Deep Learning in High Energy Physics

Examples from the LHC

Sofia Vallecorsa – June 20th, 2019
CERN openlab

A science – industry partnership to drive R&D and innovation

Evaluate state-of-the-art technologies in a challenging environment and improve them.

Test in a research environment today what will be used in many business sectors tomorrow.

Training

Dissemination and outreach
CERN

International organisation close to Geneva, straddling Swiss-French border

Founded 1954

Facilities for fundamental research in particle physics

23 member states, 1.1 B CHF budget

3’197 staff, fellows, apprentices, …

13’128 associates

“Science for peace”

1954: 12 Member States

Members: Austria, Belgium, Bulgaria, Czech republic, Denmark, Finland, France, Germany, Greece, Hungary, Israel, Italy, Netherlands, Norway, Poland, Portugal, Slovak Republic, Spain, Serbia, Sweden, Switzerland, United Kingdom

Candidate for membership: Cyprus, Slovenia

Associate members: India, Lithuania, Pakistan, Turkey, Ukraine

Observers: EC, Japan, JINR, Russia, UNESCO, United States of America

Numerous non-member states with collaboration agreements

2’531 staff members, 645 fellows, 21 apprentices

7’000 member states, 1’800 USA, 900 Russia, 270 Japan, …
The Large Hadron Collider (LHC)
The Higgs Boson
The Higgs Boson completes the Standard Model, but the Model explains only about 5% of our Universe. What is the other 95% of the Universe made of? How does gravity really work? Why there is no antimatter in nature?
Data Handling and Computation

Online Triggers and Filters

Offline Reconstruction
- Selection & reconstruction

Processed data (active tapes)

Offline Analysis
- Batch Physics Analysis
- Interactive Analysis

Event reprocessing

Offline Simulation
- Event simulation
- Geant4
ICT Infrastructure

~170 sites, 42 countries
~800k CPU cores, 600 PB of storage
2 million jobs/day
10-100 Gb links
Conclusions

DNN training and inference will likely become **important workflows** for large experiments

**Resources availability**: driving factor of the size of the problem we can solve with DNNs

Complicated network optimizations/training have **high computational cost** but…

DL development is accelerated by a **diversified community** (industry and society, applied and fundamental science)
Some background

Why? ...Big Data

*LHC is entering the Big Data era*

**Accelerators infrastructure**
- 9600 magnets for Beam Control
- 1232 superconducting dipoles for bending

**Experiments** *(detectors & physics data)*
- **330 PB of collisions data** stored by end 2018

**The computing infrastructure**
LHC data is **multi-structured, hybrid**
Why? ...New Challenges

Next generation colliders will require larger, highly granular detectors that will generate huge particle data rates $O(100 \text{ TB/s})$.
How? ... Deep Learning

DL can **recognize patterns** in large complicated data sets

Re-cast physics problems as “DL problems”

Adapt DL to HEP **requirements**

Adopting ”new” **computing models**

B. Hooberman, S.V. et al. (NIPS 2017)
Examples
Generative models

The problem:
Assume data sample follows $p_{\text{data}}$ distribution
Can we draw samples from distribution $p_{\text{model}}$ such that $p_{\text{model}} \approx p_{\text{data}}$?

A well known solution:
Assume some form for $p_{\text{model}}$ (using prior knowledge, parameterized by $\theta$)
Find the maximum likelihood estimator

$$
\theta^* = \arg \max_{\theta} \sum_{x \in D} \log(p_{\text{model}}(x; \theta))
$$

draw samples from $p_{\theta^*}$

Generative models don’t assume any prior form for $p_{\text{models}}$
Extract meaningful representation from training data
Deep Generative Models

Internal representations learned by shallow systems are simple

→ Deep Generative Models
  ▪ Allow higher levels of abstractions
  ▪ Improve generalisation and transfer

→ Multiple applications
  Discovery
  Anomaly Detection
  Planning
  Transfer Learning
Generative Models Zoo

Deep Generative Models

Fully-observed

Latent variables

NADE, eNADE
Fully-visible sigmoid belief networks
Pixel CNN/RNN
RNN Language mod.
Context tree switching

Boltzmann Machines
Discrete Markov
Random Fields
Ising, Hopfield
and Potts Models

Normal Means
Continuous
Markov Models
N-AR(p)
RNADE

Gaussian MRFs
Log-linear models

Indian buffet process
Dirichlet process mixture
Hidden Markov Model
Discrete LVM
Sparse LVMs

Cascaed Indian Buffet process
Hierarchical Dirichlet process
Sigmoid Belief Net
Deep auto-regressive networks (DARN)

Deep Gaussian processes
Recurrent Gaussian Process
GP State space model

Nonparametric
Discrete
Continuous

Non-parametric
Deep

Linear
Parametric

PCA, factor analysis
Independent components analysis
Gaussian LDS
Latent Gauss Field

Gaussian process LVM

Directed
Undirected

Pictures from Danilo Rezende’s Tutorial on Deep Generative Models
Generative Adversarial Networks

Two networks competing with each other

**Generator** generates data from random noise

Generator learning is supervised by the **discriminator** network

The forger/detective case
Forger shows its Monalisa to the detective
Detective says it is fake
Forger makes new Monalisa based on feedback
Iterate until detective is fooled
How well does it work?

Ian Goodfellow (@goodfellow_ian)

4 years of GAN progress (source: eff.org/files/2018/02/ ...)

2014  2015  2016  2017

7:26 pm - 2 Mar 2018
Performance evaluation

Deployment in scientific domains requires robust performance studies

- We need to assess the difference between model PDF and real PDF
  - Mixing and coverage (diversity)
  - Saliency
  - Mode collapse or mode dropping
  - Overfitting (has the network memorized samples?)

Need quantities that are invariant to small translation, rotation, intensity changes

- Define a way to map input into a feature space
  - Kullback-Leibler Divergence
  - Inception score, Fréchet Inception Distance
  - Maximum Mean Discrepancy
  - Structural Similarity Index

+ Physics Quantities Validation
Fast simulation in High Energy Physics

Monte Carlo simulation is a major workload in terms of computing resources.

Generative Models are a generic approach to replace expensive calculations.

Inference is faster than Monte Carlo approach.

Industry building highly optimized software, hardware, and cloud services.

Numerous R&D activities (LHC and beyond)

<table>
<thead>
<tr>
<th>Method</th>
<th>Machine</th>
<th>Time/Shower (msec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Simulation (geant4)</td>
<td>Intel Xeon Platinum 8180</td>
<td>17000</td>
</tr>
<tr>
<td>3DGAN (batch size 128)</td>
<td>Intel Xeon Platinum 8160 (TF 1.12)</td>
<td>1</td>
</tr>
</tbody>
</table>

WLCG Wall Clock time for the ATLAS experiment

S.V., ACAT 2019
Detector output as 3D image

Array of absorber material and silicon sensors

CLIC is a CERN project for a linear accelerator of electrons and positrons to TeV energies

Electromagnetic calorimeter design
- Sparse images
- Highly segmented (pixelized)
- Large dynamic range

http://cds.cern.ch/record/2254048#
3D convolutional GAN

~1M parameters
Total model Size: 3.8MB
3DGAN performance

Conditional training, Custom losses
Performance validation
Convergence and discriminator performance
Comparison to Monte Carlo
Shower Shapes, Sampling Fraction
Correlations, Sparsity, etc..
“In-house inception score”
TriForce\(^\text{(1)}\) classification and regression on GAN/GEANT4
Image Quality Analysis

\(^{\text{(1)}}\)Matt Zhang,
https://github.com/BucketOfFish/Triforce_CaloML
Generated events

*Dynamic range and sparsity*

**Agreement** to Monte Carlo across seven orders of magnitude

$E_p = 189$ GeV, $\alpha = 63^\circ$
Transfer learning: extending the energy range

- Transfer learning from 100-200 GeV pre-trained network
- Double training dataset statistics, 3 epochs training

Error on sampling fraction

2-500 GeV

Transverse shower width

Longitudinal shower width

Improved correlation description!
Sample diversity

Structural Similarity Index

Structural Similarity Index (SSIM) [4] is used to assess similarity between images.

Tipically used in denoising applications.

Measure diversity in GAN generated images.

$E=150 \text{ GeV}, \text{ orthogonal incident angle}$

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>G4 vs G4</th>
<th>GAN vs GAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1 \times 10^{-1}$</td>
<td>0.94</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>$1 \times 10^{-2}$</td>
<td>0.21</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>$1 \times 10^{-4}$</td>
<td>0.045</td>
<td>0.061</td>
<td></td>
</tr>
<tr>
<td>$1 \times 10^{-6}$</td>
<td>0.045</td>
<td>0.051</td>
<td></td>
</tr>
</tbody>
</table>

$SSIM(x, y) = [I(x, y)]^\alpha \cdot [C(x, y)]^\beta \cdot [S(x, y)]^\gamma$

SSIM as training progresses.
Other applications in fast simulation

Generative models for ALICE TPC simulation (ACAT2019)
Conditional Wasserstein GANs for fast simulation of electromagnetic showers in a CMS HGCAL prototype (IML WG 04/18)
Variational AutoEncoders to simulate ATLAS LAr calorimeter (PASC18)
Wasserstein GANs to generate high-level physics variables based on Monte Carlo ttH (superfast-simulation) (IML WG 04/18)
Particle-GAN for Full Event Simulation at the LHC (ACAT2019)
Refining Detector Simulation using Adversarial Networks (IML WG 04/18)
Model-Assisted GANs for the optimisation of simulation parameters (IML WG 04/19)
AutoEncoders & Variational AutoEncoders

AEs learn how to describe training dataset in latent space

- Data compression, dimensionality reduction (PCA) and de-noising

Variational AEs have added constraints on the encoded representations

- Learn latent model then sample from it

Many applications at the LHC
Triggers: real time event selection

We can process only a **minimal fraction** of collider data

Keep only the **interesting** events

Sophisticated studies to **optimise trigger algorithms** for specific physics processes

We don’t know what **unknown physics** looks like!
VAE for new physics mining

*Physics mining as an anomaly detection problem*

Classical strategy uses a very **loose selection**
- 1M Standard Model events per day
- Will not scale

Use **anomaly detection** tools

Train a VAE on known physics
- Monte Carlo data
- Real detector data

Run it in real time and store only “anomalies”
Selecting the unknown!

*VAE as model-independent new physics trigger*

Create a dataset of **anomalous events**
Can probe **large range of processes**
Alternative strategy, parallel to canonical approaches
Might open **new physics** directions
Pattern recognition in HEP

Particle Trajectory Reconstruction

Particle trajectory bended in a solenoid magnetic field
Need curvature to measure momentum

Particle ionize silicon sensors arranged in concentric layers
Thousands of sparse hits
Many hits are uninteresting
AutoEncoders for tracking

CNN used for activity segmentation and detector hits classification
More examples

A major challenge after LHC upgrade

Hopefield network  http://inspirehep.net/record/300646/
CNN in NOVA  https://arxiv.org/abs/1604.01444
HEP.TrkX : https://heptrkx.github.io/
TrackML : https://tinyurl.com/y84yd5hn

Quantum annealing!

Quadratic Unconstrained Binary Optimisation (QUBO) can be mapped to an Ising Hamiltonian with change of variable \{0,1\} \leftrightarrow \{-1,1\}
A quantum advantage for ML?

Quantum linear algebra is generally faster than classical counterpart

Some standard ML techniques estimate the ground state of Hamiltonians

ML algorithms have some tolerance to errors

Specific quantum techniques can be exploited to bring further improvement
Quantum Support Vector Machine

Quantum Machine Learning are among the first applications to be implemented on near-term devices

Quantum SVM for ttH (H → γγ) classification
- QSVM is simulated on IBM Qiskit simulator
- Entanglement is used to encode relationships between features

Apply PCA to input data features
- Reduced from 45 to 8, 10 or 20 (limited by number of qubits)

Running full training with quantum simulators requires large computing resources
- Memory increases with qubit, training events and complexity
Quantum GAN

Generative Adversarial Networks are among the most interesting models in classical machine learning.

- Quantum GAN would have more representational power than classical GAN.
- Different hybrid classical-quantum algorithms for generative models exist.
  - i.e. quantum Variational Auto-Encoders on D-Wave annealer.

Train a quantum GAN to generate few-pixels images.

Currently investigating two possible approaches:

- A hybrid schema with a quantum generator learning the target PDF using either a classical network or a variational quantum circuit as a discriminator (Variational Quantum Generator).
- Full quantum adversarial implementation (quGAN).
Conclusions ...again

DNN training and inference will likely become important workflows for large experiments

So what is the HEP community looking into?

- **Integrate** new frameworks in “classical” software stacks
- **Expand** a pure HTC approach to HPC and Cloud environments
- **Improve** usability and deployment
- **Long term innovation** such as Quantum Computing

DL lives in a diversified ecosystem that evolves extremely rapidly

- R&D and **collaboration** with industry and other communities is essential
Thanks!

Questions?

https://openlab.cern