Multiplication of simulated events using Machine Learning techniques

Karol Sowa, Wojciech Krupa, AGH UST Kraków

BEACH 2022, XIV International Conference on Beauty, Charm and Hyperon Hadrons, 5-11 June 2022

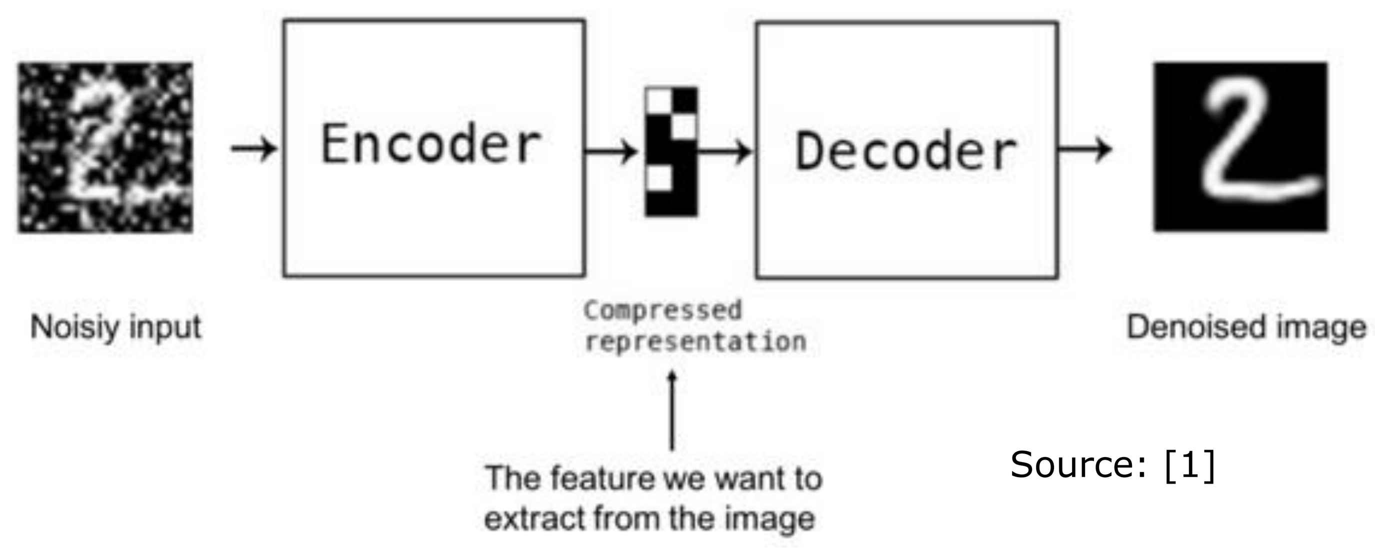
AGH

Faculty of Physics and Applied Computer Science

Variational Autoencoders (VAE)

***Autoencoder (AE)* is a specific type of neural network which learns how to** efficiently **compress** the input data, **encode** it in a lower-dimensional hidden layer (or layers) and then **reconstruct** in the best possible way. Among numerous autoencoders' applications, the most important are dimensionality reduction, feature extraction and *denoising.*

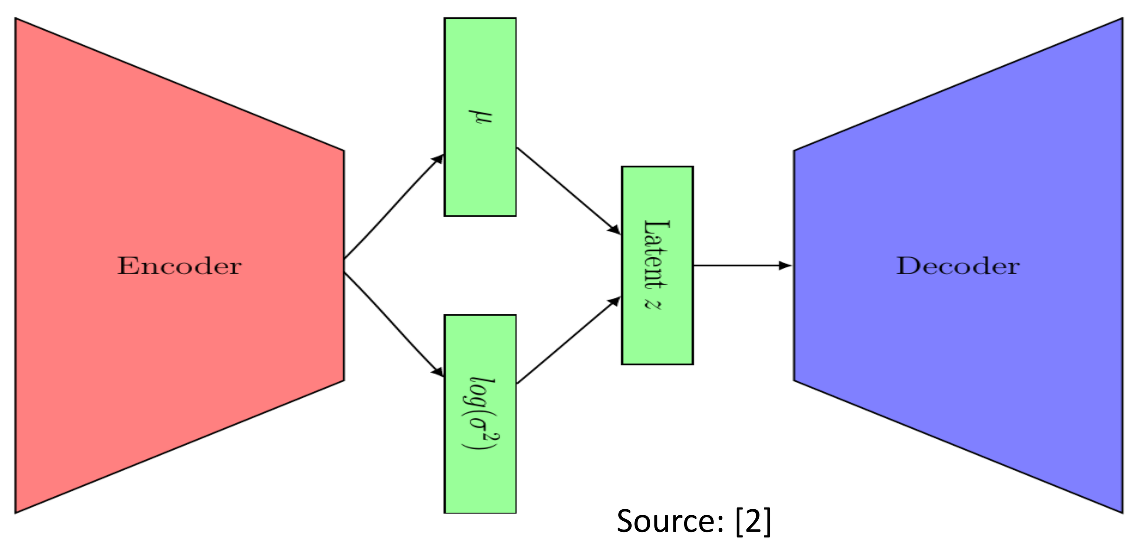
Diagram of a typical autoencoder



Source: [1]

Variational Autoencoders (VAE) are used in generating new data. They often have a **different structure** – some of their hidden layers can be higher-dimensional, as we may want to **add some noise** to the data. However, the key difference between the ordinary AE and the VAE lies in the latent space. In the case of VAE, each input data point is represented in the latent space as a **probability distribution**.

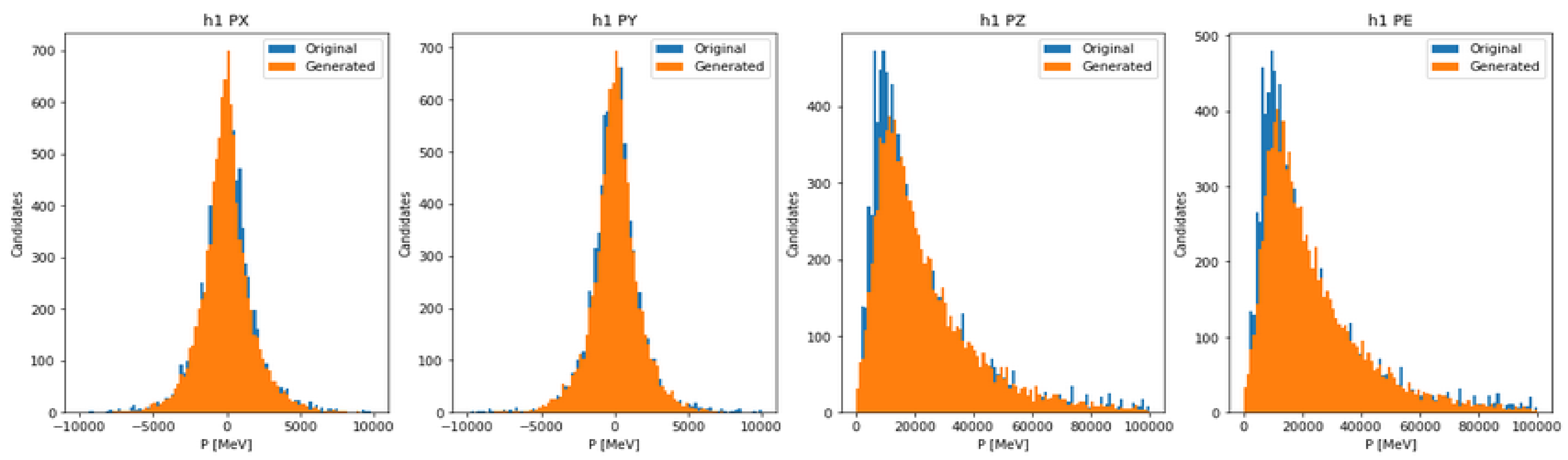
Variational autoencoder structure



Source: [2]

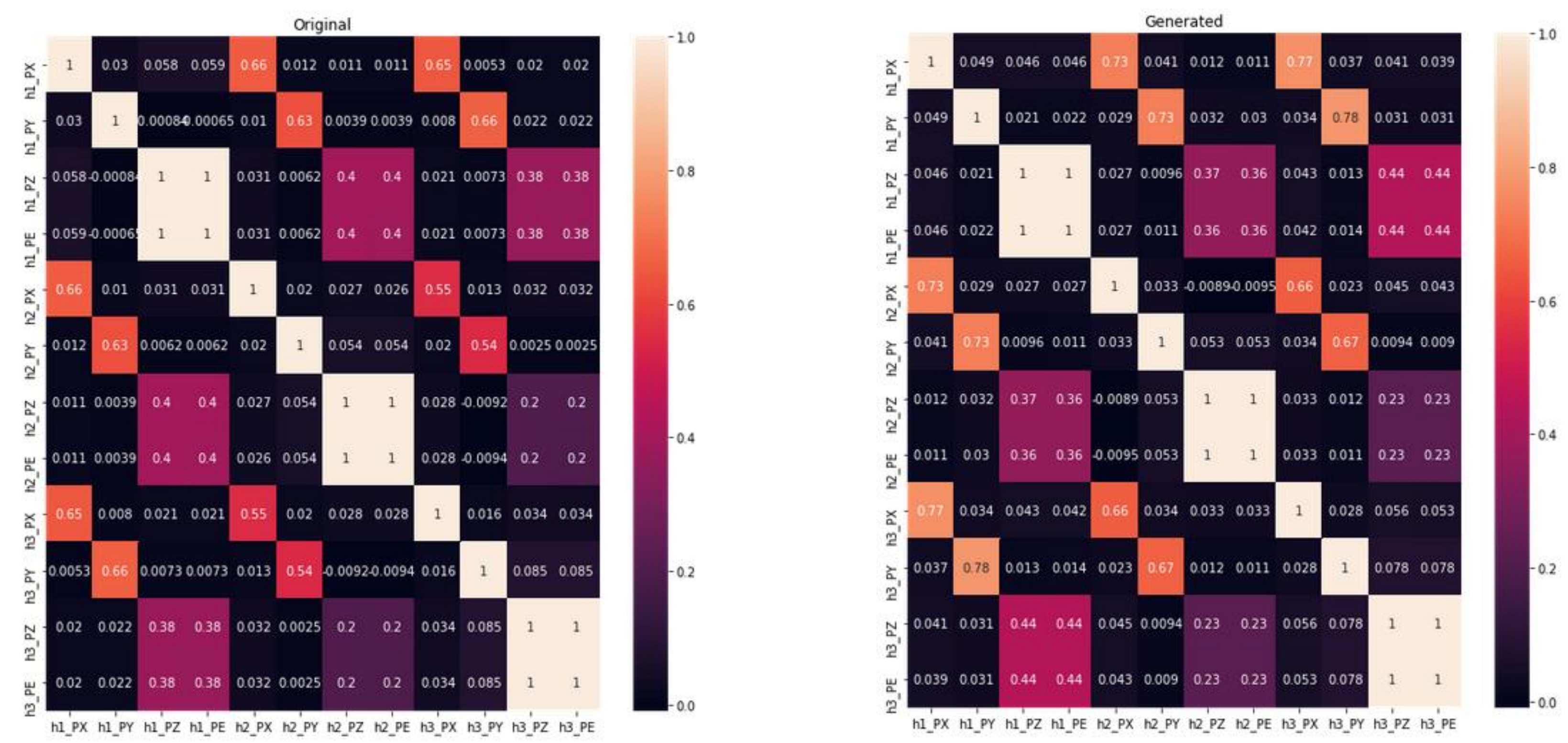
Data multiplication in HEP

The generative properties of VAEs become useful when dealing with datasets that are not big enough. In HEP, such a problem can be encountered during rare heavy meson decays analysis. VAE allows us to **use the existing samples to create new ones**. Selected results of such an approach are presented below:



D-meson 4-momentum – comparison of original samples and those generated using VAE

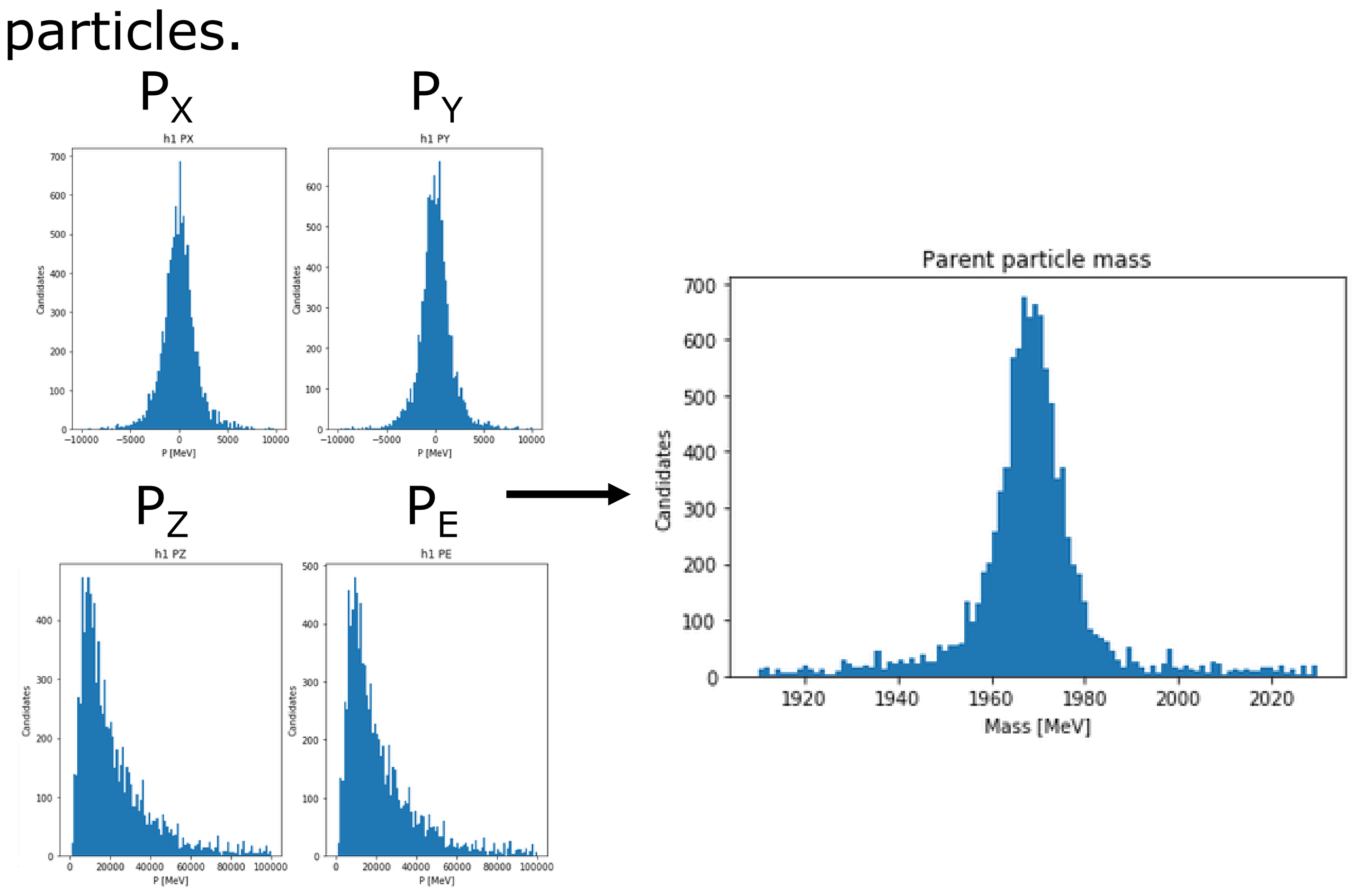
VAEs are also capable of reconstructing correlations between the variables:



Correlation matrices for original samples and those generated using VAE. 4-momenta of D meson decay final states were used as input features.

Future plans

We aim to produce large samples of multiplicated events, which would successfully reconstruct the invariant mass of the parent particles.



However, the most important aim yet to be achieved is to use multiplicated samples to significantly improve the performance of classifiers in the selection process.

References:

1) https://www.edureka.co/blog/autoencoders-tutorial/
2) https://alecokas.github.io/julia/flux/vae/2020/07/22/convolutional-vae-in-flux.html/
3) https://www.tensorflow.org/tutorials/generative/autoencoder?hl=en
4) https://towardsdatascience.com/variational-autoencoder-demystified-with-pytorch-implementation-3a06bee395ed

This research was supported in part by National Research Centre, Poland (NCN), grants No. UMO-2018/31/N/ST2/01471, UMO-2020/36/T/ST2/00168 and in part by PLGrid Infrastructure.