Conditional generation of LHCb calorimeter response

Vladislav Belavin
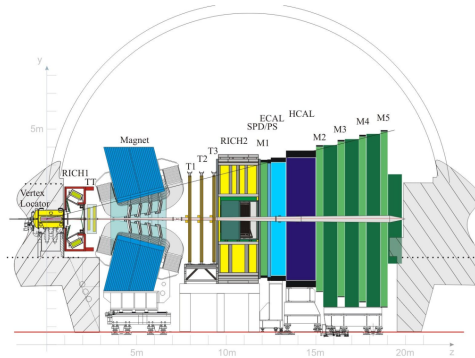
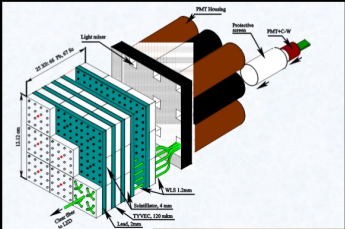National Research University Higher School of Economics

July 5, 2019
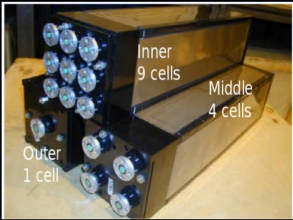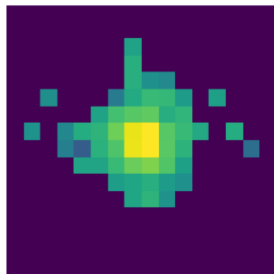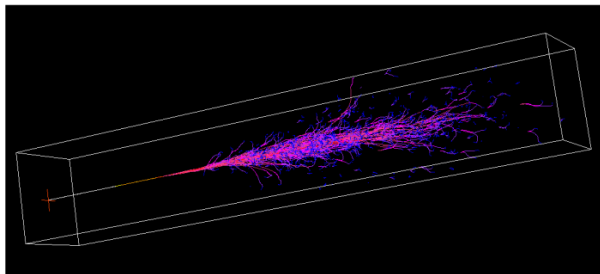
LHCb experiment

LHCb is a forward arm spectrometer with pseudorapidity range $2 < \mu < 5$ that was originally designed for the study of b-physics.

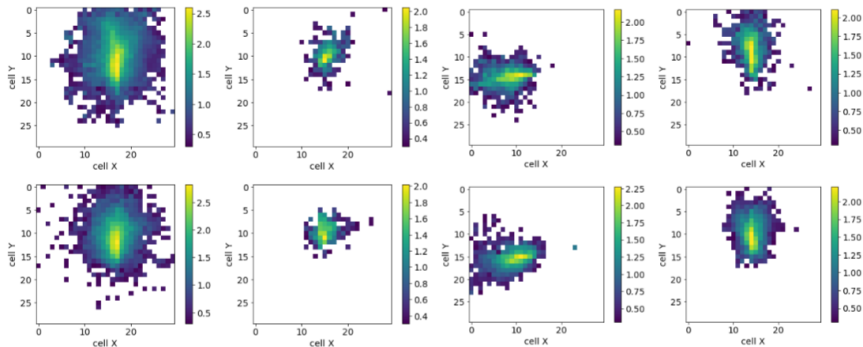SCHOOL OF DATA ANALYSIS

Electromagnetic calorimeter

Electromagnetic samples

Dataset

Dataset consists of $[30 \times 30]$ images of calorimeter responses on particles with different initial positions and momenta.

(a) $E_0 = 63.7$ GeV   (b) $E_0 = 6.5$ GeV   (c) $E_0 = 15.6$ GeV   (d) $E_0 = 15.9$ GeV

Problem statement

Train conditional generator that models following probability density:

$$p\,(\text{EnergyDeposit}|\text{ParticlePoint}, \text{ParticleMomentum}, \text{ParticlePDG})$$

SCHOOL OF DATA ANALYSIS

Metric

Minimum of two PRD scores:

Over embeddings of images:

> catches overall proximity of distributions generated and real calorimeter responses

Over a set of physical metrics:

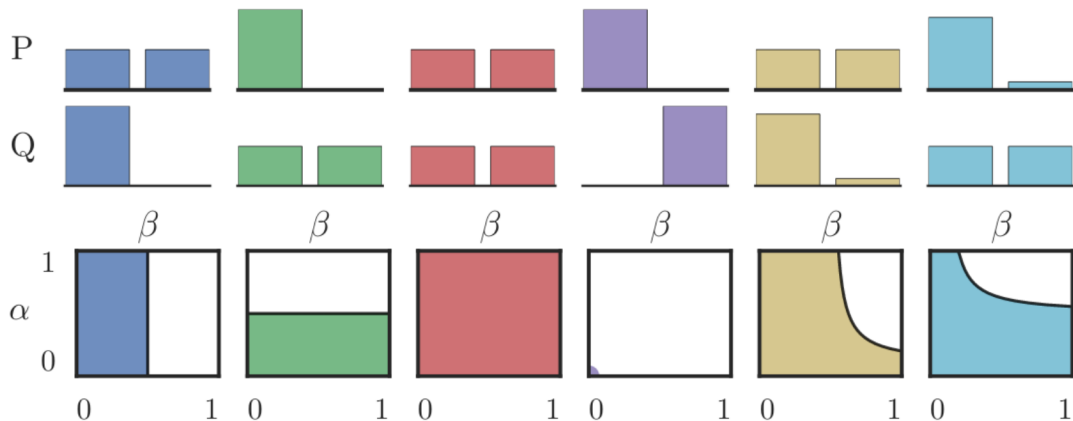> catches proximity of distributions for handmaded statistics

PRD score (1)

PRD score is a score that disentangles precision(quality of generated samples) from recall(proportion of target distribution that is covered by the generator, richness).
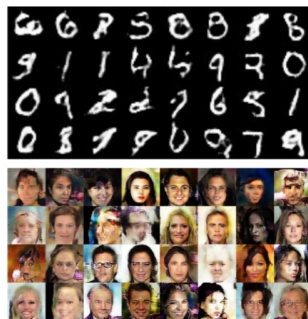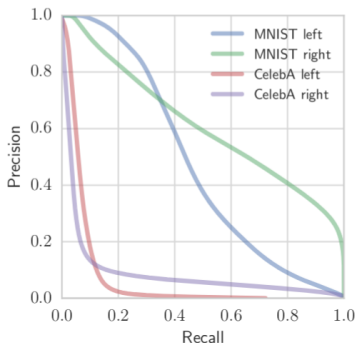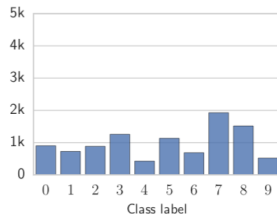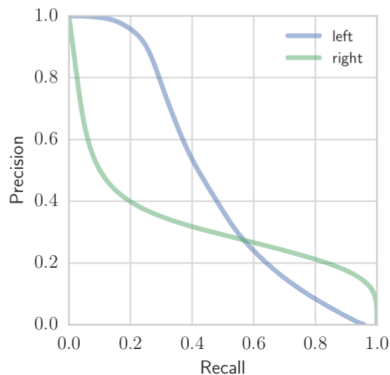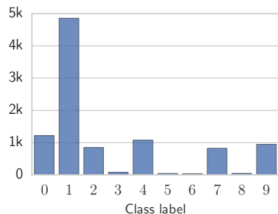
PRD score (2)

Toy example of PRD scores for pairs of Bernoulli distributions $\mathcal{P}$ and $\mathcal{Q}$.
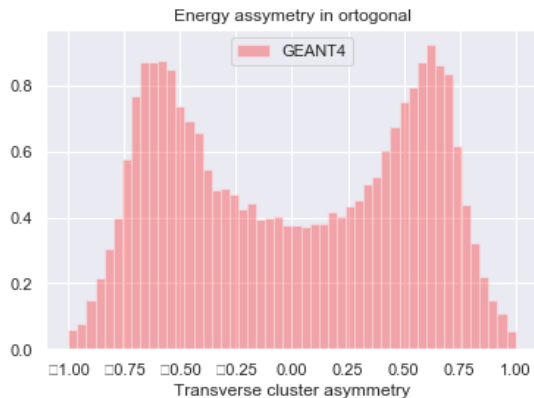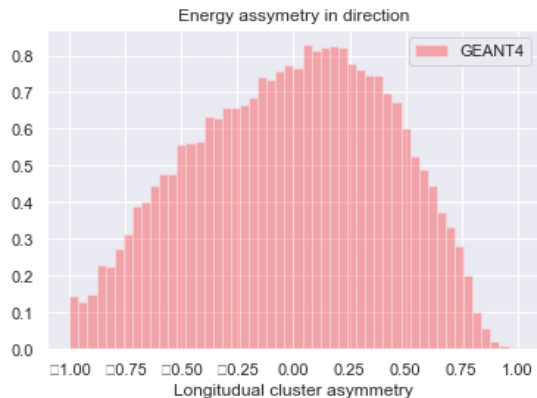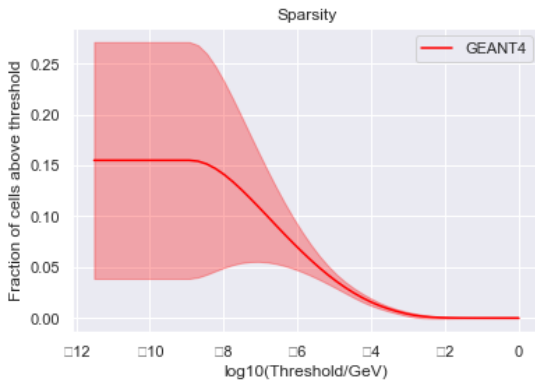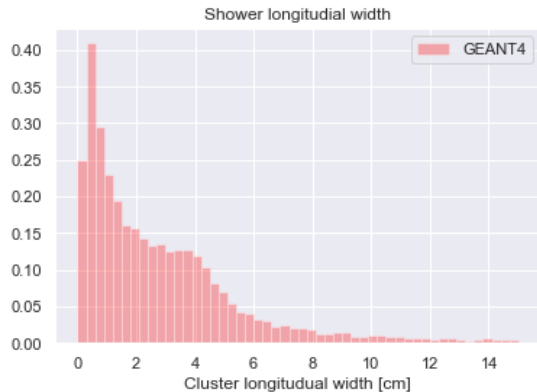
PRD score (3)

For example, imagine two generators: one can sample perfectly 6 out of 10 digits and another one can sample all digits but poorly.

Physical metrics (1)

Energy assymetry in direction

GEANT4

Longitudual cluster asymmetry

Energy assymetry in ortogonal

GEANT4

Transverse cluster asymmetry

Physical metrics (2)

Competition settings

> Data:

> images of ECAL responses
> p_x, p_y, p_z of initial particles
> x, y of initial particles
> Particle type: e^-, γ

> Goal:

> train network to generate realistic LHCb calorimeter responses

SCHOOL OF DATA ANALYSIS

Competition deadlines

Coopetition link:

`https://codalab.coresearch.club/competitions/74`

Start of competition:

05.07.2019

End of competition:

10.07.2019

Submission format

Submission should be a *.zip*-file that contains **2 files**:

> data_val_prediction.npz
> data_test_prediction.npz

Each .npz-file should contain one array of your predictions:

> EnergyDeposit

Let's go to the practice!

Starter-kit:
`https://github.com/SchattenGenie/mlhep2019_2_phase`