Detect New Physics with Deep Learning Trigger at the LHC

Zhenbin Wu (UIC)
--on behalf of the CMS Collaboration

Connect The Dots
April 20-30, 2020
The LHC Big Data Problem

Data Flow

L1 Trigger
- 40M bunch crossing per second
- Logging rate: ~100 kHz
- Non-zero suppressed RAW data rate ~1PB/s
- Zero suppressed data rate is ~20TB/s
- Coarse local reconstruction implemented on FPGA/hardware.

HLT Farm
- Logging rate ~1kHz
- Data rate ~1GB/s distributed over dozens of primary datasets
- Simplified global reconstructions implemented on CPUs.

Offline Computing
- Roughly 1GB/s data rate
- Global reconstruction fully optimized for accuracy with software implemented on CPUs.

Data Analysis
- User-written code, plots, theses, talks, etc.
- ~100 papers of 10 MB each year, less than 1kB/s
The interested physics productions are much smaller comparing to inelastic production.

- Trigger in LHC: finding a needle in a haystack scenario (anomaly)
- Event not trigger will be lost forever!
The LHC Big Data Problem

A Drastic Data Reduction

Could new physics have been discarded by the trigger already?
Model-Independent Searches in HEP

- Traditional new physics search relies on hypothesis testing with specific alternative models.
- Motivated multiple attempts for model-independent searches in high-energy physics over the years.
Anomaly Trigger

- A Model-Agnostic Trigger for anomaly events with deep learning
- Deployment at Level-1 trigger to avoid bias from upstream
- Within the resource and latency requirement of the trigger system
Autoencoders in a Nutshell

• Compression-decompression algorithm that learns to describe the a given dataset in terms of point in lower-dimension *latent space*, from which it reconstructs the original data.

• **Unsupervised learning**, used for data compression, generation, clustering, etc.

• Anomaly: any event whose decompressed output is “far” from the input, in some metric of the autoencoder loss.

\[
\text{Loss} = f(\text{input} - \text{output})
\]
CMS Phase 2 Level-1 Trigger

- Sketch of upgraded CMS Phase 2 Level-1 Trigger system
- Produce Particle Flow particles, combining Calo/Muon/Tracker information
- Produce PUPPI weight of each particles for pileup mitigation
- Outputs of each trigger systems send to Global Trigger for Level-1 decision
Example AE Model

- Train with simulated ZeroBias event at 200 pileup
- Use simulated PUPPI Jet/MET/MHT inputs (18 inputs) with preprocessing
- Model to be implemented at Global Trigger

- Activation function: ReLU
- Loss function: L1Loss
- Number of layers: 8 layers

- Model is designed with simplicity for firmware implementation and resource/latency requirement
AE Performance

- Model was trained and validated with simulated Zerobias events, no knowledge of signal during training
- Use the reconstruction loss of AE inputs and outputs as discriminator
- Inference with signal samples show the separation power
AE Implementation

- Use the hls4ml package to implement the AE model into FPGA firmware
- With additional logic for L1Loss function calculation
- Fully unroll AE with minimal latency, well within the Phase 2 Global Trigger latency budget
- With Xilinx Virtex UltraScale+ (VU9P) FPGA, the AE consumes ~10% of DSP resource, ~1% of Flip-Flops and LUTs
- AE is not sensitive to small variances of individual input features, can tolerate noisy online condition
- Trigger rate rising above a fixed budget can be prevented by additional logic in the firmware that would dynamically adapt the autoencoder threshold
Conclusions

• The LHC has an enormous potential of discovering physics beyond the Standard Model, given the unprecedented collision energy and the large variety of production mechanisms that proton-proton collisions can probe.

• We propose a model-independent anomaly detection technique, based on deep autoencoders, to identify new physics events.

• Simple AE model can be implemented at the Level-1 trigger level.

• More advanced AE model can be designed for HLT or 40MHz scouting system.

• Stay tune for the CMS Phase 2 Level-1 Trigger TDR.