

4th Inter-experiment Machine Learning Workshop

Monday 19 October 2020 - Friday 23 October 2020

Book of Abstracts

Contents

Reinforcement learning environment for deep learn physics dataset 2	1
Accelerated pixel detector tracklet finding with Graph Neural Networks on FPGAs 5 . . .	1
Matrix Element Regression with Deep Neural Networks – breaking the CPU barrier 6 . .	2
Quantum Graph Neural Networks for Track Reconstruction in Particle Physics and Beyond 7	2
UCluster: Unsupervised clustering for HEP 8	3
Teaching Machine Learning with ATLAS Open Data 9	3
Active Anomaly Detection for time-domain discoveries 10	3
AutoDQM: A Statistical Tool for Monitoring Data Quality in the CMS Detector 11	4
Design by intelligent committee: use of machine learning as a scientific advisor 12	4
Pre-Learning a Geometry Using Machine Learning to Accelerate High Energy Physics De- tector Simulations 13	5
Hyperparameter Optimisation for Machine Learning using ATLAS Grid and HPC 14 . . .	5
MLaaS4HEP: Machine Learning as a Service for HEP 16	6
SWAN: Powering CERN’s Data Analytics and Machine Learning Use cases 17	6
Reduced Precision Strategies for Deep Learning: 3DGAN Use Case 18	7
Hit-reco: ProtoDUNE denoising with DL models 19	7
Identifying jets in the Lund plane 20	8
A readily-interpretable fully-convolutional autoencoder-like algorithm for unlabelled wave- form analysis 21	8
Distributed training of graph neural network at HPC 22	8
High Fidelity Simulation of High Granularity Calorimeters with High Speed 23	9
GANplifying Event Samples 24	9
Set2Graph: Secondary Vertex finding in Jets with Neural Networks 25	10
Zero-Permutation Jet Parton Assignment 26	10

Estimating Support Size of Distribution Learnt by Generative Adversarial Networks for Particle Detector Simulation 27	11
Object condensation: one-stage grid-free multi-object reconstruction in physics detectors, graph, and image data 29	11
Quantum Generative Adversarial Networks 30	11
Accelerating GAN training using distributed tensorflow and highly parallel hardware 31	12
Invertible Networks or Partons to Detector and Back Again 32	13
DisCo: Robust Networks and automated ABCD background estimation 33	13
Lorentz Equivariant Neural Networks for Particle Physics 34	14
Improving particle-flow with deep learning 35	14
Super-resolution for calorimetry 36	15
Efficiency parametrization with Neural Networks 37	15
Graph Neural Network-based Event Classification for Measurement of the Higgs-Top Yukawa Interaction 39	16
Generative Adversarial Network for Identifying the Dark Matter Distribution of a Dwarf Spheroidal Galaxy 40	16
Graph Convolutional Operators in the the PyTorch JIT 42	17
Selective background MC simulation with graph neural networks at Belle II 43	17
General recipe to form input space for deep learning analysis of HEP scattering processes. 44	18
Foundations of a Fast, Data-Driven, Machine-Learned Simulator 45	18
Pixel Detector Background Generation using Generative Adversarial Networks at Belle II 46	19
Generative models for calorimeters response simulation - from GANs through VAE to e2e SAE 47	19
Model-Independent Detection of New Physics Signals Using Interpretable Semi-Supervised Classifier Tests 49	19
Disentangling Boosted Higgs Boson Production Modes with Machine Learning 50	20
Anomaly Awareness for BSM Searches at the LHC 52	20
Adaptive divergence for rapid adversarial optimization & (1 + epsilon)-class Classification: an Anomaly Detection Method for Highly Imbalanced or Incomplete Data Sets 53	21
Using an Optical Processing Unit for tracking and calorimetry at the LHC 55	22
Deep learning solutions for 2D calorimetric cluster reconstruction at LHCb 56	22
Simulation-Assisted Decorrelation for Resonant Anomaly Detection 57	23

Decorrelation via Disentanglement 58	23
Enhancing searches for resonances with machine learning and moment decomposition 59	24
Black-Box Optimization with Local Generative Surrogates 62	24
Fast simulation of Time Projection Chamber response at MPD using GANs 63	25
Domain Adaptation Techniques in Particle Identification for the ALICE experiment 64 . .	25
Using Machine Learning to Speed Up and Improve Detector R&D 65	26
Bayesian Neural Networks for Predictions from High Dimensional Theories 68	27
GPU and FPGA as a Service for Machine Learning Inference Accelerations 69	27
Structured models of objects, relations, and physics 72	28
End-to-End, Machine Learning-based Data Reconstruction for Particle Imaging Neutrino Detectors 73	28
Solving Inverse Problems with Invertible Neural Networks 74	29
Neural Network Pruning: from over-parametrized to under-parametrized networks 75 .	29
Deep Learning @ LHC: An ATLAS Perspective 76	29
Deep Dive on Graph Networks for Learning Simulation (Deep Mind) 79	30
Tracking GNN Walk Through 80	30
Zenseact : Deep learning and computer vision for self-driving cars 81	30
Machine Learning in Procter and Gamble 82	30
Using Topological Data Analysis to Disentangle Complex Data Sets 83	30
Introduction 84	31
CERN Knowledge Transfer 86	31
Conclusion and wrap-up 87	31
FastCaloGAN: a tool for fast simulation of the ATLAS calorimeter system with Generative Adversarial Networks 88	31

Workshop / 2

Reinforcement learning environment for deep learn physics dataset**Author:** Maciej Majewski¹¹ AGH-UST**Corresponding Authors:** maciej.witold.majewski@cern.ch, mmajewsk@cern.ch

Deep learn physics open dataset contains thousands of frames LARTPC detector data. The main problem of the dataset is semantic segmentation. This problem has been solved successfully with modified version of U-Net, as well as graph-networks. The main difficulty of this problem lays within the sparsity of data (thin tracks inside pixels, or voxels) which make it difficult to feed classical machine learning algorithms. Although the translation of a problem to reinforcement learning usually performs suboptimally compared to supervised machine learning, we propose library called LARTPC-game, which translates the deep learn physics dataset to reinforcement learning environment. It is not motivated by increased performance, but in novelty of a application of reinforcement learning in such kind of problems, as well as by a possibility of introducing a model more closely related to physics than others. This library may be used to require the model to behave exactly as a particle would behave, thus creating model of particle, directly from detector data.

Workshop / 5

Accelerated pixel detector tracklet finding with Graph Neural Networks on FPGAs**Authors:** Savannah Jennifer Thais¹; Lindsey Gray²; Markus Julian Atkinson³; Gage DeZoort¹; Isobel Ojalvo¹; Mark Neubauer³; Javier Mauricio Duarte⁴; Vesal Razavimaleki^{None}; Aneesh Heintz⁵¹ Princeton University (US)² Fermi National Accelerator Lab. (US)³ Univ. Illinois at Urbana Champaign (US)⁴ Univ. of California San Diego (US)⁵ Cornell University

Corresponding Authors: justin.gage.dezoort@cern.ch, savannah.jennifer.thais@cern.ch, javier.mauricio.duarte@cern.ch, lindsey.gray@cern.ch, isabel.rose.ojalvo@cern.ch, msn@illinois.edu, markus.atkinson@cern.ch, vrazavim@ucsd.edu, ah2263@cornell.edu

Track finding is a critical and computationally expensive step of object reconstruction for the LHC detectors. The current method of track reconstruction is a physics-inspired Kalman Filter guided combinatorial search. This procedure is highly accurate but is sequential and thus scales poorly with increased luminosity like that planned for the HL-LHC. It is therefore necessary to consider new methods for representing and reconstructing tracks.

This work makes use of Graph Neural Networks (GNNs) to explore possible improvements to track finding efficiency in the HL-LHC environment. A graph is constructed from each event by mapping hits in the pixel detector to graph nodes and constructing connecting edges using a physics-driven pre-processing. An edge-classification GNN is then used to assign physics-based probabilities to the connections. Finally, a post-processing algorithm is applied to iterate through the GNN labeled edges and form final track candidates.

We focus on a specific HL-LHC use case: tracklet finding in the innermost pixel detector using the expected Phase 2 geometry. Both ATLAS and CMS utilize inside-out track reconstruction algorithms that are seeded from the pixel detector making this a critical computing problem for HL-LHC

development. This study will provide insight into the impact these novel algorithms can have on compute-intensive and physics-critical reconstruction.

The GNN-based tracking can be further accelerated by implementing the inference network directly on FPGAs. This would improve the computational throughput of a critical reconstruction step and could allow the GNN algorithm to be used in the High-Level Trigger system.

This talk will present and compare a variety of graph construction methods, GNN architectures (GCNs, edge convolutions, and Interaction Networks), post-processing track finding (Union Find and DBScan), and data augmentations that have been explored to understand the balance between tracking accuracy and algorithmic processing efficiency. We will also present initial studies on implementing these GNNs and related algorithms on FPGAs.

Workshop / 6

Matrix Element Regression with Deep Neural Networks – breaking the CPU barrier

Authors: Florian Bury¹; Christophe Delaere^{None}

¹ *UCLouvain - CP3*

Corresponding Author: florian.bury@uclouvain.be

The Matrix Element Method (MEM) is a powerful method to extract information from measured events at collider experiments. Compared to multivariate techniques built on large sets of experimental data, the MEM does not rely on an examples-based learning phase but directly exploits our knowledge of the physics processes. This comes at a price, both in term of complexity and computing time since the required multi-dimensional integral of a rapidly varying function needs to be evaluated for every event and physics process considered. This can be mitigated by optimizing the integration, as is done in the MoMEMta package, but the computing time remains a concern, and often makes the use of the MEM in full-scale analysis unpractical or impossible. We investigate in this paper the use of a Deep Neural Network (DNN) built by regression of the MEM integral as an ansatz for analysis, especially in the search for new physics.

Workshop / 7

Quantum Graph Neural Networks for Track Reconstruction in Particle Physics and Beyond

Authors: Cenk Tuysuz¹; Bilge Demirkoz¹; Daniel Dobos²; Kristiane Sylvia Bernhard-Novotny³; Carla Sophie Rieger^{None}; Karolos Potamianos⁴; Federico Carminati⁵; Sofia Vallecorsa⁵; Fabio Fracas⁶; Jean-Roch Vlimant⁷

¹ *Middle East Technical University (TR)*

² *Lancaster University (GB)*

³ *gluoNNet*

⁴ *University of Oxford (GB)*

⁵ *CERN*

⁶ *Universita e INFN, Padova (IT)*

⁷ *California Institute of Technology (US)*

Corresponding Authors: daniel.dobos@cern.ch, jean-roch.vlimant@cern.ch, cenk.tuysuz@cern.ch, federico.carminati@cern.ch, sofia.vallecorsa@cern.ch, carieger@student.ethz.ch, fabio.fracas@cern.ch, kristiane.nowotny@cern.ch, bilge.demirkoz@cern.ch, karolos.potamianos@cern.ch

The Large Hadron Collider (LHC) at the European Organisation for Nuclear Research (CERN) will undergo an upgrade to further increase the instantaneous rate of particle collisions (luminosity) and become the High Luminosity LHC. This increase in luminosity, will yield many more detector hits (occupancy), and thus measurements will pose a challenge to track reconstruction algorithms being responsible to determine particle trajectories from those hits. Similar challenges exist in non-high energy physics (HEP) trajectory reconstruction use-cases. High occupancy, track density, complexity and fast growth exponentially increase the demand of algorithms in terms of time, memory and computing resources.

Graph Neural Networks are currently explored for HEP, but also non-HEP trajectory reconstruction applications. Use of Quantum Computers in HEP applications is also a new trend with their feature of evaluating a very large number of states simultaneously are therefore good candidates for such complex searches in large parameter and graph spaces.

In this work, we discuss the use of Parametrized Quantum Circuits (PQC) as an additional hidden layer to a previously suggested classical Graph Neural Network model (HEP.TrkX) for particle track reconstruction. The new model, which we call the Quantum Graph Neural Network (QGNN) model, is a hybrid model that can handle graph data and perform edge classification with the help of Quantum Circuits that are compatible with Noisy Intermediate Scale Quantum (NISQ) devices. We further discuss the track reconstruction as an analogous problem with flight trajectory reconstruction in aviation industry and provide a non-HEP application for the QGNN model. The initial model that can perform edge classification on HEP data was presented at the Connecting the Dots workshop.

Workshop / 8

UCluster: Unsupervised clustering for HEP

Author: Vinicius Massami Mikuni¹

¹ *Universitaet Zuerich (CH)*

Corresponding Author: vinicius.mikuni@gmail.com

In this talk I will present an unsupervised clustering (UCluster) method where a neural network is used to reduce the dimensionality of the data, while preserving the event information. The reduced representation is then clustered to a k-means friendly space with a suitable loss function. I will show how this idea can be used to unsupervised multi-class classification and anomaly detection.

Workshop / 9

Teaching Machine Learning with ATLAS Open Data

Authors: Kate Shaw¹; Meirin Oan Evans¹; Thomas James Stevenson¹

¹ *University of Sussex (GB)*

Corresponding Authors: meirin.oan.evans@cern.ch, thomas.james.stevenson@cern.ch, kate.shaw@cern.ch

Open Data are a crucial cornerstone of science. Using Open Data brings benefits such as direct access to cutting edge research, tools to promote public understanding of science and training for scientists of the future. This talk will describe the enormous potential of ATLAS Open Data and how it's used for training, courses and tutorials in machine learning, from undergraduate and postgraduate teaching to workshops and summer schools, using online resources and video tutorials. Applications could be to teach the use of neural networks and even the principles of deep learning.

Workshop / 10**Active Anomaly Detection for time-domain discoveries**

Authors: Emille Eugenia DE OLIVEIRA ISHIDA¹; Matwey Kornilov²; Konstantin Malanchev²; Maria Pruzhinskaya²; Vladimir Korolev³

Co-authors: Alina Volnova⁴; Florian Mondon⁵; Sreevarsha Sreejith⁵; Anastasia Malancheva⁶; Shubhomoy Das⁷

¹ CNRS

² Moscow State University

³ Moscow Institute of Physics and Technology

⁴ Space Research Institute of the Russian Academy of Sciences

⁵ LPC - Clermont

⁶ Cinimex

⁷ Washington State University

Corresponding Author: emille.ishida@clermont.in2p3.fr

We present the first application of adaptive machine learning to the identification of anomalies in a data set of non-periodic time series. The method follows an active learning strategy where highly informative objects are selected to be labelled. This new information is subsequently used to improve the machine learning model, allowing its accuracy to evolve with the addition of human feedback. For the case of anomaly detection, the algorithm aims to maximize the number of real anomalies presented to the expert by slightly modifying the decision boundary of a traditional isolation forest in each iteration. As a proof of concept, we apply the Active Anomaly Discovery (AAD) algorithm to light curves from the Open Supernova Catalog and compare its results to those of a static Isolation Forest (IF). For both methods, we visually inspected objects within 2% highest anomaly scores. We show that AAD was able to identify ~ 80% more true anomalies than IF. This result is the first evidence that AAD algorithms can play a central role in the search for new physics in complex datasets.

Workshop / 11**AutoDQM: A Statistical Tool for Monitoring Data Quality in the CMS Detector**

Author: Vivan Thi Nguyen¹

¹ Northeastern University (US)

Corresponding Author: vivan.nguyen@cern.ch

AutoDQM is an automated monitoring system which implements statistical tests and machine learning (ML) algorithms to compare data runs and flag anomalies for CMS data quality. It is used in conjunction with the existing Data Quality Monitoring (DQM) software to reduce the time and labor required of shifters during collision running by identifying anomalous behavior for further review from system experts. AutoDQM was used during the end of data-taking in Run 2 of the Large Hadron Collider (LHC) and is being expanded for Run 3. The tool is currently being designed to monitor the Level-1 Trigger (L1T) and all four muon sub-detectors: Cathode Strip Chambers (CSC), Drift Tube chambers (DT), Resistive Plate Chambers (RPC), and Gas Electron Multiplier chambers (GEM). To maintain the quality of collision data in future runs of the LHC, where the data rate is expected to increase, a suite of ML techniques is being developed to be used within the AutoDQM tool.

Workshop / 12

Design by intelligent committee: use of machine learning as a scientific advisor**Authors:** Darren Price¹; Stephen Burns Menary²¹ *University of Manchester and The Alan Turing Institute*² *University of Manchester***Corresponding Authors:** stephen.burns.menary@cern.ch, darren.price@cern.ch

Experimental measurements in high energy physics are primarily designed using the expert knowledge and intuition of the analysers, who define their background rejection cuts, control/signal regions and observables of interest based on their understanding of the physical processes involved. More recently, modern multivariate analysis techniques such as neural density estimation and boosted decision trees have allowed analysers to interpret their data in a way which sets highly-optimised limits on specific new physics models. However, in all experiments where data is costly, a high value is placed on model-independent measurements which characterise the data in a way which captures the salient features of the physical processes involved, but without assuming any new physics model in particular.

We demonstrate the use of neural density estimation as an “insight extractor”, capable of advising the analyser about which observables take precedence when constraining benchmark models, and further suggesting optimal selection criteria. The analyser can then design their experimental measurements ensuring that no potentially-sensitive observations are neglected, whilst the measurements themselves remain model-independent. Such an approach leverages the ability of neural networks to capture high dimensional dependencies in both the observable and parameter spaces. As an exploratory example, we consider the design of differential cross section measurements that can be used to constrain models of new phenomena, using data on the electroweak production of Z bosons in association with two jets interpreted using the Standard Model effective field theory (EFT) as a benchmark model. This approach can equally be applied in both EFT frameworks and complete beyond-the-Standard-Model theories to enhance the potential for scientific discovery.

Workshop / 13

Pre-Learning a Geometry Using Machine Learning to Accelerate High Energy Physics Detector Simulations**Authors:** Evangelos Kourlitis¹; Walter Hopkins¹; Doug Benjamin¹¹ *Argonne National Laboratory (US)***Corresponding Authors:** douglas.benjamin@cern.ch, evangelos.kourlitis@cern.ch, walter.hopkins@cern.ch

The simulation of the passage of particles through the LHC detectors occupies already more than a third of the available computing resources and it's predicted to exceed them after 2026, for the example of the ATLAS detector. Significant portion of the most prevalent simulation toolkit, Geant4, is spent to explore the geometry of the detector volume in order to calculate a particle instance fly path. Machine learning algorithms are utilized to learn the geometry beforehand in order to reduce the computational demand while the actual simulation is being produced. A high dimensional map of the geometry is constructed that guides the simulation to explore complex geometries. To achieve this a complete pipeline of data generation and storage, ML training and optimization is employed at Argonne National Laboratory computing facilities. The purpose of the work presented is to demonstrate whether the ML assisted geometry exploration can achieve accelerated simulations compared to pure Geant4 in complex geometries.

Workshop / 14

Hyperparameter Optimisation for Machine Learning using ATLAS Grid and HPC

Authors: Rui Zhang¹; Wen Guan²; Tadashi Maeno³; Fa-Hui Lin⁴; Torre Wenaus³; Alkaid Cheng¹

¹ *University of Wisconsin Madison (US)*

² *University of Wisconsin (US)*

³ *Brookhaven National Laboratory (US)*

⁴ *University of Texas at Arlington (US)*

Corresponding Authors: chi.lung.cheng@cern.ch, fahui.lin@cern.ch, wen.guan@cern.ch, rui.zhang@cern.ch, tmaeno@bnl.gov, wenaus@gmail.com

With the emerging of more and more sophisticated machine learning models in high energy physics, optimising the parameters of the models (hyperparameters) is becoming more and more crucial in order to get the best performance for physics analysis. This requires a lot of computing resources. So far, many of the training results are worked out in a personal computer or a local institution cluster, which prevents people from going to a wider search space or testing more brave ideas. We minimise the obstacle by implementing a hyperparameter optimisation (HPO) infrastructure into the ATLAS Computing Grid. Users submit one task of HPO and the Grid will take care of the optimisation procedure to return the best hyperparameter back to users. I will discuss about using High Performance Computers under this context as well.

Workshop / 16

MLaaS4HEP: Machine Learning as a Service for HEP

Authors: Luca Giommi¹; Valentin Y Kuznetsov²

Co-author: Daniele Bonacorsi³

¹ *Universita e INFN, Bologna (IT)*

² *Cornell University (US)*

³ *University of Bologna*

Corresponding Authors: daniele.bonacorsi@bo.infn.it, luca.giommi@cern.ch, valentin.kuznetsov@cern.ch

Machine Learning is increasingly used in many fields of HEP and will give its contribute in the upcoming High-Luminosity LHC (HL-LHC) program at CERN. The raising of data produced needs new approaches to train and use ML models. In this presentation we discuss the Machine Learning as a Service (MLaaS) infrastructure, that allows to read data directly in the ROOT format exploiting the World-Wide LHC Grid (WLCG) infrastructure for remote data access, and provide pre-trained models via HTTP protocol. In particular, we demonstrate the usage of MLaaS solution for a concrete physics use-case based on $t\bar{t}$ Higgs analysis. We provide some details on this particular use-case and a measure of the performances.

Workshop / 17

SWAN: Powering CERN's Data Analytics and Machine Learning Use cases

Authors: Luca Canali¹; Riccardo Castellotti¹; Prasanth Kothuri¹

¹ CERN**Corresponding Authors:** prasanth.kothuri@cern.ch, riccardo.castellotti@cern.ch, luca.canali@cern.ch

SWAN (Service for Web-based ANALysis) is CERN's general-purpose Jupyter notebook service. It offers a pre-configured, fully-fledged, and easy to use environment, integrating CERN-IT compute, GPU, storage, and analytics services, available at a simple mouse click. In this talk, we will describe the currently deployed SWAN service, as well as recent developments and service improvements that can be useful for data scientists, ML practitioners, and physicists at CERN. In particular, we will cover the use of GPUs for model training, the availability of distributed computing using Spark for data preparation, and recent work to deploy SWAN on public clouds (OCI).

Workshop / 18

Reduced Precision Strategies for Deep Learning: 3DGAN Use Case

Authors: Florian Rehm¹; Sofia Vallecorsa^{None}**Co-authors:** Vikram Saleetore ; Hans Pabst ; Adel Chaibi¹ Hochschule Coburg (DE)**Corresponding Authors:** sofia.vallecorsa@cern.ch, florian.matthias.rehm@cern.ch

Deep learning simulations are known as computational heavy with the need of a lot of memory and bandwidth. A promising approach to make deep learning more efficient and to reduce its hardware workload is to quantize the parameters of the model to lower precision. This approach results in lower execution inference time, lower memory footprint and lower memory bandwidth.

We will research the effects of low precision inference of a deep generative adversarial network [1] model which consists of a convolutional neural network. The use case is for calorimeter detector simulations of subatomic particle interactions in accelerator based high energy physics. We are comparing the inference results of the generated electron showers with the training data for different numerical bit formats and benchmark these in terms of computation and physics accuracy. The model we are quantizing, is a modified 3DGAN [2] prototype based on 2D convolutional layers. With this prototype we gained a factor 3 runtime speed up.

+

[1] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.

[2] G. r. Khattak, S. Vallecorsa, and F. Carminati, "Three dimensional energy parametrized generative adversarial networks for electromagnetic shower simulation," in 2018 25th IEEE International Conference on Image Processing (ICIP), 2018, pp. 3913–3917.

Workshop / 19

Hit-reco: ProtoDUNE denoising with DL models

Author: Marco Rossi^{None}**Corresponding Author:** marco.rossi@cern.ch

We present Hit-reco model for denoising and region of interest selection on raw simulation data from ProtoDUNE experiment. ProtoDUNE detector is hosted by CERN and it aims to test and calibrate technologies for DUNE, a forthcoming experiment in neutrino physics. Hit-reco leverages deep learning algorithms to make the first step in the reconstruction workflow, which consists in converting digital detector signals into physical high level quantities. We benchmark the artificial intelligence based approach against traditional algorithms implemented by the DUNE collaboration.

We investigate the capability of graph convolutional neural networks, while exploiting multi-GPU setups to accelerate training and inference processes.

Workshop / 20

Identifying jets in the Lund plane

Author: Frederic Alexandre Dreyer¹

¹ *University of Oxford*

Corresponding Author: frederic.dreyer@physics.ox.ac.uk

The identification of heavy particles such as top quarks or vector bosons is one of the key issues at the Large Hadron Collider. In this talk, we introduce a novel jet tagging method which relies on graph neural networks and an efficient description of the radiation patterns within a jet to optimally disentangle signatures of boosted objects from background QCD jets. We apply this framework to a number of different benchmarks, showing improved performance for Top tagging compared to existing algorithms. We study the robustness to non-perturbative and detector effects, and show how kinematic cuts in the Lund plane can mitigate overfitting of the neural network. Finally, we compare the scaling of the performance as a function of the computational cost for several methods.

Workshop / 21

A readily-interpretable fully-convolutional autoencoder-like algorithm for unlabelled waveform analysis

Authors: Benjamin Krikler¹; Sam Eriksen²; Henning Flacher¹

¹ *University of Bristol (GB)*

² *University of Bristol*

Corresponding Authors: sam.eriksen@bristol.ac.uk, b.krikler@bristol.ac.uk, henning.flacher@cern.ch

Waveform analysis is a crucial first step in the data processing pipeline for any particle physics experiment. Its accuracy, therefore, can limit the overall analysis performance although waveform analyses often face a variety of challenges, for example: overlapping ‘pile-up’ pulses, noise, non-linearities, floating baselines. Historically, many experiments have viewed template fitting as the optimal way to recover information about the underlying impulses. Here we demonstrate the connection between convolutional methods and template fitting, and adapt these approaches to overcoming typical challenges of waveform analysis.

Having set this foundation, we develop these concepts into a fully-convolutional autoencoder-like algorithm, capable of learning from unlabelled data. Being fully-convolutional this algorithm is capable of handling arbitrary length waveforms, but consequently the encoded space aims not for the dimensionality reduction of a typical autoencoder but rather sparsity in its activation. Moreover, being based on the fundamental understanding developed in the first part of the talk, this model is highly constrained, with each internal convolutional layer offering a straightforward interpretation. We finalise this talk with comparisons of the performance of this algorithm and more traditional approaches under several scenarios and consider some future potential directions.

Workshop / 22

Distributed training of graph neural network at HPC

Author: Xiangyang Ju¹

¹ *Lawrence Berkeley National Lab. (US)*

Corresponding Author: xiangyang.ju@cern.ch

Graph Neural Networks (GNN) are trainable functions that operate on a graph to learn latent graph attributes and to form a parameterized message-passing by which information is propagated across the graph, ultimately learning sophisticated graph attributes. Its application in the High Energy Physics grows rapidly in the past years, ranging from event reconstructions to data analyses, from precision measurements to the search of new physics. The size and complexity of the graphs are also growing. Because graph data structure is irregular and sparse, it imposes non-trivial computational challenges. Currently AI hardwares primarily focus on accelerating dense 1D or 2D arrays, to some extent neglecting sparse and irregular tensor calculations. In this talk, we take the GNN architecture used by the Exa.TrkX collaboration for track reconstruction and the tracking ML challenge dataset as the benchmark in evaluating distributed strategies and Artificial Intelligent (AI) accelerators. We study different AI accelerators that are either in the cloud or at a High Performance Computing center. We also study different distributed training strategies for GNN and the scalabilities of these training strategies on different AI accelerators. Finally, the talk ends with an outlook on deploying GNN for real-time data processing.

Workshop / 23

High Fidelity Simulation of High Granularity Calorimeters with High Speed

Authors: Engin Eren¹; Sascha Daniel Diefenbacher²

¹ *Deutsches Elektronen-Synchrotron DESY*

² *Hamburg University (DE)*

Corresponding Authors: sascha.diefenbacher@desy.de, engin.eren@desy.de

In this talk, we investigate the use of Generative Adversarial Networks (GANs) and a new architecture – the Bounded Information Bottleneck Autoencoder (Bib-AE) – for modeling electromagnetic showers in the central region of the Silicon-Tungsten calorimeter of the proposed International Large Detector. An accurate simulation of differential distributions including for the first time the shape of the minimum-ionizing-particle peak compared to a full GEANT4 simulation for a high-granularity calorimeter with 27k simulated channels have been achieved. Our results further strengthen the case of using generative networks for fast simulation and demonstrate that physically relevant differential distributions can be described with high accuracy. Furthermore, a detailed investigation of the latent space encoded by Bib-AE has been carried out.

Workshop / 24

GANplifying Event Samples

Authors: Anja Butter^{None}; Ben Nachman¹; Gregor Kasieczka²; Sascha Daniel Diefenbacher²; Tilman Plehn³

¹ *Lawrence Berkeley National Lab. (US)*

² *Hamburg University (DE)*

³ *Heidelberg University*

Corresponding Authors: sascha.diefenbacher@desy.de, bpnachman@lbl.gov, tilman.plehn@cern.ch, gregor.kasieczka@cern.ch, butter@thphys.uni-heidelberg.de

Generative machine learning models have been successfully applied to many problems in particle physics, ranging from event generation to fast calorimeter simulation to many more. This indicates that generative models have the potential to become a mainstay in many simulation chains. However, one question that still remains is whether a generative model can have increased statistical precision compared to the data it was trained on. I.e. whether one can meaningfully draw more samples from a generative model than the ones it was trained with. We explore this using three examples and demonstrate that generative models indeed have the capability to amplify data sets.

Based on arxiv.org/abs/2008.06545

Workshop / 25

Set2Graph: Secondary Vertex finding in Jets with Neural Networks

Authors: Eilam Gross¹; Sanmay Ganguly¹; Kyle Stuart Cranmer²; Jonathan Shlomi¹

¹ *Weizmann Institute of Science (IL)*

² *New York University (US)*

Corresponding Authors: kyle.cranmer@cern.ch, sanmay.ganguly@cern.ch, jonathan.shlomi@weizmann.ac.il, eilam.work@gmail.com

(due to slow internet connection : youtube video + recording of Q&A)

Secondary vertex finding is a crucial task for identifying jets containing heavy flavor hadron decays. Bottom jets in particular have a very distinctive topology of $b \rightarrow b \rightarrow b$ decay which gives rise to two secondary vertices with high invariant mass and several associated charged tracks.

Existing secondary vertex finding algorithms search for intersecting particle tracks, and group them into secondary vertices based on geometrical constraints. We propose an algorithm where the vertex finding step is performed with a graph neural network. Tracks are represented as objects in an unordered set, and our proposed model learns a function from this set to a graph that represents the vertex structure in the jet. We prove that models with this structure have maximal expressive power for all (continuous) set to graph functions.

We present performance metrics for evaluating vertex finding performance, and compare the performance of several different graph network architectures on a simulated dataset.

Workshop / 26

Zero-Permutation Jet Parton Assignment

Authors: Seungjin Yang¹; Jason Lee²; Inkyu Park¹; Ian James Watson³

¹ *University of Seoul, Department of Physics (KR)*

² *University of Seoul (KR)*

³ *University of Seoul*

Corresponding Authors: inkyu.park@cern.ch, ian.james.watson@cern.ch, jason.lee@cern.ch, seungjin.yang@cern.ch

For many top quark measurements, it is essential to reconstruct the top quark from its decay products. For example, the top quark pair production process in the all-jets final state has six jets initiated from daughter partons and additional jets from initial/final state radiation. Due to the many possible permutations, it is very hard to assign jets to partons. We use a deep neural network with an attention-based architecture together with a new objective function to the jet-parton assignment

problem. Our novel deep learning model and the physics-inspired objective function enable jet-parton assignment with high-dimensional data while the attention mechanism bypasses the combinatorial explosion that usually leads to intractable computational requirements. The model can also be applied as a classifier to reject the overwhelming QCD background, showing increased performance over standard classification methods.

Workshop / 27

Estimating Support Size of Distribution Learnt by Generative Adversarial Networks for Particle Detector Simulation

Authors: Kristina Jaruskova¹; Sofia Vallecorsa²

¹ *Czech Technical University in Prague*

² *CERN*

Corresponding Authors: sofia.vallecorsa@cern.ch, kristinajaruskova@gmail.com

Generative Adversarial Networks are usually used to generate images similar to the provided training data. The 3DGAN introduced in Khattak et al 2019 has the ability to simulate data from High Energy Physics detectors where each shower is represented by a three dimensional image. To evaluate the results, the generated images were compared to Monte Carlo GEANT4 simulations in terms of physics quantities where a high level of agreement was found. The question is whether the 3DGAN actually learns the target distribution. We use the Jensen-Shannon divergence to compute distances between energy depositions along different axis and we adjust the test introduced in Arora et al 2017 based on the birthday paradox to estimate the support size of the distribution learnt by the 3DGAN.

Workshop / 29

Object condensation: one-stage grid-free multi-object reconstruction in physics detectors, graph, and image data

Author: Jan Kieseler¹

¹ *CERN*

Corresponding Author: jan.kieseler@cern.ch

High-energy physics detectors, images, and point clouds share many similarities in terms of object detection. However, while detecting an unknown number of objects in an image is well established in computer vision, even machine learning assisted object reconstruction algorithms in particle physics almost exclusively predict properties on an object-by-object basis.

Traditional approaches from computer vision either impose implicit constraints on the object size or density and are not well suited for sparse detector data or rely on objects being dense and solid. The object condensation method proposed here is independent of assumptions on object size, sorting or object density, and further generalises to non-image-like data structures, such as graphs and point clouds, which are more suitable to represent detector signals. The pixels or vertices themselves serve as representations of the entire object, and a combination of learnable local clustering in a latent space and confidence assignment allows one to collect condensates of the predicted object properties with a simple algorithm.

The object condensation method is described and results on a simple object classification problem in images and as well as an application to particle flow are presented.

Workshop / 30

Quantum Generative Adversarial Networks**Author:** Su Yeon Chang¹**Co-authors:** Sofia Vallecorsa²; Steven Herbert³; Elias Fernandez-Combarro Alvarez⁴; Federico Carminati²¹ EPFL - Ecole Polytechnique Federale Lausanne (CH)² CERN³ Cambridge Quantum Computing⁴ Universidad de Oviedo (ES)**Corresponding Authors:** federico.carminati@cern.ch, steven.herbert@cambridgequantum.com, sofia.vallecorsa@cern.ch, elias.fernandez-combarro.alvarez@cern.ch, su.yeon.chang@cern.ch

In High Energy Physics (HEP), calorimeter outputs play an essential role in understanding low distance processes occurring during particle collisions. Due to the complexity of underlying physics, the traditional Monte-Carlo simulation is computationally expensive, and thus, the HEP community has suggested Generative Adversarial Networks (GAN) for fast simulation. Meanwhile, it has also been proposed that, in certain circumstances, simulation using GANs can itself be sped-up by using quantum GANs (qGANs).

Our work presents two advanced prototypes of qGAN to reproduce calorimeter outputs interpreted as pixelated images. The first model is called the dual-Parameterized Quantum Circuit (PQC) GAN, which consists of two PQCs sharing the role of a single classical generator. The first PQC learns the probability distribution over the images, while the second generates normalized pixel intensities of an individual image for each PQC input. Its application in HEP demonstrates that the model can reproduce a fixed number of images as well as their probability distribution for a reduced problem size and allows us to scale up to real calorimeter outputs.

On the other hand, the second prototype employs a Continuous Variable (CV) approach, which encodes quantum information in a continuous physical observable. The CV architecture has an advantage that it allows constructing a CV neural network (CVNN) similar to the structure of a classical fully connected layer. We built a simple binary classifier with the CVNNs to discriminate real classical data embedded in quantum states from random fake data. Following the successful results obtained in the CV classifier simulation, CV qGAN models are tested to generate calorimeter outputs with a reduced size, and their limitations are discussed.

Workshop / 31

Accelerating GAN training using distributed tensorflow and highly parallel hardware**Authors:** Renato Paulo Da Costa Cardoso¹; Sofia Vallecorsa²¹ Universidade de Lisboa (PT)² CERN**Corresponding Authors:** sofia.vallecorsa@cern.ch, renato.cardoso@cern.ch**Abstract**

Machine Learning has been used in a wide array of areas and the necessity to make it faster while still maintaining the accuracy and validity of the results is a growing problem for data scientists. This work explores the Tensorflow distributed parallel strategy approach to effectively and efficiently run a Generative Adversarial Network, GAN, model [1] in a parallel environment, as well as benchmarking different types of hardware. More specifically it will use the TensorFlow's Mirrored strategy to

parallelize a 3D GAN on multiple GPUs and use a TPU strategy to run it on Google's TPUs. The present work shows two approaches to the Tensorflow mirrored strategy, one approach uses the simplified method of parallelizing the training, where it is specified what each GPU can see, and using the built-in logic from the Tensorflow strategy it can train the model in parallel, and a second approach where it is used a custom training loop by manually setting the training process, this is by manually getting the loss, updating the gradients, and the weights of the GAN, with this, is it possible to have higher control of the training process as well as add further elements to each GPU work, increasing the overall speedup. For the TPUs we use the TPU distributed strategy present in Tensorflow, applying the same approaches as described for the mirrored strategy. This work is validated by comparing the results obtained by the original 3DGAN model as well as the Monte Carlo simulated data obtained from Geant4. It shows the run times and speed-ups obtained in both types of hardware comparing both approaches.

References

[1] G. R. Khattak, S. Vallecorsa, F. Carminati and G. M. Khan, "Particle Detector Simulation using Generative Adversarial Networks with Domain Related Constraints," 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), Boca Raton, FL, USA, 2019, pp. 28-33, doi: 10.1109/ICMLA.2019.00014.

Workshop / 32

Invertible Networks or Partons to Detector and Back Again

Authors: Marco Bellagente¹; Anja Butter^{None}; Gregor Kasieczka²; Tilman Plehn^{None}; Armand Rousselot^{None}; Ramon Winterhalder¹; Lynton Ardizzone^{None}; Ullrich Köthe^{None}

¹ Universität Heidelberg

² Hamburg University (DE)

Corresponding Authors: winterhalder@thphys.uni-heidelberg.de, md239@uni-heidelberg.de, plehn@uni-heidelberg.de, butter@thphys.uni-heidelberg.de, gregor.kasieczka@cern.ch

For simulations where the forward and the inverse directions have a physics meaning, invertible neural networks are especially useful. A conditional INN can invert a detector simulation in terms of high-level observables, specifically for ZW production at the LHC. It allows for a per-event statistical interpretation. Next, we allow for a variable number of QCD jets. We unfold detector effects and QCD radiation to a pre-defined hard process, again with a per-event probabilistic interpretation over parton-level phase space.

Workshop / 33

DisCo: Robust Networks and automated ABCD background estimation

Authors: Gregor Kasieczka¹; Ben Nachman²; Matthew Schwartz^{None}; David Shih³

¹ Hamburg University (DE)

² Lawrence Berkeley National Lab. (US)

³ Rutgers University

Corresponding Authors: gregor.kasieczka@cern.ch, dshih@physics.rutgers.edu, bpnachman@lbl.gov, schwartz@physics.harvard.edu

With the wide use of deep learning in HEP analyses, answering questions beyond the classification performance becomes increasingly important. One crucial aspect is ensuring the robustness of classifier outputs against other observables - typically an invariant mass. Superior performance in

decorrelation was so far achieved by adversarial training. We show that a simple additive term in the loss function based on a differentiable measure for independence termed distance correlation (DisCo) can achieve state-of-the-art performance while being much simpler to train. A key experimental application that relies on independent observables is the ABCD method for background estimation. We show that DisCo can be used to automatically construct a pair of powerful and independent classifiers that significantly improve performance in terms of ABCD closure, background rejection, and signal contamination.

Based on 2001.05310 and 2007.14400

Workshop / 34

Lorentz Equivariant Neural Networks for Particle Physics

Authors: Alexander Bogatskiy¹; Jan Tuzlic Offermann²; David Miller²; Marwah Roussi¹; Risi Kondor³

¹ *University of Chicago*

² *University of Chicago (US)*

³ *Flatiron Institute, University of Chicago*

Corresponding Authors: jan.offermann@cern.ch, mroussi@uchicago.edu, risi@cs.uchicago.edu, bogatsky@uchicago.edu, david.w.miller@cern.ch

We present a new set of neural network architectures, Lorentz group covariant architectures for learning the kinematics and properties of complex systems of particles. The novel design of this network, called LGN (Lorentz Group Network), implements activations as vectors that transform according to arbitrary finite-dimensional representations of the underlying symmetry group that governs particle physics, the Lorentz group. The nonlinearity of the network is based on the tensor product of representations of the Lorentz group. Consequently, the architecture is inherently covariant under Lorentz transformations and is capable of learning not only fully Lorentz-invariant objectives such as classification probabilities, but also Lorentz-covariant vector-valued objectives such as 4-momenta, while exactly respecting the action of the group. Imposing the symmetry leads to a significantly smaller ansatz (fewer learnable parameters than competing non-covariant networks), and potentially a much more interpretable model. To demonstrate the capability and performance of this network, we study the ability to classify systems of charged and neutral particles at the Large Hadron Collider resulting from the production and decay of highly energetic quarks and gluons. Specifically, we choose the benchmark task of classifying and discriminating jets formed from the hadronic decays of Lorentz-boosted massive particles from the background of light quarks and gluon jets. We show that we are able to achieve similar performance compared to other state-of-the-art neural networks trained to perform this classification task while also maintaining significantly broader generality regarding the structural origin of the physical processes involved. Moreover, we present simplified invariant and covariant architectures tailored to specific tasks with 4-vector inputs, which can be trained faster and more efficiently than the general LGN architecture.

Workshop / 35

Improving particle-flow with deep learning

Author: Sanmay Ganguly¹

Co-authors: Jonathan Shlomi¹; Eilam Gross¹; Marumi Kado²; Francesco Armando Di Bello²

¹ *Weizmann Institute of Science (IL)*

² *Sapienza Universita e INFN, Roma I (IT)*

Corresponding Authors: eilam.work@gmail.com, jonathan.shlomi@weizmann.ac.il, marumi.kado@cern.ch, sanmay.ganguly@cern.ch, francesco.armando.di.bello@cern.ch

Canonical particle flow algorithm tries to estimate neutral energy deposition in calorimeter by first performing matching between calorimeter deposits and track direction and subsequently subtracting the track momenta from the matched cluster energy deposition.

We propose a Deep Learning based method for estimating the energy fraction of individual components for each cell of the calorimeter.

We build the dataset by a toy detector (with different resolutions per calorimeter layer) using GEANT and apply image-based deep neural network models to regress the fraction of neutral energy per cell of the

detector. A comparison of the performance of several different models is carried out.

Workshop / 36

Super-resolution for calorimetry

Authors: Francesco Armando Di Bello¹; Jonathan Shlomi²

Co-authors: Lorenzo Santi¹; Marumi Kado¹; Eilam Gross²; Sanmay Ganguly²

¹ *Sapienza Universita e INFN, Roma I (IT)*

² *Weizmann Institute of Science (IL)*

Corresponding Authors: eilam.work@gmail.com, l.santi@cern.ch, francesco.armando.di.bello@cern.ch, marumi.kado@cern.ch, jonathan.shlomi@weizmann.ac.il, sanmay.ganguly@cern.ch

Super-resolution algorithms are commonly used to enhance the granularity of an imaging system beyond what can be achieved using the measuring device.

We show the first application of super-resolution algorithms using deep learning-based methods for calorimeter reconstruction using a simplified geometry consisting of overlapping showers originated by charged and neutral pions events.

The task of the presented ML algorithms is to estimate the fraction of charged and neutral energy components for each cell of the *super-calorimeter*, which represents the reconstructed calorimeter system whose granularity is up-scaled up to a factor of 4 compared to the original one. We show how the finer granularity can be used to unveil effects that would remain otherwise elusive, such as the reconstructed mass of the π^0 which is strictly connected to an unbiased estimation of the opening angle between the two photons. The performance is evaluated using several ML algorithms, including graph- and convolutional-neural networks.

Workshop / 37

Efficiency parametrization with Neural Networks

Authors: Jonathan Shlomi¹; Francesco Armando Di Bello²; Sanmay Ganguly¹; Eilam Gross¹; Marumi Kado²; Valerio Ippolito³; Guglielmo Frattari²; Chiara Badiali^{None}; Nilotpal Kakati¹

¹ *Weizmann Institute of Science (IL)*

² *Sapienza Universita e INFN, Roma I (IT)*

³ *INFN Sezione di Roma (IT)*

Corresponding Authors: nilotpal.kakati@cern.ch, jonathan.shlomi@weizmann.ac.il, francesco.armando.di.bello@cern.ch, sanmay.ganguly@cern.ch, eilam.work@gmail.com, guglielmo.frattari@cern.ch, chiara.badiali@cern.ch, valerio.ippolito@cern.ch, marumi.kado@cern.ch

An overarching issue of LHC experiments is the necessity to produce massive numbers of simulated collision events in very restricted regions of phase space. A commonly used approach to tackle the problem is the use of event weighting techniques where the selection cuts are replaced by event

weights constructed from efficiency parametrizations. These techniques are however limited by the underlying dependencies of these parametrizations which are typically not fully known and thus only partially exploited.

We propose a neural network approach to learn multidimensional ratios of local densities to estimate in an optimal fashion the efficiency. Graph neural network techniques are used to account for the high dimensional correlations between different physics objects in the event. We show in a specific toy model how this method is applicable to produce accurate efficiency maps for heavy flavor tagging classifiers in HEP experiments, including for processes on which it was not trained.

The work is based on: <https://arxiv.org/abs/2004.02665>

Workshop / 39

Graph Neural Network-based Event Classification for Measurement of the Higgs-Top Yukawa Interaction

Authors: Haichen Wang¹; Xiangyang Ju¹; Ryan Roberts¹; Shuo Han¹; Allison Lauren Xu¹; Pamela Pajarillo¹

¹ *Lawrence Berkeley National Lab. (US)*

Corresponding Authors: shuo.han@cern.ch, allisonxu@lbl.gov, pamela.pajarillo@cern.ch, benjamin.ryan.roberts@cern.ch, xiangyang.ju@cern.ch, haichen.wang@cern.ch

The measurement of the associated production of Higgs boson with a top-quark pair (ttH) at the LHC provides a direct determination of the Higgs-Top Yukawa interaction. The presence of a large number of objects in the final state makes the measurement very challenging. Multivariate Analysis methods such as Boosted Decision Trees (BDT) were used to enhance the analysis sensitivity. However, the sensitivity gain largely depends on physics-inspired input discriminating variables, which in practice require a significant amount of time to engineer carefully. In this presentation, we separate ttH signal from its main background using a Graph Neural Network (GNN), in which collision data are represented by graphs of nodes and edges. The graph representation and message passing of the GNN make the exploitation of relational information more effective between final state particles. We will report results using simulated events and compare GNN-based and BDT-based event classifiers. We will also examine how well the GNN model captures relational information in the event, which is challenging to represent in conventional BDT-based models.

Workshop / 40

Generative Adversarial Network for Identifying the Dark Matter Distribution of a Dwarf Spheroidal Galaxy

Authors: Sung Hak Lim¹; Mihoko Nojiri²; Shigeki Matsumoto³; Kohei Hayashi⁴; Shunichi Horigome³

¹ *Rutgers University*

² *Theory Center, IPNS, KEK*

³ *Kavli IPMU*

⁴ *Kavli Institute for the Physics and Mathematics of the Universe, The University of Tokyo*

Corresponding Authors: shunichi.horigome@ipmu.jp, hayakou.astro@gmail.com, shigeki.matsumoto@ipmu.jp, sunghak.lim@kek.jp, nojiri@post.kek.jp

We introduce a generative adversarial network for analyzing the dark matter distribution of a dwarf spheroidal galaxy.

The mock data generator for dwarf spheroidal galaxies in the spherically symmetric case has three

functional parameters: the number density of stars, the density of dark matter, and velocity anisotropy. The generator will be adversarially trained on a mock dataset, which contains only the line-of-sight information, to identify the dataset's unknown dark matter distribution under given velocity anisotropy.

We will explain how we implement specialized classifiers, generators cooperating with the spherical Jeans equation, and regularizers to avoid less physical solutions.

Workshop / 42

Graph Convolutional Operators in the the PyTorch JIT

Author: Lindsey Gray¹

Co-authors: Thomas Klijnsma¹; Matthias Fey²

¹ *Fermi National Accelerator Lab. (US)*

² *TU Dortmund*

Corresponding Authors: matthias.fey@tu-dortmund.de, lindsey.gray@cern.ch, thomas.klijnsma@cern.ch

The PyTorch just-in-time (jit) compiler is a powerful tool for optimizing and serializing neural network models. However, its range is limited by the subset of the python language that it is restricted to and the number of tensor operations implemented in C++. These limitations were a major blocker to using graph neural networks implemented in the geometric deep learning (GDL) library PyTorch Geometric (PyG) at scale. In particular, models being researched needed to be re-implemented, validated, and retrained before they could be deployed in inference as a service (IaaS) frameworks for wider use. To solve this, the PyG framework was extended to include an automatic analysis of user-defined convolutional operators that renders structurally identical, concrete, and jit-compatible versions of the operator. The results of these additions are GDL models that can seamlessly flow from research-focused workflows to development and deployment at scale with standard IaaS infrastructure. In this presentation we will discuss the additions made to PyG to achieve this new functionality and give examples with preliminary performance estimates of live models in the context of high energy particle physics.

Workshop / 43

Selective background MC simulation with graph neural networks at Belle II

Authors: Nikolai Hartmann¹; James Kahn²; Yannick Bross¹; Kilian Lieret³; Andreas Maximilian Lindner⁴; Thomas Kuhr¹

¹ *Ludwig Maximilians Universitat (DE)*

² *Karlsruhe Institute of Technology (KIT)*

³ *Ludwig Maximilian University Munich*

⁴ *University of Munich (LMU)*

Corresponding Authors: nikolai.marcel.hartmann@cern.ch, james.kahn@kit.edu, and.lindner@campus.lmu.de, kilian.lieret@posteo.de, thomas.kuhr@lmu.de

Searching for rare physics processes requires a good understanding of the backgrounds involved. This often requires large amounts of simulated data that are computationally expensive to produce. The Belle II collaboration is planning to collect 50 times the amount of data of its predecessor Belle. With the increase in data volume the necessary volume of simulated data increases as well. Due to aggressive event selections that enrich the signal processes of

interest, much of the simulated data is thrown away. This talk presents a method for predicting which events will be thrown away already after the computationally less expensive event generation step. This is achieved using graph neural networks applied to the simulated event decay tree. Only events selected by the neural network are passed to the resource intense detector simulation and the reconstruction step. False negatives from this selection can lead to biases in the distributions of observables for filtered events. Possible ways to mitigate this are also discussed.

Workshop / 44

General recipe to form input space for deep learning analysis of HEP scattering processes.

Authors: Lev Dudko¹; Georgii Vorotnikov¹; Petr Volkov¹; Maksim Perfilov¹

¹ *M.V. Lomonosov Moscow State University (RU)*

Corresponding Authors: lev.doudko@cern.ch, petr.volkov@cern.ch, georgii.vorotnikov@cern.ch, maksim.perfilov@cern.ch

The important step of the analysis of HEP scattering processes is the optimization of the input space for multivariate technique. We propose general recipe how to form the set of low-level observables which are sensitive to the differences in hard scattering processes at the colliders. It will be demonstrated that without any sophisticated analysis of the kinematic properties one can achieve close to optimal performance of DNN with the proposed general set of low-level observables. The proposed approach is already described in [Int.J.Mod.Phys.A 35 (2020) 21, 2050119, hep-ph:2002.09350].

Workshop / 45

Foundations of a Fast, Data-Driven, Machine-Learned Simulator

Author: Jessica N. Howard¹

Co-authors: Stephan Mandt²; Daniel Whiteson¹; Yibo Yang²

¹ *Department of Physics & Astronomy, UC Irvine*

² *Department of Computer Science, UC Irvine*

Corresponding Authors: jessica.nicole.howard@cern.ch, yibo.yang@uci.edu, daniel@uci.edu, mandt@uci.edu, jnhoward@uci.edu

We introduce a novel strategy for machine-learning-based predictive simulators, which can be trained in an unsupervised manner using observed data samples to learn a predictive model of the detector response and other difficult-to-model transformations. Particle physics detectors cannot directly probe fundamental particle collisions. Instead, statistical inference must be used to surmise information about the parameters of the latent theory space, ultimately determining the validity of the theory. High-fidelity simulations, which imitate detector response, are a crucial part of this process. However, these computationally intensive simulations have become a major bottleneck in the pursuit of discovery. Previous machine learning based solutions to this problem seek to replicate data for a fixed theory space, but do not attempt to learn a general mapping from the latent theory space to detected data. Such models will only be able to augment current simulations, limiting their scope and utility. Using Optimal Transport based machine learning techniques, we propose a method for a data-driven, physically meaningful, machine learned simulation which lays the framework for ultimately replacing the current computationally expensive simulations in particle physics.

Workshop / 46

Pixel Detector Background Generation using Generative Adversarial Networks at Belle II

Authors: Matej Srebre¹; Thomas Kuhr¹; Martin Ritter¹; Hosein Hashemi¹

¹ LMU

Corresponding Authors: gh.hashemi@physik.uni-muenchen.de, thomas.kuhr@lmu.de, martin.ritter@lmu.de, matej.srebre@physik.uni-muenchen.de

The pixel detector (PXD) is an essential part of the Belle II detector recording particle positions. Data from the PXD and other sensors allow us to reconstruct particle tracks and decay vertices. The effect of background noise on track reconstruction for measured data is emulated for simulated data by a mixture of measured background noise and easily-simulated particle decays. This model requires a large set of statistically independent PXD background noise samples in order to avoid the systematic bias of reconstructed tracks. However, data from the fine-grained PXD requires a substantial amount of storage. As an efficient way of producing background noise, we explore the idea of an on-demand PXD background generator using Generative Adversarial Networks (GANs).

Workshop / 47

Generative models for calorimeters response simulation - from GANs through VAE to e2e SAE

Authors: Kamil Rafal Deja¹; Jan Michal Dubinski¹; Piotr Wladyslaw Nowak¹; Sandro Christian Wenzel²; Tomasz Piotr Trzcinski¹

¹ Warsaw University of Technology (PL)

² CERN

Corresponding Authors: tomasz.piotr.trzcinski@cern.ch, sandro.wenzel@cern.ch, piotr.wladyslaw.nowak@cern.ch, jan.michal.dubinski@cern.ch

Simulating detectors response is a crucial task in HEP experiments. Currently employed methods, such as Monte Carlo algorithms, provide high-fidelity results at a price of high computational cost, especially for dense detectors such as ZDC calorimeter in ALICE experiment. Multiple attempts are taken to reduce this burden, e.g. using generative approaches based on Generative Adversarial Networks or Variational Autoencoders. In this talk we will present adaptation of those state-of-the-art methods for calorimeters response simulations. Although GANs and VAE are much faster, they are often unstable in training and do not allow sampling from an entire data distribution. To address these shortcomings, we introduce a novel generative model dubbed end-to-end sinkhorn autoencoder that leverages sinkhorn algorithm to explicitly align distribution of encoded real data examples and generated noise. More precisely, we extend autoencoder architecture by adding a deterministic neural network trained to map noise from a known distribution onto autoencoder latent space representing data distribution. Our method outperforms competing approaches on the challenging dataset of simulation data from Zero Degree Calorimeters of ALICE experiment in LHC, as well as standard benchmarks, such as MNIST and CelebA.

Workshop / 49

Model-Independent Detection of New Physics Signals Using Interpretable Semi-Supervised Classifier Tests

Authors: Purvasha Chakravarti¹; Mikael Kuusela²; Larry Wasserman³; Jing Lei³

¹ *Imperial College London*

² *Carnegie Mellon University (US)*

³ *Carnegie Mellon University*

Corresponding Authors: larrywasserman.cool@gmail.com, jinglei@andrew.cmu.edu, mikael.kuusela@cern.ch, p.chakravarti@imperial.ac.uk

A central goal in experimental high energy physics is to detect new physics signals that are not explained by known physics. In this work, we aim to search for new signals that appear as deviations from known Standard Model physics in high-dimensional particle physics data. To do this, we determine whether there is any statistically significant difference between the distribution of Standard Model background samples and the distribution of the experimental observations, which are a mixture of the background and a potential new signal. Traditionally, one also assumes access to a sample from a model for the hypothesized signal distribution. Here we instead investigate a model-independent method that does not make any assumptions about the signal and uses a semi-supervised classifier to detect the presence of the signal in the experimental data. We construct two test statistics using the classifier: an estimated likelihood ratio test statistic and a test based on the area under the ROC curve (AUC). Additionally, we propose a method for estimating the signal strength parameter and explore active subspace methods to interpret the proposed semi-supervised classifier in order to understand the properties of the detected signal. We investigate the performance of the methods on a data set related to the search for the Higgs boson at the Large Hadron Collider at CERN. We demonstrate that the semi-supervised tests have power comparable to the classical methods for a well-specified signal, but much higher power for an unexpected signal which might be entirely missed by the supervised tests.

Workshop / 50

Disentangling Boosted Higgs Boson Production Modes with Machine Learning

Authors: Yi-Lun Chung¹; Shih-Chieh Hsu²; Benjamin Nachman³

¹ *National Tsing Hua University (TW)*

² *University of Washington Seattle (US)*

³ *Lawrence Berkeley National Laboratory*

Corresponding Authors: bpnachman@lbl.gov, s107022801@m107.nthu.edu.tw, schsu@uw.edu

Higgs Bosons produced via gluon-gluon fusion (ggF) with large transverse momentum (p_T) are sensitive probes of physics beyond the Standard Model. However, high p_T Higgs Boson production is contaminated by a diversity of production modes other than ggF: vector boson fusion, production of a Higgs boson in association with a vector boson, and production of a Higgs boson with a top-quark pair. Combining jet substructure and event information with modern machine learning, we demonstrate the ability to focus on particular production modes. These tools hold great discovery potential for boosted Higgs bosons produced via ggF and may also provide additional information about the Higgs Boson sector of the Standard Model in extreme phase space regions for other production modes as well.

Workshop / 52

Anomaly Awareness for BSM Searches at the LHC

Author: Charanjit Kaur Khosa^{None}

Co-authors: Veronica Sanz ; Michael Soughton

Corresponding Authors: khosacharanjit@gmail.com, veronica.hirn@gmail.com

In this talk we present a new algorithm called 'Anomaly Awareness'(AA) to search for physics beyond the standard model (BSM). By making the algorithm aware of the presence of a range of different anomalies, we improve its capability to detect anomalous events, even those it had not been exposed to. As an example, we apply this method to a boosted jet topology for BSM searches at LHC and use it to uncover new resonances or EFT effects (based on arXiv:2007.14462 [cs.LG]). We will discuss AA implementation using CNNs and VAEs.

Workshop / 53

Adaptive divergence for rapid adversarial optimization & (1 + epsilon)-class Classification: an Anomaly Detection Method for Highly Imbalanced or Incomplete Data Sets

Authors: Maxim Borisyak¹; Tatiana Gaintseva¹; Andrey Ustyuzhanin¹

¹ *Yandex School of Data Analysis (RU)*

Corresponding Authors: andrey.ustyuzhanin@cern.ch, tatiana.gaintseva@cern.ch

This talk contains 2 contributions:

1) Adaptive divergence for rapid adversarial optimization.

Adversarial Optimization provides a reliable, practical way to match two implicitly defined distributions, one of which is typically represented by a sample of real data, and the other is represented by a parameterized generator. Matching of the distributions is achieved by minimizing a divergence between these distributions, and estimation of the divergence involves a secondary optimization task, which, typically, requires training a model to discriminate between these distributions. The choice of the model has its trade-off: high-capacity models provide good estimations of the divergence, but, generally, require large sample sizes to be properly trained. In contrast, low-capacity models tend to require fewer samples for training; however, they might provide biased estimations. Computational costs of Adversarial Optimization becomes significant when sampling from the generator is expensive. One of the practical examples of such settings is fine-tuning parameters of complex computer simulations. In this work, we introduce a novel family of divergences that enables faster optimization convergence measured by the number of samples drawn from the generator. The variation of the underlying discriminator model capacity during optimization leads to a significant speed-up. The proposed divergence family suggests using low-capacity models to compare distant distributions (typically, at early optimization steps), and the capacity gradually grows as the distributions become closer to each other. Thus, it allows for a significant acceleration of the initial stages of optimization. This acceleration was demonstrated on two fine-tuning problems involving Pythia event generator and two of the most popular black-box optimization algorithms: Bayesian Optimization and Variational Optimization. Experiments show that, given the same budget, adaptive divergences yield results up to an order of magnitude closer to the optimum than Jensen-Shannon divergence. While we consider physics-related simulations, adaptive divergences can be applied to any stochastic simulation.

2) (1 + epsilon)-class Classification: an Anomaly Detection Method for Highly Imbalanced or Incomplete Data Sets

Anomaly detection is not an easy problem since distribution of anomalous samples is unknown a priori. We explore a novel method that gives a trade-off possibility between one-class and two-class approaches, and leads to a better performance on anomaly detection problems with small or non-representative anomalous samples. The method is evaluated using several data sets and compared to a set of conventional one-class and two-class approaches.

Workshop / 55

Using an Optical Processing Unit for tracking and calorimetry at the LHC

Authors: David Rousseau¹; Aishik Ghosh²; Laurent Basara³; Biswajit Biswas⁴

¹ *IJCLab-Orsay*

² *Université Paris-Saclay (FR)*

³ *LAL/LRI, Université Paris Saclay*

⁴ *Centre National de la Recherche Scientifique (FR)*

Corresponding Authors: rousseau@lal.in2p3.fr, biswajit.biswas@cern.ch, laurent.basara@cern.ch, aishik.ghosh@cern.ch

Experiments at HL-LHC and beyond will have ever higher read-out rate. It is then essential to explore new hardware paradigms for large scale computations. In this work we consider the Optical Processing Units (OPU) from LightOn, which compute random matrix multiplications on large datasets in an analog, fast and economic way, fostering faster machine learning results on a dataset of reduced dimension. We consider two case studies.

1) “Event classification”: high energy proton collision at the Large Hadron Collider have been simulated, each collision being recorded as an image representing the energy flux in the detector. The task is to train a classifier to separate a Susy signal from the background. The OPU allows fast end-to-end classification without building intermediate objects (like jets). This technique is presented, compared with more classical particle physics approaches.

2) “Tracking”: high energy proton collisions at the LHC yield billions of records with typically 100,000 3D points corresponding to the trajectory of 10.000 particles. Using two datasets from previous tracking challenges, we investigate the OPU potential to solve similar or related problems in high-energy physics, in terms of dimensionality reduction, data representation, and preliminary results.

Workshop / 56

Deep learning solutions for 2D calorimetric cluster reconstruction at LHCb

Authors: Blaise Raheem Delaney¹; Joao Coelho²; Michal Mazurek³

¹ *University of Cambridge (GB)*

² *Université Paris-Saclay (FR)*

³ *National Centre for Nuclear Research (PL)*

Corresponding Authors: blaise.delaney@cern.ch, michal.kazimierz.mazurek@cern.ch, coelho@lal.in2p3.fr

Calorimetric cluster reconstruction can be performed using deep learning solutions from real-time computer vision by casting the detector readout as a two-dimensional image. The increased luminosity expected of Run III poses unprecedented challenges to shower reconstruction at LHCb. This work seeks to perform shower identification and energy regression under such conditions through both convolutional neural networks (CNNs) and graph neural networks (GNNs). To this end, we designed a CNN-based network inspired by the You Only Look Once (YOLO) architecture, capable of regressing bounding boxes and the energy associated with shower deposits. The hybrid granularity of the calorimeter modules, however, requires regularization of the pixel grid through up-sampling and breaks the translational invariance assumed by CNNs. The second approach investigated in this work addresses this issue by leveraging the ability demonstrated by GNNs to learn arbitrary detector geometries without image preprocessing. In this talk, both algorithms are validated by employing a simulated dataset loosely inspired by the LHCb electromagnetic calorimeter. Finally, a set of preliminary results using the LHCb Run III simulations is presented.

Workshop / 57

Simulation-Assisted Decorrelation for Resonant Anomaly Detection

Author: Kees Christian Benkendorfer¹

Co-authors: Ben Nachman¹; Luc Tomas Le Pottier¹

¹ *Lawrence Berkeley National Lab. (US)*

Corresponding Authors: bpnachman@lbl.gov, kees.christian.benkendorfer@cern.ch, luc.tomas.le.pottier@cern.ch

A growing number of weak- and unsupervised machine learning approaches to anomaly detection are being proposed to significantly extend the search program at the Large Hadron Collider and elsewhere. One of the prototypical examples for these methods is the search for resonant new physics, where a bump hunt can be performed in an invariant mass spectrum. A significant challenge to methods that rely entirely on data is that they are susceptible to sculpting artificial bumps from the dependence of the machine learning classifier on the invariant mass. We explore two solutions to this challenge by minimally incorporating simulation into the learning. In particular, we study the robustness of Simulation Assisted Likelihood-free Anomaly Detection (SALAD) to correlations between the classifier and the invariant mass. Next, we propose a new approach that only uses the simulation for decorrelation but the Classification without Labels (CWoLa) approach for achieving signal sensitivity. Both methods are compared using a full background fit analysis on simulated data from the LHC Olympics and are robust to correlations in the data.

Workshop / 58

Decorrelation via Disentanglement

Author: Justin Tan^{None}

Co-author: Phillip Urquijo¹

¹ *University of Melbourne*

Corresponding Author: justin.tan@coepp.org.au

Abstract

Invariance of learned representations of neural networks against certain sensitive attributes of the input data is a desirable trait in many modern-day applications of machine learning, such as precision measurements in experimental high-energy physics. We propose to use the ability of variational autoencoders to learn a disentangled latent representation to achieve the desired invariance condition. The resulting latent representation may be used for arbitrary downstream tasks without exploitation of the protected variable. We demonstrate the effectiveness of the proposed technique on a representative study of rare B decays at the Belle II experiment.

Introduction

In searches for new physics in high-energy physics, experimental analyses are primarily concerned with physical processes which are rare or hypothesized. To claim a statistically significant discovery or exclusion of new physics when studying such decays, it is necessary to maintain an appropriate signal to noise ratio. However, the naïve application of standard classification methods is liable to raise poorly understood systematic effects and ultimately degrade the significance of the final measurement.

To understand the origin of these systematic effects, we note that there are certain protected variables in experimental analyses which should remain unbiased by the analysis procedure. Variables

used to parameterize proposed models of new physics and variables used to model background contributions to the total measured event yield fall into this category. Systems responsible for separating signal from background events achieve this by sampling events with signal-like characteristics from all candidate events. If this procedure introduces sampling bias into the distribution of protected variables, this introduces systematic effects into the analysis which are difficult to characterize. Ultimately, we would like to build a classifier that makes decisions independently of certain physically important observables - such that the original distribution of these observables is preserved for any subsample of the data. This problem is commonly referred to as classifier “decorrelation” with respect to the observables of interest.

We address this task as an optimization problem of finding a representation of the observed data that is invariant to the given protected quantities. This representation should satisfy two competing criteria. Firstly, it should contain all relevant information about the data so that it may be used as a proxy for arbitrary downstream tasks, such as inference of unobserved quantities or prediction of target variables. Secondly, it should not be informative of the given protected quantities, so that downstream tasks are not influenced by these variables. If the protected quantities to be censored from the intermediate representation contain information that can improve the performance of the downstream task, it is likely that removing this information will adversely affect this task. The challenge lies in balancing both objectives without significantly compromising either requirement.

This work approaches the problem from a latent variable model perspective, in which additional unobserved variables are introduced which explain the interaction between different attributes of the observed data. These latent variables can be interpreted as a more fundamental, lower-dimensional representation of the original high-dimensional unstructured data. By appropriately constraining the structure of this latent space, we demonstrate we can isolate the influence of the protected variables into a latent subspace. This allows downstream tasks to only access a relevant subset of the learned representation without being influenced by protected attributes of the original data.

Workshop / 59

Enhancing searches for resonances with machine learning and moment decomposition

Authors: Ouail Kitouni¹; Ben Nachman²; J Michael Williams¹; Constantin Niko Weisser¹

¹ *Massachusetts Inst. of Technology (US)*

² *Lawrence Berkeley National Lab. (US)*

Corresponding Authors: mwill@mit.edu, weisser@cern.ch, ouail.kitouni@cern.ch, bpnachman@lbl.gov

A key challenge in searches for resonant new physics is that classifiers trained to enhance potential signals must not induce localized structures. Such structures could result in a false signal when the background is estimated from data using sideband methods. A variety of techniques have been developed to construct classifiers which are independent from the resonant feature (often a mass). Such strategies are sufficient to avoid localized structures, but are not necessary. We develop a new set of tools using a novel moment loss function (Moment Decomposition or Mode) which relax the assumption of independence without creating structures in the background. By allowing classifiers to be more flexible, we enhance the sensitivity to new physics without compromising the fidelity of the background estimation.

Workshop / 62

Black-Box Optimization with Local Generative Surrogates

Authors: Vladislav Belavin¹; Sergey Shirobokov²; Michael Aaron Kagan³; Andrey Ustyuzhanin¹; Atılım Güneş Baydin⁴

¹ *Yandex School of Data Analysis (RU)*² *Imperial College (GB)*³ *SLAC National Accelerator Laboratory (US)*⁴ *University of Oxford*

Corresponding Authors: atilingunes.baydin@gmail.com, sergey.shirobokov@cern.ch, vladislav.belavin@cern.ch, andrey.ustuzhanin@cern.ch, michael.aaron.kagan@cern.ch

We propose a novel method for gradient-based optimization of black-box simulators using differentiable local surrogate models. In fields such as physics and engineering, many processes are modeled with non-differentiable simulators with intractable likelihoods. Optimization of these forward models is particularly challenging, especially when the simulator is stochastic. To address such cases, we introduce the use of deep generative models to iteratively approximate the simulator in local neighborhoods of the parameter space. We demonstrate that these local surrogates can be used to approximate the gradient of the simulator, and thus enable gradient-based optimization of simulator parameters. In cases where the dependence of the simulator on the parameter space is constrained to a low dimensional submanifold, we observe that our method attains minima faster than baseline methods, including Bayesian optimization, numerical optimization, and approaches using score function gradient estimators.

Workshop / 63

Fast simulation of Time Projection Chamber response at MPD using GANs

Authors: Artem Maevskiy¹; Fedor Ratnikov²

Co-author: Alexander Zinchenko³

¹ *National Research University Higher School of Economics (RU)*² *Yandex School of Data Analysis (RU)*³ *Joint Institute for Nuclear Research (RU)*

Corresponding Authors: alexander.zinchenko@jinr.ru, fedor.ratnikov@gmail.com, artem.maevskiy@cern.ch

NICA accelerator complex is currently being assembled in JINR (Dubna) to perform studies of heavy-ion collisions and explore new regions of the QCD phase diagram. Located at one of the two interaction points of the facility, the Multi-Purpose Detector (MPD) will utilize the Time-Projection Chamber (TPC) as the main tracker of the detector's central barrel. TPC consists of a gas-filled detection volume in a uniform electric field with a 2D position-sensitive electron collection system. Combining the 2D position information with the drift time information, TPC allows to reconstruct the 3D coordinate of the original electron clusters and hence measure the charged particle's trajectory.

Accurately simulating TPC is computationally heavy. A typical heavy-ion collision event is expected to take about 25 seconds to simulate with the available resources. In this work we propose a fast-simulation model based on a Generative Adversarial Network (GAN) to generate raw TPC signals. Preliminary studies show that our model can produce high-fidelity results in under one second per collision.

Workshop / 64

Domain Adaptation Techniques in Particle Identification for the ALICE experiment

Authors: Michal Kurzynka¹; Kamil Rafal Deja¹; Georgy Kornakov²; Tomasz Piotr Trzcinski¹

¹ *Warsaw University of Technology (PL)*

² *TU Darmstadt*

Corresponding Authors: michal.kurzynka@cern.ch, tomasz.piotr.trzcinski@cern.ch, g.kornakov@gsi.de, kamil.rafal.deja@cern.ch

Classifying particle types on the basis of detectors response is a fundamental task in the ALICE experiment. Methods currently employed in this job are based on linear classifiers which are built on Monte Carlo simulation data, due to lack of labels (pdg code) in case of production data and require manual fine tuning to match latter data set distribution. This calibration is performed by highly experienced high energy physicists, often conducted as an iterative process which is time consuming. In this work, we present a proof-of-concept solution for Particle Identification (PID). The main component of this solution is a Classifier with Domain Adaptation model based on Domain Adversarial Neural Networks (DANN). Proposed model utilizes both Monte Carlo-generated and production data during training process despite the lack of labels in case of the latter data set. Such approach allows model to find such complex attributes in the common latent space, which mitigate domain shift between two data sources. Therefore, when training the model to perform classification of particles using simulation data, we do it on the basis of attributes which are valid also in real experimental data.

The main advantage of proposed model is improved classification quality on the production data, despite lack of manual calibration.

Workshop / 65

Using Machine Learning to Speed Up and Improve Detector R&D

Authors: Fedor Ratnikov¹; Denis Derkach²; Alexey Boldyrev³; Andrey Shevelev¹; Pavel Fakanov⁴

¹ *Yandex School of Data Analysis (RU)*

² *National Research University Higher School of Economics (RU)*

³ *NRU Higher School of Economics (Moscow, Russia)*

⁴ *NRU Higher School of Economics*

Corresponding Authors: andrey.shevelev@cern.ch, pavel.fakanov@gmail.com, alexey.boldyrev@cern.ch, fedor.ratnikov@gmail.com, denis.derkach@cern.ch

Design of new experiments, as well as upgrade of ongoing ones, is a continuous process in the experimental high energy physics.

Frontier R&Ds are used to squeeze the maximum physics performance using cutting edge detector technologies.

The evaluation of physics performance for a particular configuration includes sketching this configuration in Geant, simulating typical signals and backgrounds, applying reasonable reconstruction procedures, combining results in physics performance metrics.

Since the best solution is a trade-off between different kinds of limitations, a quick turn over is necessary

to evaluate physics performance for different techniques in different configurations.

Two typical problems which slow down the evaluation of physics performance for particular approaches to calorimeter detector technologies and configurations are:

- Emulating particular detector properties including raw detector response together with a signal processing chain to adequately simulate a calorimeter response for different signal and background conditions. This includes combining detector properties obtained from the general Geant simulation with properties obtained from different kinds of bench and beam tests of detector and electronics prototypes.
- Building an adequate reconstruction algorithm for physics reconstruction of the detector response which is reasonably tuned

to extract most of the performance provided by the given detector configuration.

Being approached from the first principles, both problems require significant development efforts. Fortunately, both problems may be addressed by using modern machine learning approaches, that allow combining available details of the detector techniques into corresponding higher level physics performance in a semi-automated way.

In the presentation, we discuss the use of advanced machine learning techniques to speed up and improve the precision of the detector development and optimisation cycle, with an emphasis on the experience and practical results obtained by applying this approach to optimising the electromagnetic calorimeter design as a part of the upgrade project for the LHCb detector at LHC.

Workshop / 68

Bayesian Neural Networks for Predictions from High Dimensional Theories

Author: Braden Kronheim^{None}

Co-authors: Michelle Kuchera¹; Harrison Prosper²; Alexander Karbo

¹ Davidson College

² Florida State University (US)

Corresponding Authors: brkronheim@davidson.edu, mikuchera@davidson.edu, harry.prosper@cern.ch

One of the goals of current particle physics research is to obtain evidence of physics beyond the Standard Model (BSM) at accelerators such as the Large Hadron Collider (LHC). The searches for new physics are often guided by BSM theories that depend on many unknown parameters, which makes testing their predictions computationally challenging. Bayesian neural networks (BNN) can map the parameter space of these theories to a meaningful distribution of observables. We describe a new package called TensorBNN, built on Tensorflow and Tensorflow Probability, which implements Bayesian neural networks. The utility of TensorBNN is illustrated by modeling the predictions of the phenomenological Minimal Supersymmetric Standard Model (pMSSM), a BSM theory with 19 free parameters. The predicted quantities are cross sections for arbitrary pMSSM parameter points, the mass of the associated lightest neutral Higgs boson, and the theoretical viability of the parameter points. All three quantities are modeled with average percent errors of 3.3% or less and in a time orders of magnitude shorter than the supersymmetry codes from which the results are derived [1]. The posterior densities, provided as point clouds, provide meaningful Bayesian confidence intervals for the predictions, further demonstrating the potential for machine learning to accurately model the mapping from the high dimensional spaces of BSM theories to their predictions.

[1] B. S. Kronheim, M. P. Kuchera, H. B. Prosper, A. Karbo, Bayesian Neural Networks for Fast SUSY Predictions, 2020. arXiv:2007.04506.

Workshop / 69

GPU and FPGA as a Service for Machine Learning Inference Accelerations

Authors: Yu Lou¹; Javier Mauricio Duarte²; Jeffrey Krupa^{None}; Kelvin Lin¹; Kevin Pedro³; Kyle Knoepfel⁴; Maria Acosta Flechas³; Matthew Trahms⁵; Mia Liu^{None}; Michael Wang³; Natchanon Suaysom¹; Nhan Viet Tran³; Philip Harris⁶; Scott Hauck⁷; Shih-Chieh Hsu⁸; Ta-Wei Ho⁹; Thomas Klijnsma³; Tingjun Yang³; Benjamin Hawks⁴; Burt Holzman³; Dylan Sheldon Rankin¹⁰; Jack Dinsmore^{None}

¹ *University of Washington (US)*² *Univ. of California San Diego (US)*³ *Fermi National Accelerator Lab. (US)*⁴ *Fermi National Accelerator Laboratory*⁵ *UW ACME Lab*⁶ *Unknown*⁷ *University of Washington*⁸ *University of Washington Seattle (US)*⁹ *National Tsing Hua University (TW)*¹⁰ *Massachusetts Inst. of Technology (US)*

Corresponding Authors: kelvin.deray.lin@cern.ch, philip.c.harris@cern.ch, schsu@uw.edu, maria.acosta.flechas@cern.ch, bhawks@fnal.gov, dylan.sheldon.rankin@cern.ch, natchanon.suaysom@cern.ch, tjyang@fnal.gov, mtrahms@gmail.com, nhan.viet.tran@cern.ch, louyu27@gmail.com, thomas.klijnsma@cern.ch, burt@fnal.gov, javier.mauricio.duarte@cern.ch, hauck@uw.edu, pedrok@cern.ch, ta-wei.ho@cern.ch, knoepfel@fnal.gov, mwang@fnal.gov

The data rate may surge after some planned upgrades for the high-luminosity Large Hadron Collider (LHC) and accelerator-based neutrino experiments. Since there is not enough storage to save all of the data, there is a challenging demand to process and filter billions of events in real-time. Machine learning algorithms are becoming increasingly prevalent in the particle reconstruction pipeline. Specially designed hardware can significantly accelerate the machine learning inference time compared to CPUs. Thus, we propose a heterogeneous computing framework called the Services for Optimized Network Inference on Coprocessors (SONIC) to accelerate machine learning inferences with various coprocessors. With a unified interface, the framework conveniently provides GPU as a service, using either the Nvidia Triton framework or the Microsoft Brainwave service as the backend. It also features the first open-source FPGA-as-a-service toolkit, using either our hls4ml framework or the Xilinx ML Suite as the backend. We demonstrated that our method could speed up one classification and two regression problems in the LHC experiments and ProtoDUNE-SP. By providing coprocessors as a service, our work may assist various other computing workflows across science.

Data Science Seminar / 72

Structured models of objects, relations, and physics

Corresponding Author: peterbattaglia@google.com

Plenary / 73

End-to-End, Machine Learning-based Data Reconstruction for Particle Imaging Neutrino Detectors

Corresponding Author: ktera@slac.stanford.edu

With firm evidence of neutrino oscillation and measurements of mixing parameters, neutrino experiments are entering the high precision measurement era. The detector is becoming larger and denser to gain high statistics of measurements, and detector technologies evolve toward particle imaging, essentially a hi-resolution “camera”, in order to capture every single detail of particles produced in a neutrino interaction. The forefront of such detector technologies is a Liquid Argon Time Projection Chamber (LArTPC), which is capable of recording images of charged particle tracks with breathtaking resolution. Such detailed information will allow LArTPCs to perform accurate particle identification and calorimetry, making it the detector of choice for many current and future neutrino experiments. However, analyzing hi-resolution imaging data can be challenging, requiring the development of many algorithms to identify and assemble features of the events in order to reconstruct neutrino interactions. In the recent years, we have been investigating a new approach using

deep neural networks (DNNs), a modern solution to a pattern recognition for image-like data in the field of Computer Vision. A modern DNN can be applied for various types of problems such as data reconstruction tasks including interaction vertex identification, pixel clustering, particle type and flow reconstruction. In this talk I will discuss the challenges of data reconstruction for imaging detectors, recent work and future plans for developing a full LArTPC data reconstruction chain using DNNs.

Plenary / 74

Solving Inverse Problems with Invertible Neural Networks

Corresponding Author: ullrich.koethe@iwr.uni-heidelberg.de

Interpretable models are a hot topic in neural network research. My talk will look on interpretability from the perspective of inverse problems, where one wants to infer backwards from observations to the hidden characteristics of a system. I will focus on three aspects: reliable uncertainty quantification, outlier detection, and disentanglement into meaningful features. It turns out that invertible neural networks – networks that work equally well in the forward and inverse direction – are great tools for that kind of analysis: They act as non-linear generalizations of classical methods like PCA and ICA. Examples from physics, medicine, and computer vision demonstrate the practical utility of the new method.

Plenary / 75

Neural Network Pruning: from over-parametrized to under-parametrized networks

Corresponding Author: micky.91@hotmail.com

This talk will provide an introduction to the concept of over-parametrization in neural networks and the associated benefits that have been identified from the theoretical and empirical standpoints. It will then present the practice of pruning as both a practical engineering intervention to reduce model size and a scientific tool to investigate the behavior and trainability of compressed models under the “lottery ticket hypothesis” (Frankle and Carbin, 2018). Finally, it will demonstrate features for tensor and neural network pruning in PyTorch.

Plenary / 76

Deep Learning @ LHC: An ATLAS Perspective

Corresponding Author: amir.farbin@cern.ch

6 Years after first demonstration of Deep Learning in HEP, the LHC community has explored a broad range of applications aiming for better, cheaper, faster, and easier solutions that ultimately extend the physics reach of the experiments and over come HL-LHC computing challenges. I’ll present a snapshot of where the ATLAS experiment currently stands in adoption of Deep Learning and suggest where it may go.

Walk through / 79

Deep Dive on Graph Networks for Learning Simulation (Deep Mind)

(first 10' missing on the recording, sorry)

Walk through / 80

Tracking GNN Walk Through

Corresponding Author: dtmurnane@lbl.gov

Plenary / 81

Zenseact : Deep learning and computer vision for self-driving cars

Corresponding Author: christoffer.petersson@zenuity.com

(no recording)

The mission of Zenseact is to develop a world-leading software platform for autonomous driving, with the main goal to dramatically reduce the number of traffic accidents in the world. I will discuss how we use deep learning and computer vision to reach this goal, and some of the challenges we face. I will also discuss the ongoing research collaboration between Zenseact and the hls4ml-team at CERN concerning compression and deployment techniques for fast deep learning inference

Plenary / 82

Machine Learning in Procter and Gamble

(no recording)

Procter & Gamble (P&G) is one of the oldest and largest “consumer goods” companies in the world. It is present in about 180 markets, with operations in 70 countries and almost 100 thousand employees. Machine Learning models created by the P&G Data Scientists support every aspect of this global business, from R&D, to shipment to marketing. The Data Science teams in the company have a strong representation of HEP/CERN alumni. In this talk I will review some examples of ML applied to commercial problems, emphasizing the differences between industry and academia, and the unique strengths that HEP-trained data scientist can contribute.

Plenary / 83

Using Topological Data Analysis to Disentangle Complex Data Sets

Corresponding Author: maurizio.sanarico@sdggroup.com

A recent new branch of the, currently called AI, is the Topological Data Analysis (TDA). TDA was born as an extension of algebraic topology to discrete data and, therefore, is a combination of algebraic topology, geometry, statistics and computational methods. According to E. Munch TDA comprises “a collection of powerful tools that can quantify shape and structure in data in order to answer questions from the data’s domain.” The key idea of TDA is that “data has shape and shape has meaning”, the shape can be quantified via topological signatures. Topological signatures lead to topological invariants, and such invariants enable greater understanding of the relationships in—and transformations of—real data.

There are three main streams in TDA: (1) persistent homology and its extension; (2) Mapper; (3) Morse-Smale complex analysis. We will focus on Mapper as one of the most effective and computationally efficient tools in the realm of TDA. Mapper could be seen as an extension of cluster analysis and, due to its flexibility can be adapted to a number of different applications. An introduction to Mapper with some practical real world application is the main topic of our presentation.

Plenary / 84

Introduction

Corresponding Authors: gian.michele.innocenti@cern.ch, rousseau@lal.in2p3.fr, lorenzo.moneta@cern.ch, andrea.wulzer@cern.ch, riccardo.torre@cern.ch, simon.akar@cern.ch, pietro.vischia@cern.ch

Plenary / 86

CERN Knowledge Transfer

Corresponding Authors: han.dols@cern.ch, nicholas.ziogas@cern.ch

Workshop / 87

Conclusion and wrap-up

Workshop / 88

FastCaloGAN: a tool for fast simulation of the ATLAS calorimeter system with Generative Adversarial Networks

Author: Michele Faucci Giannelli¹

¹ INFN e Università Roma Tor Vergata (IT)

Corresponding Author: michele.faucci.giannelli@cern.ch

Building on the recent success of deep learning algorithms, Generative Adversarial Networks (GANs) are exploited for modelling the response of the ATLAS detector calorimeter of different particle types; simulating calorimeter showers for photons, electrons and pions over a range of energies

(between 256 MeV and 4 TeV) in the full detector η range. The properties of showers in single-particle events and of jets in di-jets events are compared with full detector simulation performed by GEANT4. The good performance of FastCaloGAN demonstrates the potential of GANs to perform a fast calorimeter simulation for the ATLAS experiment.