CaloChallenge Workshop

Report dei Contributi

ID contributo: **1** Tipo: **non specificato**

Score-based Generative Models for Calorimeter Shower Simulation

martedì 30 maggio 2023 16:30 (30 minuti)

Diffusion generative models are a new class of generative algorithms that have been shown to produce realistic images even in high dimensional spaces, currently surpassing other state-of-the-art models for different benchmark categories and applications. In this work we introduce CaloScore, a score-based generative model for collider physics applied to calorimeter shower generation. Three different diffusion models are investigated using the Fast Calorimeter Simulation Challenge 2022 dataset. CaloScore is the first application of a score-based generative model in collider physics and is able to produce high-fidelity calorimeter images for all datasets, providing an alternative paradigm for calorimeter shower simulation.

Autori principali: NACHMAN, Benjamin (Lawrence Berkeley National Laboratory); MIKUNI, Vinicius (LBNL)

Relatore: MIKUNI, Vinicius (LBNL)

Classifica Sessioni: Diffusion

ID contributo: **2** Tipo: **non specificato**

CaloMan: Fast generation of calorimeter showers with density estimation on learned manifolds

martedì 30 maggio 2023 15:00 (30 minuti)

The efficient simulation of particle propagation and interaction within the detectors of the Large Hadron Collider (LHC) is of primary importance for precision measurements and new physics searches. The most computationally expensive simulations involve calorimeter showers, which will become ever more costly and high-dimensional as the Large Hadron Collider moves into its High Luminosity era. The computational costs can be heavily reduced by replacing (parts of) the simulation pipeline with generative networks. We thus propose to model calorimeter showers, first by learning a lower-dimensional manifold structure with an auto-encoder, and to then perform density estimation on this manifold with a normalising flow. Our approach, lies on the notion that the seemingly high-dimensional data of HEP experiments, actually has a much lower intrinsic dimensionality. In machine learning, this is known as the manifold hypothesis, which states that high-dimensional data is supported on low-dimensional manifolds. By reducing the dimensionality of the data we enable fast training and generation without compromising accuracy.

Autori principali: CATERINI, Anthony (layer 6); ROSS, Brendan (layer 6); LOAIZA-GANEM, Gabriel (layer 6); REYES GONZALEZ, Humberto Alonso (Istituto Nazionale di Fisica Nucleare); CRESSWELL, Jesse (layer 6); LETIZIA, Marco (Istituto Nazionale di Fisica Nucleare)

Relatore: REYES GONZALEZ, Humberto Alonso (Istituto Nazionale di Fisica Nucleare)

Classifica Sessioni: VAEs

ID contributo: **3** Tipo: **non specificato**

Denoising Diffusion Models for High Fidelity Calorimeter Simulation

martedì 30 maggio 2023 17:00 (30 minuti)

Denoising diffusion models have recently become state of the art in the ML community because of their stable training procedure and ability to generate high quality images in reasonable computation times. We employ diffusion models for the task of generating calorimeter showers within the context of the CaloChallenge. Our diffusion models are based on 3D cylindrical convolutional networks, which take advantage of symmetries of the underlying data representation. For dataset 1, which has a basic cylindrical geometry but irregular binning between the different layers, we employ a differentiable embedding procedure that learns a reversible mapping from the original data format to a regular geometry on which cylindrical convolutions can be applied. We find our diffusion approach is able to generate high quality showers for all three datasets, achieving classifier AUC scores of ~0.7 or better.

Autori principali: PEDRO, Kevin (Fermilab); AMRAM, Oz (Fermilab)

Relatore: AMRAM, Oz (Fermilab)

Classifica Sessioni: Diffusion

ID contributo: **4** Tipo: **non specificato**

(inductive) CaloFlow

mercoledì 31 maggio 2023 09:30 (30 minuti)

We apply CaloFlow to GEANT4 showers of Dataset 1, producing high-fidelity samples with a sampling time of less than 0.1ms per shower. We validated the fidelity of the samples using multiple metrics, including a classifier metric. To generalize CaloFlow to the higher dimensional Datasets 2 and 3, we propose a new approach called Inductive CaloFlow. This approach involves training the flow on the pattern of energy deposition in both the current and previous layer of a GEANT4 event. Inductive CaloFlow can efficiently generate new events for large calorimeter geometries and reproduces GEANT4-like events with high fidelity. With both approaches, a teacher-student pairing was used to increase sampling speed without significant loss of sample quality.

Autori principali: KRAUSE, Claudius (Heidelberg University); SHIH, David (Rutgers University); PANG, Ian (Rutgers University); BUCKLEY, Matthew (Rutgers University)

Relatore: PANG, Ian (Rutgers University)

Classifica Sessioni: Normalised Flow

ID contributo: **6** Tipo: **non specificato**

CaloPointFlow - Generating Calorimeter Showers as Point Clouds

mercoledì 31 maggio 2023 11:00 (30 minuti)

In particle physics, precise simulations of the interaction processes in calorimeters are essential for scientific discovery. However, accurate simulations using GEANT4 are computationally very expensive and pose a major challenge for the future of particle physics. In this study, we apply the CaloPointFlow model, a novel generative model based on normalizing flows, to fast and high-fidelity calorimeter shower generation. We use the CaloPointFlow model, an adapted version of the PointFlow model for 3D shape generation, to generate calorimeter showers using point clouds that exploit the sparsity and leverage the geometry of the data. We preprocess the voxelized datasets of the Fast Calorimeter Simulation Challenge 2022 to point clouds and apply the CaloPointFlow model to all three datasets without any adaptation. Furthermore, we evaluate the performance of our model on metrics such as energy resolution, longitudinal and transverse shower profiles, and shower shapes, and compare it with GEANT4. We demonstrate that our model can produce realistic and diverse samples with a sampling time of around 30 million single 4D points per minute. However, the model also has some limitations, such as its inability to capture the point-to-point correlation and its generation of multiple points per cell, which are in contradiction to the data. To address these issues, we propose a novel method that uses a second sampling step to compute the marginal likelihoods of each cell being hit and sample the energies accordingly. We also discuss some ideas on how to handle the point-to-point correlations in future work. The main strengths of our model are its ability to handle diverse datasets, its fast and stable convergence, and its highly efficient point production.

Autori principali: Dr. KRÜCKER, Dirk (DESY); Prof. BORRAS, Kerstin (DESY / RWTH Aachen); SCHNAKE, Simon (DESY / RWTH Aachen)

Relatore: SCHNAKE, Simon (DESY / RWTH Aachen)

Classifica Sessioni: Normalised Flow

ID contributo: **7** Tipo: **non specificato**

CANADA-GAN (acronym and name subject to possible future change, please excuse)

We propose the application of our previously published Cross AtteNtion meAn-fielD mAtching CANADA-GAN for generating particle showers in high-granularity datasets. Results are presented for dataset 2 and 3. Point cloud generative models are known to benefit from higher granularity, making these datasets well-suited for high-granularity calorimeters. Although the regular architecture of the detector encourages the use of operations that capitalize on this regularity, we argue that such assumptions are over-idealistic for real physics applications. In our approach, the primary advantage of the regular architecture is the simplified dequantization process, which allows our model to potentially perform well on more complex and realistic detectors.
The proposed model is trained in an adversarial paradigm and demonstrates promising performance on marginal distributions (E, z, α, R) distributions, as well as energy-weighted distributions. The adversarial training paradigm helps to capture the intricate structure of particle showers and generate realistic samples. However, further research is required to develop more sophisticated metrics to accurately assess the model's performance in generating particle showers and handling the voxelization transformation.

Autore principale: KÄCH, Benno (Deutsches-Elektronen Synchrotron (DESY))

Coautore: Dr. MELZER-PELLMANN, Isabell-Alissandra (DESY); SCHAM, Moritz; SCHNAKE, Simon (DESY / RWTH Aachen)

Relatore: KÄCH, Benno (Deutsches-Elektronen Synchrotron (DESY))

Classifica Sessioni: GANs

ID contributo: **8** Tipo: **non specificato**

Generating Accurate Showers in Highly Granular Calorimeters Using Normalizing Flows

mercoledì 31 maggio 2023 10:00 (30 minuti)

Normalizing flows are a type of generative models that can be trained directly by minimizing the negative log-likelihood. It has been shown that flows can accurately model showers in low complexity calorimeters. We show how normalizing flows can be improved and adapted to accurately model showers in calorimeters with significantly higher complexity. One of the key points here is to move away from dense layers to convolutional layers. We show our results on datasets 2 and 3 of the CaloChallenge.

Autori principali: KRAUSE, Claudius (Heidelberg University); SHIH, David (Rutgers University); GAEDE, Frank (DESY); Prof. KASIECZKA, Gregor (University of Hamburg); Dr. DIEFENBACHER, Sascha (University of Hamburg); BUSS, Thorsten (University of Hamburg)

Relatore: BUSS, Thorsten (University of Hamburg)

Classifica Sessioni: Normalised Flow

ID contributo: **9** Tipo: **non specificato**

A containerized solution for fast calorimeter simulation

The simulation of electromagnetic and hadronic interactions in calorimeters is a very demanding process, both in terms of time and computing resources. A novel technique based on Generative Adversarial Networks (GANs) may benefit from a more efficient use of computing resources, although initial training could be computationally demanding. Nowadays and in the near future we expect to have more computing facilities available, albeit they may have different setup (different OS and schedulers, number of available CPUs per computing node, presence of computing accelerators⋯).
In this contribution we present a model of calorimeter simulation based on GANs which trains simultaneously two neural networks: a generator which, given Geant4-simulated data as training sample, aims at generating sample the most similar to it and a discriminator which, fed data from both the training sample and the generator-produced Geant4-like sample, tries distinguishing actual Geant4 data from generator-produced data. When finished training, our model simulates calorimeter response as similarly as possible to Geant4, but much faster. In this contribution, training data are the CaloChallenge public ones. We also present a general framework based on containers to ensure full portability of the code through different computing clusters: in this way, our framework may run not only on HEP-specific resources, but also on supercomputing clusters providing further and more powerful resources.

Autori principali: CORCHIA, Federico Andrea Guillaume (Istituto Nazionale di Fisica Nucleare); FAUCCI GIANNELLI, Michele (Istituto Nazionale di Fisica Nucleare); FRANCHINI, Matteo (Istituto Nazionale di Fisica Nucleare); RINALDI, Lorenzo (Istituto Nazionale di Fisica Nucleare)

Relatori: CORCHIA, Federico Andrea Guillaume (Istituto Nazionale di Fisica Nucleare); RINALDI, Lorenzo (Istituto Nazionale di Fisica Nucleare)

Classifica Sessioni: GANs

ID contributo: **10** Tipo: **non specificato**

Latent Generative Models for Calo Simulation with VQ-VAE

martedì 30 maggio 2023 14:00 (30 minuti)

Simulation of calorimeter response is important for modern high energy physics experiments. With increasingly large and high granularity design of calorimeters, the computational cost of conventional MC-based simulation of each particle-material interaction is becoming a major bottleneck.

We propose a new generative model based on a two-stage generative model which is similar to recently latent diffusion models. The first stage model aims only to compress the calorimeter response into a regularized latent space based on Vector Quantized Variational Autoencoders (VQ-VAE). The second stage model handles generative sampling in the latent representation, conditional on phase space parameters such as pT. This second stage can be trained independently of the first, and we demonstrate prior models based on RNNs and diffusion.

For the Calo Challenge Dataset 1, our demonstration model achieves a speedup of more than 10^4 over GEANT4. The modeling of energy deposition and shower shape is comparable to other approaches such as CaloFlow, with substantially fewer parameters and a factor of ~2 speedup.

For Datasets 2, we have designed a fully convolutional architecture, employing novel operations which maintain equivariance w.r.t. the cylindrical geometry of the data. This combined with the two-stage modeling approach shows promising generation quality with substantially faster training and approximately 2x faster inference time relative to CaloScore.

In addition to cylindrical convolution and two-stage VQ-VAE training, we also introduce a novel, self-supervised proxy model that can be used as a perceptual loss function to help with training any AutoEncoder-based model. While some challenges remain in achieving ultra high fidelity suitable for certain physics applications, we present several innovative techniques that may help solve the puzzle of fast and accurate calorimeter simulation.

Autori principali: LIU, Qibin (TDLI., Shanghai JiaoTong University); Dr. SHIMMIN, Chase (Yale University); LIU, Xiulong (University of Washington); Prof. SHLIZERMAN, Eli (University of Washington); Prof. HSU, Shih-Chieh (University of Washington)

Relatore: LIU, Qibin (TDLI., Shanghai JiaoTong University)

Classifica Sessioni: VAEs

ID contributo: **11** Tipo: **non specificato**

CaloClouds: Fast Geometry-Independent Highly-Granular Calorimeter Simulation

martedì 30 maggio 2023 17:30 (30 minuti)

Simulating showers of particles in highly-granular detectors is a key frontier in the application of machine learning to particle physics.
Achieving high accuracy and speed with generative machine learning models would enable them to augment traditional simulations and alleviate a major computing constraint.
This work achieves a major breakthrough in this task by for the first time directly generating a point-cloud of $O(1000)$ space points with energy depositions in the detector in 3D-space without relying on a fixed-grid structure. This is made possible by two key innovations: i) using recent improvements in generative modelling we apply a diffusion model and ii) an initial even higher-resolution point-cloud of up to 40000 so-called GEANT4 steps which are subsequently down-sampled to the desired number of up to 6000 space points.
We showcase the performance of this approach using the specific example of simulating photon showers in the planned electromagnetic calorimeter of the International Large Detector (ILD) and achieve overall good modelling of physically relevant distributions.

Autori principali: KOROL, Anatolii (Deutsches Elektronen-Synchrotron (DESY)); EREN, Engin (Deutsches Elektronen-Synchrotron DESY); BUHMANN, Erik (University of Hamburg); GAEDE, Frank (DESY); KASIECZKA, Gregor (University of Hamburg); KRÜGER, Katja (DESY); MCKEOWN, Peter (DESY); DIEFENBACHER, Sascha (Berkeley Laboratory); KORCARI, William (University of Hamburg)

Relatore: KOROL, Anatolii (Deutsches Elektronen-Synchrotron (DESY))

Classifica Sessioni: Diffusion

ID contributo: **12** Tipo: **non specificato**

Bus at piazza Marconi

martedì 30 maggio 2023 08:30 (30 minuti)

Bus will leave at 9:00 for Villa Mondragone

ID contributo: **13** Tipo: **non specificato**

Registration

martedì 30 maggio 2023 09:30 (30 minuti)

ID contributo: **14** Tipo: **non specificato**

Introduction

martedì 30 maggio 2023 10:00 (5 minuti)

Relatore: FAUCCI GIANNELLI, Michele (Istituto Nazionale di Fisica Nucleare)

ID contributo: **15** Tipo: **non specificato**

Welcome from INFN Tor Vergata

martedì 30 maggio 2023 10:05 (5 minuti)

Relatore: DI CIACCIO, Anna (Istituto Nazionale di Fisica Nucleare)

ID contributo: **16** Tipo: **non specificato**

ML in industry

martedì 30 maggio 2023 10:15 (20 minuti)

Relatore: PALAZZO, Serena (CS)

ID contributo: **17** Tipo: **non specificato**

CaloShowerGAN

martedì 30 maggio 2023 11:00 (30 minuti)

Autore principale: ZHANG, Rui

Coautore: FAUCCI GIANNELLI, Michele (Istituto Nazionale di Fisica Nucleare)

Relatori: FAUCCI GIANNELLI, Michele (Istituto Nazionale di Fisica Nucleare); ZHANG, Rui

Classifica Sessioni: GANs

ID contributo: **18** Tipo: **non specificato**

MDMA-GAN

martedì 30 maggio 2023 11:30 (30 minuti)

Relatore: KÄCH, Benno (Deutsches-Elektronen Synchrotron (DESY))

Classifica Sessioni: GANs

ID contributo: **19** Tipo: **non specificato**

BolognaGAN: a containerized solution

martedì 30 maggio 2023 12:00 (30 minuti)

Relatore: CORCHIA, Federico Andrea Guillaume (Istituto Nazionale di Fisica Nucleare)

Classifica Sessioni: GANs

ID contributo: **20** Tipo: **non specificato**

Discussion

martedì 30 maggio 2023 12:30 (30 minuti)

Classifica Sessioni: GANs

ID contributo: **21** Tipo: **non specificato**

Bus to dinner

martedì 30 maggio 2023 19:00 (30 minuti)

ID contributo: **22** Tipo: **non specificato**

Bus at piazza Marconi

mercoledì 31 maggio 2023 08:30 (30 minuti)

ID contributo: **23** Tipo: **non specificato**

Going into production, FastCalo in the bigger picture

mercoledì 31 maggio 2023 12:00 (20 minuti)

Relatore: FAUCCI GIANNELLI, Michele (Istituto Nazionale di Fisica Nucleare)

ID contributo: **24** Tipo: **non specificato**

CaloChallenge result

mercoledì 31 maggio 2023 12:20 (20 minuti)

Relatore: KRAUSE, Claudius (Heidelberg University)

ID contributo: **25** Tipo: **non specificato**

Discussion

mercoledì 31 maggio 2023 12:40 (50 minuti)

ID contributo: **26** Tipo: **non specificato**

Bus depart to Frascati

mercoledì 31 maggio 2023 14:40 (20 minuti)

ID contributo: **27** Tipo: **non specificato**

CaloScore

Relatore: MIKUNI, Vinicius (LBNL)

Classifica Sessioni: GANs

ID contributo: **28** Tipo: **non specificato**

Discussion

mercoledì 31 maggio 2023 11:30 (30 minuti)

Classifica Sessioni: Normalised Flow

ID contributo: **29** Tipo: **non specificato**

CERN-Geneva VAE

martedì 30 maggio 2023 14:30 (30 minuti)

Relatore: SALAMANI, Dalila (CERN)

Classifica Sessioni: VAEs

ID contributo: **30** Tipo: **non specificato**

Discussion

martedì 30 maggio 2023 15:30 (30 minuti)

Classifica Sessioni: VAEs

ID contributo: **31** Tipo: **non specificato**

Discussion

martedì 30 maggio 2023 18:00 (30 minuti)

Classifica Sessioni: Diffusion

ID contributo: **32** Tipo: **non specificato**

Welcome from Fellini

martedì 30 maggio 2023 10:10 (5 minuti)

Relatore: BANDIERA, LAURA (Istituto Nazionale di Fisica Nucleare)