
SCUOLA NORMALE SUPERIORE

Anomaly detection Forum Meeting: mpp group, CERN

Syed Anwar Ul Hasan

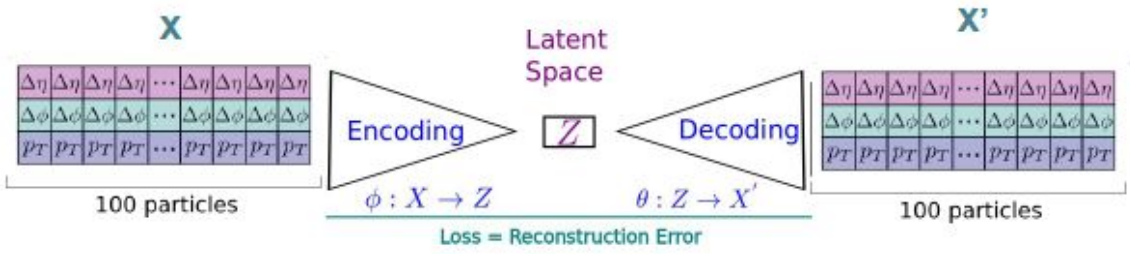
(Postdoc fellow: Scuola Normale Superiore di Pisa)

11th November, 2021

Recap: ONNX inference on CPU with Jet-level VAE - I

- Jet-level VAE model (Tensorflow)

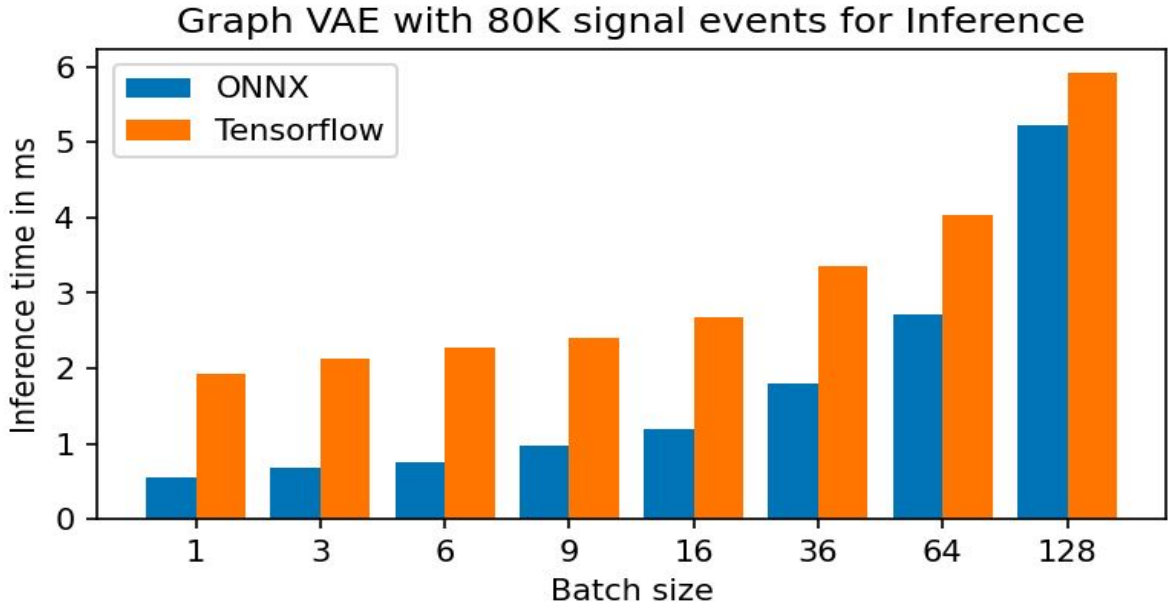
  Input: Particle list (η, φ, pt), Jet1 & Jet2



- Model conversion to onnx model *(tf2onnx)*
- Dijet test data:
  - test_sample = [‘GtoWW15br’, ‘GtoWW15na’]

Sensitivity study: Model inference (CPU) for Graph-VAE model with Dijet events data - I

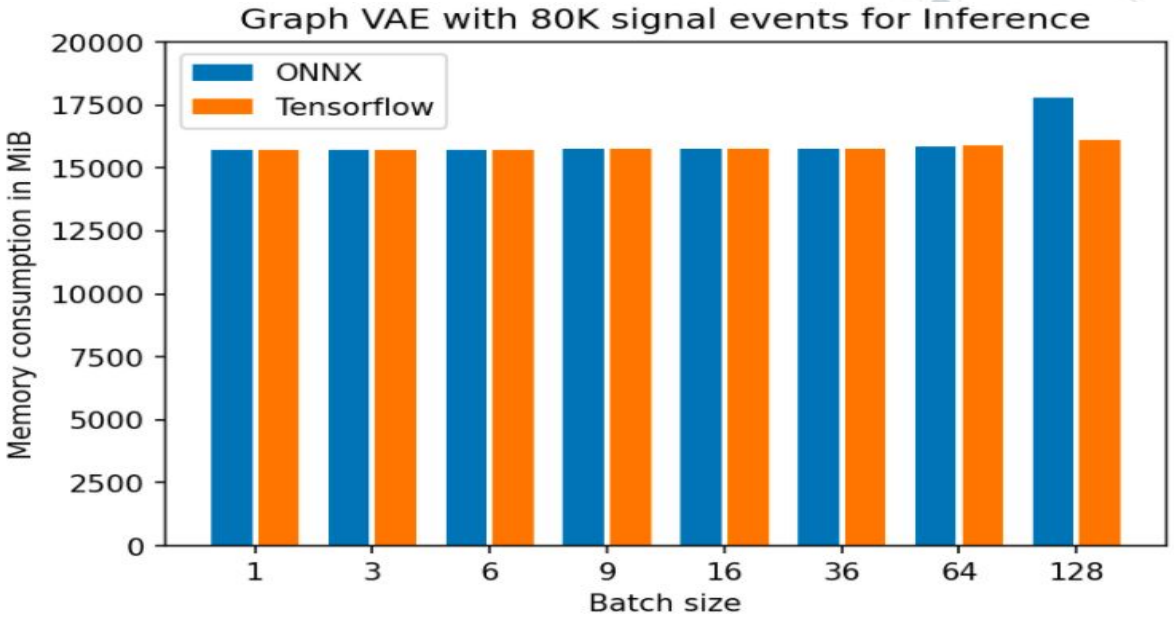
Difference in Inference time between ONNX and TF saturates with increase in batch size



Sensitivity Analysis: Model inference (CPU) for Graph-VAE model with Dijet events data - II

Memory Footprint stays constant with varying batch size, similar behavior for ONNX and TF

Need to check on memory footprint vs. batch size with dense (parameter space) and highly accurate GNN VAE model



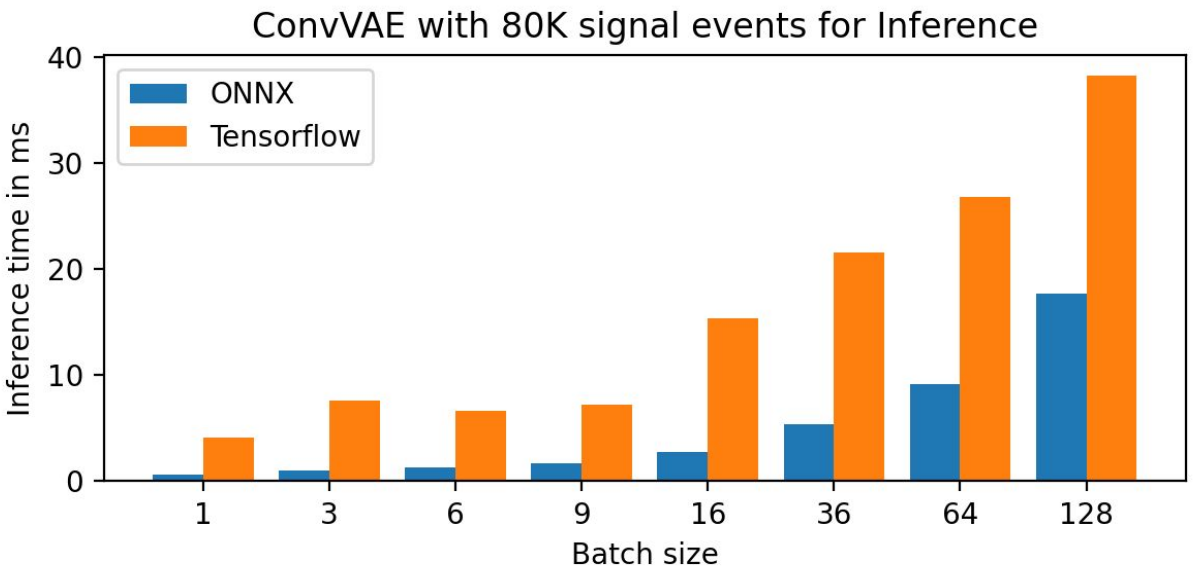
Sensitivity study: Model inference (CPU) for Conv-VAE - I

- With input data (signal) stored at EOS drive and converted into TF Dataset object

Difference in Inference time between ONNX and TF shrinks quickly with varying batch size

maximal gain at batch_size: 1

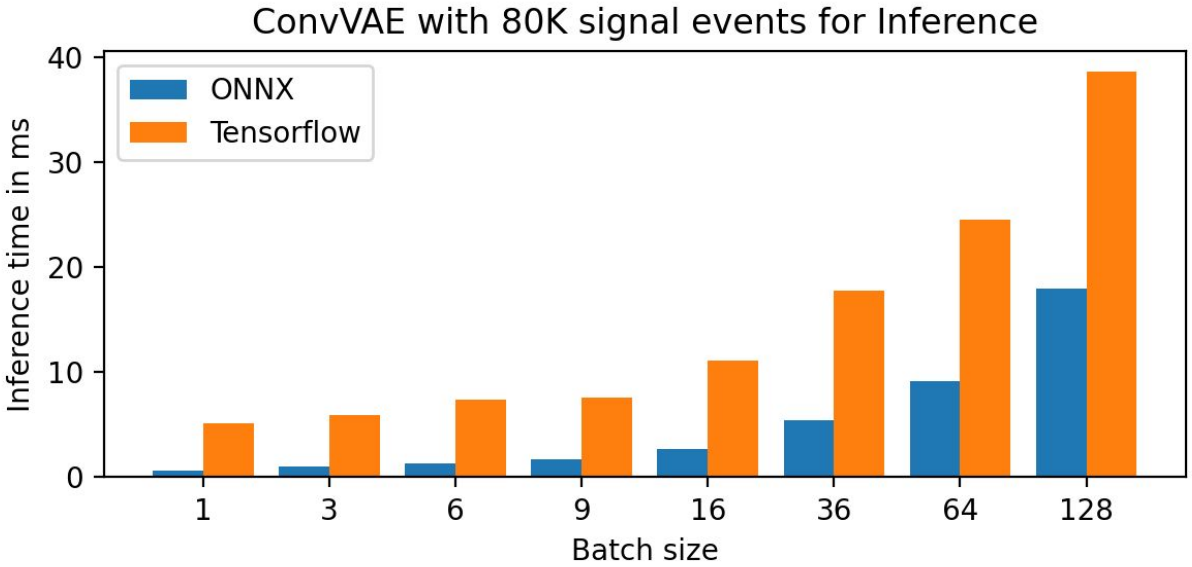
RUN 1



Sensitivity study: Model inference (CPU) for Conv-VAE - II

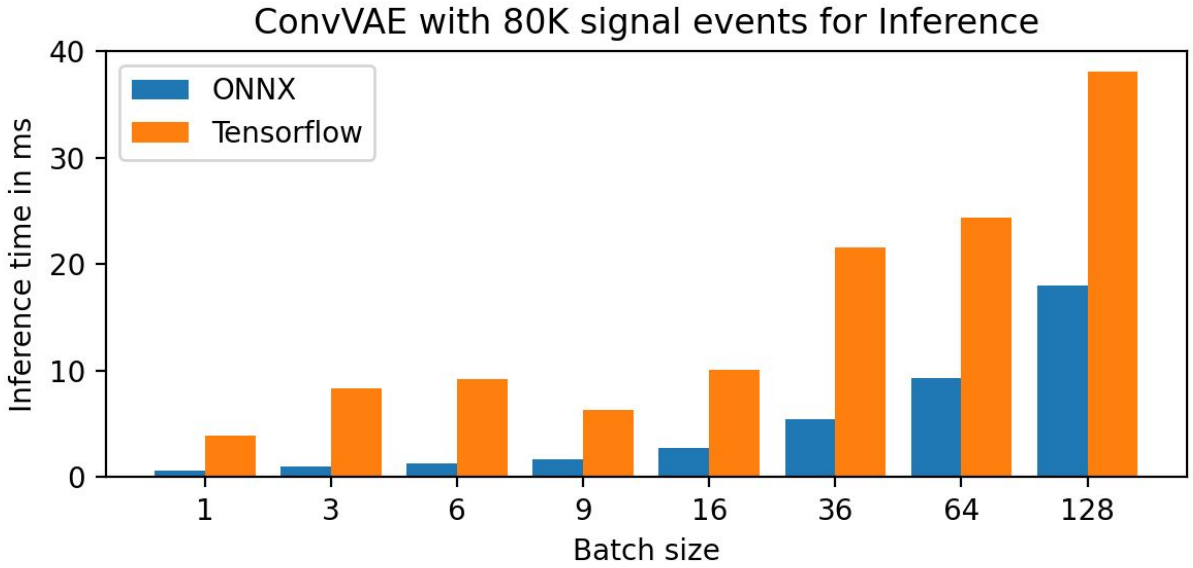
Both Run 1 and Run 2 show similar inference results, reaffirms TF and ONNX runtime working correctly

RUN 2



Sensitivity study: Model inference (CPU) for Conv-VAE - III

- New training and saving the model. Generating a new ONNX model file for inference. This is done to assess robustness and validate inference results

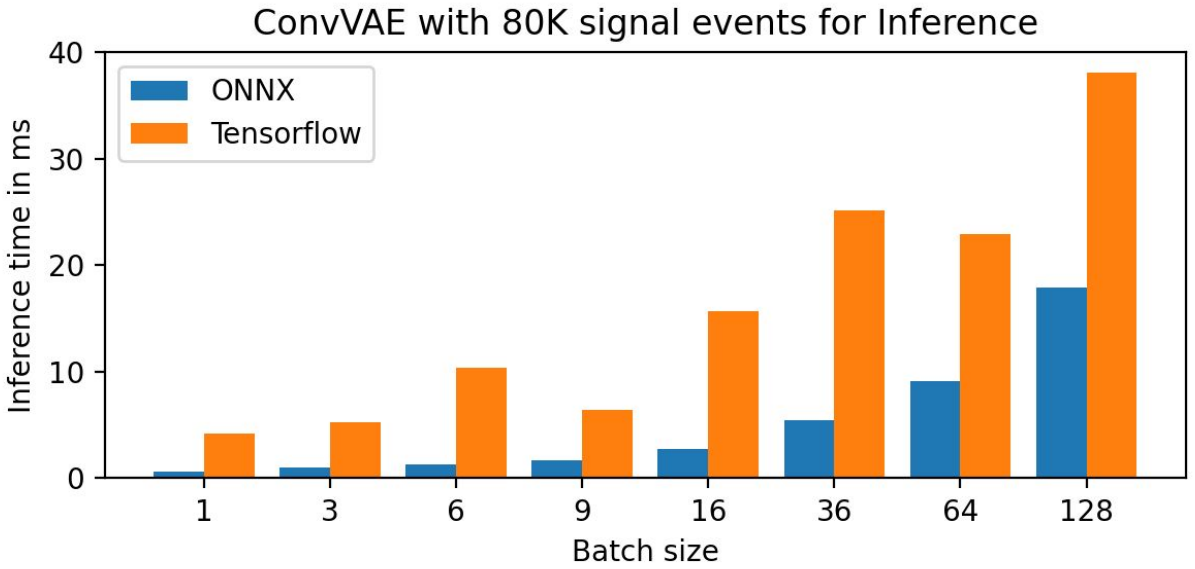


Sensitivity study: Model inference (CPU) for Conv-VAE - IV

- With input data (signal) stored at local directory @ Mustafar machine and converted into TF Dataset object

Both input data at EOS and local disk show similar looking inference results as inference is done on batched TF Dataset Object

RUN 1



Sensitivity study: Model inference (CPU) for Conv-VAE - V

- With input data (signal) stored at local directory @ Mustafar machine and converted into TF Dataset object

RUN 2