1st AI-INFN User Forum

Report dei Contributi

https://agenda.infn.it/e/40489

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

Welcome and introduction

martedì 11 giugno 2024 14:30 (10 minuti)

Autori principali: RONCHIERI, Elisabetta (Istituto Nazionale di Fisica Nucleare); BARBETTI, Matteo (INFN CNAF); ANDERLINI, Lucio (Istituto Nazionale di Fisica Nucleare)

Relatore: ANDERLINI, Lucio (Istituto Nazionale di Fisica Nucleare)

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

INFN-CNAF: status and perspectives

martedì 11 giugno 2024 14:40 (20 minuti)

Autore principale: DELL'AGNELLO, Luca (Istituto Nazionale di Fisica Nucleare)

Relatore: DELL'AGNELLO, Luca (Istituto Nazionale di Fisica Nucleare)

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

Final remarks and closing

mercoledì 12 giugno 2024 12:15 (15 minuti)

Autori principali: RONCHIERI, Elisabetta (Istituto Nazionale di Fisica Nucleare); BARBETTI, Matteo (INFN CNAF); ANDERLINI, Lucio (Istituto Nazionale di Fisica Nucleare)

Relatore: RONCHIERI, Elisabetta (Istituto Nazionale di Fisica Nucleare)

ID contributo: **5** Tipo: **non specificato**

AI-based approach for provider selection in the INDIGO PaaS Orchestration System of INFN Cloud

mercoledì 12 giugno 2024 09:15 (25 minuti)

INFN Cloud provides scientific communities supported by the Institute with a federated Cloud infrastructure and a dynamic portfolio of services based on the needs of the supported use cases. The federative middleware of INFN Cloud is based on the INDIGO PaaS orchestration system, consisting of interconnected open-source microservices. Among these, the INDIGO PaaS Orchestrator receives high-level deployment requests and coordinates the process of creating deployments on the IaaS platforms provided by federated providers.

In the default configuration, the INDIGO PaaS Orchestrator determines the provider to submit the deployment creation request to from an ordered list of providers, selection based on the user's group affiliation. This list is provided by the Cloud Provider Ranker service, which applies a ranking algorithm using a restricted set of metrics related to deployments and defined Service Level Agreements for providers. The INDIGO PaaS Orchestrator submits the deployment to the first provider in the list and, in case of failure, scales to the next provider until the list is exhausted.

This contribution presents the activity aimed at improving the ranking system and optimizing resource usage through an approach based on the use of artificial intelligence techniques. In this context, significant preparatory work was carried out to identify the most meaningful metrics, as well as the sources from which to retrieve these metrics. The subsequent dataset preparation allowed us to study the case in detail, identifying and comparing different artificial intelligence techniques. The proposed approach involves creating two models: one for predictive classification of deployment success/failure and one for deployment creation time regression. A linear combination of the output of the two models, along with training on recent and mobile time windows, allows for the definition of an ordered list of providers that the orchestrator can use for deployment submission.

Autore principale: GIOMMI, Luca (Istituto Nazionale di Fisica Nucleare)

Relatore: GIOMMI, Luca (Istituto Nazionale di Fisica Nucleare)

Classifica Sessioni: Wednesday morning: Part I

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

Hyperparameter Optimization for Deep Learning Models Using High Performance Computing

mercoledì 12 giugno 2024 11:25 (25 minuti)

Clusters counting in a drift chamber represents the most promising breakthrough in particle identification (PID) techniques in particle physics experiments. In this paper, neural network models, such as the Long Short-Term Memory (LSTM) Model and Convolutional Neural Network (CNN) Model, are trained using various hyperparameters like loss functions, activation functions, different numbers of neurons, batch sizes, and varying numbers of epochs etc. These models are trained for a two-step reconstruction algorithm, which involves peak finding and clusterization. For the peak finding algorithm, a trained Long Short-Term Memory (LSTM) model is used to discriminate between ionization signals (primary and secondary peaks) and noise in the waveform, addressing a classification problem. Concurrently, a Convolutional Neural Network model is utilized to determine the number of primary ionization clusters based on the detected peaks, dealing with a regression problem. The trained models (LSTM and CNN) are applied to the simulations of particles traversing a gas mixture made of 90% Helium (He) and 10% Isobutane (C4H10) filling drift tubes with the same geometry as the ones used for the beam test at CERN in 2023 of the prototype of the IDEA detector for FCC. The simulation parameters included a cell size of 1.5 cm, a sampling rate of 1.2 GHz, a time window of 2000 ns, 5000 events, a number of ionization clusters with a mean value of 25.25, a number of electrons per cluster with a mean value of 1.468, and momentum of pi- meson particles ranging from 4 to 180GeV/c.

Autori principali: DIACONO, Domenico (Politecnico di Bari and Istituto Nazionale di Fisica Nucleare Bari, Italy); GRANCAGNOLO, Francesco (Istituto Nazionle di Fisica Nucleare Lecce, Italy); ZHAO, Guang (Institute of High Energy Physics, 19B Yuquan Road, Beijing, 100049, Beijing, China); WU, Linghui (Institute of High Energy Physics, 19B Yuquan Road, Beijing, 100049, Beijing, China); ABBRESCIA, Marcello (Politecnico di Bari and Istituto Nazionale di Fisica Nucleare Bari, Italy); DONG, Mingyi (Institute of High Energy Physics, 19B Yuquan Road, Beijing, 100049, Beijing, China); ANWAR, Muhammad Numan (Politecnico di Bari and Istituto Nazionale di Fisica Nucleare Bari, Italy); DE FILIPPIS, Nicola (Politecnico di Bari and Istituto Nazionale di Fisica Nucleare Bari, Italy); SUN, Shengsen (Institute of High Energy Physics, 19B Yuquan Road, Beijing, 100049, Beijing, China)

Relatore: ANWAR, Muhammad Numan (Politecnico di Bari and Istituto Nazionale di Fisica Nucleare Bari, Italy)

Classifica Sessioni: Wednesday morning: Part II

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

Multi-scale cross attention transformer encoder for τ lepton pair invariant mass reconstruction

mercoledì 12 giugno 2024 11:00 (25 minuti)

With the observation of the Standard Model Higgs boson (H) by the CMS and ATLAS experiments in 2012, the last missing elementary particle predicted by the Standard Model of particle physics was found. Since then, extensive measurements in various decay channels of the Higgs boson have been performed. One of them is the decay into a pair of τ leptons. It is the decay channel of the Higgs boson into fermions with the second-largest branching fraction, only surpassed by the decay into a pair of bottom quarks.
In such analyses, the reconstructed invariant mass of the di-τ pair is an important discriminant to separate signal (H) from background events, which can be reconstructed starting from the decay products of each τ lepton. However, the presence of neutrinos in the final state determines a lack in terms of energy which leads to an underestimation of the invariant mass itself.
The proposed work consists in a Deep Learning (DL) model that, instead of just regressing the mass, estimates the full four-vector of each τ of the system before decay for a high-resolution reconstruction of the invariant mass and therefore retrieve information about the kinematics of the parent particle.
The implemented model is a multi-scale cross attention transformer encoder (a DL model, born for Natural Language Processing tasks and now demonstrating its power as a universal architecture). This multi-modal network can extract information from the substructure of the di-τ products and the kinematics of the reconstructed taus through self-attention transformer layers. The learned information is subsequently integrated to improve regression performance using an additional transformer encoder with cross-attention heads. $H \rightarrow \tau\tau$ process together with the main backgrounds are used as a benchmark to measure the performance of the new algorithm with respect to the currently used one in CMS (SVFit).

Autore principale: CAMAGNI, Valentina (Universita & INFN, Milano-Bicocca (IT))

Coautore: BRIVIO, Francesco (Istituto Nazionale di Fisica Nucleare); DINI, Paolo (Istituto Nazionale di Fisica Nucleare); GOVONI, Pietro (Istituto Nazionale di Fisica Nucleare); GENNAI, Simone (Istituto Nazionale di Fisica Nucleare)

Relatore: CAMAGNI, Valentina (Universita & INFN, Milano-Bicocca (IT))

Classifica Sessioni: Wednesday morning: Part II

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

Use of a UNet network for the identification of cavities inside mines

martedì 11 giugno 2024 15:25 (25 minuti)

Muon radiography is a technique that utilizes muons from cosmic rays to investigate otherwise hard-to-reach environments. This technique offers several advantages, including the absence of accelerators to generate particles that interact with the target under examination. It is a non-invasive technique, both for humans and the observed object. Furthermore, due to the muons' ability to penetrate dense materials over long distances, it is suitable for studying large structures such as mines, hills, and pyramids, as well as highly dense materials like radioactive waste containers and blast furnaces.
This study presents an application of this technique to the Temperino mines.
Typically, cavity detection using muon radiography is based on a visual method. This work demonstrates how, through the use of a UNet neural network, cavities can be detected and delineated. This allows you to define a percentage of precision in the detection of these voids, moving from subjective to objective identification. Additionally, more detailed information about their shape and size is available.
The combination of muon radiography with the use of a UNet neural network demonstrates its significant potential as a tool for subsurface exploration and geological studies, providing a more accurate and reliable approach for detecting and characterizing voids.

Autore principale: Sig. PACCAGNELLA, Andrea (Istituto Nazionale di Fisica Nucleare Firenze)

Relatore: Sig. PACCAGNELLA, Andrea (Istituto Nazionale di Fisica Nucleare Firenze)

Classifica Sessioni: Tuesday afternoon: Part I

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

Benchmarking image segmentation on AMD-Xilinx FPGAs

martedì 11 giugno 2024 15:50 (25 minuti)

In the context of scientific computing there is a growing interest towards DNNs (Deep Neural Networks), which are being used in several applications, spanning from medical images segmentation and classification, to the on-line analysis of experimental data.
In addition to GPUs, FPGAs are also emerging as compute accelerators promising higher energy-efficiency and lower latency, for the inference phase of such DNNs.
In this talk, we introduce a 2D UNet medical image segmentation application as an use case; then we focus on the implementation of its inference phase on the FPGA; and finally we compare its performance on an AMD-Xilinx Alveo U250 FPGA, with its performance on Intel CPUs and NVIDIA GPU accelerators, in terms of: accuracy, time-to-solution and energy-efficiency.

Autori principali: Sig.na SISINI, Valentina (INFN and University of Ferrara); MIOLA, Andrea (Istituto Nazionale di Fisica Nucleare, University of Ferrara); CALORE, Enrico (Istituto Nazionale di Fisica Nucleare); SCHIFANO, Sebastiano Fabio (INFN and University of Ferrara); Prof. ZAMBELLI, Cristian (University of Ferrara)

Relatore: Sig.na SISINI, Valentina (INFN and University of Ferrara)

Classifica Sessioni: Tuesday afternoon: Part I

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

Quantum Machine learning frameworks for charged particle tracking

mercoledì 12 giugno 2024 11:50 (25 minuti)

With the advent of the High Luminosity LHC era, which is expected to significantly increase the amount of data collected by the detectors, the computational complexity of the particle tracking problem is expected to increase.

Conventional algorithms suffer from scaling problems. In our work, we represent charged particle tracks as a graph data structure and we are investigating the use of quantum machine learning techniques to see if we can gain a quantum advantage in the reconstruction of tracks.

Finding the optimal combination of classical machine learning tools and quantum libraries is challenging, especially since most quantum tools are still in the developing phase and they are not stable.

We report on our experience in testing quantum machine learning frameworks, such as Jax, Pennylane, and IBM Qiskit, eventually using GPUs as accelerators. Finally, we give an outlook on the expected performance in terms of scalability of accuracy and efficiency of the particle tracking problem.

Autori principali: CAPPELLI, Laura (Istituto Nazionale di Fisica Nucleare); ARGENTON, Matteo (Istituto Nazionale di Fisica Nucleare)

Coautore: BOZZI, Concezio (Istituto Nazionale di Fisica Nucleare); CALORE, Enrico (Istituto Nazionale di Fisica Nucleare); SCHIFANO, Sebastiano (Istituto Nazionale di Fisica Nucleare); AMITRANO, Valentina (Istituto Nazionale di Fisica Nucleare)

Relatore: CAPPELLI, Laura (Istituto Nazionale di Fisica Nucleare)

Classifica Sessioni: Wednesday morning: Part II

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

Virtual Painting recoloring using Vision Transformer on Deep Embedded X-Ray Fluoresce synthetic dataset

martedì 11 giugno 2024 17:10 (25 minuti)

In the last few years, the rise of deep learning techniques has affected also the field on physics-based imaging applied to cultural heritage. One possible application of such techniques is the virtual digital restoration of pictorial artworks.

Two main problems we face when exploring such landscape are

1. The small dataset sizes (due to the slow pace of such analysis, as well as Intellectual Property issues)
2. The huge dimensionality of each datapoint (due to the fact that physics-based techniques produces spectral datacubes)

We address those issues by creating a (huge) synthetic dataset and then embedding it unsupervisingly in a metric space, using an ad-hoc trained Deep (Variational) Embedding model, exploiting the hidden statistical relations of each pixel spectra.

Starting from such dataset, it is possible to train supervisingly standard Computer Vision models (such as U-ResNets or Vision Transformers) to try to assign an human readable RGB color to spectral datacubes, thus performing virtual digital restoration.

This task could be relevant, for example, in the context of detached frescoes, where colour legibility may have been (partially) lost, but physical signals may still be found.

Autore principale: BOMBINI, Alessandro (Istituto Nazionale di Fisica Nucleare)

Relatore: BOMBINI, Alessandro (Istituto Nazionale di Fisica Nucleare)

Classifica Sessioni: Tuesday afternoon: Part II

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

First Stages on Spectral Classification using Synthetic Datasets.

martedì 11 giugno 2024 17:35 (25 minuti)

X-ray fluorescence (XRF) has been extensively utilized across various disciplines for an extended period. Despite the long-standing availability of fundamental physical parameters associated with the process, the extraction of elemental information from spectra remains predominantly a manual or algorithmic endeavor, given the spectral variations and inherent complexity in different fields of application. This fact limits the range of applicability of AI.
To address this challenge, we generate plausible look a like macro XRFs spectra out of fundamental X-ray parameters, skipping the intermediate intricate features, treating them as perturbations. Our method features independent components adaptable to diverse fields, instrumentation, and sample types. We generate a large dataset using this approach to train a Fully Connected Neural Network to classify elemental lines in the 0-30 KeV range, from a predetermined selection of elements. We then test the Neural Network on completely experimental data from Aerosol measures.

Autore principale: GARCIA-AVELLO BOFIAS, Fernando (Istituto Nazionale di Fisica Nucleare)

Coautore: Dr. BOMBINI, Alessandro (INFN)

Relatore: GARCIA-AVELLO BOFIAS, Fernando (Istituto Nazionale di Fisica Nucleare)

Classifica Sessioni: Tuesday afternoon: Part II

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

HERD data classification

martedì 11 giugno 2024 16:45 (25 minuti)

The growing demand for GPUs has led to rapid development of machine learning research techniques in all areas of science, including High Energy Physics. We present a study focused on the classification task of simulated electrons and protons detected by the HERD detector. HERD is a high-energy cosmic-ray detector based on a deep three-dimensional electromagnetic calorimeter, proposed to be installed on the Chinese Space Station. The main scientific objectives of HERD include detecting dark matter particles, studying cosmic ray composition, and observing high energy gamma rays.
Our classification task is based on data from Monte Carlo simulations of proton and electron particle showers in the HERD electromagnetic calorimeter, with energies ranging from 100 GeV to 20 TeV. We have two datasets, one composed of three-dimensional images, and the other from their 2-dimensional projections.
Our approach is inspired by the Inception neural network, a very deep convolutional neural network that achieved state-of-the-art performance in the ImageNet Large Scale Visual Recognition Challenge 2015 when combined with residual connections.

Autore principale: TABARRONI, Luca (Istituto Nazionale di Fisica Nucleare)

Coautore: FORMATO, Valerio (RM2)

Relatore: TABARRONI, Luca (Istituto Nazionale di Fisica Nucleare)

Classifica Sessioni: Tuesday afternoon: Part II

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

Enhancing Nodule segmentation Utilizing Attention U-Net: Insights from LUNA-16 Dataset

martedì 11 giugno 2024 15:00 (25 minuti)

Abstract
Lung cancer remains a significant global health challenge, with early detection playing a critical role in improving patient outcomes. Computed Tomography (CT) imaging has become a cornerstone in the early diagnosis and staging of lung cancer, allowing for the detection of pulmonary nodules that may indicate malignancy. However, accurately segmenting and characterizing these nodules from CT scans presents substantial challenges due to variations in size, shape, appearance, and the presence of noise and artifacts. In this study, we propose a comprehensive approach to enhance nodule segmentation in lung CT scans, utilizing the great challenge of Lung Nodule Analysis 2016 (LUNA-16) dataset. We employ a pre-trained U-net based architecture from the Lung Quant (Lizzi et al., 2023) algorithm for robust segmentation of lung regions in CT scans. The pre-trained U-Net model accurately delineates lung boundaries, providing a reliable basis for subsequent nodule segmentation.
Following lung segmentation, we compare the performance of standard U-Net with an attention-enhanced variant, Attention U-Net, for nodule segmentation within the segmented lung regions. Attention mechanisms dynamically highlight informative regions, improving segmentation accuracy, particularly in challenging cases.
Experiments are conducted on the LUNA-16 dataset, comprising CT scans from multiple institutions and acquisition protocols. Quantitative evaluations compare U-Net and Attention U-Net performance using metrics such as Dice coefficient, sensitivity, specificity, and false positive rate. Our results showcase exceptional performance, with a Dice score of 73% for nodule segmentation and 90% for lung segmentation. These findings underscore the superiority of Attention U-Net over conventional U-Net, demonstrating higher accuracy in nodule segmentation and a reduction in false positives. This study presents a comprehensive approach to enhance nodule segmentation in lung CT scans, leveraging deep learning architectures and pre-processing techniques on the LUNA-16 dataset.

Lizzi, F., Postuma, I., Brero, F., Cabini, R. F., Fantacci, M. E., Lascialfari, A., Oliva, P., Rinaldi, L., & Retico, A. (2023). Quantification of pulmonary involvement in COVID-19 pneumonia: an upgrade of the LungQuant software for lung CT segmentation. European Physical Journal Plus, 138(4). https://doi.org/10.1140/epjp/s13360-023-03896-4

Autori principali: ZAFARANCHI, Arman (Istituto Nazionale di Fisica Nucleare); LIZZI, Francesca (Istituto Nazionale di Fisica Nucleare); RETICO, Alessandra (Istituto Nazionale di Fisica Nucleare); SCAPICCHIO, Camilla (Istituto Nazionale di Fisica Nucleare); FANTACCI, Maria Evelina (Istituto Nazionale di Fisica Nucleare)

Relatore: ZAFARANCHI, Arman (Istituto Nazionale di Fisica Nucleare)

Classifica Sessioni: Tuesday afternoon: Part I

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

Leveraging RAG Architecture for Effective Email Response Automation: a CNAF Tier-1 User Support use case

mercoledì 12 giugno 2024 10:05 (25 minuti)

CNAF provides computing resources to over 60 scientific communities and supports over 1700 active users through its *User Support* (US) department. US handles daily emails and tickets to help users in employing effectively computing resources and using latest software technologies. Since 2003, CNAF hosts the main INFN computing center, one of WLCG Tier-1.

The primary challenge is to handle user queries and support requests related to the hardware and software technologies employed by the various experiments at CNAF. Users often require expert assistance via email to troubleshoot issues, optimize code performances, or leverage specialized features of the Tier-1 infrastructure.

To tackle this problem, we propose the development of a *Retrieval Augmented Generation* (RAG) model tailored specifically for automating responses to support-related emails. The RAG model will leverage advanced *Natural Language Processing* techniques to understand and generate accurate responses based on the context retrieved and the user queries given.

Through the LangChain framework [1], we have built a vector database containing information from the Tier-1 User Guide [2]. Leveraging an efficient retrieval system [3] and following the RAG pipeline, portions of the User Guide have been prompted directly into a *Large Language Model* (LLM) to address user queries. This approach has aimed to enhance the responsiveness and accuracy of the LLM through specific domain knowledge encoded within vector spaces and cutting-edge semantic similarity models, grasping human-like responses through *Foundation Models*.

References

[1] Applications that can reason. Powered by LangChain, https://www.langchain.com

[2] INFN-CNAF Tier-1 User Guide, https://l.infn.it/t1guide

[3] Wen Li *et al.*, "Approximate Nearest Neighbor Search on High Dimensional Data - Experiments, Analyses, and Improvement."IEEE Trans. Knowl. Data Eng. **32** (2020) 1475

Autori principali: TRASHAJ, Alberto (Università di Bologna); BARBETTI, Matteo (INFN CNAF); RONCHIERI, Elisabetta (Istituto Nazionale di Fisica Nucleare); PELLEGRINO, Carmelo (Istituto Nazionale di Fisica Nucleare); CESINI, Daniele (Istituto Nazionale di Fisica Nucleare); GIUGLIANO, Carmen (Istituto Nazionale di Fisica Nucleare); LATTANZIO, Daniele (Istituto Nazionale di Fisica Nucleare); MORGANTI, Lucia (Istituto Nazionale di Fisica Nucleare); PASCOLINI, Alessandro (Istituto Nazionale di Fisica Nucleare); RENDINA, Andrea (Istituto Nazionale di Fisica Nucleare); SHTIMMERMAN, Aksieniia (Istituto Nazionale di Fisica Nucleare)

Relatore: TRASHAJ, Alberto (Università di Bologna)

Classifica Sessioni: Wednesday morning: Part I

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

Transformer-based models for scientific text classification

mercoledì 12 giugno 2024 09:40 (25 minuti)

The transformer model, introduced by Google in 2017, has become renowned in natural language processing (NLP). It represents a significant advancement completely departing from the mechanisms of Recurrent Neural Networks and Convolutional Neural Networks.
The features that contribute to the superior performance of transformers in NLP tasks include self-attention, multi-head attention, and positional encoding. The attention mechanism enables the model to extract dependency relationships between words, whereas the positional encoding extracts information about the position in the sequence of words.
Systematic literature reviews (SLR) help scientists to identify, select, access, and synthesize studies relevant to a specific topic. One crucial phase in conducting an SLR involves screening a vast number of articles sourced from various databases. This screening process is particularly challenging given the number of papers to be reviewed. For this reason, it is essential to perform this activity automatically with text classification.
In this context, we have employed transformer-based models to discern the primary topics covered in papers relevant to specific research areas, and group them into clusters. The primary aim is to develop an automated procedure that aids the scientific community in their search efforts by extracting pertinent information from vast amounts of documents.
During the study, we have used a range of transformer implementations, predominantly deriving from BERT, tailored to specific use cases including COVID-19, rehabilitation, and physics. Within the context of COVID-19, we have deployed models like BioBERT and PubMedELECTRA, achieving remarkable performance, particularly with PubMedBERT-large achieving a 0.9021 F1-score using 5-fold cross-validation.

Autori principali: CANAPARO, Marco (Istituto Nazionale di Fisica Nucleare); RONCHIERI, Elisabetta (Istituto Nazionale di Fisica Nucleare); TODESCHINI, Sofia Camilla (Università di Bologna); ZURLO, Giovanni (Università di Giovanni)

Relatore: ZURLO, Giovanni (Università di Giovanni)

Classifica Sessioni: Wednesday morning: Part I