6th Inter-experiment Machine Learning Workshop

Monday, 29 January 2024 - Friday, 2 February 2024

CERN

Book of Abstracts

Contents

Contributed Talks / 1

Longitudinal Beam Diagnostics and Phase Space Reconstruction in the LHC Using ML

Authors: Alexandre Lasheen[1]; Georges Trad[1]; Helga Timko[1]; Konstantinos Iliakis[1]; Simon Albright[1]; Theodoros Argyropoulos[1]

[1] *CERN*

Corresponding Authors: helga.timko@cern.ch, simon.albright@cern.ch, alexandre.lasheen@cern.ch, theodoros.argyropoulos@cern.ch, georges.trad@cern.ch, konstantinos.iliakis@cern.ch

Accurate knowledge of longitudinal beam parameters is essential for optimizing the performance and operational efficiency of particle accelerators like the Large Hadron Collider (LHC). However, conventional methods to determine them, such as fitting techniques and tracking-based longitudinal tomography, are time-consuming and limited to analyzing data from a few bunches only. To address this, we propose the development of a machine learning (ML) model that leverages the existing high-resolution measurements of longitudinal bunch profiles and utilizes an encoder-decoder architecture to achieve two primary objectives. Firstly, it efficiently extracts the physical beam parameters, such as injection errors, bunch length, and bunch intensity, eliminating the need for computationally expensive fitting methods. Secondly, it reconstructs the longitudinal beam distribution. The ML model is designed to operate in real-time, enabling online monitoring of multi-bunch beams. This application demonstrates the potential of ML-techniques in enhancing beam diagnostics and allowing more precise control of large particle accelerators.

Contributed Talks / 3

DeepTreeGANv2: Iterative Pooling of Point Clouds

Author: Moritz Scham[1]

Co-authors: Benno Kach [1]; Dirk Krucker [1]; Kerstin Borras [2]; Simon Schnake [1]

[1] *Deutsches Elektronen-Synchrotron (DE)*

[2] *DESY / RWTH Aachen University*

Corresponding Author: moritz.scham@cern.ch

In High Energy Physics, detailed and time-consuming simulations are used for particle interactions with detectors. To bypass these simulations with a generative model, the generation of large point clouds in a short time is required, while the complex dependencies between the particles must be correctly modelled. Particle showers are inherently tree-based processes, as each particle is produced by the decay or detector interaction of a particle of the previous generation.
In this work, we present a significant extension to DeepTreeGAN, featuring a critic, that is able to aggregate such point clouds iteratively in a tree-based manner. We show that this model can reproduce complex distributions, and we evaluate its performance on the public JetNet 150 and CaloChallange datasets.
Two papers under review are available here:
https://cernbox.cern.ch/s/WoNDFxdOiKpg0fG

Contributed Talks / 4

Out-of-Distribution Multi-set Generation with Context Extrapolation for Amortized Simulation and Inverse Problems

Author: Hosein Hashemi[1]

[1] *LMU Munich*

Addressing the challenge of Out-of-Distribution (OOD) multi-set generation, we introduce YonedaVAE, a novel equivariant deep generative model inspired by Category Theory, motivating the Yoneda-Pooling mechanism. This approach presents a learnable Yoneda Embedding to encode the relationships between objects in a category, providing a dynamic and generalizable representation of complex relational data sets. YonedaVAE introduces a self-distilled set generator, capable of zero-shot creating sets with variable inter-category and intra-category cardinality, facilitated by the new Adaptive Top-p Sampling. We demonstrate that YonedaVAE can produce new point clouds with cardinalities well beyond the training data and achieve context extrapolation. Trained on low luminosity ultra-high-granularity data of PXD at Belle II, YonedaVAE can generate high luminosity valid signatures with the correct intra-event correlation without exposure to similar data during training. Being able to generalize to out-of-distribution samples, YonedaVAE stands as a valuable method for extrapolative multi-set generation tasks in scientific discovery, including de novo protein design, Drug Discovery, and simulating geometry-independent detector responses beyond experimental limits.

Would you like to be considered for an oral presentation?:

Yes

Poster Session / 5

Uncertainty-aware Machine Learning for Proton Therapy Range Verification with a Digital Tracking Calorimeter

Author: Alexander Schilling[1]

[1] *University of Applied Sciences Worms*

Corresponding Author: alexander.georg.schilling@cern.ch

Proton therapy is highly sensitive to range uncertainties due to the nature of the dose deposition of charged particles. To ensure treatment quality, range verification methods can be used to verify that the individual spots in a pencil beam scanning treatment fraction match the treatment plan. We employ uncertainty-aware deep neural networks to predict the Bragg peak depth in an anthropomorphic phantom based on secondary charged particle detection in a silicon pixel telescope designed for proton computed tomography. The subsequently predicted Bragg peak positions, along with their uncertainties, are compared to the treatment plan to determine the quality of the treatment fraction.

The range verification model is a multi-task multilayer perceptron, predicting the beam range in water as well as the Bragg peak position in the patient. The two task losses are weighed against each other by automatically learning their respective homoscedastic uncertainties. Along with the predicted values, we additionally predict the aleatoric uncertainty. The epistemic uncertainty is estimated with Monte Carlo Dropout. The Bragg peak is predicted with a mean absolute error of 1.1 mm and the predicted uncertainties are sufficiently accurate to derive the quality of the treatment fraction.

Would you like to be considered for an oral presentation?:

Contributed Talks / 6

Differentiable Vertex Fitting for Jet Flavour Tagging

Authors: Ines Ochoa[1]; Jonathan Shoemaker[2]; Michael Kagan[2]; Rachel Emma Clarke Smith[2]; Ruben Miguel De Almeida Inacio[1]

[1] *LIP - Laboratorio de Instrumentação e Física Experimental de Partículas (PT)*

[2] *SLAC National Accelerator Laboratory (US)*

Corresponding Authors: rachel.smith@cern.ch, ruben.inacio@tecnico.ulisboa.pt, jshoe@slac.stanford.edu, michael.aaron.kagan@cern.ch, ines.ochoa@cern.ch

We propose a differentiable vertex fitting algorithm that can be used for secondary vertex fitting, and that can be seamlessly integrated into neural networks for jet flavour tagging. Vertex fitting is formulated as an optimization problem where gradients of the optimized solution vertex are defined through implicit differentiation and can be passed to upstream or downstream neural network components for network training. More broadly, this is an application of differentiable programming to integrate physics knowledge into neural network models in high energy physics. We demonstrate how differentiable secondary vertex fitting can be integrated into larger transformer-based models for flavour tagging and improve heavy flavour jet classification.

Would you like to be considered for an oral presentation?:

Poster Session / 7

Apples to Apples in Jet Quenching

Authors: Guilherme Milhano[1]; João A. Gonçalves[2]

[1] *LIP-Lisbon & CERN TH*

[2] *LIP - IST*

Corresponding Authors: guilherme.milhano@cern.ch, jgoncalves@lip.pt

Progress in the theoretical understanding of parton branching dynamics that occurs within an expanding QGP relies on detailed and fair comparisons with experimental data for reconstructed jets. Such validation is only meaningful when the computed object, be it analitically or via event generation, accounts for the complexity of experimentally reconstructed jets. The reconstruction of jets in heavy ion collisions involves a, necessarily imperfect, subtraction of the large and fluctuating background: reconstructed jets always include background contamination. The identification of jet quenching effects, that is modifications of the branching dynamics by interaction with QGP leading to changes on jet observables, should be done against a baseline that accounts for possible background contamination on unmodified jets. In practical terms, jet quenching effects are only those not present in samples of vacuum jets that have been embedded in a realistic heavy-ion background and where subtraction has been carried out analogously to that in the heavy ion case and as close as possible to what is done experientally. Using the extensively validated JEWEL event generator, we will present an extensive survey of the sensitivity to background effects of commonly used jet observables. Further, we will assess the robustness of Machine Learning studies aimed at classifying jets according to their degree of modification by the QGP, e.g [1], to a reference where background contamination is accounted for.

Poster Session / 8

Particle Transformer for tau lepton pair invariant mass reconstruction for the HH→bbtautau CMS analysis

Author: Valentina Camagni[1]

Co-authors: Pietro Govoni [1]; Simone Gennai [1]; Marco Paganoni [1]

[1] *Universita & INFN, Milano-Bicocca (IT)*

Corresponding Author: v.camagni1@campus.unimib.it

One of the most interesting channels to probe theories beyond the Standard Model at LHC, is the production of a new massive particle, that decays into pairs of Higgs Bosons which, in turn, decay into a pair of b-quarks and a pair of tau leptons. A fundamental discriminant variable to separate HH signal from the backgrounds is the invariant mass of the di-tau system, which can be reconstructed starting from the decay products of each tau lepton. In order to reconstruct the di-tau system in the best possible way, special techniques are needed as the presence of neutrinos from tau decay do not allow for a complete reconstruction of the event. To this end, a Transformer-based architecture (Particle Transformer - ParT) has been implemented to estimate the four-momentum of the neutrinos involved in the decay for a high-resolution reconstruction of the corresponding invariant mass. The regression has been subsequently combined with the classification task for the discrimination of signal events from background ones. ParT can be seen as a graph neural network in which each node consists of a visible particle of the decay's final state while the links represent the connections between them, modeled through pair-wise features. This last input help to enhance the attention mechanism (core of the Transformer class models), which has the purpose of learning the interactions between particles, and improve the model's explainability. ParT showed better results with respect to the most common used algorithm in CMS (SVFit), which is, in addition, extremely CPU consuming.

Poster Session / **9**

Decorrelation using Optimal Transport

Authors: Johnny Raine[1]; Malte Algren[1]; Tobias Golling[1]

[1] *Universite de Geneve (CH)*

Corresponding Authors: tobias.golling@unige.ch, malte.algren@unige.ch, johnny.raine@cern.ch

Novel decorrelation method using Convex Neural Optimal Transport Solvers (Cnots) that is able to decorrelate a continuous feature space against protected attributes with optimal transport. We demonstrate how well it performs in the context of jet classification in high energy physics, where classifier scores are desired to be decorrelated from the mass of a jet.

Contributed Talks / **10**

Reinforcement Learning Algorithms for Charged Particle Tracking with Applications in Proton Computed Tomography

Author: Tobias Kortus[None]

Co-authors: Nicolas Gauger ; Ralf Keidel [1]

[1] *Fachhochschule Worms (DE)*

Corresponding Author: kortus@ztt.hs-worms.de

Deep learning, especially graph neural networks, significantly improved tracking performances in modern particle detectors while reducing runtimes compared to previous state of the art approaches. However, training neural networks requires significant amount of labeled data, usually acquired by

performing complex particle simulations. We present first studies of leveraging deep reinforcement learning (RL) and constrained multi-agent reinforcement learning (MARL) as ground-truth free alternatives to supervised learning.

Instead of minimizing a loss function based on ground-truth, we optimize by trial-and-error behavior policies, acting as approximations to the full combinatorial optimization problem, maximizing the physical plausibility of sampled track candidates. Our approaches works on graph-structured data, capturing track hypotheses through edge connections between particles in the detector layers. We demonstrate, on simulated data for a particle detector used for proton computed tomography, the high potential as well as the competitiveness of RL for both single-agent as well as multi-agent settings.

Partially based on: T. Kortus, R. Keidel, N. R. Gauger, on behalf of the Bergen pCT collaboration, "Towards Neural Charged Particle Tracking in Digital Tracking Calorimeters with Reinforcement Learning," in IEEE Transactions on Pattern Analysis and Machine Intelligence, doi: 10.1109/TPAMI.2023.3305027.

Poster Session / **11**

Particle identification with Machine Learning for ALICE ITS2

Author: Giorgio Alberto Lucia[1]

[1] *Universita e INFN Torino (IT)*

Corresponding Author: giorgio.lucia@edu.unito.it

Particle identification (PID) plays a pivotal role in numerous measurements performed by the ALICE Collaboration. Various ALICE detectors offer PID information through complementary experimental techniques. The former ALICE Inner Tracking System 1 (ITS1) provided PID information by measuring the specific energy loss of low-momentum charged particles during LHC Run 1 and Run 2. The upgraded ITS (ITS2) has higher granularity and lower material budget, resulting in a significant improvement of the spatial resolution. To cope with the high interaction rates of the LHC Run 3 (500 kHz and 50 kHz for pp and Pb-Pb collisions respectively), the ITS2 features a digital readout and therefore it lacks the capability to measure directly the energy loss. However, the cluster topology of the signal left by a charged particle traversing the layers of the ITS2 can be interpreted as a proxy of the particle energy loss.

In this contribution, a novel approach to perform particle identification with the ALICE ITS2 is presented. A Boosted Decision Tree (BDT) algorithm is employed to exploit the combination of cluster topology and the particle momentum to infer the particle species. Monte Carlo simulations are used to validate this technique which has been tested on the data samples collected at the beginning of the LHC Run 3. The internal parameters of the BDTs are optimised to enhance the separation between the different particle species by employing the Optuna package. In this way, a remarkable separation among different particle species at low momentum is achieved.

Finally, Machine Learning (ML) offers a viable avenue for PID determination between the Z = 1 and Z = 2 particles, independently on their momentum.
A study on the PID capabilities of the Muon Forward Tracker (MFT), a detector that shares the same technology as the ITS2, was performed. Using a ML-based method, it could be possible to distinguish Z > 1 (anti-)nuclei from Minimum Ionizing Particles by adopting unsupervised learning algorithms.

Poster Session / **12**

Electron identification with a convolutional neural network

Authors: Bruna Pascual[1]; Dominique Godin[1]; Hoang Dai Nghia Nguyen[2]; Jean-Francois Arguin[1]; Olivier Denis[1]

[1] *Universite de Montreal (CA)*

[2] *Université de Montreal (CA)*

Corresponding Authors: olivier.denis.1@umontreal.ca, bruna.pascual.dias@cern.ch, jean.francois.arguin@gmail.com, hoang.dai.nghia.nguyen@cern.ch, dominique.godin@cern.ch

The identification of electrons plays an important role in a large fraction of the physics analyses performed at ATLAS. An improved electron identification algorithm is presented that is based on a convolutional neural network (CNN). The CNN utilizes the images of the deposited energy in the calorimeter cells around the reconstructed electron candidates for each of the electromagnetic and hadronic calorimeter layers. In addition, the CNN algorithm utilizes as input features the same high-level variables that are used by the likelihood (LLH) and deep neural network (DNN) algorithms currently used in ATLAS, as well as the information of up to five inner detector tracks that are matched to an electron candidate during its reconstruction. The CNN algorithm results in a significant improvement in identification performance, corresponding for example to an improvement in background rejection of factors of about 3 to 10 with respect to the LLH algorithm for its "Loose" working point, depending on the pseudorapidity and transverse momentum of the electron candidate.
Reference:
Electron Identification with a Convolutional Neural Network in the ATLAS Experiment, ATLAS Collaboration

Poster Session / 13

Model-independent strategy for New Physics seartch at LHC using Anomaly Detection algorithms

Author: Louis Vaslin[1]

[1] *High Energy Accelerator Research Organization (JP)*

Corresponding Author: lovaslin@post.kek.jp

The search for rare New Physics signals is one of the main challenges that are addressed in LHC experiments. Classical search strategies rely on signal hypothesis and simulation to optimize the sensitivity of an analysis. However, relying on hypothesis and simulation has drawbacks. To address this problem, I propose a model independent strategy for heavy resonance searches.

This strategy relies on a novel Machine Learning based anomaly detection techniques 1. The model is composed of an Auto-Encoder architecture coupled with a discriminant network. The two networks are trained adversarially in a GAN-like setting. The objective is to use the discriminant loss as constraint on the Auto-Encoder in addition to the usual reconstruction error. The model is trained in order to identify any signal as an anomaly while providing for a suitable data-driven background model. This background modeling feature relies on the use of the Distance Correlation (DisCo) regularization term as well as event reweighting.

In order to assess the significance of any potential resonance, an enhanced implementation of the BumpHunter algorithm is used [2]. This implementation provides a method to evaluate efficiently the global p-value of a localized deviation in the data even if the statistics is limited in the test statistic distribution. The test performed on the data provided for the LHC Olympics 2020 challenge [3] shows promising results, my method being able to correctly identify a hidden signal with up to 3 sigma significance.

1 L. Vaslin, V. Bara, J. Donini. GAN-AE: an anomaly detection algorithm for New Physics search in LHC data. *Eur. Phys. J. C* 83, 1008 (2023). doi:10.1140/epjc/s10052-023-12169-4
[2] L. Vaslin, S. Calvet, V. Bara, J. Donini. pyBumpHunter: A model independent bump hunting tool in Python for high energy physics analyses. *SciPost Phys. Codebases* 15 (2023), doi:10.21468/SciPostPhysCodeb.15
[3] G. Kasieczka et al. The LHC Olympics 2020 a community challenge for anomaly detection

in high energy physics. *Report on Progress in Physics*, 84(12):124201, dec 2021. doi:10.1088/1361-6633/ac36b9

Poster Session / 14

Topological separation of dielectron signals using machine learning in Pb–Pb collisions with ALICE

Author: Jerome Jung[1]

[1] *Goethe University Frankfurt (DE)*

Corresponding Author: jung@ikf.uni-frankfurt.de

Dielectrons are an exceptional tool to study the evolution of the medium created in heavy-ion collisions. In central collisions, the energy densities are sufficient to create a quark-gluon plasma (QGP). At LHC energies, the dominant background process for the measurements of thermal e^+e^- pairs originating from the QGP are correlated HF hadron decays which dominate the dielectron yield for invariant masses above 1.1 GeV/c^2. Their contribution is modified in the medium compared to elementary collisions to an unknown extent, leading to large uncertainties in the subtraction of known hadronic sources.

Alternatively, a topological separation can be utilised to disentangle them from the contribution of thermal dielectrons originating from the primary vertex.
As machine learning (ML) algorithms have achieved state-of-the-art performance in a variety of high-energy physics analyses, deep neural networks (DNNs) can be applied to capture the complex multidimensional correlations in the tracking parameters to identify these pairs.

In this poster, a DNN to classify dielectron sources based on their decay topology with the ALICE detector will be presented for simulated Pb–Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. Their performance will be compared to the established analysis on the distance-of-closes approach (DCA) to the primary vertex.
Finally, the way these ML techniques could be incorporated in future dielectron analysis will be discussed.

Poster Session / 16

DeepCore2.0: Convolutional Neural Network for Tracking in the Core of High-Energy Jets

Author: Hichem Bouchamaoui[1]

Co-authors: Jim Olsen [1]; Nicholas Joseph Haubrich [1]; Soohyun Yoon [1]

[1] *Princeton University (US)*

Corresponding Author: bouchamaouihichem@gmail.com

Tracking of charged particles in dense environments, especially in the core of high transverse-momentum (p_T) jets, presents a growing challenge with increasing LHC luminosity. Despite the CMS phase-1 pixel detector upgrade, and dedicated cluster splitting and pattern recognition algorithms like JetCore, there is still significant room for improvement. Limiting the computation time for track reconstruction represents an additional challenge as the number of proton-proton interactions per crossing (pileup) increases. DeepCore is a machine learning algorithm designed to improve track seeding in the core of high-p_T jets in the presence of increased pileup. In this talk, we summarize recent improvements to DeepCore optimized in the context of a hybrid JetCore+DeepCore model, leading to a significant increase in track reconstruction efficiency relative to JetCore alone

for particles with p_T above 10 GeV. This improved algorithm, referred to as DeepCore2.0, also leads to a reduction in overall computation time for track reconstruction, with further reduction possible in the future.

Poster Session / 17

A data-driven and model-agnostic approach to solving combinatorial assignment problems in searches for new physics

Authors: Anthony Badea[1]; Javier Montejo Berlingen[2]

[1] *University of Chicago (US)*

[2] *The Barcelona Institute of Science and Technology (BIST) (ES)*

Corresponding Authors: anthony.badea@cern.ch, javier.montejo.berlingen@cern.ch

We present a novel approach to solving combinatorial assignment problems in particle physics without the need to introduce prior knowledge or assumptions about the particles' decay. The correct assignment of decay products to parent particles is achieved in a model-agnostic fashion by introducing a novel neural network architecture, Passwd-ABC, which combines a custom layer based on attention mechanisms and dual autoencoders. We demonstrate how the network, trained purely on background events in an unsupervised setting, is capable of reconstructing correctly hypothetical new particles regardless of their mass, decay multiplicity and substructure, and produces simultaneously an anomaly score that can be used to efficiently suppress the background. This model allows to extend the suite of searches for localized excesses to include non-resonant particle pair production where the reconstruction of the two resonant masses is thwarted by combinatorics. Based on https://arxiv.org/abs/2309.05728.

Contributed Talks / 19

Accelerating the search for mass bumps using the Data-Directed Paradigm

Authors: Ali El Moussaouy[1]; Amit Shkuri[2]; Bruna Pascual[1]; Emile Baril[1]; Etienne Dreyer[2]; Eva Mayer[3]; Fannie Bilodeau[1]; Georges Azuelos[1]; Jean-Francois Arguin[1]; Julien Noce Donini[3]; Nilotpal Kakati[2]; Samuel Calvet[4]; Shikma Bressler[2]

[1] *Universite de Montreal (CA)*

[2] *Weizmann Institute of Science (IL)*

[3] *Université Clermont Auvergne (FR)*

[4] *LPC Clermont-Ferrand, IN2P3 / UCA (FR)*

Corresponding Authors: fannie.bilodeau@cern.ch, eva.mayer@cern.ch, ali.el.moussaouy@cern.ch, etienne.dreyer@cern.ch, georges.azuelos@cern.ch, shikma.bressler@cern.ch, emile.baril@cern.ch, amit.shkuri@cern.ch, nilotpal.kakati@cern.ch, samuel.pierre.jean.calvet@cern.ch, julien.donini@cern.ch, bruna.pascual.dias@cern.ch, jean.francois.arguin@gmail.com

The Data-Directed paradigm (DDP) is a search strategy for efficiently probing new physics in a large number of spectra with smoothly-falling SM backgrounds. Unlike the traditional analysis strategy, DDP avoids the need for a simulated or functional-form based background estimate by directly predicting the statistical significance using a convolutional neural network trained to regress the log-likelihood-based significance. In this way, a trained network is used to identify mass bumps directly on data. By saving a considerable amount of time, this approach has the potential to expand the discovery reach by checking many unexplored regions. The method has shown good performance when finding various beyond standard model particles in simulation data. A description of the method and recent developments will be presented.

Contributed Talks / 20

Electron and Proton Classification with AMS ECAL Using Convolutional Vision Transformers and Domain Adaptation

Authors: Berk Turk[1]; Bilge Demirkoz[1]; Emre Akbas[1]; Zhili Weng[2]; Gulce Karagoz[1]

[1] *Middle East Technical University (TR)*

[2] *Massachusetts Inst. of Technology (US)*

Corresponding Authors: berk.turk@cern.ch, zhili.weng@cern.ch, emre.akbas@cern.ch, bilge.demirkoz@cern.ch, gulce.karagoz@cern.ch

Alpha Magnetic Spectrometer (AMS-02) is a precision high-energy cosmic-ray experiment on the ISS operating since 2011 and has collected more than 228 billion particles. Among them, positrons are important to understand the particle nature of dark matter. Separating the positrons from cosmic background protons is challenging above 1 TeV. Therefore, we use state-of-the-art convolutional and transformer models, CoAtNet and Convolutional Vision Transformer (CvT), that employ the shower signals from the ECAL to classify the electrons/positrons in the dominant cosmic proton background. We created sets of electrons, positrons, and protons events from the ISS data and Monte Carlo Simulation in the energy range between 0.2-2 TeV by applying various data quality cuts on reconstructed variables obtained from the subdetectors. Initially, since ECAL showers are not tunned in the AMS MC, our MC trained models show a lower proton rejection on the ISS data. To accommodate the difference between the training and test domain distributions, we implemented domain adaptation with the CoAtNet and CvT to mitigate this dataset bias/domain shift. We also trained domain adaptation with a set of well-reconstructed 1 electron charge ISS events without electron/proton labels at TeV energy order as the target dataset. We evaluated the models between 1-2 TeV energy using ISS and MC events with the proton rejection vs. electron efficiency and proton rejection vs. energy at near 90% electron efficiency plots. We performed experiments using various training and validation dataset combinations and other hyperparameters with the CvT and CoAtNet. Among them, the best models are obtained with the 1-2TeV MC events as training data and half of the labeled 1-2 TeV ISS events as validation data. Using domain adaptation with the CoAtNet, we obtained a maximum proton rejection at 88% electron efficiency on the ISS data. We also rejected all of the MC protons at higher than 99.8% electron efficiency with both CvT and CoAtNet. At 90% electron efficiency, the proton rejection power of the CvT and CoAtNet is 5 and 7 times higher than the proton rejection power of the AMS's Boosted Decision Tree and ECAL Likelihood Estimator for MC events in the 1-2 TeV range.

Would you like to be considered for an oral presentation?:

Contributed Talks / 21

Training and optimisation of large transformer models at CERN: an ATLAS case study on Kubeflow

Author: Maxence Draguet[1]

[1] *University of Oxford (GB)*

Corresponding Author: maxence.draguet@physics.ox.ac.uk

Heavy flavour jets underpin a large part of the ATLAS physics programme, such as analyses of Higgs boson decays to quarks and super-symmetry searches with b-jets. The algorithms for identifying jets originating from b- and c-quarks are instrumental in these efforts, with the recently introduced GN2 model 1 showing remarkable improvements in tagging efficiency. Given its complexity and data demands, high-performance GPU clusters are essential for training GN2. Unfortunately, many within the collaboration lack such resources, emphasising the need for equitable project access. Additionally, the performance of GN2 can be improved through further optimisation of its hyperparameters,

an even more computationally demanding task that can be automated with frameworks like Katib. Addressing these two challenges, the ATLAS flavour tagging group is assessing training on CERN IT's Kubeflow infrastructure for machine learning (ml.cern.ch) backed by Kubernetes. This talk will showcase a framework for CERN users to utilise these resources and present the initial results of this effort. Furthermore, the talk will highlight a cutting-edge approach to optimise hyperparameters using µTransfer [2], a deep learning technique recently developed to optimise large language models by zero-shot transferring the performance from lower-complexity models to their full-complexity equivalent. While centred on an ATLAS use case, the methodology presented will be relevant to any collaboration employing advanced ML.

1 ATLAS Collaboration, Public plots for MC/MC and Data/MC comparisons of Dl1d and GN1, and simulation performance of GN2, (2023), https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2023-01/

[2] Yang, Greg, et al., Tensor programs v: Tuning large neural networks via zero-shot hyperparameter transfer. arXiv preprint arXiv:2203.03466 (2022).

Contributed Talks / 22

A Deep Generative Model for Hadronization

Authors: Adam Kania[None]; Andrzej Konrad Siodmok[1]; Ben Nachman[2]; Jay Chan[3]; Xiangyang Ju[2]

[1] *Jagiellonian University (PL)*

[2] *Lawrence Berkeley National Lab. (US)*

[3] *Lawrence Berkeley National Laboratory*

Corresponding Authors: ad.kania@student.uj.edu.pl, xiangyang.ju@cern.ch, jay.chan@cern.ch, benjamin.philip.nachman@cern.ch, a.siodmok@cern.ch

Hadronization is a critical step in the simulation of high-energy particle and nuclear physics experiments. As there is no first principles understanding of this process, physically-inspired hadronization models have a large number of parameters that are fit to data. We propose an alternative approach that uses deep generative models, which are a natural replacement for classical techniques, since they are more flexible and may be able to improve the overall precision. We first demonstrate using neural networks to emulate specific hadronization when trained using the inputs and outputs of classical methods. A protocol is then developed to fit a deep generative hadronization model in a realistic setting, where we only have access to a set of hadrons in data. Finally, we build a deep generative hadronization model that includes both kinematic (continuous) and flavor (discrete) degrees of freedom. Our approach is based on Generative Adversarial Networks and we show the performance within the context of the cluster model within the Herwig event generator.

Poster Session / 23

Improving the line-segment tracking algorithm with machine learning for the High Luminosity LHC

Authors: Avi Yagil[1]; Balaji Venkat Sathia Narayanan[1]; Gavin Niendorf[2]; Jonathan Guiang[1]; Manos Vourliotis[1]; Matevz Tadel[1]; Mayra Silva[3]; Peter Elmer[4]; Peter Wittich[2]; Philip Chang[5]; Slava Krutelyov[1]; Tres Reid[2]; Yanxi Gu[1]

[1] *Univ. of California San Diego (US)*

[2] *Cornell University (US)*

[3] *University of Florida*

[4] *Princeton University (US)*

[5] *University of Florida (US)*

Corresponding Authors: michael.george.reid@cern.ch, vyacheslav.krutelyov@cern.ch, philip.chang@cern.ch, a.yagil@cern.ch, yanxi.gu@cern.ch, peter.elmer@cern.ch, jonathan.guiang@cern.ch, mayrasilva@ufl.edu, emmanouil.vourliotis@cern.ch, gavinniendorf@gmail.com, matevz.tadel@cern.ch, balaji.venkat.sathia.narayanan@cern.ch, wittich@cornell.edu

In this work, we present a study on how machine learning (ML) can be used to enhance charged particle tracking algorithms. In particular, we focus on the line-segment-based tracking (LST) algorithm that we have designed to be naturally parallelized and vectorized on modern processors. LST has been developed specifically for the Compact Muon Solenoid (CMS) Experiment at the LHC, towards the High Luminosity LHC (HL-LHC) upgrade, and we have shown excellent efficiency and performance results, leveraging a full simulation of the CMS detector. At the same time, promising ML solutions, mainly Graph Neural Networks (GNNs), for charged particle tracking have been emerging, based initially on the simplified TrackML dataset. Preliminary results from these studies suggest that parts of LST could be improved by ML. Thus, a thorough study of exactly how and where this might be done is described. First, a lightweight neural network is used in place of explicitly defined track-quality selections. This neural network recovers a significant amount of efficiency for displaced tracks, reduces false positives, and has little-to-no impact on the throughput. These results clearly establish that ML can be used to improve LST without penalty. Next, exploratory studies of GNN track-building algorithms are described, where LST is used to create the input graph. Then, an edge-classifier GNN is trained, and the efficiency of the resultant edge scores is compared with LST. These GNN studies provide insights into the practicality and performance of using more ambitious ML algorithms for HL-LHC tracking at the CMS Experiment.

Poster Session / 24

Deep learning for the rare top decay t→sW at the LHC

Authors: Ian James Watson[1]; Inkyu Park[2]; Jason Lee[3]; Jeewon Heo[2]; Seungjin Yang[4]; Woojin Jang[2]; Youn Jung Roh[2]

[1] *University of Seoul*

[2] *University of Seoul, Department of Physics (KR)*

[3] *University of Seoul (KR)*

[4] *Kyung Hee University (KR)*

Corresponding Authors: jason.lee@cern.ch, woojin.jang@cern.ch, inkyu.park@cern.ch, ian.james.watson@cern.ch, youn.jung.roh@cern.ch, seungjin.yang@cern.ch, jeewon.heo@cern.ch

The Cabibbo Kobayashi Maskawa (CKM) matrix describes the flavor-changing quark interactions. Vts is the matrix element that describes the coupling between the top and strange quark, has not been directly measured. A direct measurement of |Vts| can be performed by identifying the strange jets from top decays. The strange jet tagging problem is challenging due to the similarity of strange jets and light jets. When tagging the strange jet decaying from top quarks, both the jet properties and topology of the event can be considered. For this task, we employ a deep-learning model, SAJA (Self-Attention for Jet Assignment), based on the self-attention mechanism that can utilize all event topology and jet properties. The SAJA model finds the jets decaying from t→sW in the Dileptonic top pair production using the whole event information.

Poster Session / 25

ML-assisted reconstruction of hadron-collider events with mini-jets

Authors: Daniel Britzger[1]; Josef Modestus Murnauer[2]; Roman Kogler[3]; Stefan Kluth[2]

[1] *Max-Planck-Institut für Physik München*

[2] *Max Planck Society (DE)*

[3] *DESY (DE)*

Corresponding Authors: skluth@mpp.mpg.de, roman.kogler@cern.ch, josef@mpp.mpg.de, daniel.britzger@cern.ch

The task of reconstructing physical observables from recorded experimental data in hadron collider events is a common challenge in LHC data analysis. Experimental measurements, such as hits in tracking detectors and signals in calorimeters, are combined into particle-flow objects, such as jets, muons, electrons, and missing transverse energy. However, reconstructing key observables related to the dynamics of particles created in hard collisions, like top-quarks, weak bosons (W, Z), or the Higgs boson, is intricate due to combinatorial ambiguities, tagging inefficiencies, acceptance losses, pile-up, and other experimental effects.

In this study, we propose a novel approach to reconstruct hadron collider events by utilizing mini-jets as the sole reconstructed objects, along with a machine-learning algorithm to determine the desired observables. These mini-jets, obtained with a distance measure of R=0.1, condense the full information from all particles in an event into a manageable size both experimentally and computationally. We demonstrate that a deep neural network can directly regress observables related to intermediate W bosons or top quarks, as well as particle-level jets with larger R and dressed leptons. This methodology surpasses classical reconstruction algorithms, offering a more efficient and generic event reconstruction for future LHC analyses.

Contributed Talks / 26

Unweighted event generation with matrix element surrogates

Authors: Daniel Maître[None]; Frank Siegert[None]; Henry Truong[None]; Steffen Schumann[None]; Timo Janssen[None]

Corresponding Author: timo.janssen@theorie.physik.uni-goettingen.de

We show that employing a sophisticated neural network emulation of QCD multijet matrix elements based on dipole factorisation can lead to a drastic acceleration of unweighted event generation in high-multiplicity LHC production processes. We incorporate these emulations as fast and accurate surrogates in a two-stage rejection sampling algorithm within the SHERPA Monte Carlo that yields unbiased unweighted events suitable for phenomenological analyses and post-processing in experimental workflows, e.g. as input to a time-consuming detector simulation. For the computational cost of unweighted events we achieve a reduction by factors between 16 and 350 for the considered channels. We also show how this technique can be used for NLO calculations with emulated loop amplitudes.

SciPost Phys. 15, 107 (2023), arXiv:2301.13562

Contributed Talks / 28

End-to-end Reconstruction Algorithm for Highly Granular Calorimeters

Authors: Philipp Zehetner[1]; Jan Kieseler[2]; Shah Rukh Qasim[3]; Dolores Garcia[2]; Kenneth Long[4]

[1] *Ludwig Maximilians Universitat (DE)*

[2] *CERN*

[3] *University of Zurich (CH)*

[4] *Massachusetts Inst. of Technology (US)*

Corresponding Authors: jan.kieseler@cern.ch, philipp.zehetner@cern.ch, shah.rukh.qasim@cern.ch, maria.dolores.garcia.marti@ce kenneth.long@cern.ch

We present an end-to-end reconstruction algorithm for highly granular calorimeters that includes track information to aid the reconstruction of charged particles. The algorithm starts from calorimeter hits and reconstructed tracks, and outputs a coordinate transformation in which all shower objects are well separated from each other, and in which clustering becomes trivial. Shower properties such as particle ID and energy are predicted from representative points within showers. This is achieved using an extended version of the object condensation loss, a graph segmentation technique that allows the clustering of a variable number of showers in every event while simultaneously performing regression and classification tasks. The backbone is an architecture based on a newly-developed translation-equivariant version of GravNet layers. These dynamically build learnable graphs from input data to exchange information along their edges. The model is trained on data from a simulated detector that matches the complexity of the CMS high-granularity calorimeter (HGCAL), for which it can be retrained in the future.

Poster Session / 29

Machine learning for enhanced measurement of Higgs boson production cross section via vector boson fusion in H->WW* with the ATLAS detector

Author: Metea Castilleja Marr[1]

[1] *Simon Fraser University (CA)*

Corresponding Author: metea.castilleja.marr@cern.ch

In this analysis, we apply modern machine learning techniques to the $H \to WW^* \to e\nu\mu\nu$ decay channel using data from the ATLAS detector collected during Run-2 of the LHC to precisely measure the total cross sections of both gluon-gluon Fusion (ggF) and Vector Boson Fusion (VBF) Higgs production modes. The detailed results can be found in the 2023 publication "Measurements of higgs boson production by gluon-gluon fusion and vector-boson fusion using $H \to WW^* \to e\nu\mu\nu$ decays in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector."

VBF is the second most prevalent Higgs production mode at the LHC, contributing approximately 7\% to the total production of Higgs bosons. In VBF processes, a quark from each of the colliding protons emits a W or Z boson, which interact to produce a Higgs boson. VBF events always contain at least two strong forward jets due to the quarks that remain after W or Z boson emission, providing a clear and identifiable signature in the detector.

We implement a Deep Neural Network (DNN) with Keras and TensorFlow to distinguish VBF event candidates from other signal and background processes. This DNN undergoes training on various kinematic variables involving leptons, jets, missing transverse energy, and other parameters associated with VBF topology. It acts as a robust classifier, effectively identifying VBF Higgs boson production amidst alternative Higgs production modes and background processes. Future enhancements to the analysis include a 2-dimensional DNN capable of classifying not only VBF processes but also ggF processes, the dominant Higgs production mode at the LHC.

Poster Session / 30

Machine learning approaches for parameter reweighting in MC samples of top quark production in CMS

Author: Valentina Guglielmi[1]

Co-authors: Katerina Lipka [1]; Simone Amoroso [1]

[1] *Deutsches Elektronen-Synchrotron (DE)*

Corresponding Author: valentina.guglielmi@cern.ch

In particle physics, Monte Carlo (MC) event generators are needed to compare theory to the measured data. Many MC samples have to be generated to account for theoretical systematic uncertainties, at a significant computational cost. Therefore, the MC statistic becomes a limiting factor for most measurements and the significant computational cost of these programs a bottleneck in most physics analyses. In this contribution, the Deep neural network using Classification for Tuning and Reweighting (DCTR) approach is evaluated for the reweighting of two systematic uncertainties in MC simulations of top quark pair production within the CMS experiment. DCTR is a method, based on a Deep Neural Network (DNN) technique, to reweight simulations to different model parameters by using the full kinematic information in the event. This methodology avoids the need for simulating the detector response multiple times by incorporating the relevant variations in a single sample.

Contributed Talks / 31

The DL Advocate: Playing the devil's advocate with hidden systematic uncertainties

Authors: Aleksandr Iniukhin[1]; Andrea Mauri[2]; Andrei Golutvin[3]; Andrey Ustyuzhanin[4]; Guillermo Hijano Mendizabal[5]; Nicola Serra[5]; Patrick Haworth Owen[5]

[1] *National Research University Higher School of Economics (RU)*

[2] *Imperial College (GB)*

[3] *Imperial College London*

[4] *Yandex School of Data Analysis (RU)*

[5] *University of Zurich (CH)*

Corresponding Authors: guillermo.hijano@cern.ch, nicola.serra@cern.ch, andrey.goloutvin@cern.ch, patrick.haworth.owen@cern.ch, andrey.ustyuzhanin@cern.ch, a.mauri@cern.ch, aleksandr.iniukhin@cern.ch

We propose a new method based on machine learning to play the devil's advocate and investigate the impact of unknown systematic effects in a quantitative way. This method proceeds by reversing the measurement process and using the physics results to interpret systematic effects under the Standard Model hypothesis. We explore this idea with two alternative approaches, one relies on a combination of gradient descent and optimisation techniques, the other employs reinforcement learning. We illustrate the potentiality of the presented method by considering two examples, firstly the case of a branching fraction measurement of the decay of a b-hadron, secondly the determination of the P_5' angular observable in $B^0 \rightarrow \mu^+\mu^-$ decays. Based on https://arxiv.org/abs/2303.15956

Would you like to be considered for an oral presentation?:

Poster Session / 32

Generating parton-level events from reconstructed events with Conditional Normalizing Flows

Author: Adrian Antonio Petre[1]

[1] *ETH Zurich (CH)*

Corresponding Author: adrian.antonio.petre@cern.ch

We introduce a new approach using generative machine learning to sample meaningful generator-level events given reconstructed events in the CMS detector. Our method combines Transformers

and Normalizing Flows to tackle the challenge of integrating the Matrix Element Method with importance sampling. We propose using a Transformer network to analyze the full reconstructed event and extract latent information, which is then used to condition a Normalizing Flow network. This approach enables the generation of probable sets of partons that are compatible with observed objects. We demonstrate the performance of our approach on a complex final state, like ttH(bb) in the semileptonic decay channel, and discuss potential applications.

Would you like to be considered for an oral presentation?:

Poster Session / 38

BSM models and parameter inference via an n-channel 1D-CNN

Authors: Harri Waltari[1]; Max Fusté Costa[2]; Prashant Singh[1]; Rikard Enberg[1]; Stefano Moretti[1]; Yong Sheng Koay[None]

[1] *Uppsala University*

[2] *DESY*

Corresponding Authors: yongsheng.koay@physics.uu.se, prashant.singh@scilifelab.uu.se, harri.waltari@physics.uu.se, rikard.enberg@physics.uu.se, stefano.moretti@physics.uu.se

In this study, we focused on inferring BSM models and their parameters from the kinematic distributions of collider signals via an n-channel 1D-Convolutional Neural Network (1D-CNN). As new physics may influence two or more observables, a 2D-CNN approach might not always be the best option. Alternatively, one can use a simple MLP with any number of observables via an event-by-event approach. However, this may not be the best in inferring global parameters that influence the entire dataset. Instead, we opt for an n-channel 1D-CNN that allows simultaneous inference from n observables while still inferring from distributions. To exemplify our approach, we applied our method on a two-component dark matter model using mono-jet and mono-Z signals. It is used to distinguish between different dark matter fields and also to infer their spin and mass. Given the importance of having sufficient training data for ML and also the computational expenses associated with simulations, we introduced a data augmentation technique. This involves modifying our architecture to be multiheaded and also including auxiliary information as additional inputs. Beyond conserving computational resources, this also improves our method's robustness against variations in signal regions.

Contributed Talks / 39

Modeling N_{ch} distributions and p_{T} spectra in high-energy pp collisions with DNNs

Author: Maria Alejandra Calmon Behling[1]

Co-authors: Henner Buesching [1]; Jerome Jung [1]; Mario Kruger [1]

[1] *Goethe University Frankfurt (DE)*

Corresponding Authors: mario.kruger@cern.ch, jung@ikf.uni-frankfurt.de, henner.buesching@cern.ch, maria.calmon@cern.ch

During the data-taking campaigns Run 1 and Run 2 of the Large Hadron Collider (LHC), the ALICE collaboration collected a large amount of proton-proton (pp) collisions across a variety of center-of-mass energies ($\sqrt{s}$). This extensive dataset is well suited to study the energy dependence of particle production. Deep neural networks (DNNs) provide a powerful regression tool to capture underlying

multidimensional correlations inherent in the data. In this contribution, DNNs are used to parameterize recent ALICE measurements of charged-particle multiplicity (N_{ch}) distributions and transverse momentum (p_T) spectra. The model architectures are defined and validated using a Bayesian-Optimization hyperparameter search on PYTHIA simulations for a wide $\sqrt{s}$ range and then trained on the ALICE data. An ensemble method is used to predict the observables of interest, extrapolating the measurements towards higher N_{ch} and p_T values as well as to unmeasured $\sqrt{s}$ from 0.5 to 100 TeV. We demonstrate that the predicted p_T spectra can serve as a reference for future heavy-ion measurements, e.g. the O–O campaign planned in LHC Run 3, where no dedicated pp data-taking at the same $\sqrt{s}$ is currently foreseen.

Would you like to be considered for an oral presentation?:

Poster Session / 41

Systematic Effects in Jet Tagging Performance for the ATLAS Detector

Authors: Ben Nachman[1]; Daniel Whiteson[2]; Kevin Thomas Greif[2]; Michael James Fenton[2]

[1] *Lawrence Berkeley National Lab. (US)*

[2] *University of California Irvine (US)*

Corresponding Authors: benjamin.philip.nachman@cern.ch, kgreif@uci.edu, daniel.whiteson@cern.ch, michael.james.fenton@cern

Machine learning based jet tagging techniques have greatly enhanced the sensitivity of measurements and searches involving boosted final states at the LHC. However, differences between the Monte-Carlo simulations used for training and data lead to systematic uncertainties on tagger performance. This talk presents the performance of boosted top and W boson taggers when applied on data sets containing systematic variations that approximate some of these differences. The taggers are shown to have differing sensitivity to the systematic variations, with the most powerful taggers showing the largest sensitivity. This trend presents obstacles for the further deployment of machine learning techniques at the LHC, and an open challenge for the HEP-ML community.

Contributed Talks / 43

Masked particle modelling

Authors: Johnny Raine[1]; Lukas Alexander Heinrich[2]; Matthew Leigh[3]; Michael Kagan[4]; Rita Osadchy[5]; Samuel Byrne Klein[1]; Tobias Golling[1]

[1] *Universite de Geneve (CH)*

[2] *Technische Universitat Munchen (DE)*

[3] *University of Geneva*

[4] *SLAC National Accelerator Laboratory (US)*

[5] *University of Haifa*

Corresponding Authors: tobias.golling@unige.ch, lukas.heinrich@cern.ch, matthew.leigh@cern.ch, johnny.raine@cern.ch, samuel.byrne.klein@cern.ch, michael.aaron.kagan@cern.ch

The Bert pretraining paradigm has proven to be highly effective in many domains including natural language processing, image processing and biology. To apply the Bert paradigm the data needs to be described as a set of tokens, and each token needs to be labelled. To date the Bert paradigm has not been explored in the context of HEP. The samples that form the data used in HEP can be described as a set of particles (tokens) where each particle is represented as a continuous vector. We explore different approaches for discretising/labelling particles such that the Bert pretraining can be

performed and demonstrate the utility of the resulting pretrained models on common downstream HEP tasks.

Poster Session / 44

Energy-based graph autoencoders for semivisible jet tagging in the Lund representation

Authors: Annapaola De Cosa[1]; Christoph Frederik Ribbe[1]; Florian Eble[1]; Roberto Seidita[1]

[1] *ETH Zurich (CH)*

Corresponding Authors: christoph.frederik.ribbe@cern.ch, roberto.seidita@cern.ch, annapaola.de.cosa@cern.ch, florian.eble@cern.ch

Semivisible jets are are a novel signature arising in Hidden Valley (HV) extensions of the SM with a confining interaction [1]. Originating from a double shower and hadronization process and containing undetectable dark bound states, semivisible jets are expected to have a substantially different radiation pattern compared to SM jets.

Unsupervised machine learning allows to learn the showering pattern of SM jets from data and successfully tag semivisible jets without relying on assumptions on the showering dynamics of the HV interaction [2]. Lund trees [3] are a natural representation of hadronic jets, encoding the full showering history. We show how a graph autoencoder can succesfully learn the Lund tree structure of SM jets and tag semivisible jets as anomalies. We furthermore propose a novel training workflow that extends the normalized autoencoder architecture [4, 5] to graph networks, allowing to suppress out-of-distribution reconstruction in a fully signal-agnostic fashion by constraining the low-reconstruction-error phase space to match the support of the training data.

Poster Session / 45

Unsupervised tagging of semivisible jets with energy-based autoencoders in CMS

Author: Florian Eble[1]

Co-authors: Annapaola De Cosa [1]; Roberto Seidita [1]

[1] *ETH Zurich (CH)*

Corresponding Author: florian.eble@cern.ch

A particularly interesting application of autoencoders (AE) for High Energy Physics is their use as anomaly detection (AD) algorithms to perform a signal-agnostic search for new physics. This is achieved by training AEs on standard model physics and tagging potential new physics events as anomalies. The use of an AE as an AD algorithm relies on the assumption that the network better reconstructs examples it was trained on than ones drawn from a different probability distribution, i.e. anomalies. Using the search for non resonant production of semivisible jets as a benchmark, we demonstrate the tendency of AEs to generalize beyond the dataset they are trained on, hindering their performance. We show how normalized AEs, specifically designed to suppress this effect, give a sizable boost in performance. We further propose a different loss function and using the Energy Mover's Distance as a metric to reach the optimal performance in a fully signal-agnostic way.

Contributed Talks / 46

Parametrising profiled likelihoods with neural networks

Authors: Humberto Reyes-González[1]; Jack Y. Araz[2]; Wolfgang Waltenberger[3]

[1] *University of Genoa*

[2] *Jefferson Lab*

[3] *Austrian Academy of Sciences (AT)*

Corresponding Authors: wolfgang.waltenberger@gmail.com, harg.zeppelin@gmail.com, jackaraz@jlab.org

Full statistical models encapsulate the complete information of an experimental result, including the likelihood function given observed data. Since a few years ago ATLAS started publishing statistical models that can be reused via the pyhf framework; a major step towards fully publishing LHC results. In the case of fast Simplified Model Spectra based reinterpretation we are often only interested in the profiled likelihood given a signal strength. However, their computation using pyhf take the order of seconds per parameter point, slowing down SMS reinterpretation by orders of magnitude. Thus, to fully leverage from the precision obtained from full statistical models without compromising speed, we propose to learn the profiled likelihood functions with Neural Networks (NNs). We show that such functions can be well described with simple NNs, published in the ONNX format, and easily used by different reinterpretation tools.

Contributed Talks / 47

Reinforcement learning for automatic data quality monitoring in HEP experiments

Authors: Julian Garcia Pardinas[1]; Lorenzo Del Pianta[None]; Maximilian Janisch[None]; Nicola Serra[2]; Olivia Jullian Parra[1]; Suzanne Klaver[3]; Thomas Lehéricy[None]

[1] *CERN*

[2] *University of Zurich (CH)*

[3] *Nikhef National institute for subatomic physics (NL)*

Corresponding Authors: thomas.lehericy@math.uzh.ch, julian.garcia.pardinas@cern.ch, maximilian.janisch@math.uzh.ch, lorenzo.del.pianta.perez@cern.ch, suzanne.klaver@cern.ch, nicola.serra@cern.ch, olivia.jullian.parra@cern.ch

The usage of modern ML techniques to automate the search for anomalies in collider physics is a very active and prolific field. Typical cases are the search for signatures of physics beyond the Standard Model and the identification of problems in the detector systems that would lead to bad-quality data, unusable for physics data analysis. We are interested in the second type of task, which can also be referred to as data-quality monitoring. In large experimental collaborations, this kind of anomaly detection usually relies on large pools of rotating shifters, taken from within the members of the collaboration. Great benefits can be gained by the partial automation of those tasks, in terms of both an increased efficiency for the collection of good data and a reduction in the need for associated person power.

Besides the usual challenges in the detection of anomalies with ML, additional difficulties arise in situations where the nominal experimental conditions are rapidly changing, for example during the period of commissioning of a new detector. In such a case, the algorithms need to be continuously retrained in an efficient manner. Additionally, if the optimisation goal doesn't only look at the data-collection efficiency but includes human factors (for example, trying to reduce the need for redundant shifter actions), the definition of an adequate loss is not trivial. To face these extra challenges, we propose the application of Reinforcement Learning techniques with human feedback to the task of data-quality monitoring.

In this contribution, we describe a simplified simulated setup designed to study the automation of data-quality monitoring in two regimes, "online" and "offline". The "online" one deals with the problem of spotting problems in a detector while data is being collected, aiming at a prompt fixing that will increase the future data-collection efficiency. The "offline" one focuses on the problem of

classifying data that has already been collected as usable or unusable. We present the progress on the application of RL algorithms in those regimes, discuss the performance achieved and identify future lines of work.

Would you like to be considered for an oral presentation?:

Yes

Contributed Talks / 48

Re-simulation-based self-supervision for representation learning

Authors: Benedikt Maier[1]; Jeffrey Krupa[2]; Maurizio Pierini[3]; Michael Kagan[4]; Nathaniel Sherlock Woodward[5]; Philip Coleman Harris[5]

[1] *KIT - Karlsruhe Institute of Technology (DE)*

[2] *Massachusetts Institute of Technology*

[3] *CERN*

[4] *SLAC National Accelerator Laboratory (US)*

[5] *Massachusetts Inst. of Technology (US)*

Corresponding Authors: jeffrey.krupa@cern.ch, maurizio.pierini@cern.ch, michael.aaron.kagan@cern.ch, nathaniel.sherlock.woodward@cern, benedikt.maier@cern.ch, philip.coleman.harris@cern.ch

Self-Supervised Learning (SSL) is at the core of training modern large ML models, providing a scheme for learning powerful representations in base models that can be used in a variety of downstream tasks. However, SSL training strategies must be adapted to the type of training data, thus driving the question: what are powerful SSL strategies for collider physics data? In the talk, we present a novel re-simulation-based SSL (RS3L) strategy wherein we develop a method of "re-simulation" to drive data augmentation for contrastive learning. We show how a RS3L-trained base model can learn powerful representations that can be used for downstream discrimination tasks and can help mitigate uncertainties.

Contributed Talks / 49

Advances in developing deep neural networks for finding primary vertices in proton-proton collisions at the LHC

Authors: Michael David Sokoloff[1]; Sara Shinde[1]; Simon Akar[1]

Co-authors: Elliott Kauffman [2]; Henry Fredrick Schreiner [3]; Michael Peters ; Mohamed Elashri [4]; Rocky Bala Garg [5]; Will Tepe

[1] *University of Cincinnati (US)*

[2] *Princeton University (US)*

[3] *Princeton University*

[4] *University of Cincinnati*

[5] *Stanford University (US)*

Corresponding Authors: shindesc@mail.uc.edu, michael.david.sokoloff@cern.ch, simon.akar@cern.ch

We have been studying the use of deep neural networks (DNNs) to identify and locate primary vertices (PVs) in proton-proton collisions at the LHC. Earlier work focused on finding primary vertices in simulated LHCb data using a hybrid approach that started with kernel density estimators (KDEs) derived from the ensemble of charged track parameters heuristically and predicted "target histogram" proxies from which PV positions are extracted. We have demonstrated that using a UNet architecture performs indistinguishably from a "flat" convolutional neural network model. We have recently developed an "end-to-end" tracks-to-hists DNN that predicts target histograms directly from track parameters using simulated LHCb data that provides better performance (a lower false positive rate for the same high efficiency) than the best KDE-to-hists model studied. This DNN also provides better efficiency than the default heuristic algorithm for the same low false positive rate.

We are currently instantiating the end-to-end tracks-to-hists DNN within the software stack for Allen, LHCb's GPU-resident, first-level software trigger. In this context we are studying the evolution of the tracks-to-hists DNN performances after various levels of pruning and quantization. We will show that high-level performances are maintained even after substantial reduction of model use of compute resource.

Would you like to be considered for an oral presentation?:

Yes

Poster Session / 50

Conditional normalizing flows for correcting simulations

Authors: Caio Cesar Daumann[1]; Davide Valsecchi[2]; Jan Lukas Spaeh[1]; Johannes Erdmann[3]; Massimiliano Galli[2]; Mauro Donega[4]

[1] *Rheinisch Westfaelische Tech. Hoch. (DE)*

[2] *ETH Zurich (CH)*

[3] *RWTH Aachen University*

[4] *E*

Corresponding Authors: johannes.erdmann@cern.ch, donega@cern.ch, janlukas.spaeh@physik.rwth-aachen.de, davide.valsecchi@cern.ch, massimiliano.galli@cern.ch, caio.daumann99@gmail.com

Simulated events are key ingredients for almost all high-energy physics analyses. However, imperfections in the configuration of the simulation often result in mis-modelling and lead to discrepancies between data and simulation. Such mis-modelling often must be taken into account by correction factors accompanied by systematical uncertainties, that can compromise the sensitivity of measurements and searches.

To address this issue, we propose to use normalizing flows, a powerful technique for learning the underlying distributions of input data. We employ a conditional normalizing flow that utilizes as conditions kinematic variables together with a boolean that differentiates between simulation and data. By training the flows with both simulation and data and mapping both distributions to the same latent space, the flow can learn both underlying distributions and map between them using the latent space as an intermediary. Thus, one can map simulated events to the latent space, and then swap the conditional boolean. Due to the invertibility of the flows, one can then map from the latent space to the data space, yielding the corrected values for the variables. The most important innovation of this method is that a single flow is sufficient to morph a multidimensional distribution into another.

We demonstrate that the proposed architecture can transform simulated events into data using a toy example. The toy distributions are inspired by physical distributions, where the variables are generated conditioned on kinematic variables, and the resulting distributions exhibit correlations between themselves. These correlations and kinematic distributions differ for simulation and data distributions. The distributions include non-continuous functions, which are handled with suitable

transformations. We assess the quality of the corrections by training a classifier on the toy data and the corrected simulated events.

Contributed Talks / 54

Finetuning Foundation Models for Joint Analysis Optimization

Authors: Lukas Alexander Heinrich[1]; Matthias Vigl[1]; Nicole Michelle Hartman[2]

[1] *Technische Universitat Munchen (DE)*

[2] *TUM (DE)*

Corresponding Authors: nicole.hartman@cern.ch, matthias.vigl@cern.ch, lukas.heinrich@cern.ch

Most searches at the LHC employ an analysis pipeline consisting of various discrete components, each individually optimized and later combined to provide relevant features used to discriminate SM background from potential signal. These are typically high-level features constructed from particle four-momenta. However, the combination of individually optimized tasks doesn't guarantee an optimal performance on the final analysis objective. In this study, we show how an analysis would benefit from adopting an end-to-end ML optimization approach. Specifically, we investigate the impact of jointly optimizing particle identification and signal vs background discrimination exploiting the transformer-based ParT architecture [arXiv:2202.03772] as foundation model, showing the effectiveness of finetuning in the case of multi jets final states with CMS open data [DOI:10.7483/OPENDATA.CMS.JGJX.MS7Q].

Contributed Talks / 55

Accelerating Graph-Based Tracking with Symbolic Regression

Authors: Carlo Schiavi[1]; Francesco Armando Di Bello[1]; Nathalie Soybelman[2]

Co-author: Eilam Gross [2]

[1] *INFN e Universita Genova (IT)*

[2] *Weizmann Institute of Science (IL)*

Corresponding Authors: francesco.armando.di.bello@cern.ch, nathalie.soybelman@cern.ch, carlo.schiavi@cern.ch

Tracking, the reconstruction of particle trajectories from hits in the inner detector is a computationally intensive task due to the large combinatorics of detector signals. Recent efforts have proven that ML techniques can be successfully applied to the tracking problem, extending and improving the conventional methods based on feature engineering. However, the inference of complex networks can be too slow to be used in the trigger system. Quantising the network and deploying it on an FPGA is feasible but challenging and highly non-trivial. An efficient alternative can employ symbolic regression (SR), which already proved its performance in replacing a dense neural network for jet classification. We propose a novel approach that uses SR to replace a graph-based neural network. Using a simplified toy example, we substitute each network block with a symbolic function, preserving the graph structure of the data and enabling message passing. This approach significantly speeds up inference on a CPU without sacrificing much accuracy.

Contributed Talks / 56

Conditional Set-to-Set Generation for Fast Simulation using Diffusion and Graph-to-Graph Translation

Authors: Dmitrii Kobylianskii[1]; Nathalie Soybelman[1]; Nilotpal Kakati[1]

Co-authors: Eilam Gross [1]; Etienne Dreyer [1]

[1] *Weizmann Institute of Science (IL)*

Corresponding Authors: dmitrii.kobylianskii@cern.ch, nathalie.soybelman@cern.ch, nilotpal.kakati@cern.ch

Simulating particle physics data is a crucial yet computationally expensive aspect of analyzing data at the LHC. Typically, in fast simulation methods, we rely on a surrogate calorimeter model with a subsequent reconstruction algorithm to generate a set of reconstructed objects. This work demonstrates the potential to generate these reconstructed objects in one shot, effectively replacing both the calorimeter simulation and reconstruction steps. Our primary goal in this set-to-set generation is to accurately replicate the detector's resolution and the properties of the reconstructed objects. Building on the success of our previous slot-attention-based model, we introduce two innovative approaches to improve this task and evaluate their performance using a realistic dataset. This dataset incorporates a realistic detector simulation and a machine learning-based reconstruction algorithm. In the first approach, we enhance the slot-attention mechanism with a state-of-the-art graph diffusion model. This entails starting with a noisy graph and progressively eliminating noise conditioned on the truth particle set, ultimately generating the reconstructed objects.
The second approach involves graph refinement, directly converting the set of truth particles into the set of reconstructed objects. These approaches outperform our previous baseline regarding both accuracy and resolution of predicted particle properties.

Would you like to be considered for an oral presentation?:

Poster Session / **57**

Simulation-based inference in the search for CP violation in leptonic WH production

Authors: Marta Fernandes Da Silva[1]; Ricardo Barrué[2]

Co-authors: Ines Ochoa [1]; Patricia Conde Muino [1]; Rui Santos [3]; Valerio Dao [4]

[1] *LIP - Laboratorio de Instrumentação e Física Experimental de Partículas (PT)*

[2] *LIP - Laboratorio de Instrumentacao e Fisica Experimental de Partículas*

[3] *ISEL and CFTC-UL*

[4] *CERN*

Corresponding Authors: ricardo.barrue@cern.ch, ines.ochoa@cern.ch, marta.fernandes.da.silva@cern.ch

Beyond the Standard Model (BSM) sources of CP violation are one of the required ingredients for solving the matter-antimatter puzzle. Simulation-based inference methods hold the promise of allowing the estimation of optimal observables or likelihood ratios without requiring approximations (e.g. of the effect of shower and hadronization), ensuring a high sample efficiency through the use of high-dimensional data and simulator information.

In this work, started in arXiV:2308.02882, we focus on leptonic WH production, a difficult channel due to the presence of a (unobservable) neutrino. We explore the use of such methods and benchmark them against kinematic and angular observables commonly used in experimental analyses.

This work aims at informing analysis strategies for Run 3 and beyond, with the ultimate goal of extracting the best sensitivity to BSM physics out of LHC data.

Would you like to be considered for an oral presentation?:

Contributed Talks / 58

the Fair Universe project and the HiggsML Uncertainty Challenge

Authors: Aishik Ghosh[1]; Ben Nachman[2]; Ben Thorne[3]; Christopher Harris[4]; Daniel Whiteson[1]; David Rousseau[5]; Elham E Khoda[6]; Ihsan Ullah[7]; Isabelle Guyon[None]; Paolo Calafiura[2]; Peter Nugent[3]; Ragansu Chakkappai[8]; Sascha Diefenbacher[2]; Shih-Chieh Hsu[9]; Steven Farrell[10]; Wahid Bhimji[None]; Yuan-Tang Chou[6]

[1] *University of California Irvine (US)*

[2] *Lawrence Berkeley National Lab. (US)*

[3] *LBNL*

[4] *Unknown*

[5] *IJCLab-Orsay*

[6] *University of Washington (US)*

[7] *ChaLearn*

[8] *Université Paris-Saclay (FR)*

[9] *University of Washington Seattle (US)*

[10] *Lawrence Berkeley National Laboratory*

Corresponding Authors: sfarrell@lbl.gov, elham.e.khoda@cern.ch, christopher.harris@cern.ch, wbhimji@lbl.gov, pcalafiura@lbl.gov, benjamin.philip.nachman@cern.ch, guyon@clopinet.com, sdiefenbacher@lbl.gov, aishik.ghosh@cern.ch, schsu@uw.edu, rousseau@lal.in2p3.fr, ragansu.chakkappai@cern.ch, daniel.whiteson@cern.ch, yuan-tang.chou@cern.ch

The Fair Universe project is building a large-compute-scale AI ecosystem for sharing datasets, training large models and hosting challenges and benchmarks. Furthermore, the project is exploiting this ecosystem for an AI challenge series focused on minimizing the effects of systematic uncertainties in High-Energy Physics (HEP), and on predicting accurate confidence intervals. This talk will describe the challenge platform we have developed that builds on the open-source benchmark ecosystem Codabench to interface it to the NERSC HPC center and its Perlmutter system with over 7000 A100 GPUs.

This presentation will also tease the first of our Fair Universe public challenges hosted on this platform, the Fair Universe: HiggsML Uncertainty Challenge, which will apply to be a NeurIPS 2024 competition. Participants will be presented a large training dataset corresponding to H to tau tau cross section measurement at the Large Hadron Collider. They should design an analysis technique able to not just measure the signal strength but to provide a confidence interval, which correct coverage will be evaluated automatically from pseudo-experiments. The confidence interval should include statistical uncertainty and also systematic uncertainties (concerning detector calibration, background levels etc⋯). It is expected that advanced analysis techniques that are able to control the impact of systematics will perform best.

A hackathon that took place in Nov 2023 during the AI and the Uncertainty Challenge in Fundamental Physics Workshop in Paris (see presentation and conclusion) has enabled us to validate the platform and the robustness of the ranking with a simplified prototype of the competition.

The Codabench/NERSC platform allows for hosting challenges also from other communities, and we also intend to make our benchmark designs available as templates so similar efforts can be easily launched in other domains.

Would you like to be considered for an oral presentation?:

Yes

Contributed Talks / 59

Attention to the strengths of physics interactions: Enhanced Deep Learning Event Classification for Particle Physics Experiments

Authors: Polina Moskvitina[1]; Clara Nellist[2]; Rob Verheyen[None]; Roberto Ruiz De Austri[None]; Sascha Caron[1]; Zhongyi Zhang[3]

[1] *Nikhef National institute for subatomic physics (NL)*

[2] *University of Amsterdam and Nikhef (NL)*

[3] *NIKHEF*

Corresponding Authors: polina.moskvitina@cern.ch, robverheyen@gmail.com, shadowquark.yu@gmail.com, rruiz@ific.uv.es, sascha.caron@cern.ch, clara.nellist@cern.ch

A major task in particle physics is the measurement of rare signal processes. These measurements are highly dependent on the classification accuracy of these events in relation to the huge background of other Standard Model processes. Reducing the background by a few tens of percent with the same signal efficiency can already increase the sensitivity considerably.

This study demonstrates the importance of adding physical information (and inductive biases) to these architectures. In addition to the information previously proposed for jet tagging, we add particle measures for energy-dependent particle-particle interaction strengths as predicted by the Feynman rules of the Standard Model (SM). Our work includes this information into different methods for classifying events, in particular Boosted Decision Trees, Transformer Architectures (Particle Transformer) and Graph Neural Networks (Particle Net). We find that the integration of physical information into the attention matrix (transformers) or edges (graphs) notably improves background rejection by 10% to 30% over baseline models (Particle Net), with about 10% of this improvement directly attributable to what we call the SM interaction matrix.

Poster Session / 61

Transformers for Particle Track Reconstruction and Hit Clustering

Authors: Nadezhda Dobreva[None]; Sascha Caron[1]; Yue Zhao[None]; Zef Wolffs[1]

[1] *Nikhef National institute for subatomic physics (NL)*

Corresponding Authors: sascha.caron@cern.ch, nadetodobreva@gmail.com, yue.zhao@surf.nl, zwolffs@nikhef.nl

Track reconstruction is a crucial part of High Energy Physics (HEP) experiments. Traditional methods for the task scale poorly, making machine learning and deep learning appealing alternatives. Following the success of transformers in the field of language processing, we investigate the feasibility of training a Transformer to translate detector signals into track parameters. We study and compare different architectures. Firstly, an autoregressive Transformer model with the original encoder-decoder architecture which reconstructs a particle's trajectory given a few initial hits. Secondly, an encoder-only architecture used as a classifier, producing a class label for each hit in an event, given pre-defined bins within the track parameter space. Lastly, and encoder-only model with the purpose of regressing track parameter values for each hit in an event, followed by clustering.

The Transformer models are benchmarked on simplified datasets generated by the recently developed simulation framework REDuced VIrtual Detector (REDVID) as well as a subset of the TrackML data. The preliminary results of the proposed models show promise for the application of these deep learning techniques on more realistic data for particle reconstruction.

This work has been previously presented at the following conferences: Connecting The Dots 2023 (https://indico.cern.ch/event/1252748/contributions/5521505/), NNV 2023 (https://indico.nikhef.nl/event/4510/contribu and ML4Jets2023 (https://indico.cern.ch/event/1253794/contributions/5588602/).

Poster Session / 65

Search for long-lived heavy neutral leptons using a displaced jet tagger

Authors: CMS Collaboration[None]; Matthias Komm[1]

[1] *Deutsches Elektronen-Synchrotron (DE)*

Corresponding Author: matthias.komm@cern.ch

A search for long-lived heavy neutral leptons (HNLs) is presented, which considers the hadronic final state and coupling scenarios involving all three lepton generations in the 2-20 GeV HNL mass range for the first time. A central feature of the analysis is a novel jet tagger, based on a deep neural network (DNN), that has been developed to identify displaced jets from an HNL decay using various features of the jet and its constituent particles without explicitly requiring the reconstruction of displaced vertices. The DNN uses a domain adaptation technique to ensure accurate performance of the resulting classifier in data. A broad range of HNL lifetimes are probed by parametrizing the DNN as a function of the HNL displacement in the laboratory frame. Contributions from background processes are determined from data. No excess of events in data over the expected background is observed. Upper limits on the HNL production cross section are derived as functions of the HNL mass and the three coupling strengths to each lepton generation and presented as exclusion limits in the coupling-mass plane, as lower limits on the HNL lifetime, and on the HNL mass.

Would you like to be considered for an oral presentation?:

No

Simultaneous calibration of jet energy and mass using DNN in ATLAS

Author: Pierre Antoine Delsart[1]

[1] *LPSC/CNRS (Grenoble, FR)*

Corresponding Author: delsart@in2p3.fr

The energy and mass measurements of jets are crucial tasks for the Large Hadron Collider experiments. This paper presents a new calibration method to simultaneously calibrate these quantities for large-radius jets measured with the ATLAS detector using a deep neural network (DNN). To address the specificities of the calibration problem, special loss functions and training procedures are employed, and a complex network architecture, which includes feature annotation and residual connection layers, is used. The DNN-based calibration is compared to the standard numerical approach in an extensive series of tests. The DNN approach is found to perform significantly better in almost all of the tests and over most of the relevant kinematic phase space. In particular, it consistently improves the energy and mass resolutions, with a 30% better energy resolution obtained for transverse momenta pT>500 GeV.

Would you like to be considered for an oral presentation?:

Yes

Thoughts about Machine Learning

Corresponding Author: juergen.schmidhuber@kaust.edu.sa

I'll discuss past, present, future, and far future of machine learning and artificial intelligence.

Mini-bio. The New York Times headlined: "When A.I. Matures, It May Call Jürgen Schmidhuber 'Dad'." Since age 15, his main goal has been to build a self-improving A.I. smarter than himself, then retire. His lab's deep learning artificial neural networks based on ideas published in the "Annus Mirabilis" 1990-1991 have revolutionised machine learning and A.I. By 2017, they were on over 3 billion smartphones, and used billions of times per day, for Facebook's automatic translation, Google's speech recognition, Google Translate, Apple's Siri & QuickType, Amazon's Alexa, etc. He pioneered generative adversarial networks (1990, now widely used), artificial curiosity, Transformers with linearized self-attention (1991 - Transformers are the basis of the famous ChatGPT), and meta-learning machines that learn to learn (since 1987). Today, the most cited neural networks all build on work done in his labs. Elon Musk tweeted: "Schmidhuber invented everything." He is recipient of numerous awards, Director of the AI Initiative at KAUST in KSA, Scientific Director of the Swiss AI Lab IDSIA, Adj. Prof. of A.I. at Univ. Lugano, and Co-Founder & Chief Scientist of the company NNAISENSE. He is a frequent keynote speaker at major events, and advising various governments on A.I. strategies.

Would you like to be considered for an oral presentation?:

Keynotes ML Research / 69

Modeling epistemics with machine learning

How do we model and reason with machine learning? Machine learning models typically do not build upon mechanistic assumptions. I will discuss important concepts required to gauge machine-learning model outputs, in particular controling their uncertainty and wether they support causal reasonning. I will also discuss implicit modeling (inductive biases), contrasting tree-based models with neural networks.

Would you like to be considered for an oral presentation?:

Keynotes Physics + ML / 73

Collider Physics and ML

Corresponding Author: gregor.kasieczka@cern.ch

Keynotes Physics + ML / 89

Improving data-driven model predictions using physics in the CERN accelerator complex

Corresponding Author: francesco.maria.velotti@cern.ch

Most of limitations to fully exploit the CERN injector complex comes from problems that have a rather known formalism, but usually, due to interplay of many other factors make the analytical or

numerical solutions almost impossible. In this context, machine learning is playing a key role allowing us to produce empirical models to estimate the accelerator behaviour to the desired accuracy for its control and optimisation. In some cases though, data availability is limited or there is the need to extrapolate beyond the training domain. In this context, Physics Informed Machine Learning may hold the key more powerful models. This talk will introduce some of the ongoing work and highlight possible paths for the future.

Would you like to be considered for an oral presentation?:

Keynotes Physics + ML / 90

ML and Molecular Physics

Corresponding Author: jonas.koehler@microsoft.com

Would you like to be considered for an oral presentation?:

CERN Data Science Seminar / 91

Problem solving as a translation task

Neural architectures designed for machine translation can be used to solve problems of mathematics, by considering that solving amounts to translating the problem, a sentence in some mathematical language, into its solution, another sentence in mathematical language. Presenting examples from symbolic and numerical mathematics, and theoretical physics, I show how such techniques can be applied to develop AI for Science, and help understand the inner workings of language models.

Organised as a Data Science Seminar (Indico page)
Video conference link

Would you like to be considered for an oral presentation?:

92

Closeout

Keynotes Physics + ML / 102

ML and Quantum Physics

Corresponding Author: e.greplova@tudelft.nl

Would you like to be considered for an oral presentation?:

Keynote Industry / 103

Retrieval Augmented Language Models

Corresponding Author: casey@contextual.ai

Would you like to be considered for an oral presentation?:

Tutorials / 104

Thinking like Transformers

Corresponding Author: gail.weiss@epfl.ch

Transformers - the purely attention based NN architecture - have emerged as a powerful tool in sequence processing. But how does a transformer think? When we discuss the computational power of RNNs, or consider a problem that they have solved, it is easy for us to think in terms of automata and their variants (such as counter machines and pushdown automata). But when it comes to transformers, no such intuitive model is available.

In this tutorial I will present a programming language, RASP (Restricted Access Sequence Processing), which we hope will serve the same purpose for transformers as finite state machines do for RNNs. In particular, we will discuss the transformer architecture, identify its base components, and abstract them into a small number of primitives which we will then compose into a small programming language: RASP. We will go through some example programs in the language, and discuss how a given RASP program relates to the transformer architecture.

Would you like to be considered for an oral presentation?:

Poster Session / 105

The DL Advocate: Playing the devil's advocate with hidden systematic uncertainties - Poster

Authors: Aleksandr Iniukhin[1]; Andrea Mauri[2]; Andrei Golutvin[3]; Andrey Ustyuzhanin[4]; Guillermo Hijano Mendizabal[5]; Nicola Serra[5]; Patrick Haworth Owen[5]

[1] *National Research University Higher School of Economics (RU)*

[2] *Imperial College (GB)*

[3] *Imperial College London*

[4] *Yandex School of Data Analysis (RU)*

[5] *University of Zurich (CH)*

Corresponding Authors: aleksandr.iniukhin@cern.ch, a.mauri@cern.ch, andrey.ustyuzhanin@cern.ch, patrick.haworth.owen@cern.ch, andrey.goloutvin@cern.ch, nicola.serra@cern.ch, guillermo.hijano@cern.ch

We propose a new method based on machine learning to play the devil's advocate and investigate the impact of unknown systematic effects in a quantitative way. This method proceeds by reversing the measurement process and using the physics results to interpret systematic effects under the Standard Model hypothesis. We explore this idea with two alternative approaches, one relies on a combination of gradient descent and optimisation techniques, the other employs reinforcement learning. We illustrate the potentiality of the presented method by considering two examples, firstly the case of a

branching fraction measurement of the decay of a b-hadron, secondly the determination of the P_5' angular observable in $B^0 \to \mu^+\mu^-$ decays. Based on https://arxiv.org/abs/2303.15956

Would you like to be considered for an oral presentation?:

Poster Session / 106

Longitudinal Beam Diagnostics and Phase Space Reconstruction in the LHC Using ML - Poster

Authors: Alexandre Lasheen[1]; Georges Trad[1]; Helga Timko[1]; Konstantinos Iliakis[1]; Simon Albright[1]; Theodoros Argyropoulos[1]

[1] *CERN*

Corresponding Authors: konstantinos.iliakis@cern.ch, georges.trad@cern.ch, theodoros.argyropoulos@cern.ch, alexandre.lasheen@cern.ch, simon.albright@cern.ch, helga.timko@cern.ch

Accurate knowledge of longitudinal beam parameters is essential for optimizing the performance and operational efficiency of particle accelerators like the Large Hadron Collider (LHC). However, conventional methods to determine them, such as fitting techniques and tracking-based longitudinal tomography, are time-consuming and limited to analyzing data from a few bunches only. To address this, we propose the development of a machine learning (ML) model that leverages the existing high-resolution measurements of longitudinal bunch profiles and utilizes an encoder-decoder architecture to achieve two primary objectives. Firstly, it efficiently extracts the physical beam parameters, such as injection errors, bunch length, and bunch intensity, eliminating the need for computationally expensive fitting methods. Secondly, it reconstructs the longitudinal beam distribution. The ML model is designed to operate in real-time, enabling online monitoring of multi-bunch beams. This application demonstrates the potential of ML-techniques in enhancing beam diagnostics and allowing more precise control of large particle accelerators.

Would you like to be considered for an oral presentation?:

Poster Session / 107

A Deep Generative Model for Hadronization - Poster

Authors: Adam Kania[None]; Andrzej Konrad Siodmok[1]; Ben Nachman[2]; Jay Chan[3]; Xiangyang Ju[2]

[1] *Jagiellonian University (PL)*

[2] *Lawrence Berkeley National Lab. (US)*

[3] *Lawrence Berkeley National Laboratory*

Corresponding Authors: a.siodmok@cern.ch, benjamin.philip.nachman@cern.ch, jay.chan@cern.ch, xiangyang.ju@cern.ch, ad.kania@student.uj.edu.pl

Hadronization is a critical step in the simulation of high-energy particle and nuclear physics experiments. As there is no first principles understanding of this process, physically-inspired hadronization models have a large number of parameters that are fit to data. We propose an alternative approach that uses deep generative models, which are a natural replacement for classical techniques, since they are more flexible and may be able to improve the overall precision. We first demonstrate using neural networks to emulate specific hadronization when trained using the inputs and outputs of classical methods. A protocol is then developed to fit a deep generative hadronization model in a realistic setting, where we only have access to a set of hadrons in data. Finally, we build a deep generative hadronization model that includes both kinematic (continuous) and flavor (discrete) degrees of

freedom. Our approach is based on Generative Adversarial Networks and we show the performance within the context of the cluster model within the Herwig event generator.

Would you like to be considered for an oral presentation?:

Poster Session / **108**

Accelerating the search for mass bumps using the Data-Directed Paradigm - Poster

Authors: Ali El Moussaouy[1]; Amit Shkuri[2]; Bruna Pascual[1]; Emile Baril[1]; Etienne Dreyer[2]; Eva Mayer[3]; Fannie Bilodeau[1]; Georges Azuelos[1]; Jean-Francois Arguin[1]; Julien Noce Donini[3]; Nilotpal Kakati[2]; Samuel Calvet[4]; Shikma Bressler[2]

[1] *Universite de Montreal (CA)*

[2] *Weizmann Institute of Science (IL)*

[3] *Université Clermont Auvergne (FR)*

[4] *LPC Clermont-Ferrand, IN2P3 / UCA (FR)*

Corresponding Authors: jean.francois.arguin@gmail.com, bruna.pascual.dias@cern.ch, julien.donini@cern.ch, samuel.pierre.jean.calvet@cern.ch, nilotpal.kakati@cern.ch, amit.shkuri@cern.ch, emile.baril@cern.ch, shikma.bressler@cern.ch, georges.azuelos@cern.ch, etienne.dreyer@cern.ch, ali.el.moussaouy@cern.ch, eva.mayer@cern.ch, fannie.bilodeau@cern.ch

The Data-Directed paradigm (DDP) is a search strategy for efficiently probing new physics in a large number of spectra with smoothly-falling SM backgrounds. Unlike the traditional analysis strategy, DDP avoids the need for a simulated or functional-form based background estimate by directly predicting the statistical significance using a convolutional neural network trained to regress the log-likelihood-based significance. In this way, a trained network is used to identify mass bumps directly on data. By saving a considerable amount of time, this approach has the potential to expand the discovery reach by checking many unexplored regions. The method has shown good performance when finding various beyond standard model particles in simulation data. A description of the method and recent developments will be presented.

Would you like to be considered for an oral presentation?:

Poster Session / **109**

Advances in developing deep neural networks for finding primary vertices in proton-proton collisions at the LHC - Poster

Authors: Michael David Sokoloff[1]; Sara Shinde[1]; Simon Akar[1]

Co-authors: Elliott Kauffman [2]; Henry Fredrick Schreiner [3]; Michael Peters ; Mohamed Elashri [4]; Rocky Bala Garg [5]; Will Tepe

[1] *University of Cincinnati (US)*

[2] *Princeton University (US)*

[3] *Princeton University*

[4] *University of Cincinnati*

[5] *Stanford University (US)*

Corresponding Authors: simon.akar@cern.ch, michael.david.sokoloff@cern.ch, shindesc@mail.uc.edu

We have been studying the use of deep neural networks (DNNs) to identify and locate primary vertices (PVs) in proton-proton collisions at the LHC. Earlier work focused on finding primary vertices in simulated LHCb data using a hybrid approach that started with kernel density estimators (KDEs) derived from the ensemble of charged track parameters heuristically and predicted "target histogram" proxies from which PV positions are extracted. We have demonstrated that using a UNet architecture performs indistinguishably from a "flat" convolutional neural network model. We have recently developed an "end-to-end" tracks-to-hists DNN that predicts target histograms directly from track parameters using simulated LHCb data that provides better performance (a lower false positive rate for the same high efficiency) than the best KDE-to-hists model studied. This DNN also provides better efficiency than the default heuristic algorithm for the same low false positive rate.

We are currently instantiating the end-to-end tracks-to-hists DNN within the software stack for Allen, LHCb's GPU-resident, first-level software trigger. In this context we are studying the evolution of the tracks-to-hists DNN performances after various levels of pruning and quantization. We will show that high-level performances are maintained even after substantial reduction of model use of compute resource.

Would you like to be considered for an oral presentation?:

Yes

Poster Session / **110**

DeepTreeGANv2: Iterative Pooling of Point Clouds - Poster

Author: Moritz Scham[1]

Co-authors: Benno Kach [1]; Dirk Krucker [1]; Kerstin Borras [2]; Simon Schnake [1]

[1] *Deutsches Elektronen-Synchrotron (DE)*

[2] *DESY / RWTH Aachen University*

Corresponding Author: moritz.scham@cern.ch

In High Energy Physics, detailed and time-consuming simulations are used for particle interactions with detectors. To bypass these simulations with a generative model, the generation of large point clouds in a short time is required, while the complex dependencies between the particles must be correctly modelled. Particle showers are inherently tree-based processes, as each particle is produced by the decay or detector interaction of a particle of the previous generation.
In this work, we present a significant extension to DeepTreeGAN, featuring a critic, that is able to aggregate such point clouds iteratively in a tree-based manner. We show that this model can reproduce complex distributions, and we evaluate its performance on the public JetNet 150 and CaloChallange datasets.
Two papers under review are available here:
https://cernbox.cern.ch/s/WoNDFxdOiKpg0fG

Would you like to be considered for an oral presentation?:

Poster Session / **111**

Out-of-Distribution Multi-set Generation with Context Extrapolation for Amortized Simulation and Inverse Problems - Poster

Author: Hosein Hashemi[1]

[1] *LMU Munich*

Addressing the challenge of Out-of-Distribution (OOD) multi-set generation, we introduce Yone-daVAE, a novel equivariant deep generative model inspired by Category Theory, motivating the Yoneda-Pooling mechanism. This approach presents a learnable Yoneda Embedding to encode the relationships between objects in a category, providing a dynamic and generalizable representation of complex relational data sets. YonedaVAE introduces a self-distilled set generator, capable of zero-shot creating sets with variable inter-category and intra-category cardinality, facilitated by the new Adaptive Top-p Sampling. We demonstrate that YonedaVAE can produce new point clouds with cardinalities well beyond the training data and achieve context extrapolation. Trained on low luminosity ultra-high-granularity data of PXD at Belle II, YonedaVAE can generate high luminosity valid signatures with the correct intra-event correlation without exposure to similar data during training. Being able to generalize to out-of-distribution samples, YonedaVAE stands as a valuable method for extrapolative multi-set generation tasks in scientific discovery, including de novo protein design, Drug Discovery, and simulating geometry-independent detector responses beyond experimental limits.

Would you like to be considered for an oral presentation?:

Yes

Poster Session / **112**

Differentiable Vertex Fitting for Jet Flavour Tagging - Poster

Authors: Ines Ochoa[1]; Jonathan Shoemaker[2]; Michael Kagan[2]; Rachel Emma Clarke Smith[2]; Ruben Miguel De Almeida Inacio[1]

[1] *LIP - Laboratorio de Instrumentação e Física Experimental de Partículas (PT)*

[2] *SLAC National Accelerator Laboratory (US)*

Corresponding Authors: ines.ochoa@cern.ch, rachel.smith@cern.ch, michael.aaron.kagan@cern.ch, jshoe@slac.stanford.edu, ruben.inacio@tecnico.ulisboa.pt

We propose a differentiable vertex fitting algorithm that can be used for secondary vertex fitting, and that can be seamlessly integrated into neural networks for jet flavour tagging. Vertex fitting is formulated as an optimization problem where gradients of the optimized solution vertex are defined through implicit differentiation and can be passed to upstream or downstream neural network components for network training. More broadly, this is an application of differentiable programming to integrate physics knowledge into neural network models in high energy physics. We demonstrate how differentiable secondary vertex fitting can be integrated into larger transformer-based models for flavour tagging and improve heavy flavour jet classification.

Would you like to be considered for an oral presentation?:

Poster Session / **113**

Reinforcement Learning Algorithms for Charged Particle Tracking with Applications in Proton Computed Tomography - Poster

Author: Tobias Kortus[None]

Co-authors: Nicolas Gauger ; Ralf Keidel [1]

[1] *Fachhochschule Worms (DE)*

Corresponding Author: kortus@ztt.hs-worms.de

Deep learning, especially graph neural networks, significantly improved tracking performances in modern particle detectors while reducing runtimes compared to previous state of the art approaches. However, training neural networks requires significant amount of labeled data, usually acquired by performing complex particle simulations. We present first studies of leveraging deep reinforcement learning (RL) and constrained multi-agent reinforcement learning (MARL) as ground-truth free alternatives to supervised learning.

Instead of minimizing a loss function based on ground-truth, we optimize by trial-and-error behavior policies, acting as approximations to the full combinatorial optimization problem, maximizing the physical plausibility of sampled track candidates. Our approaches works on graph-structured data, capturing track hypotheses through edge connections between particles in the detector layers. We demonstrate, on simulated data for a particle detector used for proton computed tomography, the high potential as well as the competitiveness of RL for both single-agent as well as multi-agent settings.

Partially based on: T. Kortus, R. Keidel, N. R. Gauger, on behalf of the Bergen pCT collaboration, "Towards Neural Charged Particle Tracking in Digital Tracking Calorimeters with Reinforcement Learning," in IEEE Transactions on Pattern Analysis and Machine Intelligence, doi: 10.1109/TPAMI.2023.3305027.

Would you like to be considered for an oral presentation?:

Poster Session / 114

Electron and Proton Classification with AMS ECAL Using Convolutional Vision Transformers and Domain Adaptation - Poster

Authors: Berk Turk[1]; Bilge Demirkoz[1]; Emre Akbas[1]; Zhili Weng[2]; Gulce Karagoz[1]

[1] *Middle East Technical University (TR)*

[2] *Massachusetts Inst. of Technology (US)*

Corresponding Authors: gulce.karagoz@cern.ch, bilge.demirkoz@cern.ch, emre.akbas@cern.ch, zhili.weng@cern.ch, berk.turk@cern.ch

Alpha Magnetic Spectrometer (AMS-02) is a precision high-energy cosmic-ray experiment on the ISS operating since 2011 and has collected more than 228 billion particles. Among them, positrons are important to understand the particle nature of dark matter. Separating the positrons from cosmic background protons is challenging above 1 TeV. Therefore, we use state-of-the-art convolutional and transformer models, CoAtNet and Convolutional Vision Transformer (CvT), that employ the shower signals from the ECAL to classify the electrons/positrons in the dominant cosmic proton background. We created sets of electrons, positrons, and protons events from the ISS data and Monte Carlo Simulation in the energy range between 0.2-2 TeV by applying various data quality cuts on reconstructed variables obtained from the subdetectors. Initially, since ECAL showers are not tunned in the AMS MC, our MC trained models show a lower proton rejection on the ISS data. To accommodate the difference between the training and test domain distributions, we implemented domain adaptation with the CoAtNet and CvT to mitigate this dataset bias/domain shift. We also trained domain adaptation with a set of well-reconstructed 1 electron charge ISS events without electron/proton labels at TeV energy order as the target dataset. We evaluated the models between 1-2 TeV energy using ISS and MC events with the proton rejection vs. electron efficiency and proton rejection vs. energy at near 90% electron efficiency plots. We performed experiments using various training and validation dataset combinations and other hyperparameters with the CvT and CoAtNet. Among them, the best models are obtained with the 1-2TeV MC events as training data and half of the labeled 1-2 TeV ISS events as validation data. Using domain adaptation with the CoAtNet, we obtained a maximum proton rejection at 88% electron efficiency on the ISS data. We also rejected all of the MC protons at higher than 99.8% electron efficiency with both CvT and CoAtNet. At 90% electron efficiency, the

proton rejection power of the CvT and CoAtNet is 5 and 7 times higher than the proton rejection power of the AMS's Boosted Decision Tree and ECAL Likelihood Estimator for MC events in the 1-2 TeV range.

Would you like to be considered for an oral presentation?:

Poster Session / 115

Training and optimisation of large transformer models at CERN: an ATLAS case study on Kubeflow - Poster

Author: Maxence Draguet[1]

[1] *University of Oxford (GB)*

Corresponding Author: maxence.draguet@physics.ox.ac.uk

Heavy flavour jets underpin a large part of the ATLAS physics programme, such as analyses of Higgs boson decays to quarks and super-symmetry searches with b-jets. The algorithms for identifying jets originating from b- and c-quarks are instrumental in these efforts, with the recently introduced GN2 model 1 showing remarkable improvements in tagging efficiency. Given its complexity and data demands, high-performance GPU clusters are essential for training GN2. Unfortunately, many within the collaboration lack such resources, emphasising the need for equitable project access. Additionally, the performance of GN2 can be improved through further optimisation of its hyperparameters, an even more computationally demanding task that can be automated with frameworks like Katib. Addressing these two challenges, the ATLAS flavour tagging group is assessing training on CERN IT's Kubeflow infrastructure for machine learning (ml.cern.ch) backed by Kubernetes. This talk will showcase a framework for CERN users to utilise these resources and present the initial results of this effort. Furthermore, the talk will highlight a cutting-edge approach to optimise hyperparameters using μTransfer 2, a deep learning technique recently developed to optimise large language models by zero-shot transferring the performance from lower-complexity models to their full-complexity equivalent. While centred on an ATLAS use case, the methodology presented will be relevant to any collaboration employing advanced ML.

1 ATLAS Collaboration, Public plots for MC/MC and Data/MC comparisons of Dl1d and GN1, and simulation performance of GN2, (2023), https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2023-01/

2 Yang, Greg, et al., Tensor programs v: Tuning large neural networks via zero-shot hyperparameter transfer. arXiv preprint arXiv:2203.03466 (2022).

Would you like to be considered for an oral presentation?:

Poster Session / 116

Unweighted event generation with matrix element surrogates - Poster

Authors: Daniel Maître[None]; Frank Siegert[None]; Henry Truong[None]; Steffen Schumann[None]; Timo Janssen[None]

Corresponding Author: timo.janssen@theorie.physik.uni-goettingen.de

We show that employing a sophisticated neural network emulation of QCD multijet matrix elements based on dipole factorisation can lead to a drastic acceleration of unweighted event generation in high-multiplicity LHC production processes. We incorporate these emulations as fast and accurate

surrogates in a two-stage rejection sampling algorithm within the SHERPA Monte Carlo that yields unbiased unweighted events suitable for phenomenological analyses and post-processing in experimental workflows, e.g. as input to a time-consuming detector simulation. For the computational cost of unweighted events we achieve a reduction by factors between 16 and 350 for the considered channels. We also show how this technique can be used for NLO calculations with emulated loop amplitudes.

SciPost Phys. 15, 107 (2023), arXiv:2301.13562

Would you like to be considered for an oral presentation?:

Poster Session / 117

End-to-end Reconstruction Algorithm for Highly Granular Calorimeters - Poster

Authors: Philipp Zehetner[1]; Jan Kieseler[2]; Shah Rukh Qasim[3]; Dolores Garcia[2]; Kenneth Long[4]

[1] *Ludwig Maximilians Universitat (DE)*

[2] *CERN*

[3] *University of Zurich (CH)*

[4] *Massachusetts Inst. of Technology (US)*

Corresponding Authors: kenneth.long@cern.ch, maria.dolores.garcia.marti@cern.ch, shah.rukh.qasim@cern.ch, philipp.zehetner@cern.ch, jan.kieseler@cern.ch

We present an end-to-end reconstruction algorithm for highly granular calorimeters that includes track information to aid the reconstruction of charged particles. The algorithm starts from calorimeter hits and reconstructed tracks, and outputs a coordinate transformation in which all shower objects are well separated from each other, and in which clustering becomes trivial. Shower properties such as particle ID and energy are predicted from representative points within showers. This is achieved using an extended version of the object condensation loss, a graph segmentation technique that allows the clustering of a variable number of showers in every event while simultaneously performing regression and classification tasks. The backbone is an architecture based on a newly-developed translation-equivariant version of GravNet layers. These dynamically build learnable graphs from input data to exchange information along their edges. The model is trained on data from a simulated detector that matches the complexity of the CMS high-granularity calorimeter (HGCAL), for which it can be retrained in the future.

Would you like to be considered for an oral presentation?:

Poster Session / 118

Modeling N_{ch} distributions and p_{T} spectra in high-energy pp collisions with DNNs - Poster

Author: Maria Alejandra Calmon Behling[1]

[1] *Goethe University Frankfurt (DE)*

Corresponding Author: maria.calmon@cern.ch

During the data-taking campaigns Run 1 and Run 2 of the Large Hadron Collider (LHC), the ALICE collaboration collected a large amount of proton-proton (pp) collisions across a variety of center-of-mass energies ($\sqrt{s}$). This extensive dataset is well suited to study the energy dependence of particle

production. Deep neural networks (DNNs) provide a powerful regression tool to capture underlying multidimensional correlations inherent in the data. In this contribution, DNNs are used to parameterize recent ALICE measurements of charged-particle multiplicity (N_{ch}) distributions and transverse momentum (p_T) spectra. The model architectures are defined and validated using a Bayesian-Optimization hyperparameter search on PYTHIA simulations for a wide $\sqrt{s}$ range and then trained on the ALICE data. An ensemble method is used to predict the observables of interest, extrapolating the measurements towards higher N_{ch} and p_T values as well as to unmeasured $\sqrt{s}$ from 0.5 to 100 TeV. We demonstrate that the predicted p_T spectra can serve as a reference for future heavy-ion measurements, e.g. the O–O campaign planned in LHC Run 3, where no dedicated pp data-taking at the same $\sqrt{s}$ is currently foreseen.

Would you like to be considered for an oral presentation?:

Poster Session / **119**

Masked particle modelling - Poster

Authors: Johnny Raine[1]; Lukas Alexander Heinrich[2]; Matthew Leigh[3]; Michael Kagan[4]; Rita Osadchy[5]; Samuel Byrne Klein[1]; Tobias Golling[1]

[1] *Universite de Geneve (CH)*

[2] *Technische Universitat Munchen (DE)*

[3] *University of Geneva*

[4] *SLAC National Accelerator Laboratory (US)*

[5] *University of Haifa*

Corresponding Authors: michael.aaron.kagan@cern.ch, samuel.byrne.klein@cern.ch, johnny.raine@cern.ch, matthew.leigh@cern.ch, lukas.heinrich@cern.ch, tobias.golling@unige.ch

The Bert pretraining paradigm has proven to be highly effective in many domains including natural language processing, image processing and biology. To apply the Bert paradigm the data needs to be described as a set of tokens, and each token needs to be labelled. To date the Bert paradigm has not been explored in the context of HEP. The samples that form the data used in HEP can be described as a set of particles (tokens) where each particle is represented as a continuous vector. We explore different approaches for discretising/labelling particles such that the Bert pretraining can be performed and demonstrate the utility of the resulting pretrained models on common downstream HEP tasks.

Would you like to be considered for an oral presentation?:

Poster Session / **120**

Parametrising profiled likelihoods with neural networks - Poster

Authors: Humberto Reyes-González[1]; Jack Y. Araz[2]; Wolfgang Waltenberger[3]

[1] *University of Genoa*

[2] *Jefferson Lab*

[3] *Austrian Academy of Sciences (AT)*

Corresponding Authors: jackaraz@jlab.org, harg.zeppelin@gmail.com, wolfgang.waltenberger@gmail.com

Full statistical models encapsulate the complete information of an experimental result, including the likelihood function given observed data. Since a few years ago ATLAS started publishing statistical

models that can be reused via the pyhf framework; a major step towards fully publishing LHC results. In the case of fast Simplified Model Spectra based reinterpretation we are often only interested in the profiled likelihood given a signal strength. However, their computation using pyhf take the order of seconds per parameter point, slowing down SMS reinterpretation by orders of magnitude. Thus, to fully leverage from the precision obtained from full statistical models without compromising speed, we propose to learn the profiled likelihood functions with Neural Networks (NNs). We show that such functions can be well described with simple NNs, published in the ONNX format, and easily used by different reinterpretation tools.

Would you like to be considered for an oral presentation?:

Poster Session / 121

Reinforcement learning for automatic data quality monitoring in HEP experiments - Poster

Authors: Julian Garcia Pardinas[1]; Lorenzo Del Pianta[None]; Maximilian Janisch[None]; Nicola Serra[2]; Olivia Jullian Parra[1]; Suzanne Klaver[3]; Thomas Lehéricy[None]

[1] CERN

[2] University of Zurich (CH)

[3] Nikhef National institute for subatomic physics (NL)

Corresponding Authors: olivia.jullian.parra@cern.ch, nicola.serra@cern.ch, suzanne.klaver@cern.ch, lorenzo.del.pianta.perez@cern.ch, maximilian.janisch@math.uzh.ch, julian.garcia.pardinas@cern.ch, thomas.lehericy@math.uzh.ch

The usage of modern ML techniques to automate the search for anomalies in collider physics is a very active and prolific field. Typical cases are the search for signatures of physics beyond the Standard Model and the identification of problems in the detector systems that would lead to bad-quality data, unusable for physics data analysis. We are interested in the second type of task, which can also be referred to as data-quality monitoring. In large experimental collaborations, this kind of anomaly detection usually relies on large pools of rotating shifters, taken from within the members of the collaboration. Great benefits can be gained by the partial automation of those tasks, in terms of both an increased efficiency for the collection of good data and a reduction in the need for associated person power.

Besides the usual challenges in the detection of anomalies with ML, additional difficulties arise in situations where the nominal experimental conditions are rapidly changing, for example during the period of commissioning of a new detector. In such a case, the algorithms need to be continuously retrained in an efficient manner. Additionally, if the optimisation goal doesn't only look at the data-collection efficiency but includes human factors (for example, trying to reduce the need for redundant shifter actions), the definition of an adequate loss is not trivial. To face these extra challenges, we propose the application of Reinforcement Learning techniques with human feedback to the task of data-quality monitoring.

In this contribution, we describe a simplified simulated setup designed to study the automation of data-quality monitoring in two regimes, "online" and "offline". The "online" one deals with the problem of spotting problems in a detector while data is being collected, aiming at a prompt fixing that will increase the future data-collection efficiency. The "offline" one focuses on the problem of classifying data that has already been collected as usable or unusable. We present the progress on the application of RL algorithms in those regimes, discuss the performance achieved and identify future lines of work.

Would you like to be considered for an oral presentation?:

Yes

Poster Session / **122**

Re-simulation-based self-supervision for representation learning - Poster

Authors: Benedikt Maier[1]; Jeffrey Krupa[2]; Maurizio Pierini[3]; Michael Kagan[4]; Nathaniel Sherlock Woodward[5]; Philip Coleman Harris[5]

[1] *KIT - Karlsruhe Institute of Technology (DE)*

[2] *Massachusetts Institute of Technology*

[3] *CERN*

[4] *SLAC National Accelerator Laboratory (US)*

[5] *Massachusetts Inst. of Technology (US)*

Corresponding Authors: philip.coleman.harris@cern.ch, benedikt.maier@cern.ch, nathaniel.sherlock.woodward@cern.ch, michael.aaron.kagan@cern.ch, maurizio.pierini@cern.ch, jeffrey.krupa@cern.ch

Self-Supervised Learning (SSL) is at the core of training modern large ML models, providing a scheme for learning powerful representations in base models that can be used in a variety of downstream tasks. However, SSL training strategies must be adapted to the type of training data, thus driving the question: what are powerful SSL strategies for collider physics data? In the talk, we present a novel re-simulation-based SSL (RS3L) strategy wherein we develop a method of "re-simulation" to drive data augmentation for contrastive learning. We show how a RS3L-trained base model can learn powerful representations that can be used for downstream discrimination tasks and can help mitigate uncertainties.

Would you like to be considered for an oral presentation?:

Poster Session / **123**

Finetuning Foundation Models for Joint Analysis Optimization - Poster

Authors: Lukas Alexander Heinrich[1]; Matthias Vigl[1]; Nicole Michelle Hartman[2]

[1] *Technische Universitat Munchen (DE)*

[2] *TUM (DE)*

Corresponding Authors: lukas.heinrich@cern.ch, matthias.vigl@cern.ch, nicole.hartman@cern.ch

Most searches at the LHC employ an analysis pipeline consisting of various discrete components, each individually optimized and later combined to provide relevant features used to discriminate SM background from potential signal. These are typically high-level features constructed from particle four-momenta. However, the combination of individually optimized tasks doesn't guarantee an optimal performance on the final analysis objective. In this study, we show how an analysis would benefit from adopting an end-to-end ML optimization approach. Specifically, we investigate the impact of jointly optimizing particle identification and signal vs background discrimination exploiting the transformer-based ParT architecture [arXiv:2202.03772] as foundation model, showing the effectiveness of finetuning in the case of multi jets final states with CMS open data [DOI:10.7483/OPENDATA.CMS.JGJX.MS7Q].

Would you like to be considered for an oral presentation?:

Poster Session / **124**

the Fair Universe project and the HiggsML Uncertainty Challenge - Poster

Authors: Aishik Ghosh[1]; Ben Nachman[2]; Ben Thorne[3]; Christopher Harris[4]; Daniel Whiteson[1]; David Rousseau[5]; Elham E Khoda[6]; Ihsan Ullah[7]; Isabelle Guyon[None]; Paolo Calafiura[2]; Peter Nugent[3]; Ragansu Chakkappai[8]; Sascha Diefenbacher[2]; Shih-Chieh Hsu[9]; Steven Farrell[10]; Wahid Bhimji[None]; Yuan-Tang Chou[6]

[1] *University of California Irvine (US)*

[2] *Lawrence Berkeley National Lab. (US)*

[3] *LBNL*

[4] *Unknown*

[5] *IJCLab-Orsay*

[6] *University of Washington (US)*

[7] *ChaLearn*

[8] *Université Paris-Saclay (FR)*

[9] *University of Washington Seattle (US)*

[10] *Lawrence Berkeley National Laboratory*

Corresponding Authors: ragansu.chakkappai@cern.ch, yuan-tang.chou@cern.ch, daniel.whiteson@cern.ch, rousseau@lal.in2p3.fr, schsu@uw.edu, aishik.ghosh@cern.ch, sdiefenbacher@lbl.gov, guyon@clopinet.com, benjamin.philip.nachman@cern.ch, pcalafiura@lbl.gov, wbhimji@lbl.gov, christopher.harris@cern.ch, elham.e.khoda@cern.ch, sfarrell@lbl.gov

The Fair Universe project is building a large-compute-scale AI ecosystem for sharing datasets, training large models and hosting challenges and benchmarks. Furthermore, the project is exploiting this ecosystem for an AI challenge series focused on minimizing the effects of systematic uncertainties in High-Energy Physics (HEP), and on predicting accurate confidence intervals. This talk will describe the challenge platform we have developed that builds on the open-source benchmark ecosystem Codabench to interface it to the NERSC HPC center and its Perlmutter system with over 7000 A100 GPUs.

This presentation will also tease the first of our Fair Universe public challenges hosted on this platform, the Fair Universe: HiggsML Uncertainty Challenge, which will apply to be a NeurIPS 2024 competition. Participants will be presented a large training dataset corresponding to H to tau tau cross section measurement at the Large Hadron Collider. They should design an analysis technique able to not just measure the signal strength but to provide a confidence interval, which correct coverage will be evaluated automatically from pseudo-experiments. The confidence interval should include statistical uncertainty and also systematic uncertainties (concerning detector calibration, background levels etc···). It is expected that advanced analysis techniques that are able to control the impact of systematics will perform best.

A hackathon that took place in Nov 2023 during the AI and the Uncertainty Challenge in Fundamental Physics Workshop in Paris (see presentation and conclusion) has enabled us to validate the platform and the robustness of the ranking with a simplified prototype of the competition.

The Codabench/NERSC platform allows for hosting challenges also from other communities, and we also intend to make our benchmark designs available as templates so similar efforts can be easily launched in other domains.

Would you like to be considered for an oral presentation?:

Yes

Poster Session / **125**

Attention to the strengths of physics interactions: Enhanced Deep Learning Event Classification for Particle Physics Experiments - Poster

Authors: Polina Moskvitina[1]; Clara Nellist[2]; Rob Verheyen[None]; Roberto Ruiz De Austri[None]; Sascha Caron[1]; Zhongyi Zhang[3]

[1] *Nikhef National institute for subatomic physics (NL)*

[2] *University of Amsterdam and Nikhef (NL)*

[3] *NIKHEF*

Corresponding Authors: clara.nellist@cern.ch, sascha.caron@cern.ch, rruiz@ific.uv.es, shadowquark.yu@gmail.com, robverheyen@gmail.com, polina.moskvitina@cern.ch

A major task in particle physics is the measurement of rare signal processes. These measurements are highly dependent on the classification accuracy of these events in relation to the huge background of other Standard Model processes. Reducing the background by a few tens of percent with the same signal efficiency can already increase the sensitivity considerably.

This study demonstrates the importance of adding physical information (and inductive biases) to these architectures. In addition to the information previously proposed for jet tagging, we add particle measures for energy-dependent particle-particle interaction strengths as predicted by the Feynman rules of the Standard Model (SM). Our work includes this information into different methods for classifying events, in particular Boosted Decision Trees, Transformer Architectures (Particle Transformer) and Graph Neural Networks (Particle Net). We find that the integration of physical information into the attention matrix (transformers) or edges (graphs) notably improves background rejection by 10% to 30% over baseline models (Particle Net), with about 10% of this improvement directly attributable to what we call the SM interaction matrix.

Would you like to be considered for an oral presentation?:

Tutorials / 126

Quantum machine learning

Corresponding Author: sofia.vallecorsa@cern.ch

Would you like to be considered for an oral presentation?:

127

Workshop welcome and introduction

Corresponding Author: julian.garcia.pardinas@cern.ch