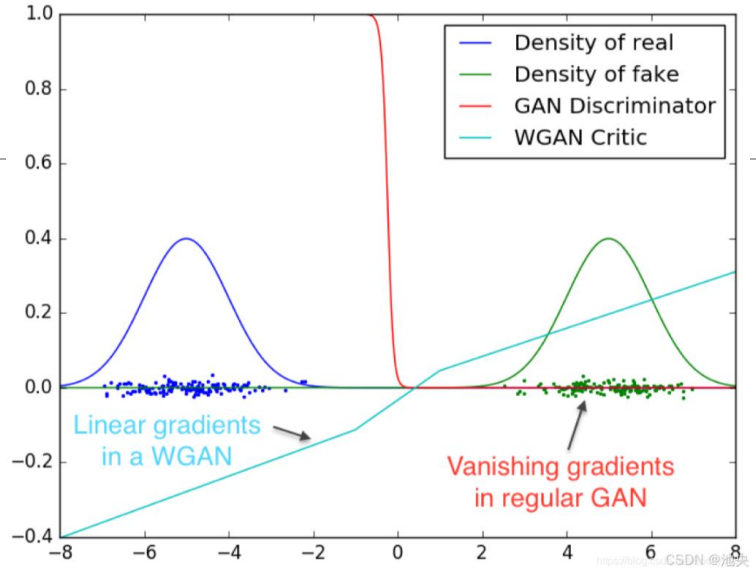
wGAN

arXiv:1701.07875

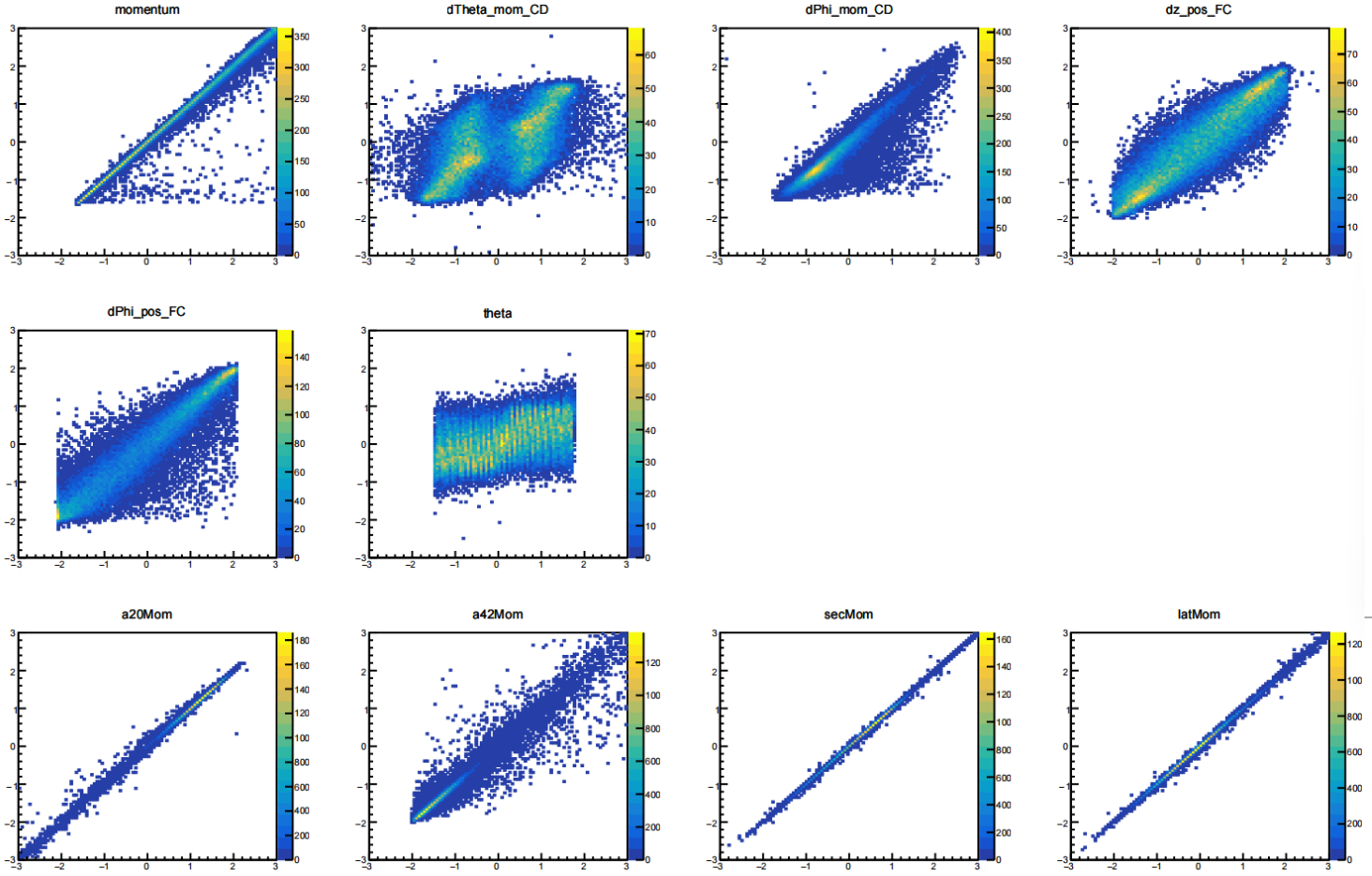
- Wasserstein GAN, an improved GAN
  - To solve instability and gradient vanishing
    - Especially if two distributions are non-overlapping.
  - Replace Discriminator with a Critic
    - Replace JS divergence with the Wasserstein distance, print scores instead of probabilities
    - Lipschitz Constraint: via weight clipping or gradient penalty (wGAN-GP)
  - Loss for D(C) and G:
    - $L_{\text{critic}} = \mathbb{E}_{x \sim P_r}[f_w(x)] - \mathbb{E}_{z \sim P_z}[f_w(G(z))]$
    - $L_G = -\mathbb{E}_{z \sim P_z}[f_w(G(z))]$

K-Lipschitz Constraint:

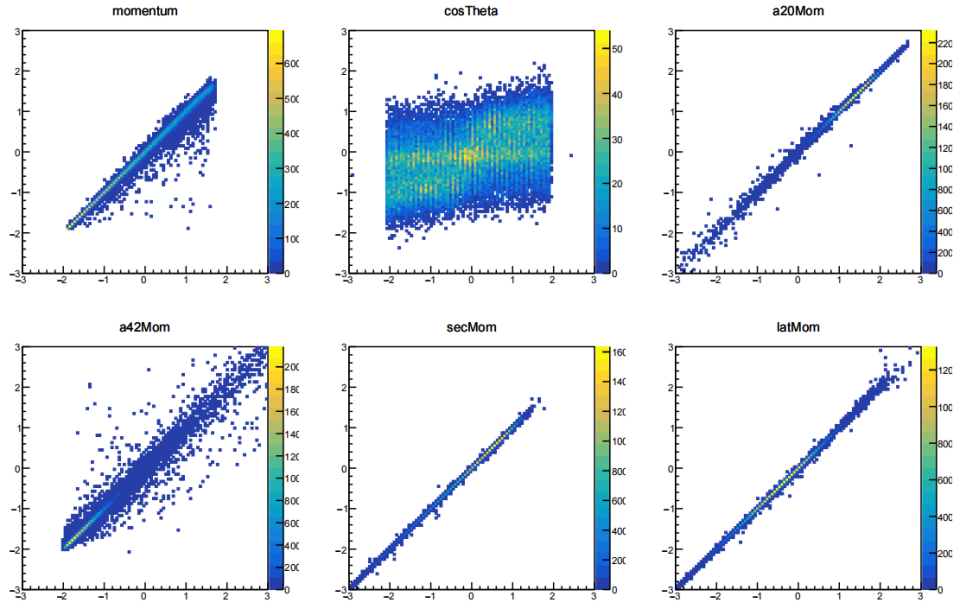
$$|f(x_1) - f(x_2)| \leq K|x_1 - x_2|$$



Regressor

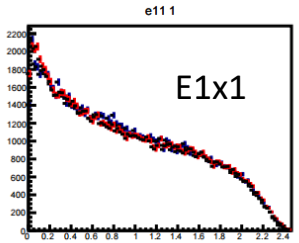


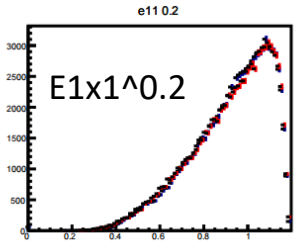
e^+

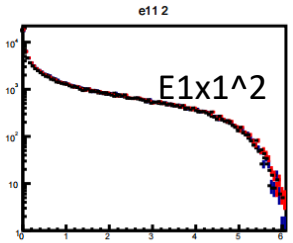


γ

Pre-gen result



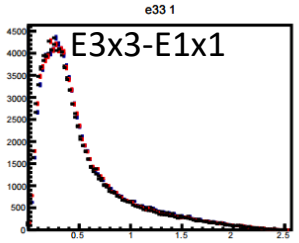


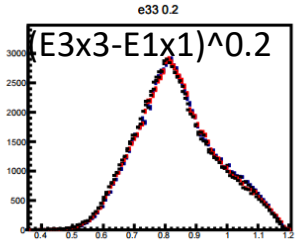


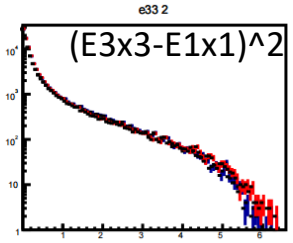
E3x3

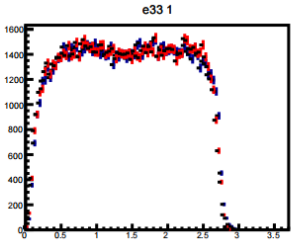
E3x3^0.2

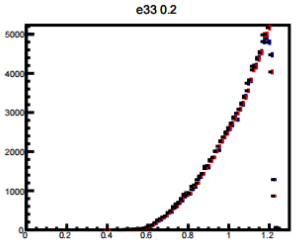
E3x3^2

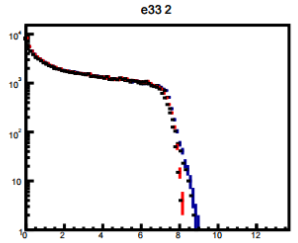


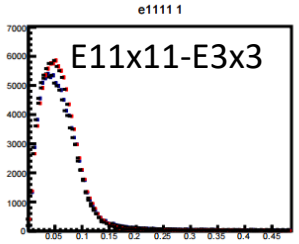


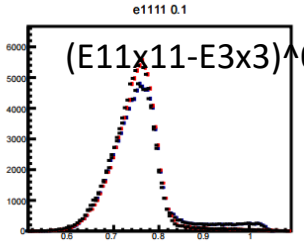


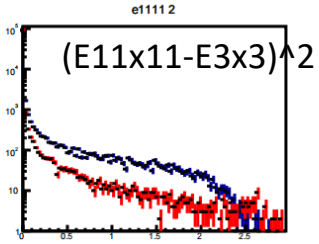


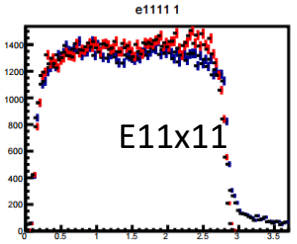


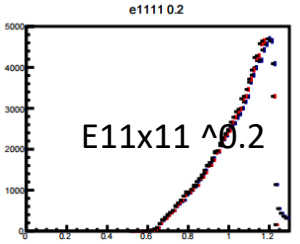


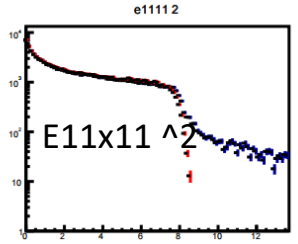








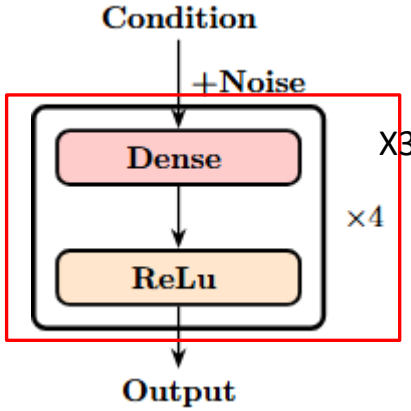




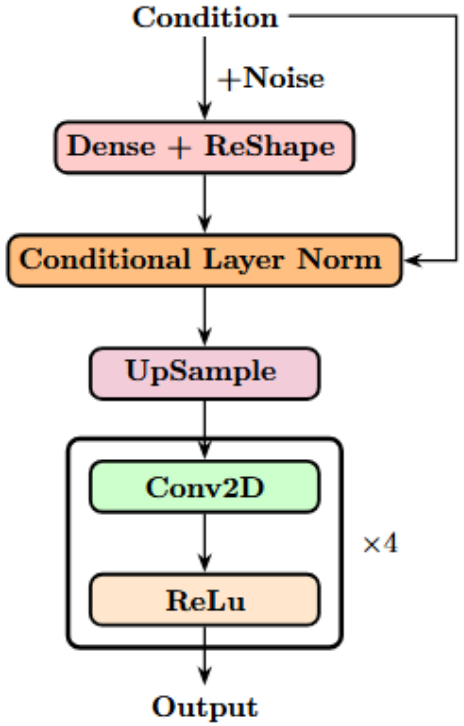
Use momentum to replace E11x11 in more_condition

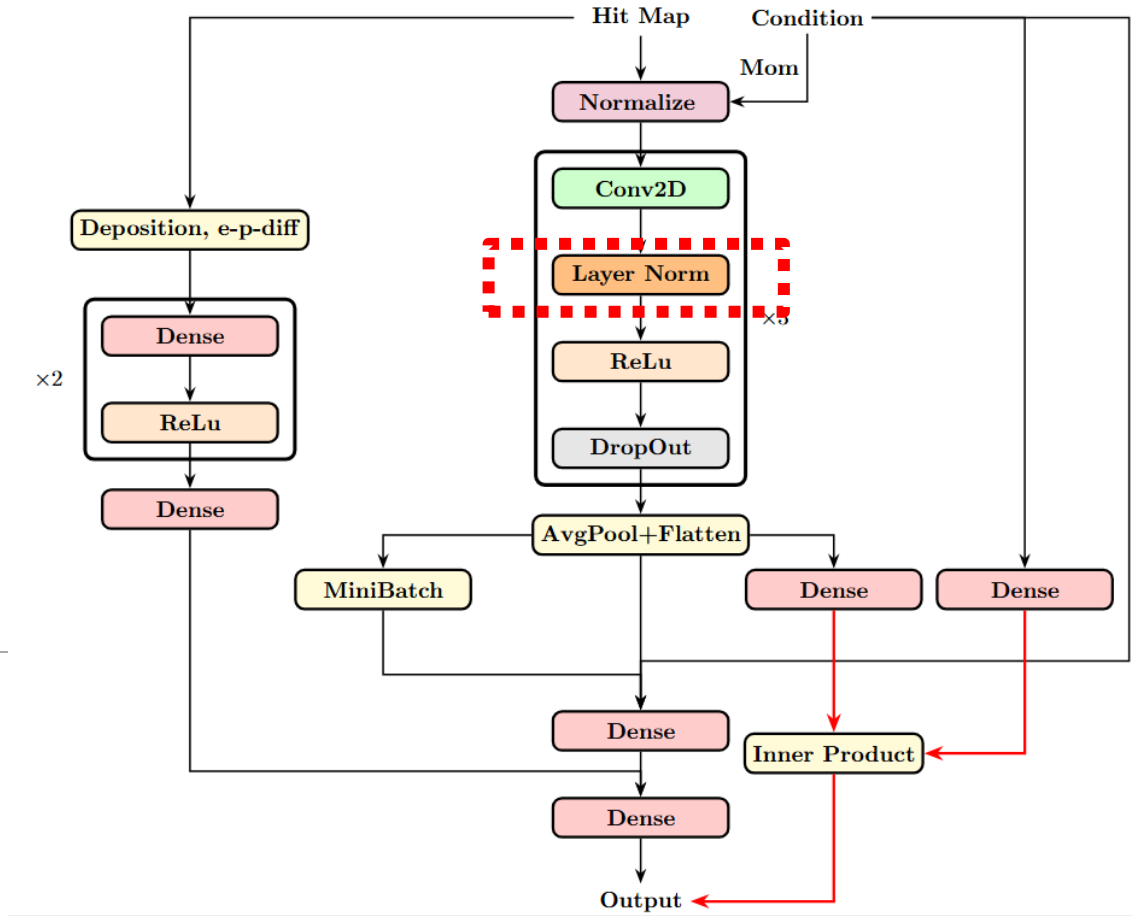
Current models

- Pre-gen, Generator, Discriminator



X3 expert





Pre-gen

Generator

Discriminator

cGAN with projection D

- Keep the previous condition-injection

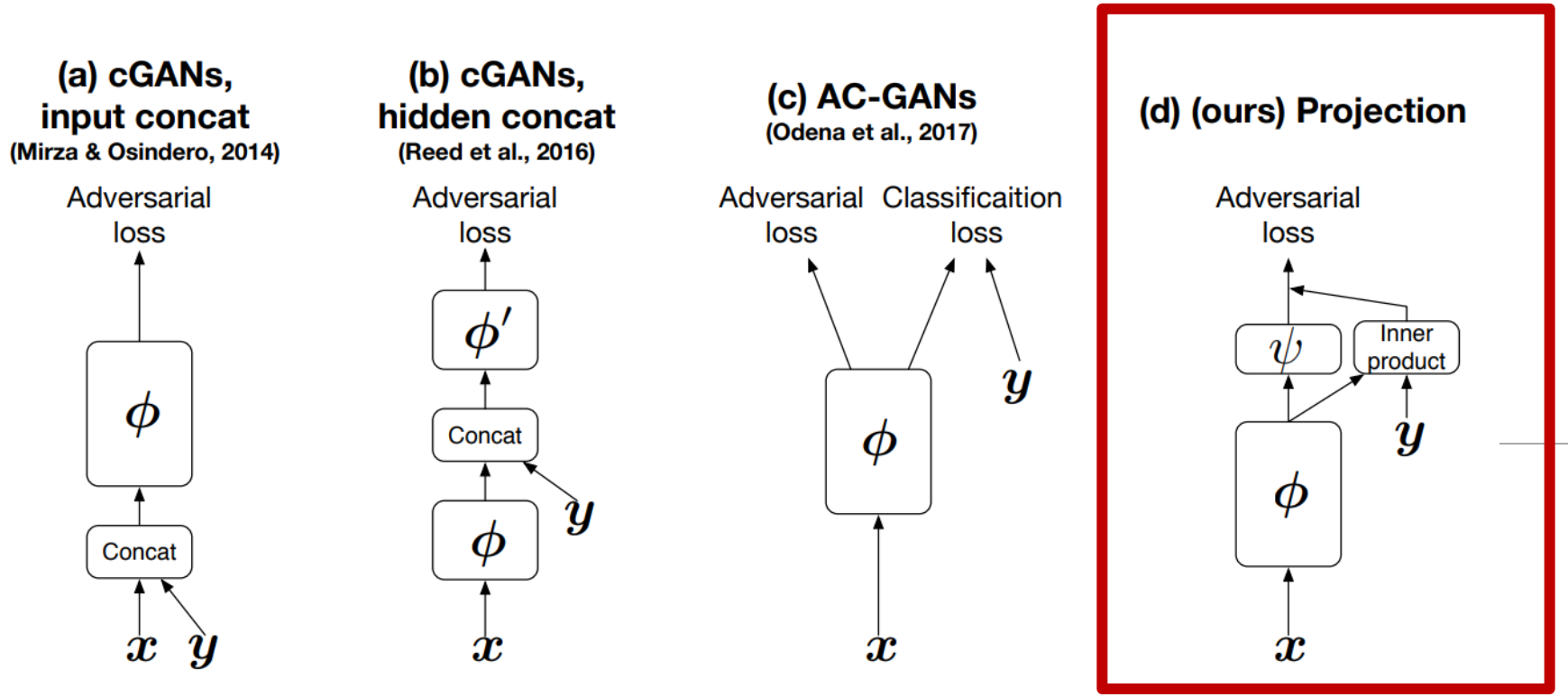
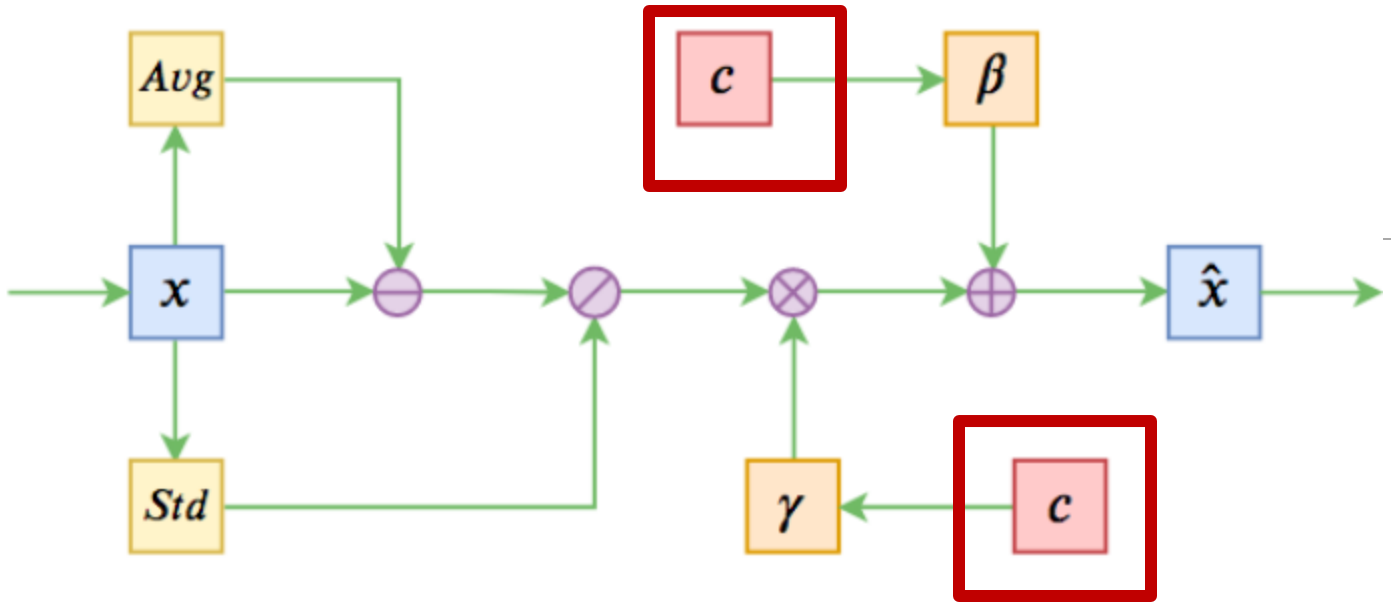


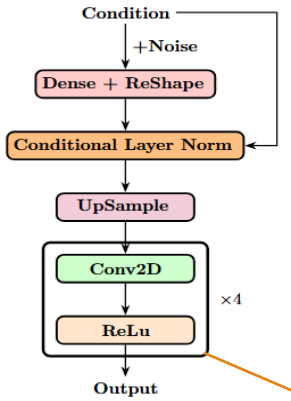
Figure 1: Discriminator models for conditional GANs

Conditional Layer Normalization

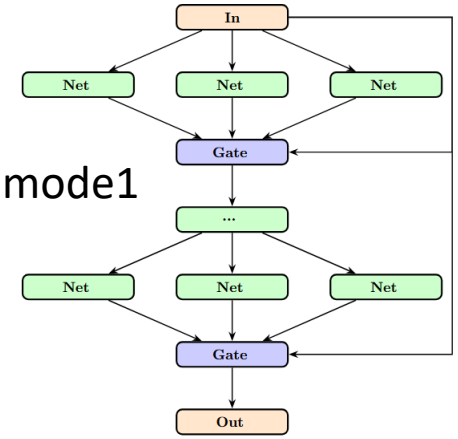
$$LN(x) = \gamma\left(\frac{x-\mu}{\sigma}\right) + \beta$$

$$CLN(x) = \gamma'\left(\frac{x-\mu}{\sigma}\right) + \beta'$$
$$\gamma' = \gamma + w_\gamma * c$$
$$\beta' = \beta + w_\gamma * c$$

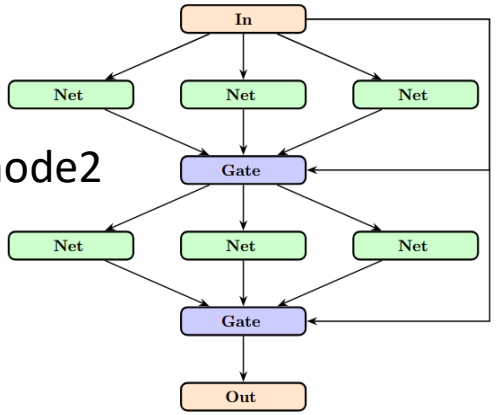




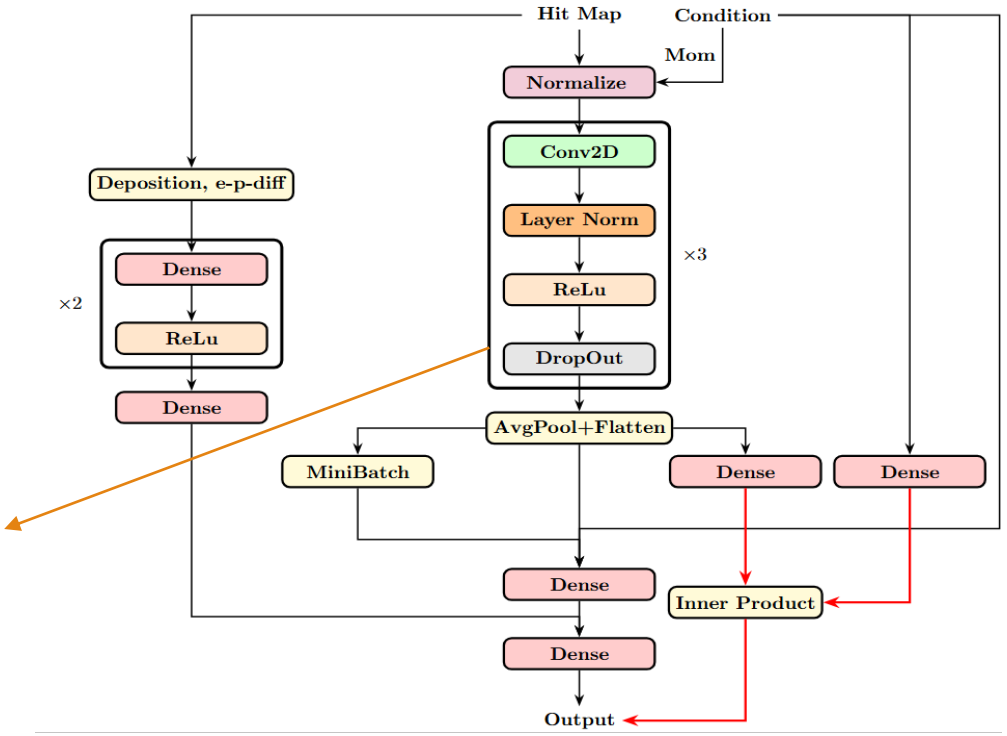
Generator



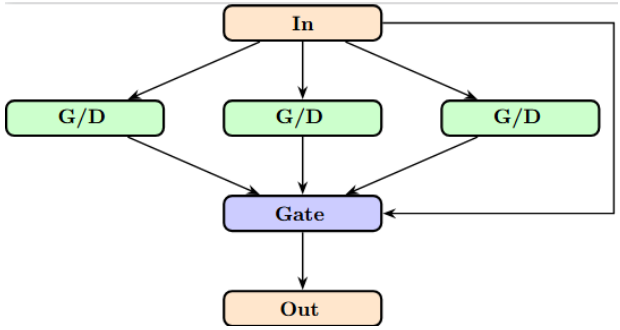
mode1



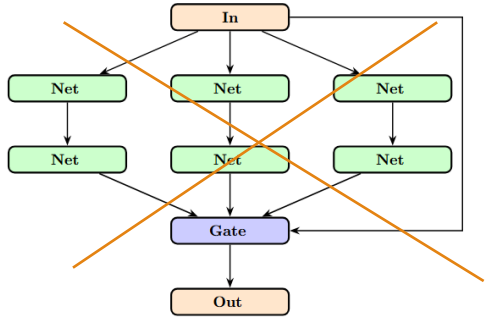
mode2

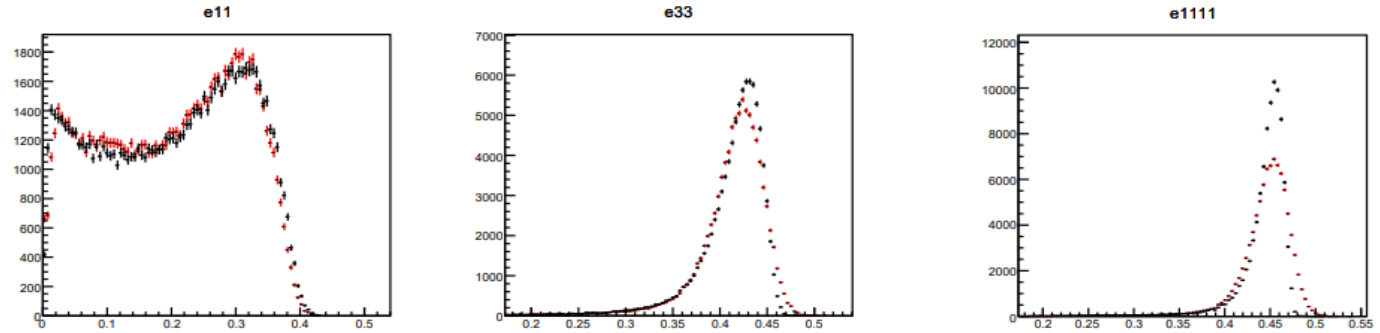


Discriminator

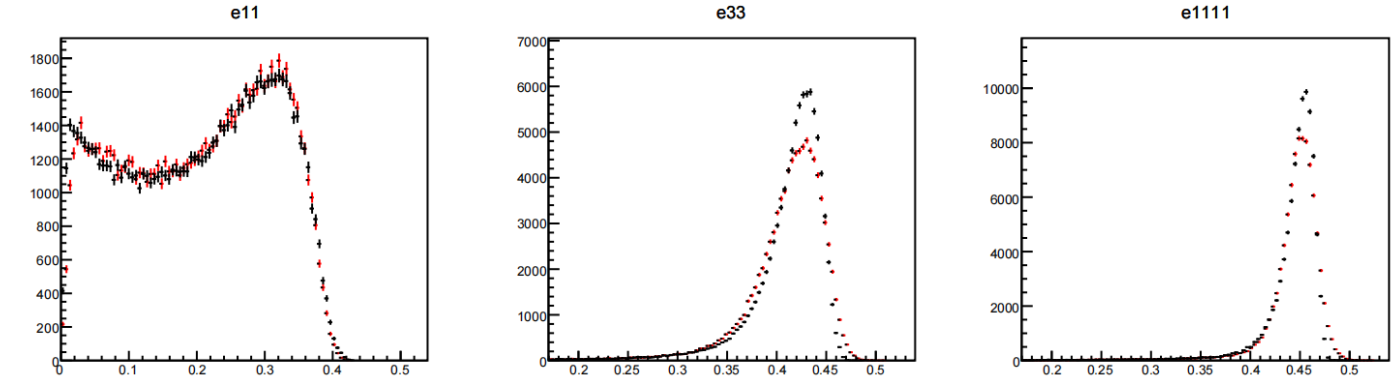


差别应该不大





pre-gen_truth in training



pre-gen in training