Weekly Report

CMS DQM-ML4DC
Patomporn (Jab)
22 July 2019

Outline

- My Pointless Experiment
  - Extended AE
    - Sparse + Contractive AE
    - Sparse + Variational AE
    - Contractive + Variational AE
    - The Standard AE
  - Results
- Simplest prototype of Malfunction Spotter

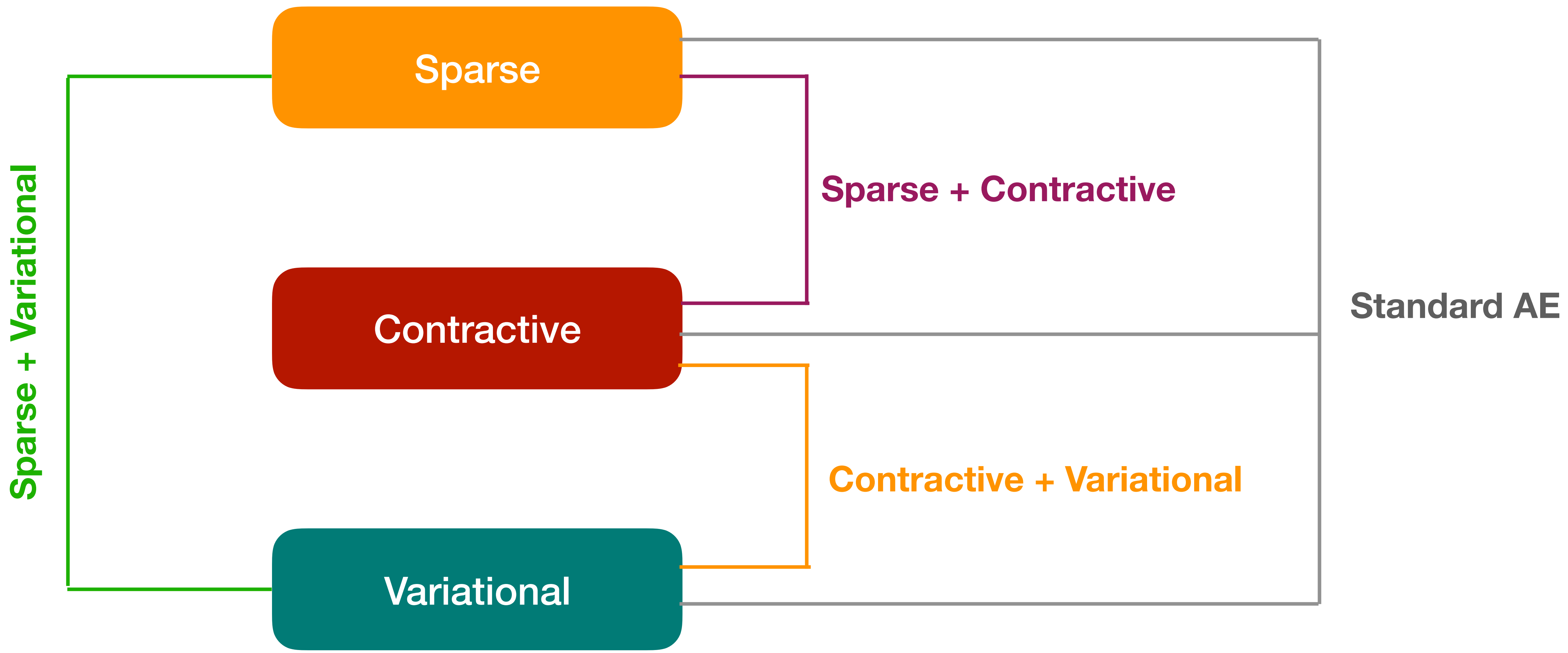
My Pointless Experiment

Extended Model

- Simple idea: could we combine those technique ?
- Add independent constrain into the same model and then we could consider loss function as

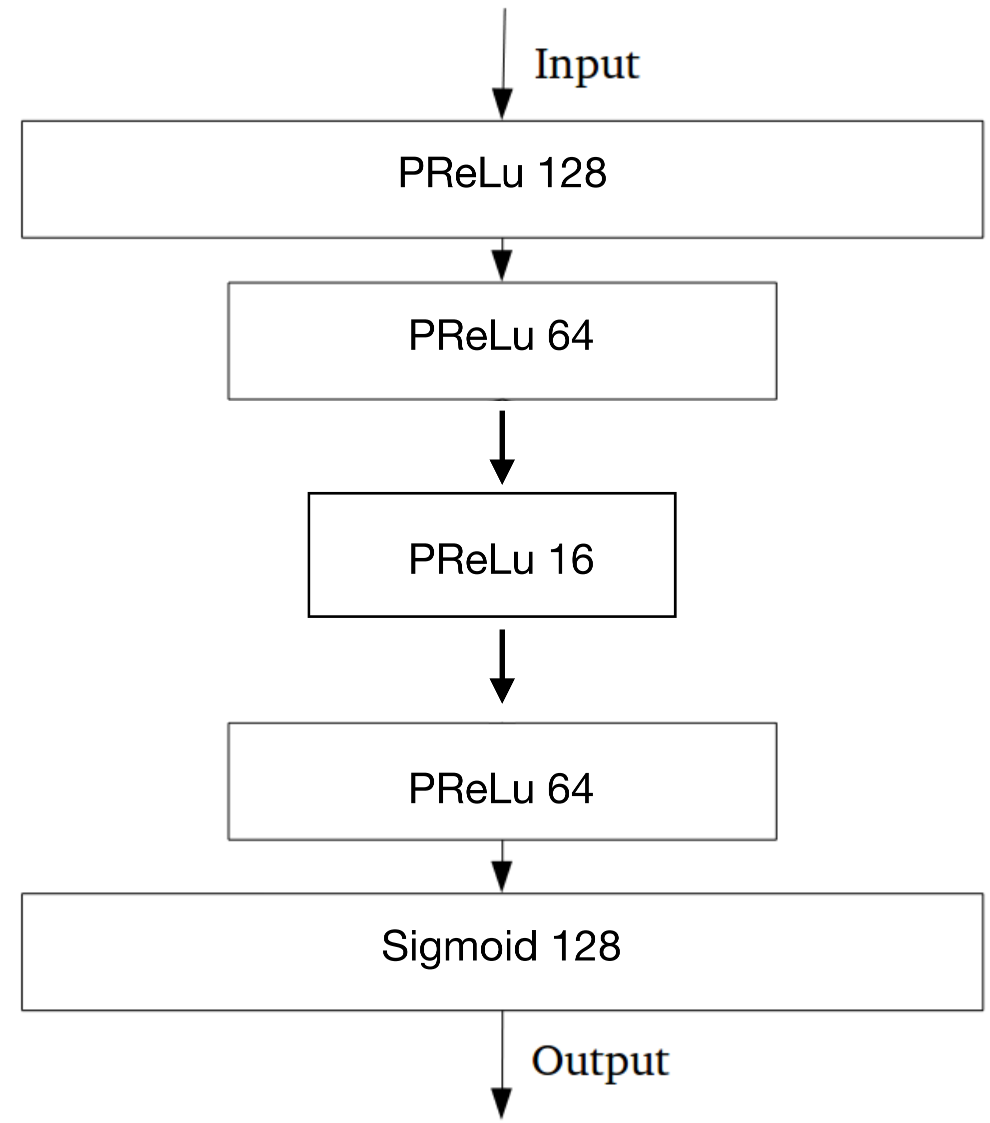
$$\mathcal{L}_{\mathrm{tot}} \equiv \mathcal{L}_{\mathrm{MSE}} + \sum_i \mathcal{L}_{\mathrm{Constrain}_i}$$

Combination of Constrain



Sparse + Contractive AE

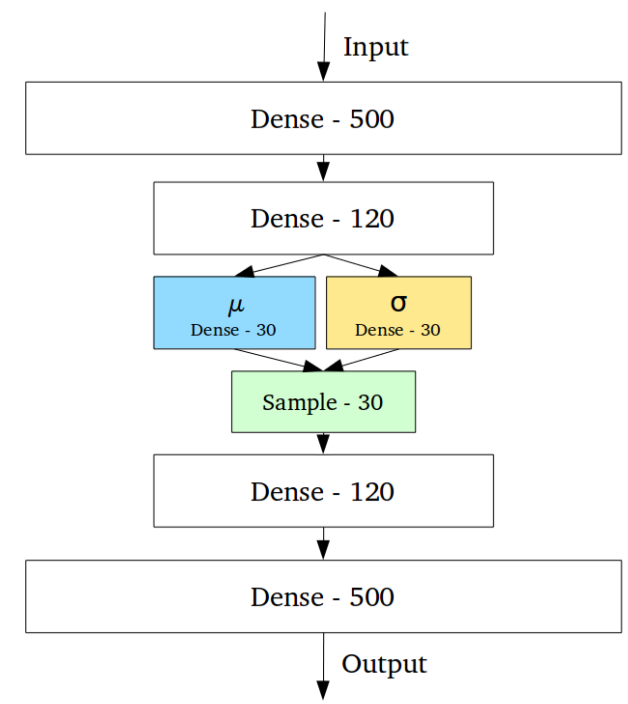
- Combine
  - Prevent overfitting
  - Prevent variation in data



$$\mathcal{L}_{\text{tot}} \equiv \frac{1}{N}\sum_{i}^{N}|x_i - \tilde{x}_i|^2 + \lambda_{\text{s}}\sum_{j}||w_j|| + \lambda_{\text{c}}||J_h(x)||^2$$

Sparse + Variational AE

- Combine
  - Prevent overfitting
  - Remove discontinuity in encoding space

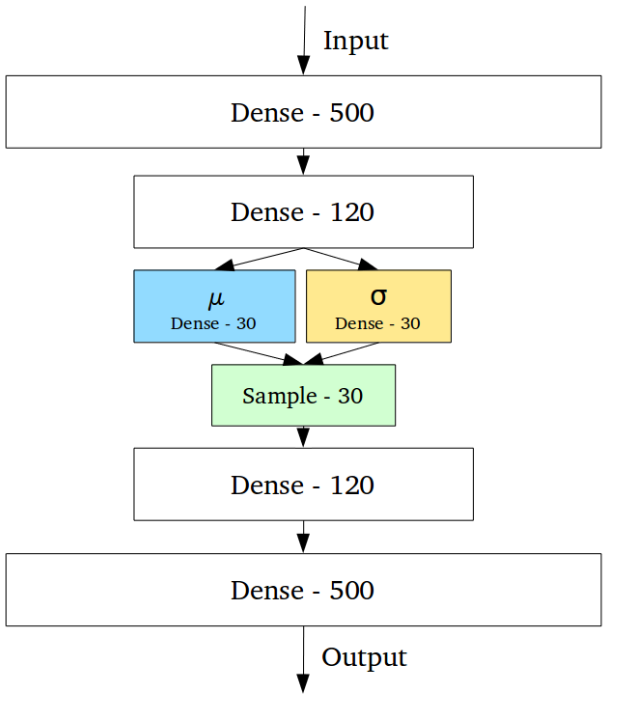


$$\mathcal{L}_{\mathrm{tot}} \equiv \frac{1}{N}\sum_{i}^{N} |x_i - \tilde{x_i}|^2 + \lambda_{\mathrm{s}} \sum_{j} ||w_j|| + \mathcal{D}_{\mathrm{KL}}(p|q)$$

[1] https://towardsdatascience.com/intuitively-understanding-variational-autoencoders-1bfe67eb5daf

Contractive + Variational AE

- Combine
  - Prevent variation in data
  - Remove discontinuity in encoding space

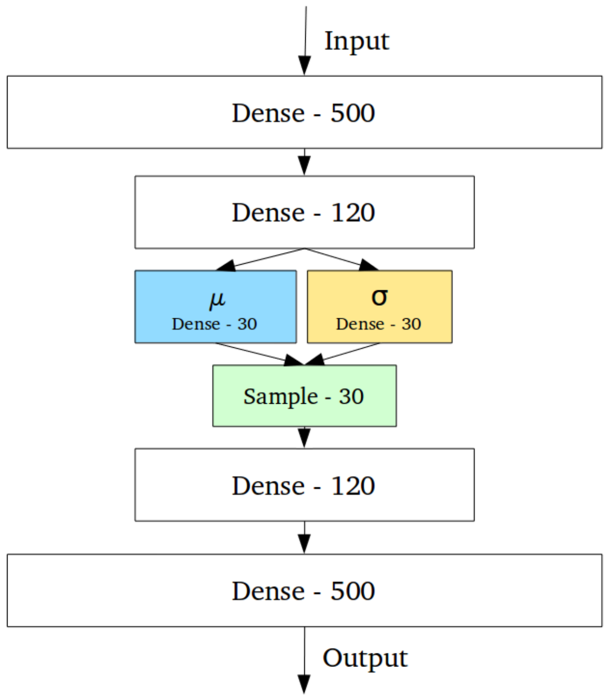


$$\mathcal{L}_{\mathrm{tot}} \equiv \frac{1}{N}\sum_{i}^{N}|x_i - \tilde{x}_i|^2 + \lambda_{\mathrm{c}}||J_h(x)||^2 + \mathcal{D}_{\mathrm{KL}}(p|q)$$

[1] https://towardsdatascience.com/intuitively-understanding-variational-autoencoders-1bfe67eb5daf

The Standard AE

- Combine
  - Prevent overfitting
  - Prevent variation in data
  - Remove discontinuity in encoding space

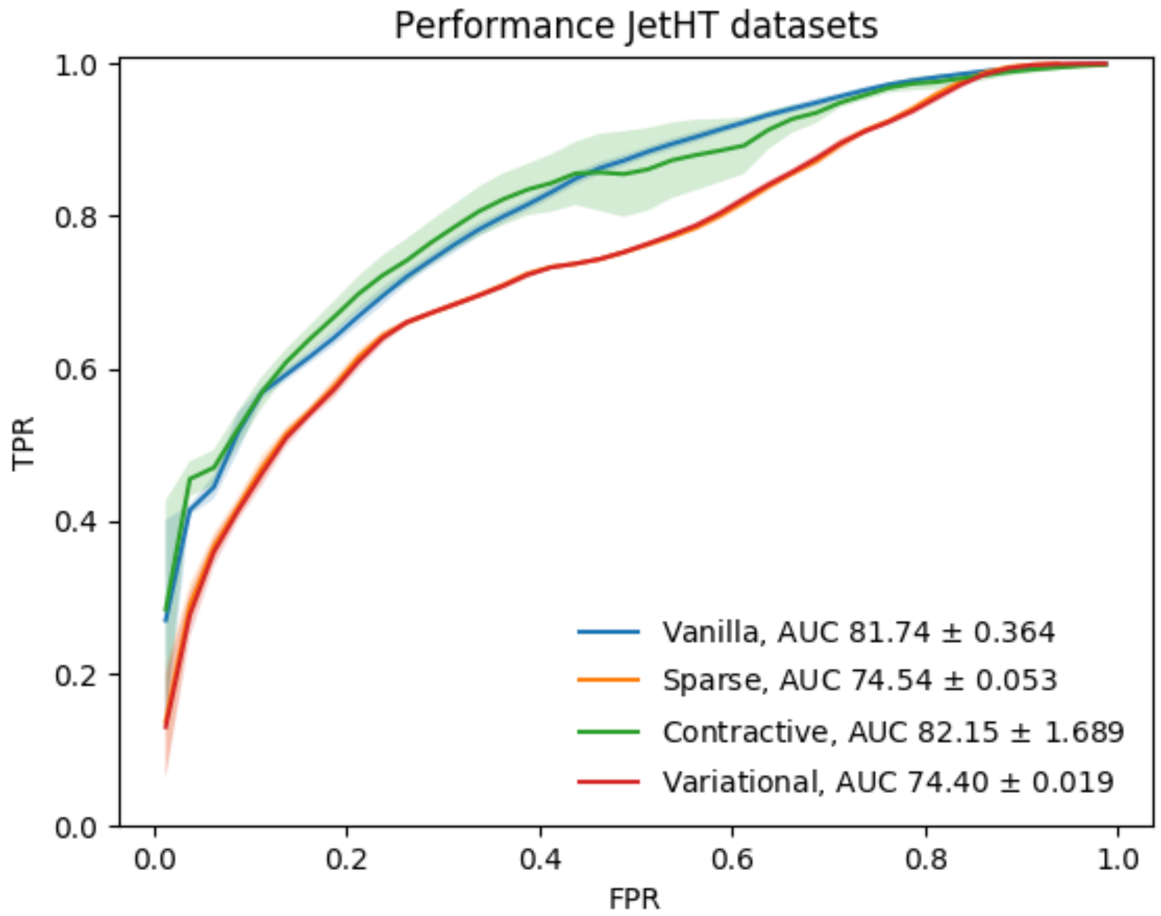


$$\mathcal{L}_{\mathrm{tot}} \equiv \frac{1}{N}\sum_{i}^{N} |x_i - \tilde{x}_i|^2 + \lambda_{\mathrm{s}} \sum_{j} ||w_j|| + \lambda_{\mathrm{c}} ||J_h(x)||^2 + \mathcal{D}_{\mathrm{KL}}(p|q)$$

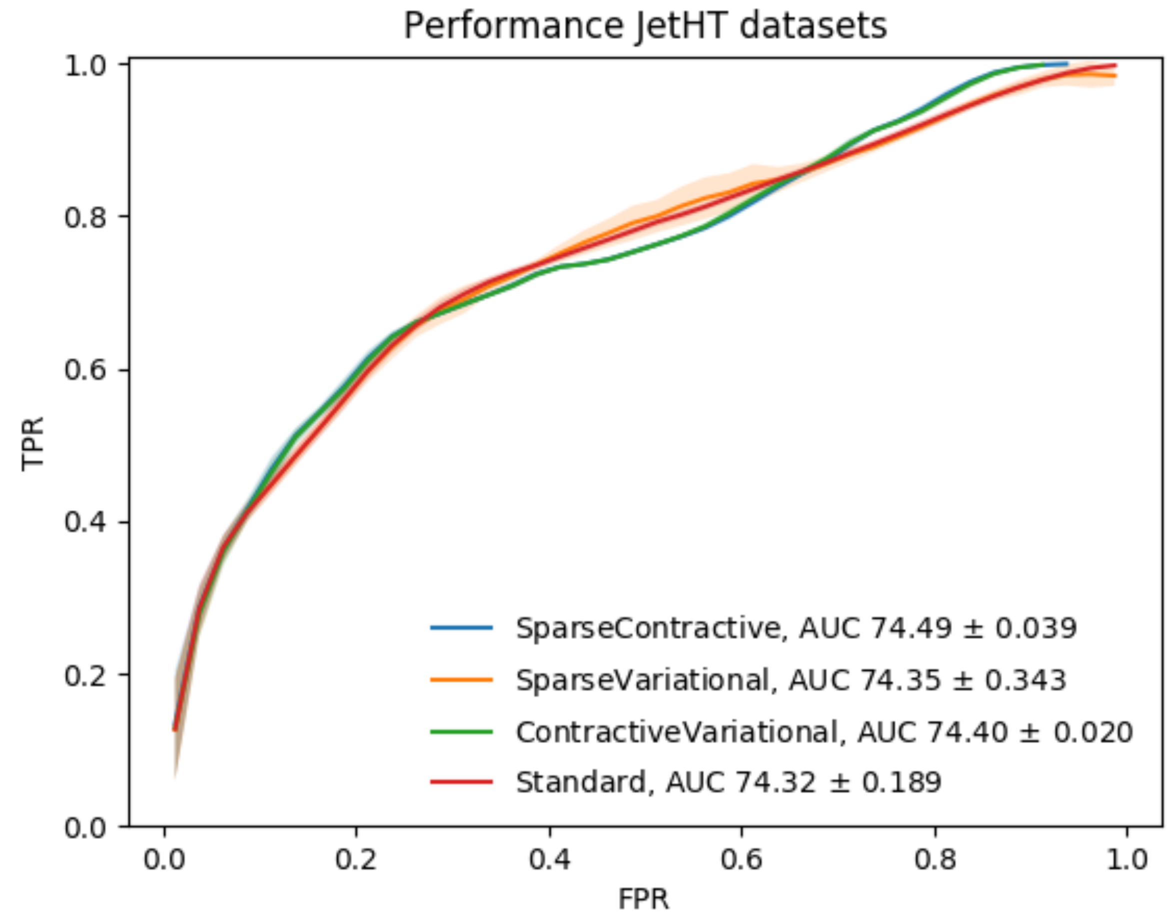
[1] https://towardsdatascience.com/intuitively-understanding-variational-autoencoders-1bfe67eb5daf

Results

JetHT

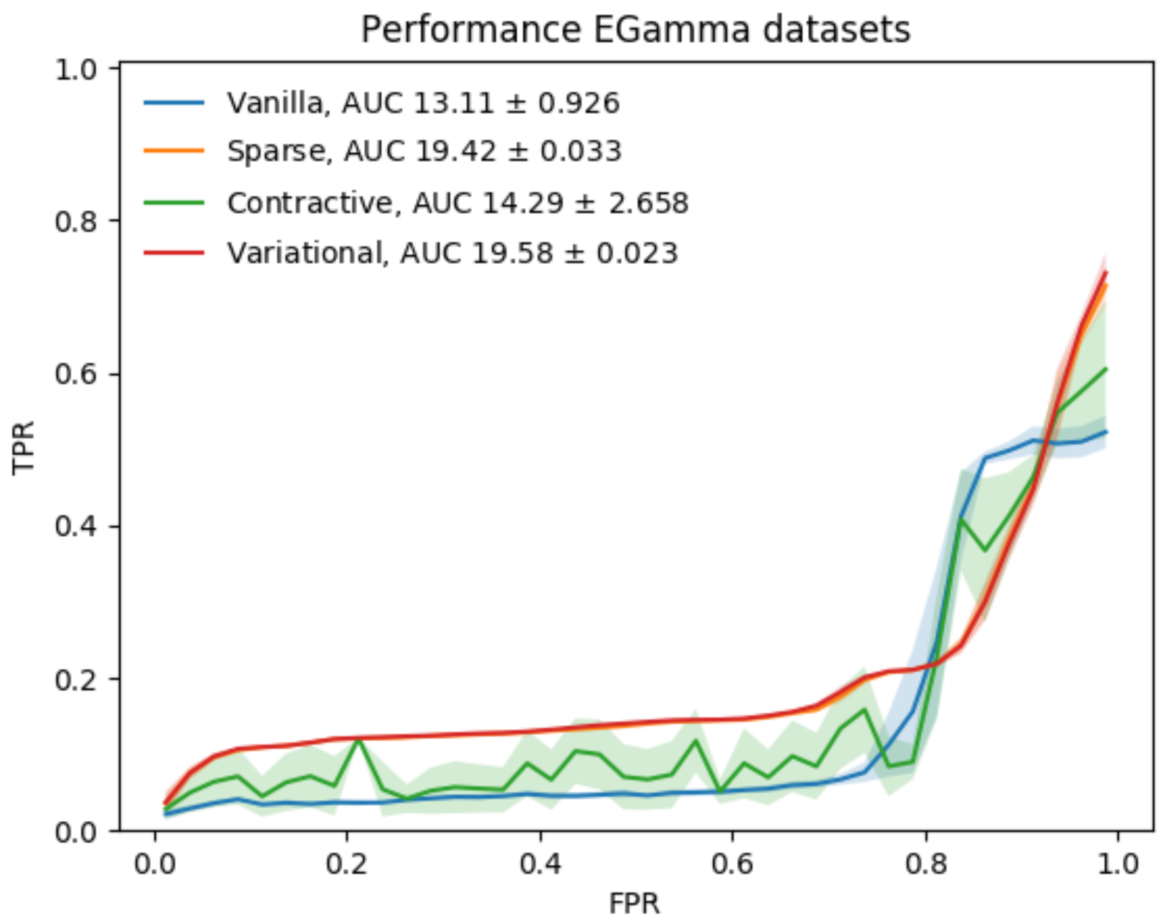


Original AE

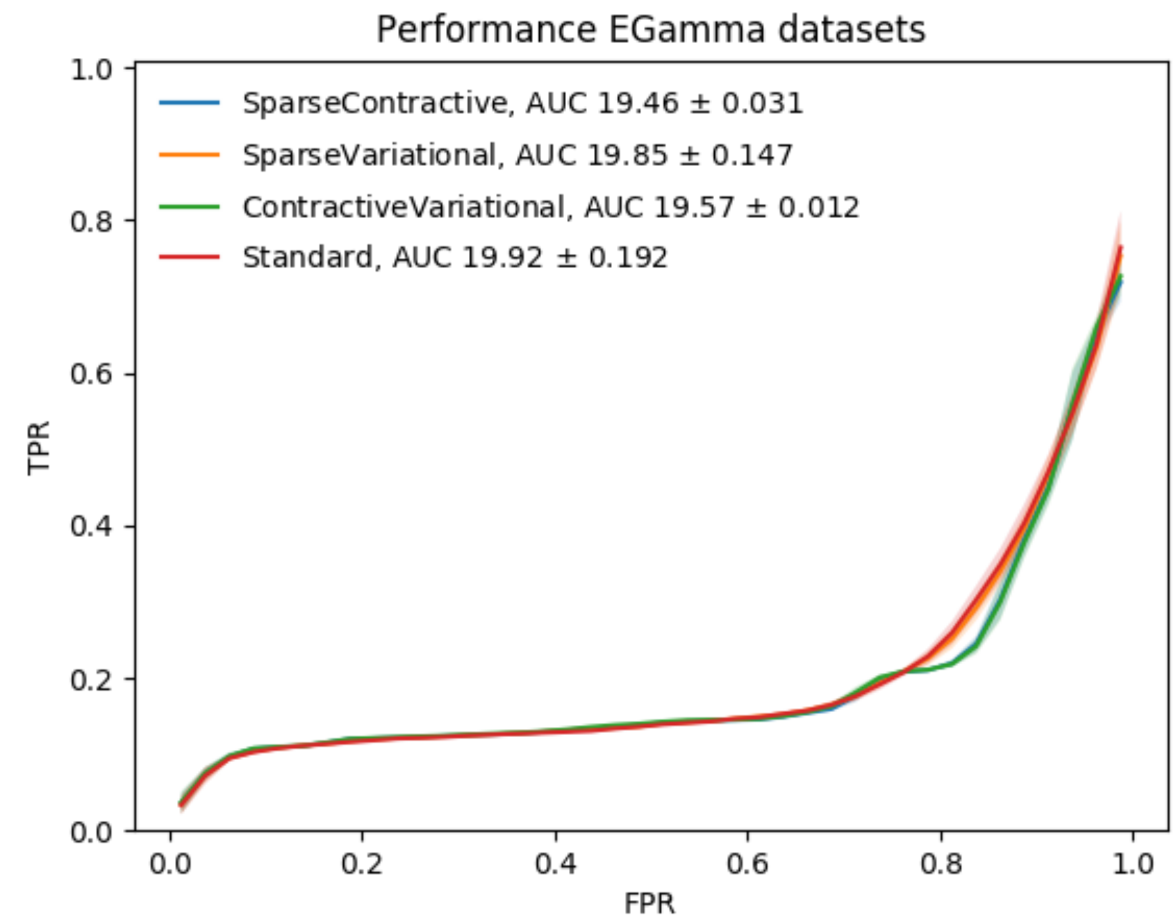


Extended AE

EGamma

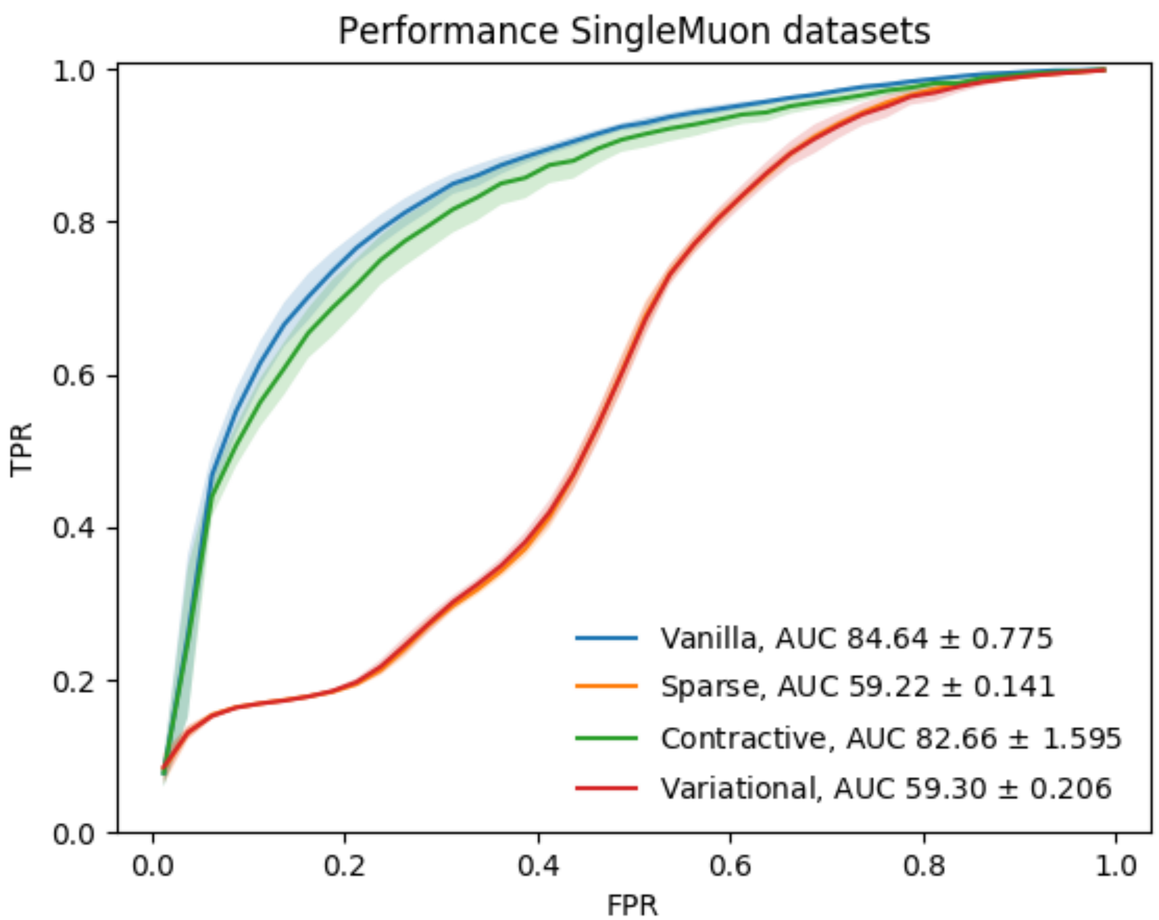


Original AE

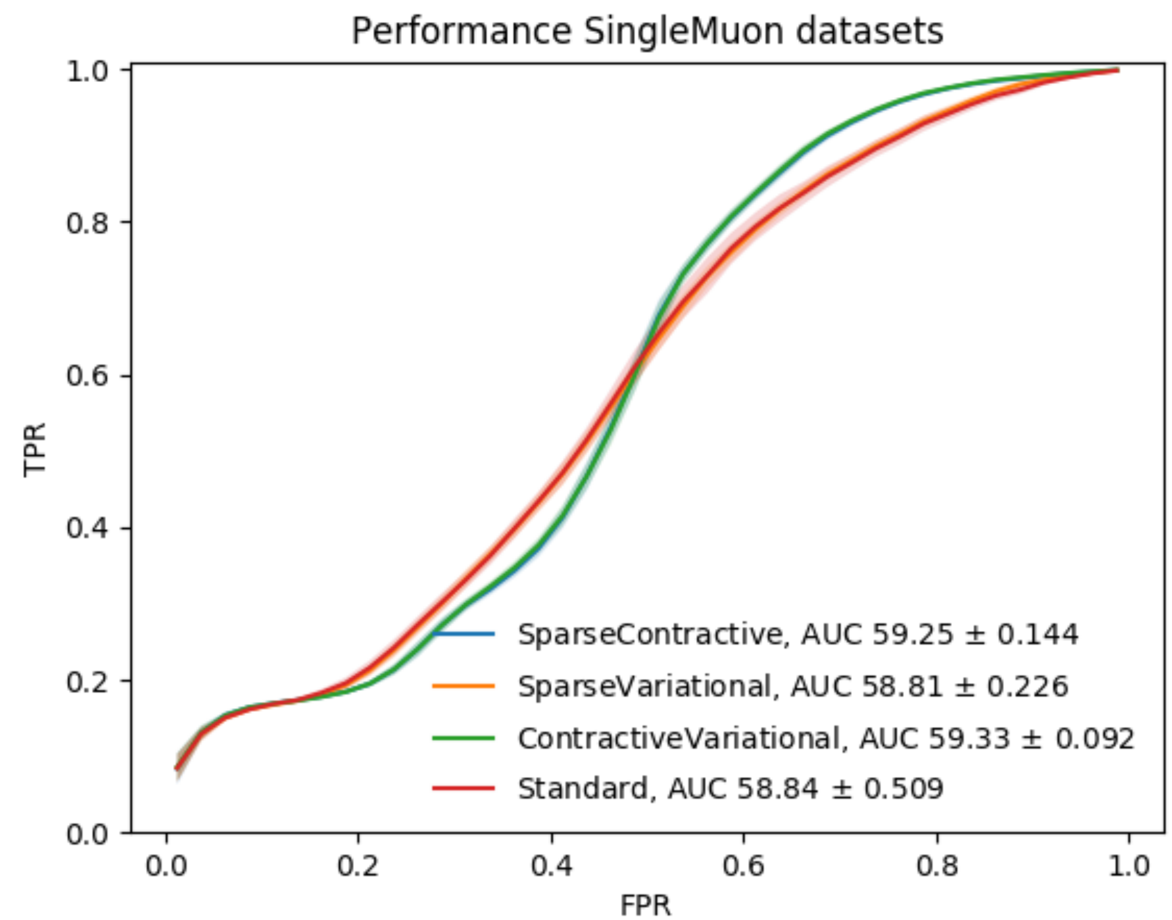


Extended AE

SingleMuon

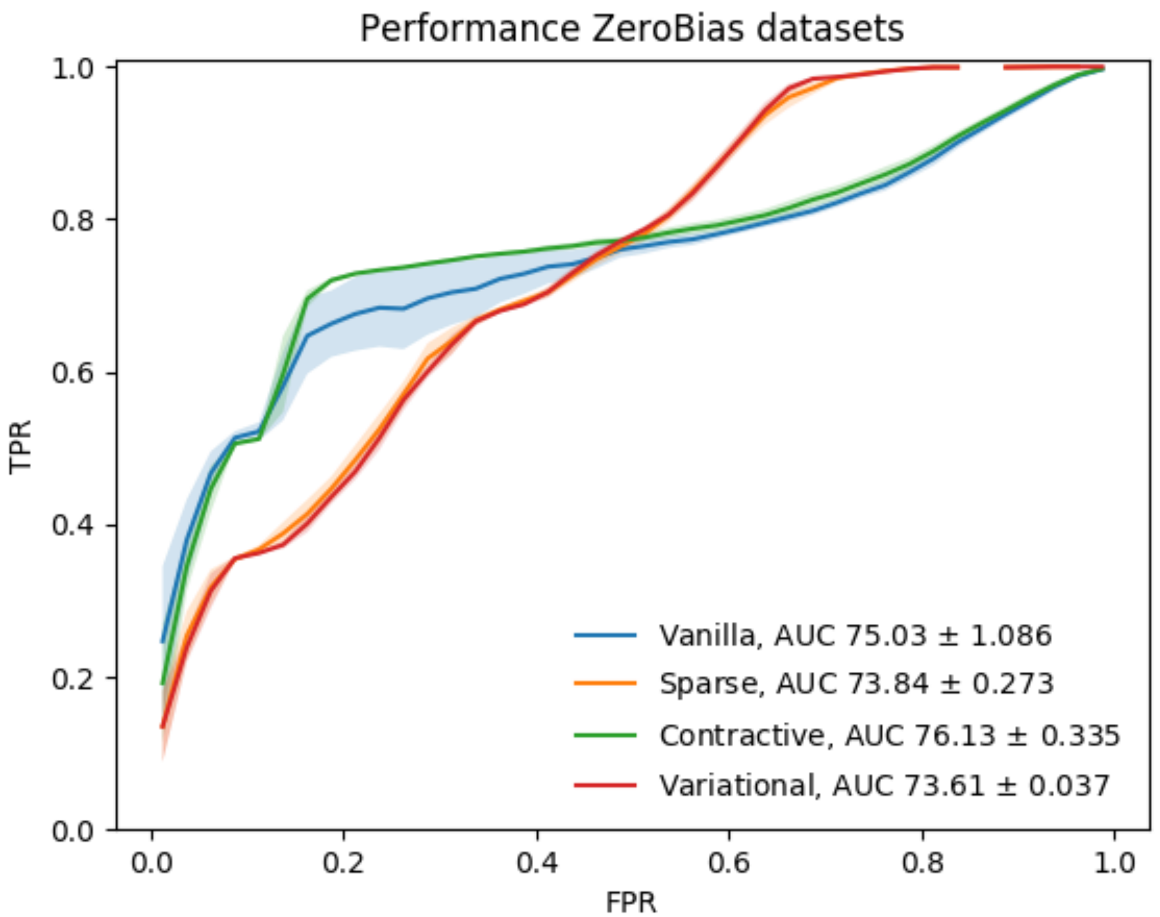


Original AE

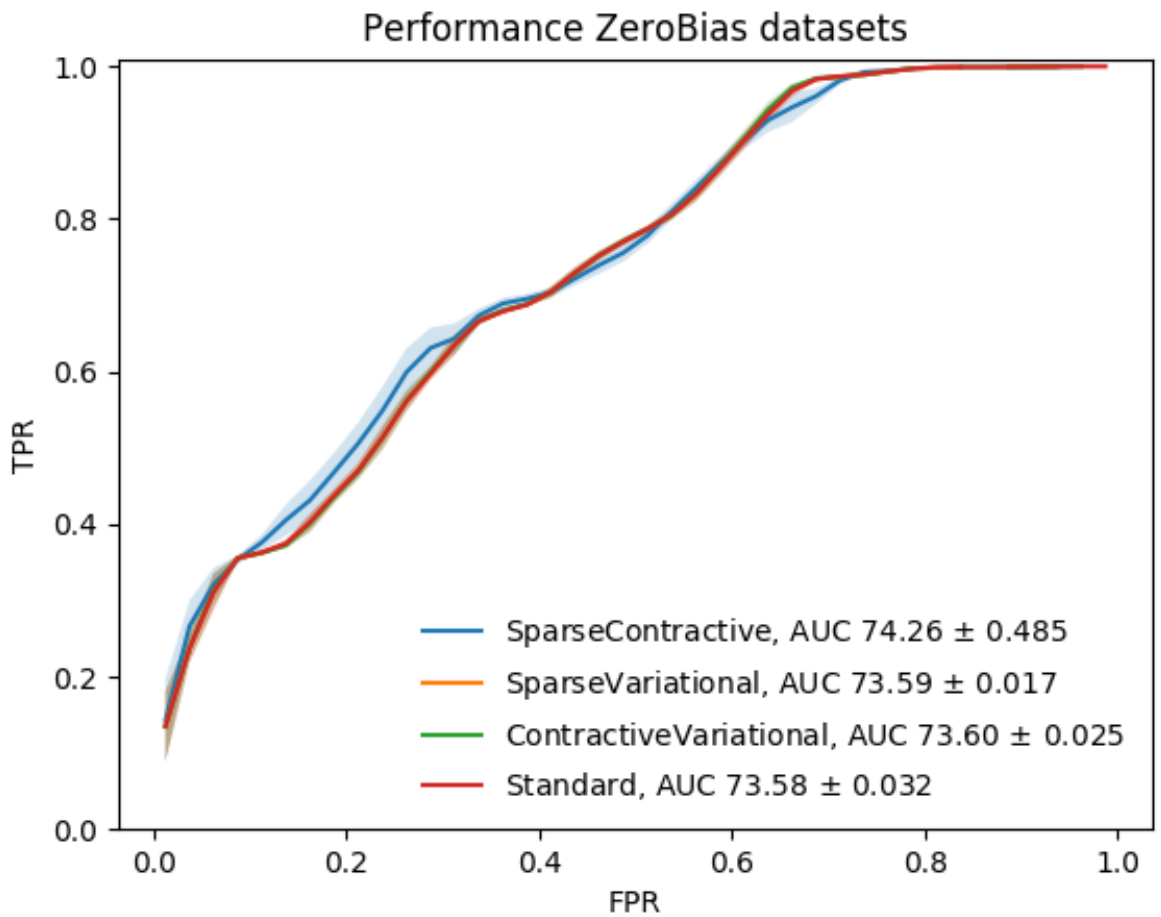


Extended AE

ZeroBias



Original AE



Extended AE

To sum up

- Summary:
  - Too many constrain is fairly more robust but the accuracy still unremarkable
- To do:
  - Re-evaluate ROC after all of Failure Scenario has been process

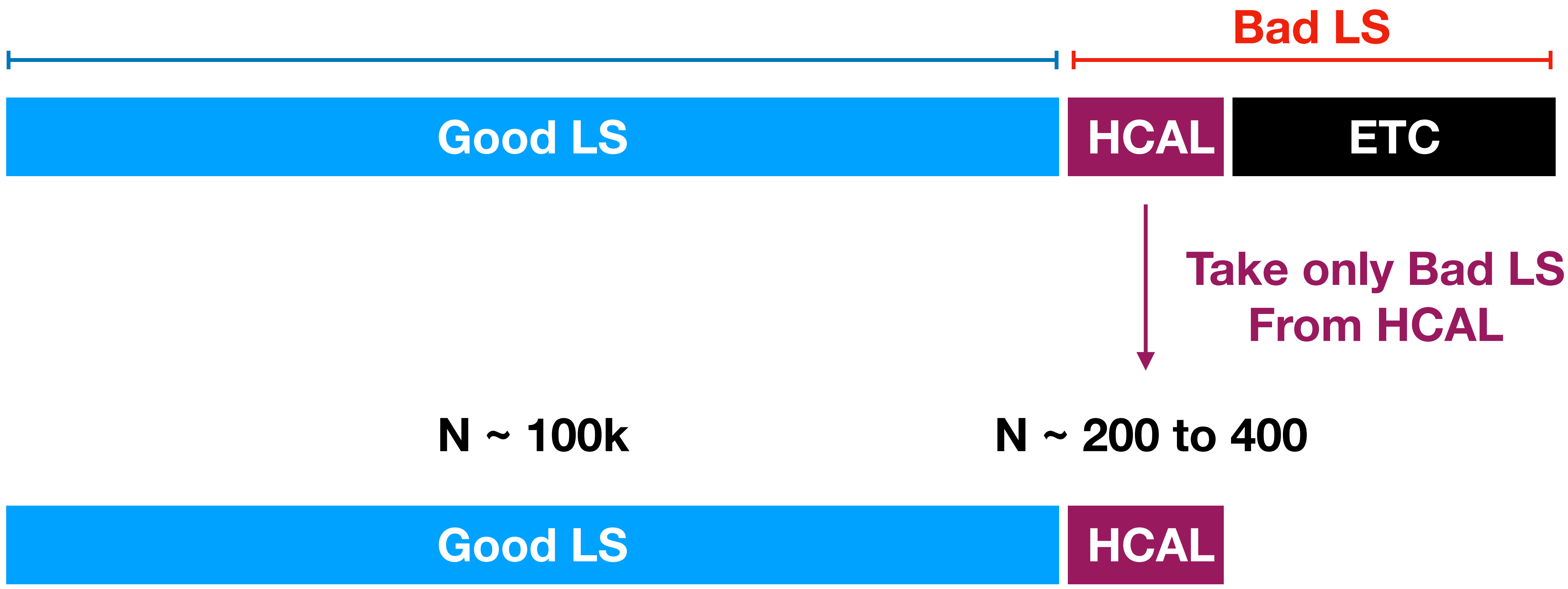
Simplest Prototype of The New Approach

Simplest prototype of Malfunction Spotter

- Simpler Objective: Spot the HCAL malfunction from 4 channels of 2018 datasets
- Supervised
- Input (x): Physics objects
- Label (y): HCAL status => GOOD or ELSE (NOT GOOD)
    - Combine all of those status
      [‘hcal-hb’, ‘hcal-he', ‘hcal-hf’, ‘hcal-hcal’]
    - In the future, we could extend to identify ECAL and so on.

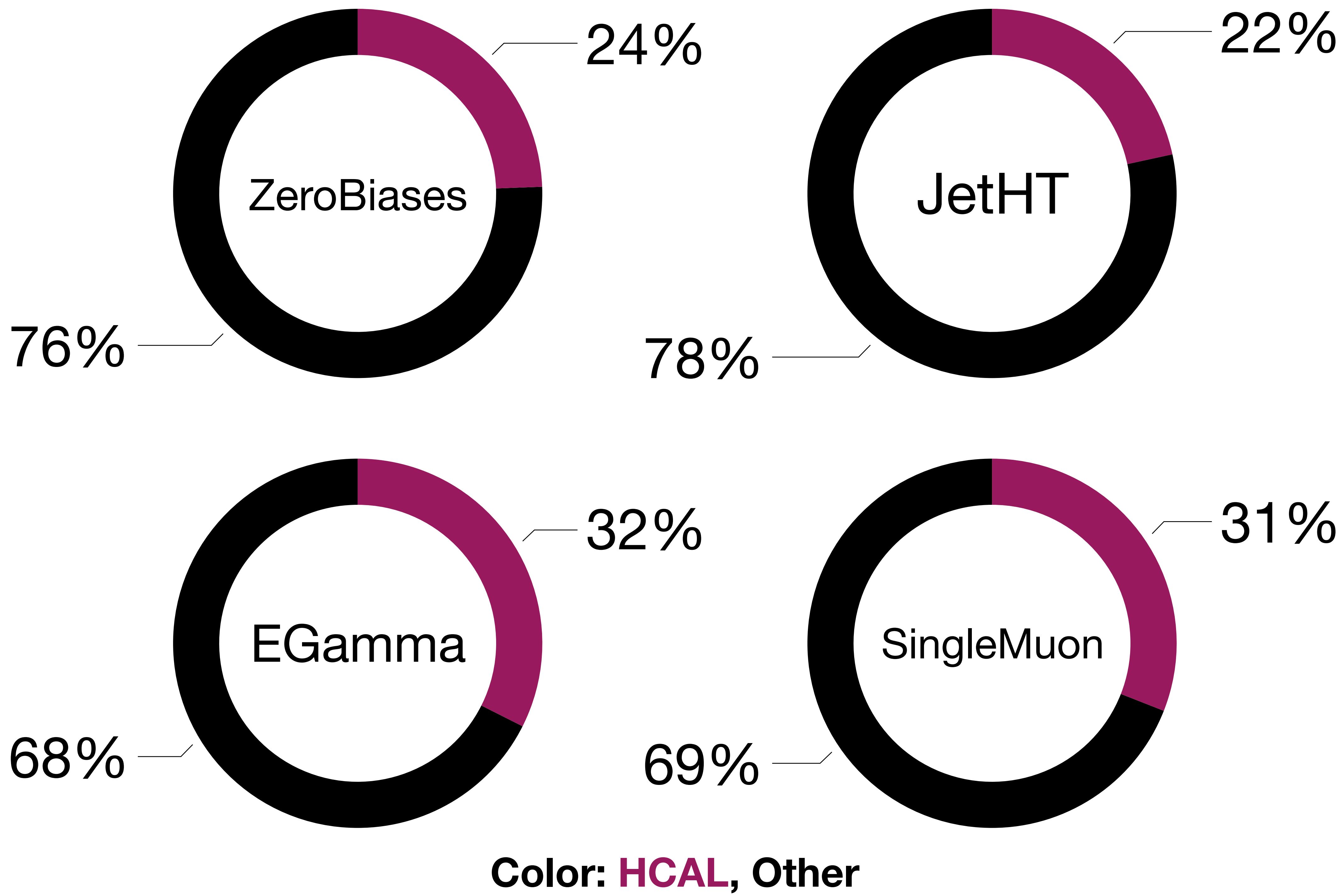
Dataset preparation

- Extract the sub-detector status in LS granularity from Fabio's API

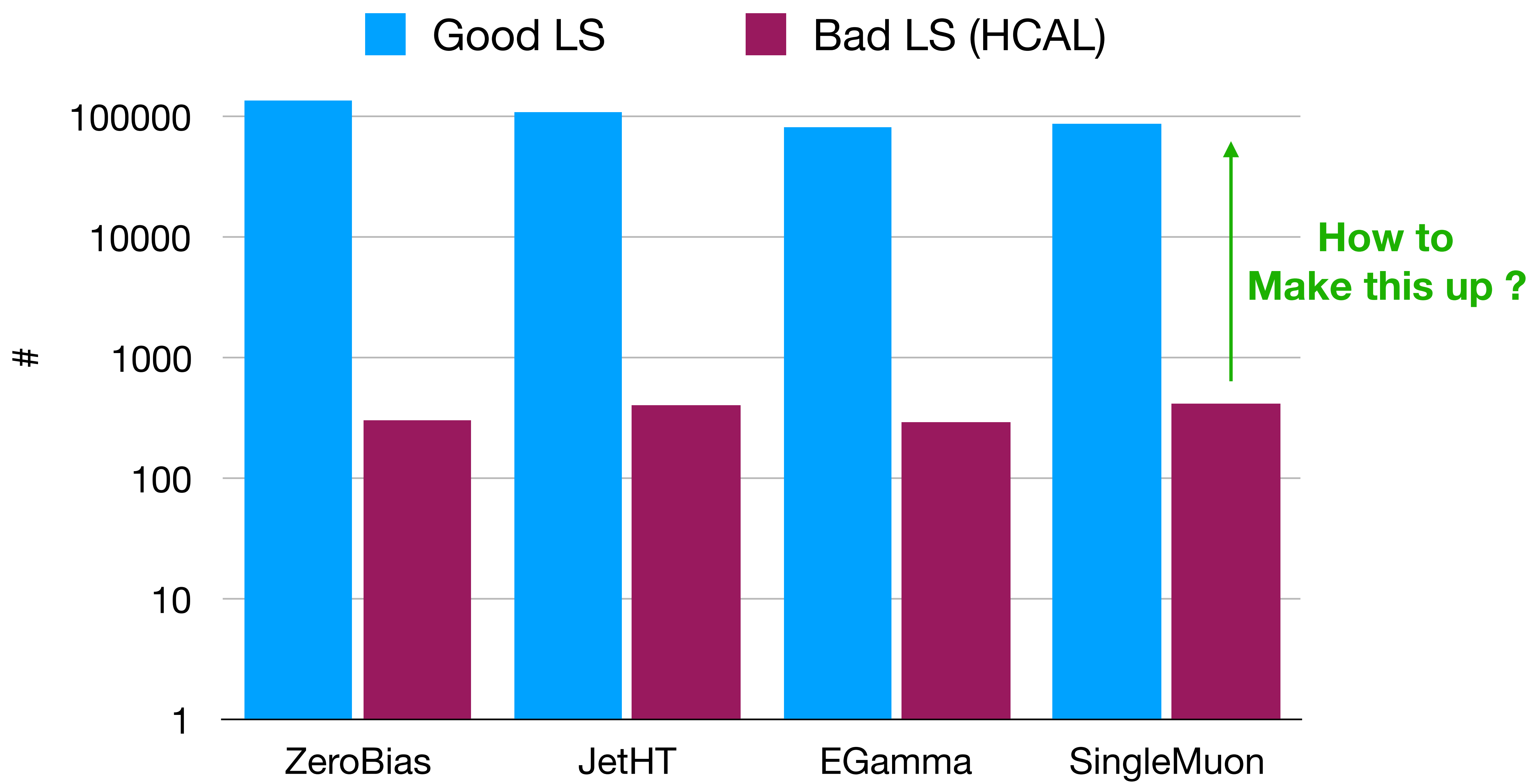


Imbalanced Class! We have a very few bad LS but plenty of good LS

Malfunction of Bad LS in 2018 dataset



Imbalanced Class



Make it Pseudo-balanced class

- Our Choices

  1.Replicate bad LS

  2.Replicate bad LS with gaussian noise

- Any comment for this idea are totally welcome