Sparse Data Generation with VAE

Breno Orzari, Javier Duarte, Jean-Roch Vlimant, Mary Touranakou, Maurizio Pierini, Raghav Kansal and Thiago Tomei

IML Working Group meeting
23/11/20

Approaches

We have been experimenting with two generative model approaches for sparse data generation. Our work regarded to the use of:

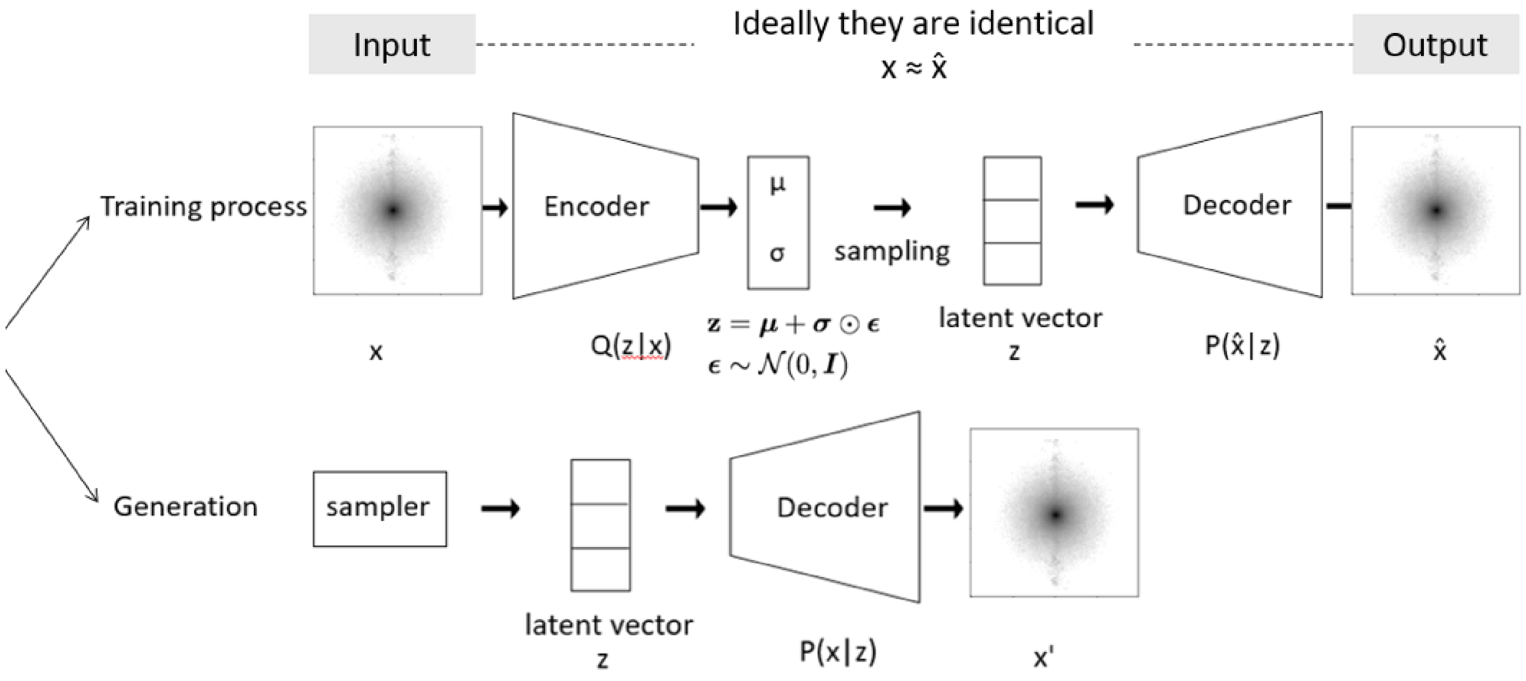
I. a variational autoencoder (VAE),

II. and a graph-based generative adversarial network (GAN) (Raghav's talk).

This presentation will be focused on the VAE approach for sparse data generation.

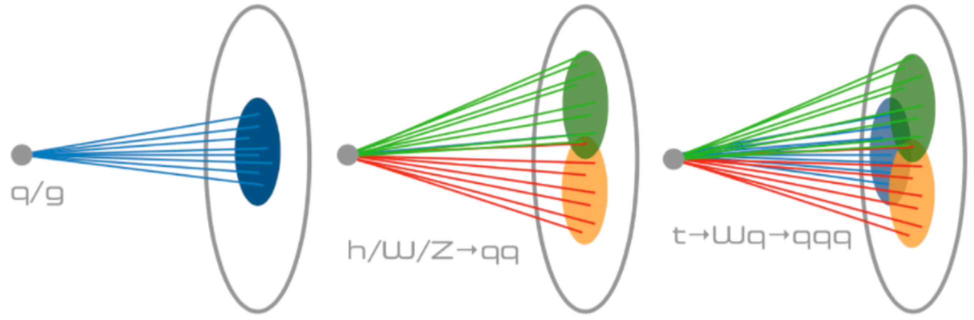
VAE for sparse data generation

We test a variational autoencoder (VAE) architecture for reconstructing and generating jets.



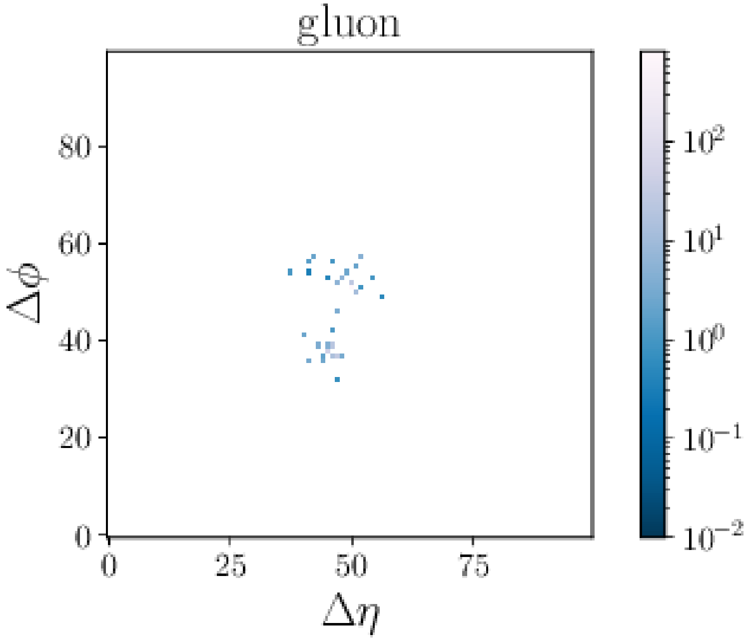
Dataset

- The dataset used (JEDI-net paper Moreno, Eric A. et al, arXiv:1908.05318) consists of high-momentum jets originating from gluons, light quarks, Z bosons, W bosons and top quarks.
- We utilize only the gluon jets dataset (~ 177K jets) for the VAE, splitting the data into training (70%), validation (15%) and testing (15%) subsets.



Pictorial representations of the different jet categories existing in the dataset (arXiv:1908.05318).

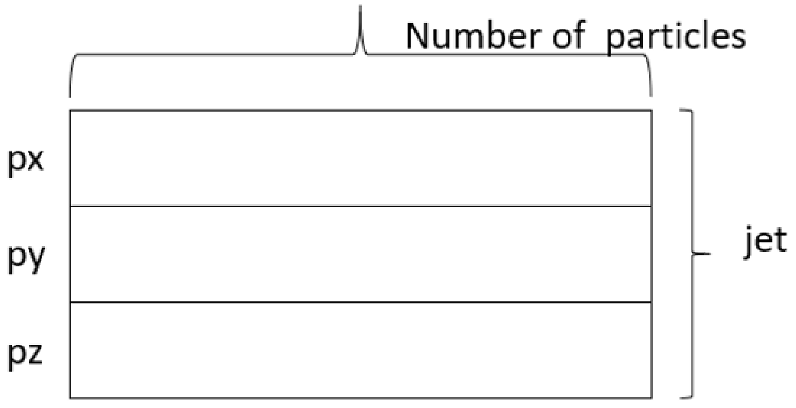
Dataset



Graphical representation of a gluon jet (arXiv:1908.05318)

- Jets can be characterized as sparse sets of items (particles) that are intrinsically unordered.
- Although, sometimes, an ordering might be given to the data (e.g. ordering particles by decreasing p_T), it is also important to preserve its permutation invariance (depending on application-specific requirements).

Dataset



Structural representation of a gluon jet

In our VAE study, each jet is represented as a list of 100 particles with 3 features $p_{x,}$ p_y and p_z (particle momentum in cartesian coordinates; worked better than η, φ and $p_{T,}$). In cases, where less than 100 particles are present in the jet, zero-padding is applied for non-existent particles up to 100.

- We apply feature-dependent standardization such that each feature ($p_{x,}$ p_y and p_z) has zero mean and unit variance.

Loss function

$$L^{VAE} = L_{reco} + \beta D_{KL}$$

where

$$D_{KL}(\underbrace{q_\phi(z|x)}_{\sim N(\mu,\, \sigma)} \,||\, \underbrace{p_\theta(z)}_{\sim N(0,\, 1)})$$

The loss function of a VAE consists of two terms:

1. The reconstruction loss (e.g. traditionally a generic loss function such as the MSE or Cross-entropy between the output and the input) that penalizes the network for producing outputs (reconstructed inputs) different from the inputs.

2. The Kullback-Leibler (KL) divergence used as a loss function between the encoder's distribution $q_\varphi(z|x)$ and the $p_\theta(z)$ that optimizes the probability distribution parameters (μ and σ) to closely resemble those of the target distribution.

Reconstruction loss function

- We consider the use of a permutation-invariant Nearest Neighbour Distance (NND) known as the Chamfer loss (arXiv:1906.02795) for the reconstruction loss.

$$L_{reco} = \sum_i min[d_{eucl}(X_i, \hat{X})]^2 + \sum_i min[d_{eucl}(X, \hat{X}_i)]^2$$

- We train a VAE using the MSE for the reconstruction loss and then, we compare the results with a VAE trained with the Chamfer loss.
- Our goal is to show that training a VAE with the Chamfer reconstruction loss provides similar results to a VAE trained with an MSE reconstruction loss, whereas the Chamfer loss preserves the permutation invariance.

Reconstruction loss function

- To impose physics constraints for our domain-specific application, we further modify the reconstruction loss by adding two extra terms, the jet mass and the jet p_T, to enforce the model to learn the jet kinematics.
- The jet mass and the jet p_T (input and reconstructed) are computed from the sum of the momenta of the particles in the jet.

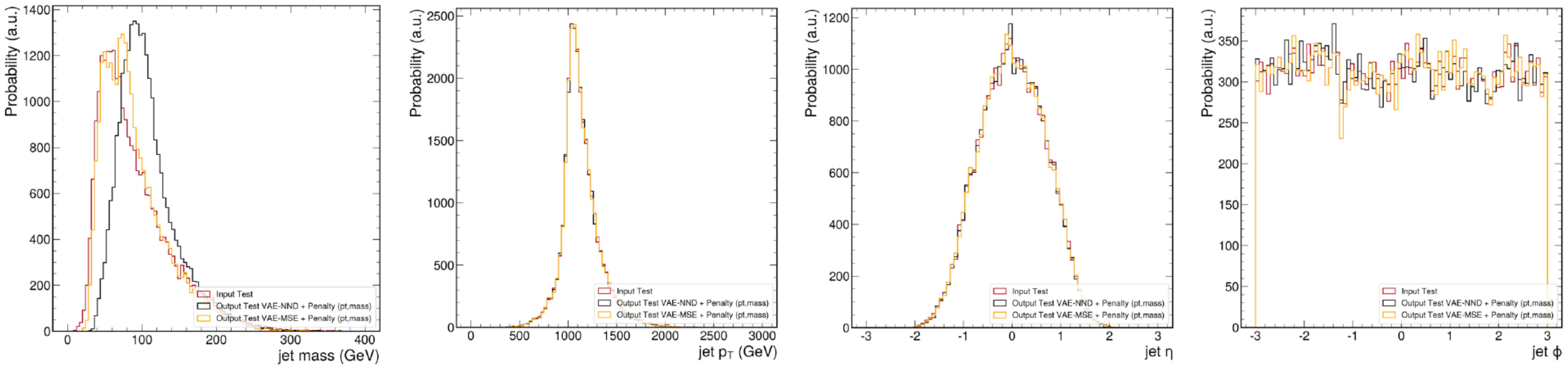
$$L_{reco} = \sum_i min[d_{eucl}(X_i, \hat{X})]^2 + \sum_i min[d_{eucl}(X, \hat{X}_i)]^2$$

$$+ \sum_j [d_{eucl}(p_T^{jet}, \hat{p}_T^{jet})]^2 + \sum_j [d_{eucl}(m^{jet}, \hat{m}^{jet})]^2$$

Analysis

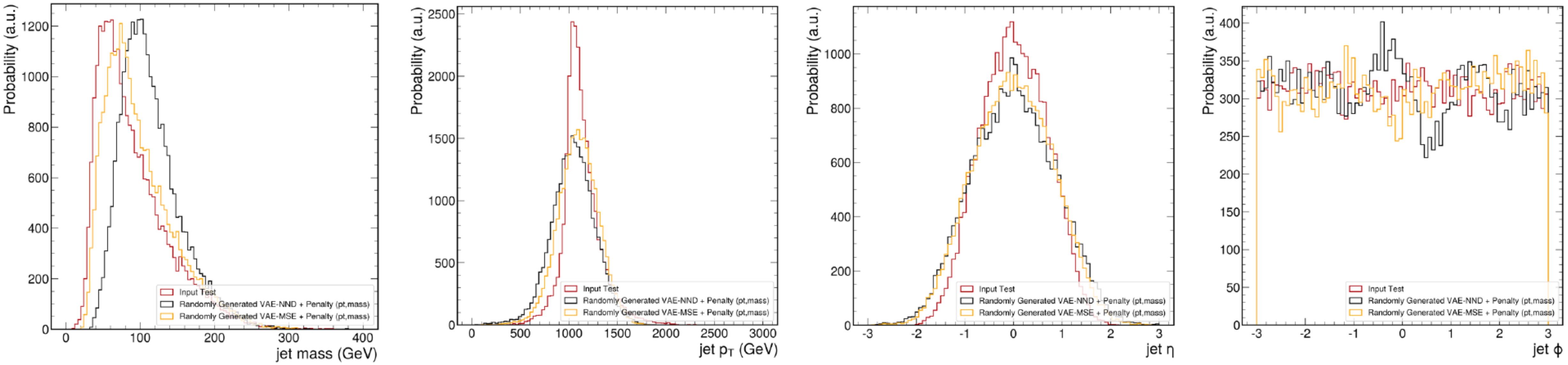
- We first train the VAE and measure its reconstruction performance. In order to do so, we compare:
  - the distributions of jet features such as jet mass, jet momentum, jet eta, etc. between input and reconstructed (output) jets. The jet features are computed from the VAE input and output of the jet constituents' four-momenta.
- We then, use the decoder of the trained VAE as a generator of jet constituents when given an input z of Gaussian sampled latent variables. In the same respect, we compare:
  - the distributions of jet features between input and randomly generated output (gaussian sampled) jets. The jet features are computed from the VAE input and the randomly generated output of the jet constituents' four-momenta.

Results – Reconstruction



- We observe that in terms of reconstruction, MSE performs better than the Chamfer loss for the jet mass.
- Both the MSE and the Chamfer reconstruction loss function provide similar performance for other jet features.
  - In that respect, it is important to highlight the capability of learning jet features that do not directly get into the loss function like the jet pseudorapidity η and jet azimuthal angle φ.

Results - Generation



- In terms of generation, we observe a lower agreement between input and randomly generated output both for the MSE and the Chamfer loss VAE model.
- MSE performs better than the Chamfer loss for the jet mass, although still with lower accuracy than in the reconstruction.
- The results are very similar for the rest of the jet features where MSE and the Chamfer loss seem to have similar performance.

Jets Generation

- We approximate the latent distribution (prior $p_\theta(z)$) as a Gaussian distribution, but from the VAE generation results, it seems that this might not be the best distribution to use (probably too simplistic).
- beta-KLD values did not prove adequate to acquire good results with generation.
- Idea of **Normalizing Flows** (arXiv:1908.09257) : apply a transformation on the latent space to acquire a more appropriate, complex distribution to sample from.
- Next step: improving the VAE decoder’s performance on jet generation by applying Normalizing Flows for learning the prior.

Summary

- We trained a variational autoencoder (VAE) on a sparse dataset with a permutation-invariant loss function.
- We compared its performance with the performance achieved when training the same VAE model with a standard MSE loss.
- For lower β-KLD values, we can acquire high accuracy in reconstructing sparse data from a compressed low dimensional representation of it.

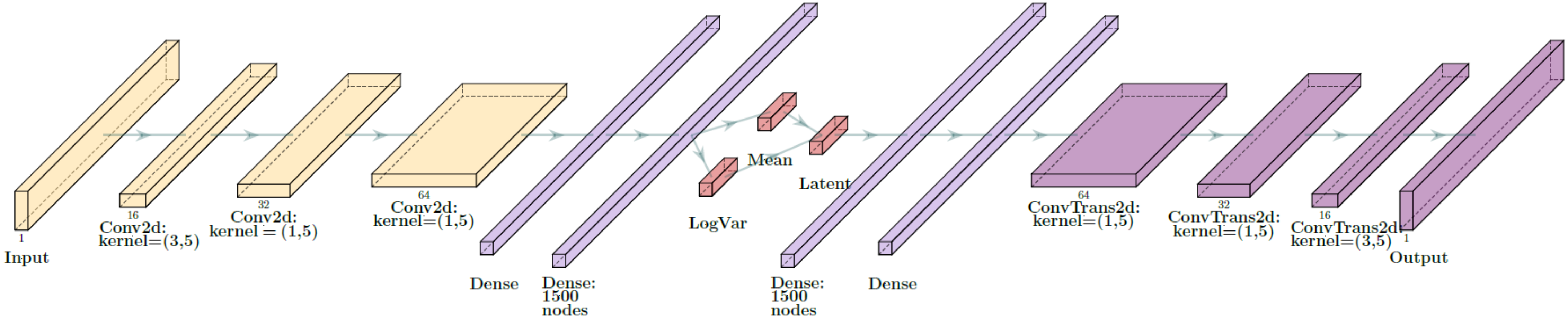
  → β can be further optimized to acquire better reconstruction than shown in the figures above – we observed this behavior at the very beginning of our studies before finding a beta trade-off between reconstruction and generation.
- When the VAE is used as a generator, the two loss functions provide similar performance.
- We wish to highlight the potential of the Chamfer loss for physics-specific applications on sparse datasets as, in contrary to generic loss functions that are order-dependent (like MSE), that one preserves the permutation-invariance.

Next steps

- Splitting work into two packages:
  1. Improving the VAE decoder's performance on generation by applying Normalizing Flows (arXiv:1908.09257) for distribution learning.
  2. Improving VAE reconstruction by looking into the possibility of modeling jet substructure (e.g. EFPs) and investigating possible applications.
- Moving into a permutation-invariant architecture along with the permutation-invariant loss (e.g. graph neural networks, energy flow networks, interaction networks, etc.) for the construction of a fully permutation-invariant model.

Backup

VAE architecture

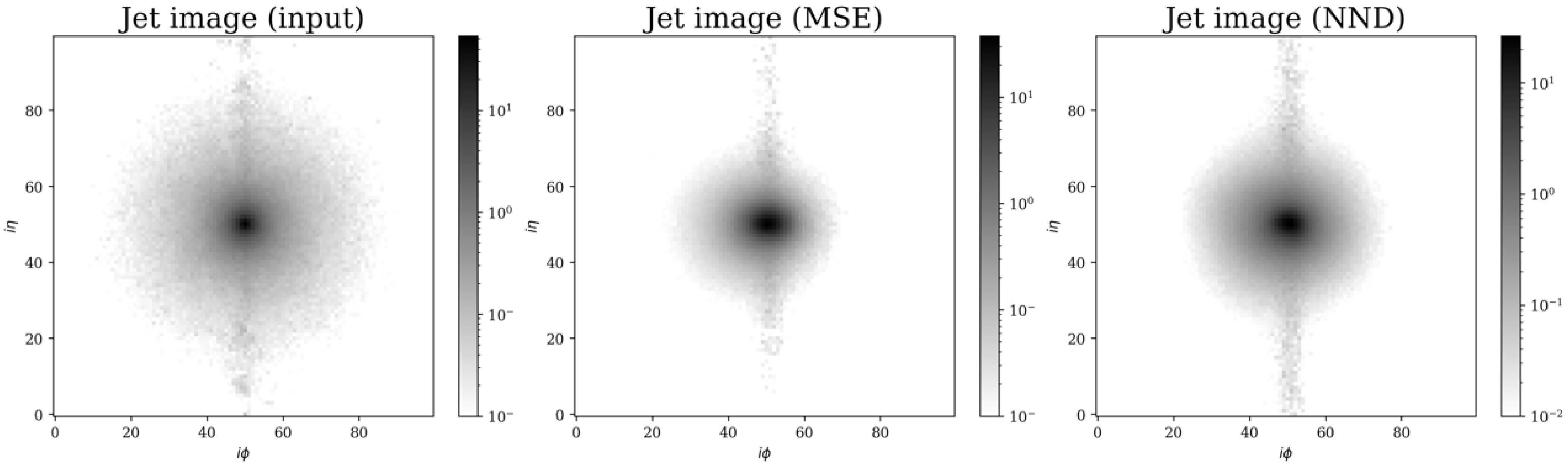


- ReLU is being used as the activation function on all layers except for the last layer where linear activation is used.

Implementation details

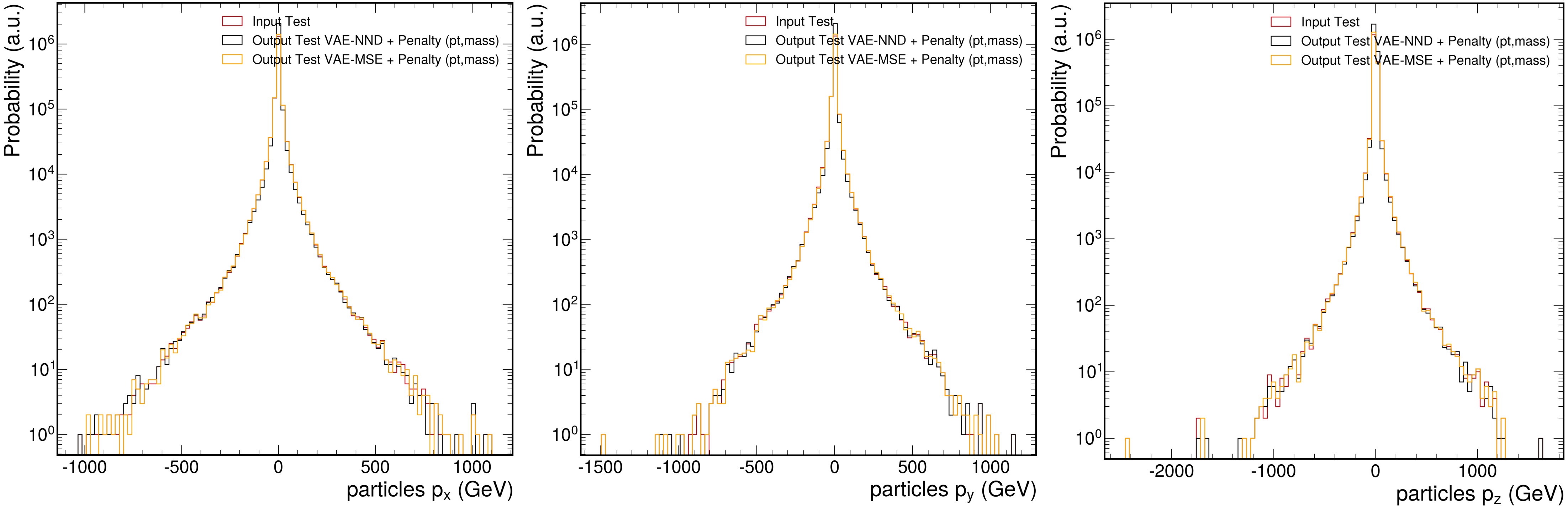
- Models implemented in Pytorch
- Adam optimizer
- Learning rate: 0.001
- Latent dimension: 20
- Early stopping used; patience: 80 epochs
- Physics constraints introduced in the reconstruction loss to force the VAE to learn high-level jet features are useful for our domain-specific application.
- The beta regularization term is used on the KL Divergence to weight the trade-off between reconstruction and generation of the VAE. Beta is being optimized per model.

Qualitative Analysis



- We plot the average gluon jet image defined as an image of 100x100 pixels in the rapidity-azimuthal plane where each pixel reflects the sum of particle p_T in that pixel.
- In this way, we can qualitatively compare the reconstruction performance of the VAE models trained with MSE and Chamfer loss.

Reconstruction results



Comparison of the distributions of the particles' features p_x, p_y, p_z between input and reconstructed (output) particles for the two VAE models.