Convolutional Autoencoders for Anomaly Detection in the L1 Trigger

Sierra Weyhmiller

Supervisor: Maurizio Pierini

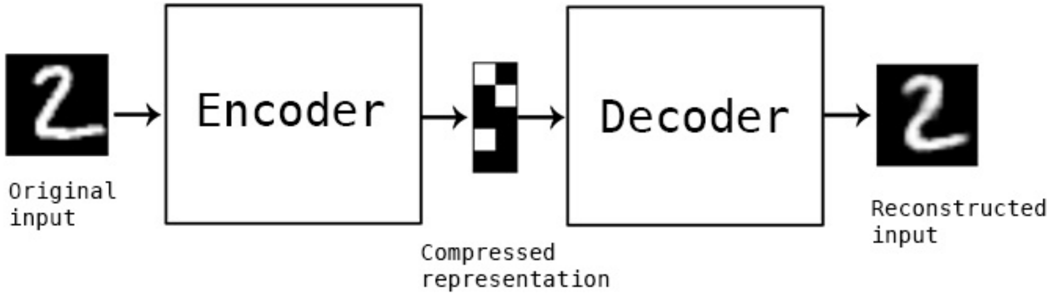
CMG Meeting 21/8

Introduction

- Main Ideas
- Prior Work
- Data
- Basic Algorithm and Results
- Network Compression
- Future Work
- Summary

Anomalies and Autoencoders

- Anomalous events may be new physics candidates
- Deep learning in real-time
- Model independent method requiring high rejection rate for low trigger rate
- Autoencoders
  - Encode input in smaller dimensional space
  - Anomalous events will fail encode/decode flow
  - Anomalous data have high loss
- Convolutional autoencoder
  - Build "image" from event object
  - Convolution learns small, meaningful features



Prior Work

- Threshold test for identifying anomalous events in the LHC's High Level Trigger (HLT)
  - Variational autoencoders trained on SM events
  - Mixed events -> categorize BSM events as anomalous
  - Resources - GPU
  - Latency O(1 ms)
- hls4ml: Fast inference of deep neural networks in FPGAs
  - Jet classifier
  - Resources - FPGA
  - Latency ~ 100 ns

L1 Trigger Restrictions

- Need an online algorithm
  - Trigger system might miss anomalies
- L1 Trigger
  - Limited resources (FPGA)
  - Low latency requirements (O(1 µs))

Data - SM

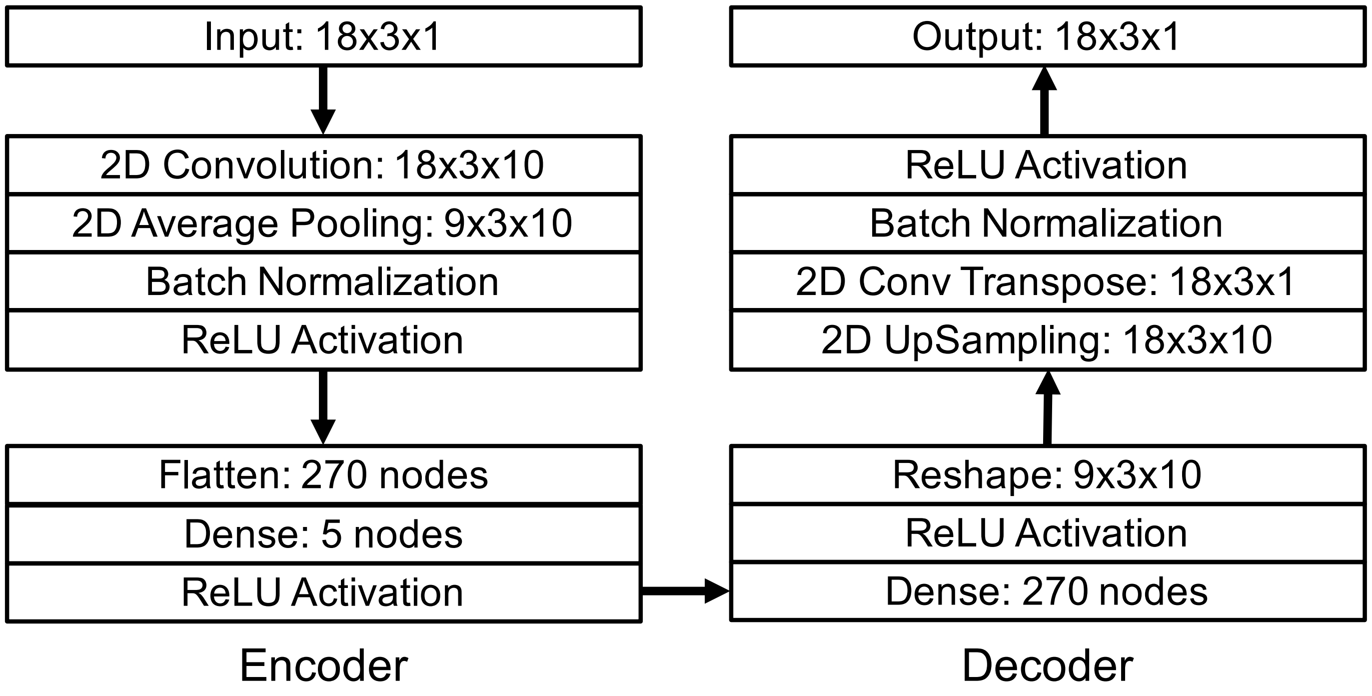
- Trained on SM, tested on BSM
- SM dominated by QCD
  - For simplicity, SM sample is 3.8 million QCD events
  - Ntuples with up to 10 jets, 4 muons, and 4 electrons
  - p_T, η, ϕ
  - Array of size [3.8M, 18, 3]

Data - BSM

- 4 sample BSM events
  - 291k VBF -> H -> invisible (VBF)
  - 50k Z’ -> ZH, MZ’ = 1TeV (ZH 1000)
  - 49k Z’ -> ZH, MZ’ = 0.8TeV (ZH 800)
  - 50k Z’ -> ZH, MZ’ = 0.6TeV (ZH 600)

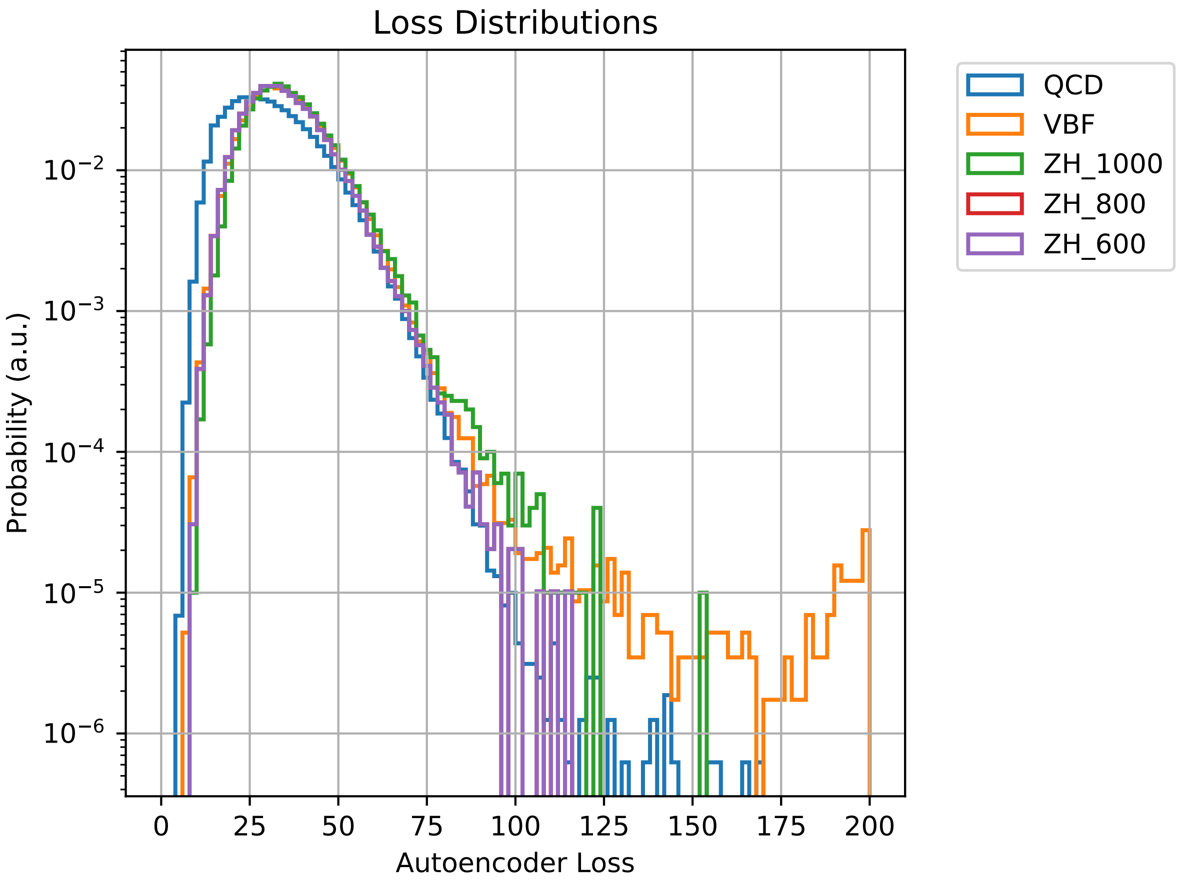
Basic Algorithm

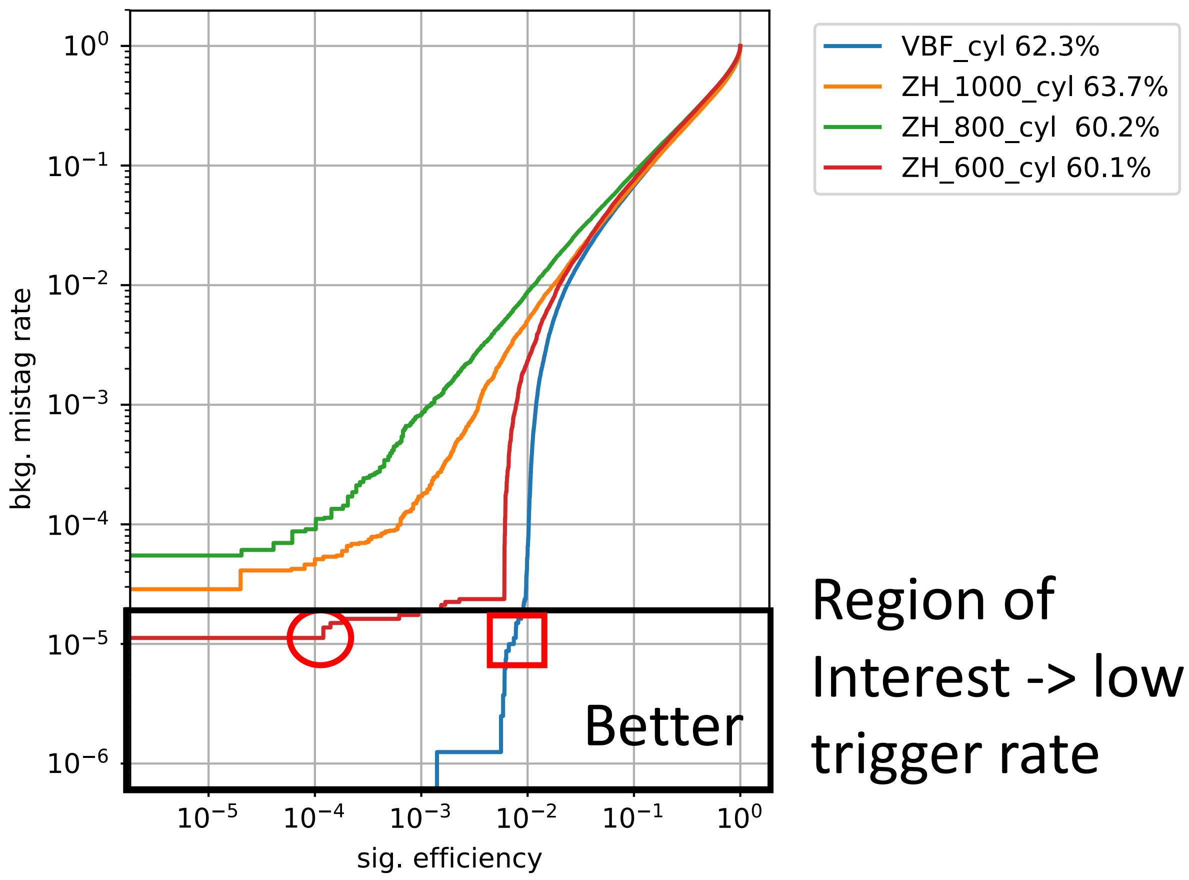
- Standardized Data
- Training, validation, test - 3:1:1 ratio
- Developed in Keras with TensorFlow
- MSE loss, Adam optimizer
- Alternative architectures -> no significant improvement
  - Latent Space Sizes
  - Coordinate Systems



Results

59.7 kFLOPS ~ 28 ns





Region of Interest -> low trigger rate

10X Signal Enhancement

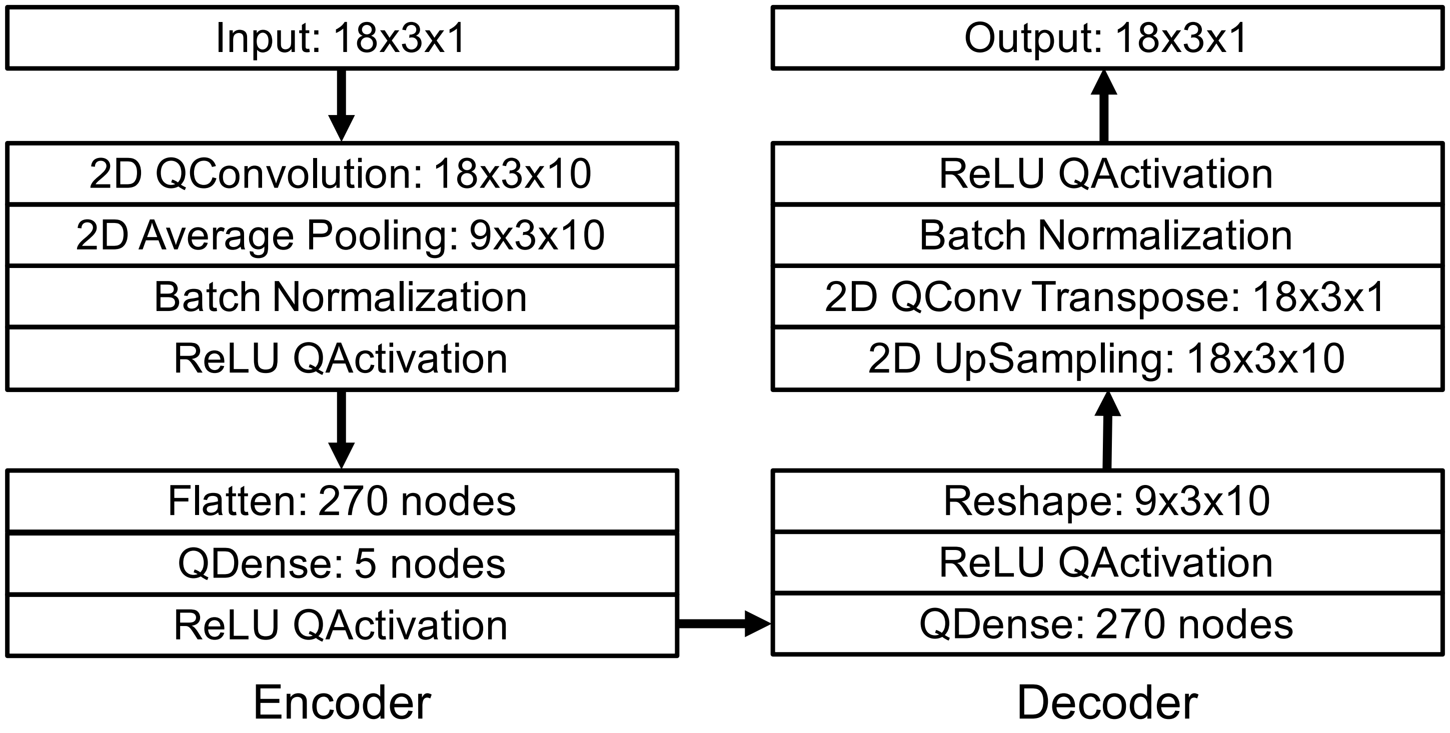
1000X Signal Enhancement

Loss Distribution

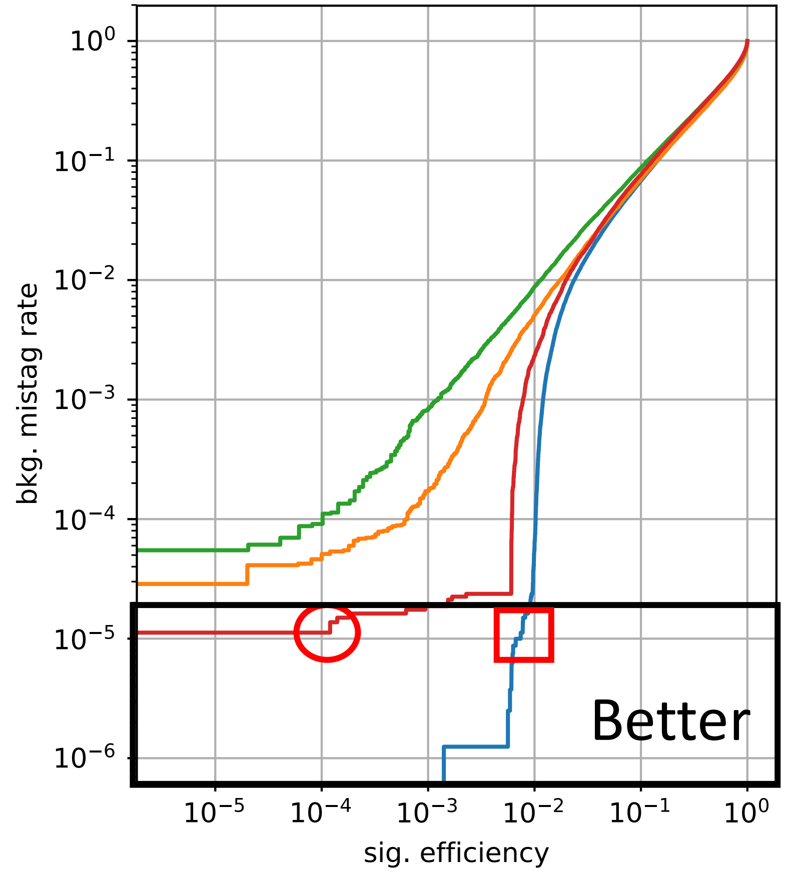
ROC Curve

Quantized Algorithm

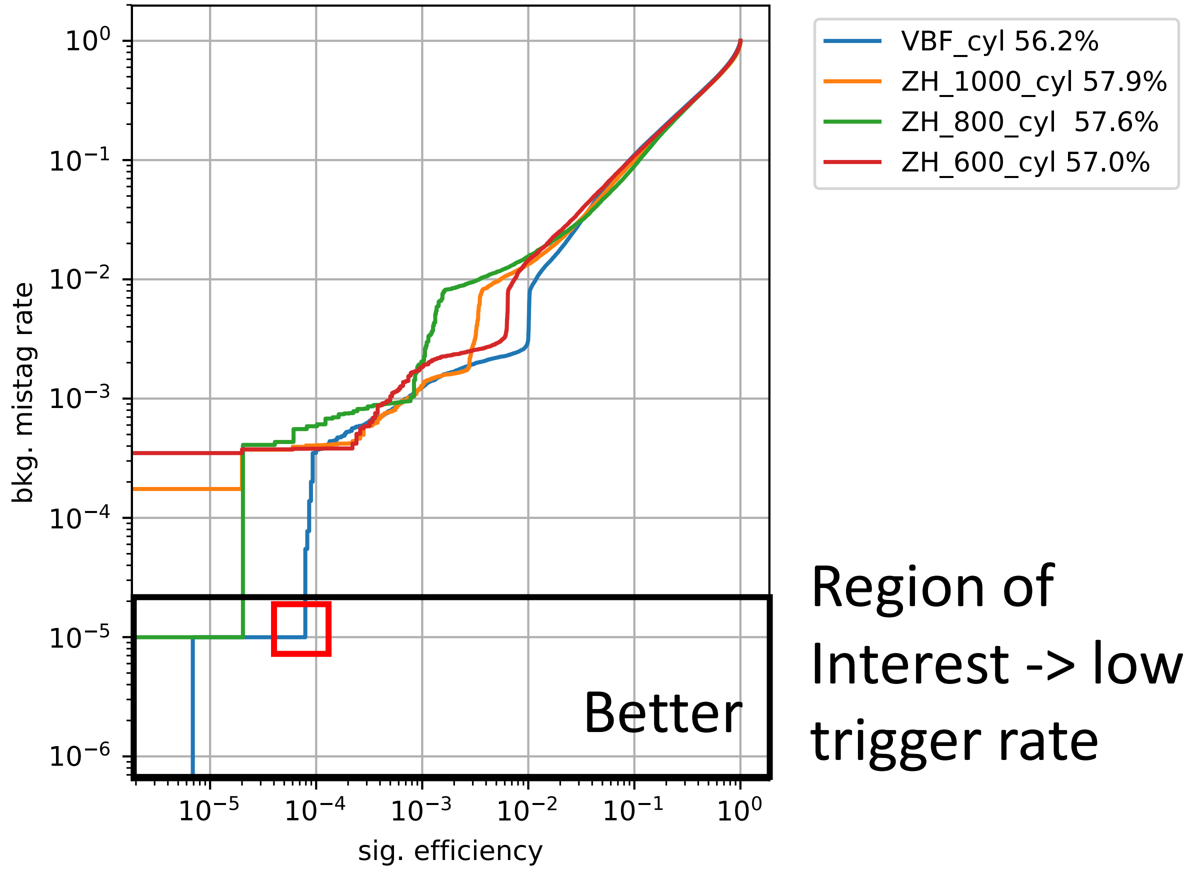
- Further reduction of latency and resource usage
- QKeras – compressed algorithm through quantization
- Basic algorithm with quantized layers – 8 bits
  - Convolutional
  - Dense
  - Activation



Results – ROC Curves



Region of Interest -> low trigger rate



Region of Interest -> low trigger rate

10X Signal Enhancement

1000X Signal Enhancement

32 Bits, 85.5 nJ

Factor of 15 reduction in energy

10X Signal Enhancement

8 Bits, 5.64 nJ

Future

- Test on Run3 data in FPGA for resource and latency usage
  - hls4ml
- Deploy in L1 trigger

Summary

- Anomaly detection algorithm in the L1 trigger
- Real-time Machine Learning
- Model-independent method with high background rejection rate
- Convolutional autoencoder
  - trained on SM, tested on BSM
- Quantized layers to reduce resources and latency
- Resources and latency post-synthesis still being calculated

Thanks!

- MPP Team
- Boston University Study Abroad
- Notre Dame Study Abroad
- Notre Dame Glynn Family Honors Program