

An FPGA implementation of Variational Autoencoders for anomaly detection in HEP (Remote presentation)

Thursday, 23 March 2023 11:30 (30 minutes)

If new physics does exist at the scales investigated by the Large Hadron Collider (LHC) at CERN, it is more elusive than expected.

Finding interesting results may be challenging using conventional methods, usually based on model-dependent hypothesis testing, without substantially increasing the number of analyses.

Thus, standard signal-driven search strategies could fail in reaching new results, and unsupervised machine learning techniques could fill this critical gap.

Such applications, running in the trigger system of the LHC experiments, could spot anomalous events that would otherwise go unnoticed, enhancing the LHC's scientific capabilities.

The most basic unsupervised machine learning technique is an autoencoder (AE) with a bottleneck. It is constructed using a network that translates a high-dimensional data representation onto itself to create an average or typical entity.

Standard autoencoders are recommended for unsupervised jet classification, but they are known to have problems in more general applications.

The AE learns to compress and rebuild the training data very well, but when new, untrained data is run through the trained AE, it will produce a considerable loss or reconstruction error.

By using the AE, it is possible to search for data that differs significantly from training data or even training data that is a small subclass of anomalous instances.

In addition, the AE fails if the anomalous data is structurally simpler than the dominant class because the AE can encode simpler data with fewer features more efficiently.

It is possible to overcome the disadvantages of AE by substituting a different classification measure for the reconstruction error.

A possible alternative approach to the reconstruction error in the case of Variational Autoencoders (VAEs) is to derive a metric from the latent space embedding.

The work will consist of implementing a VAE model targeted at FPGA (Field Programmable Gate Array) hardware architecture in order to determine the best latency and resource consumption without sacrificing model accuracy.

Models will be optimized for classification between anomalous jets and QCD jets images, in an unsupervised setting by training solely on the QCD background.

A comparison will be made between the reconstruction error and a latent space metric to determine the best anomaly detection score that enhances the separation of the two classes.

The goal of the model is to reconstruct the input data information as accurately as possible.

Additionally, because of the design of the VAEs architecture, the high-dimensional data representation is transformed into a compressed lower-dimensional latent distribution during the encoding stage.

Subsequently, the decoder learns stochastic modelling and aims to generate input-like data by sampling from the latent distribution.

The information about each dataset instance hidden in the high-dimensional input representation should be present in the latent space after training and the model can return the shape parameters describing the probability density function of each input quantity given a point in the compressed space.

With this application at the LHC, it could ideally be possible to classify the jets and even find anomalies using this latent space representation.

A companion tool based on High-Level Synthesis (HLS) named HLS4ML will implement deep learning models in FPGAs.

Furthermore, a compression and quantization optimization of neural networks will reduce the model size, latency, and energy consumption.

We expect VAE will find many uses in science, outperforming classical or standard deep learning baselines and even being able to solve challenges for physics beyond the Standard Model that were previously unsolvable. Ideally applicable in a wide spectrum of signal background discrimination through anomaly detection, this application is expected to produce excellent results in a variety of fields.

Primary author: Mr VALENTE, Lorenzo (University of Bologna)

Co-authors: LORUSSO, Marco (Alma Mater Studiorum - University of Bologna); ANZALONE, Luca (University of Bologna); BONACORSI, Daniele (University of Bologna)

Presenter: Mr VALENTE, Lorenzo (University of Bologna)

Session Classification: Physics and Engineering Applications

Track Classification: Track 1: Physics and Engineering Applications