Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

DeepCosmoNet

Deep Learning for Cosmic Web Analysis

Luca Naso, Koexai S.r.l

Spoke 3 Progetti Bandi a Cascata, 24/09, 2024

DeepCosmoNet
Deep Learning for Cosmic Web Analysis

Tematica: 1

Visualizzazione Scientifica con supporto di Intelligenza Artificiale

Sotto-tematica: d)

Sviluppo e applicazione degli algoritmi di deep learning per lo sfruttamento di dati da distribuzioni 3D point-clouds non strutturate in voxel (e.g. simulazioni/dati cosmologici). Ottimizzazione di reti deep con tecniche NAS per rappresentare e analizzare distribuzioni 3D point-cloud. Visualizzazione, analisi e interpretazione dello spazio latente per la generazione di distribuzioni point-clouds in diversi scenari cosmologici e/o astrofisici.

Project Overview: challenge and proposed solution

Today, **cosmology** relies heavily on **numerical simulations**, especially N-body simulations, to study the large-scale structure of the universe. These simulations are not only computationally intensive but also require significant time for analysis. Specifically, identifying components of the Cosmic Web structure (such as halos, voids, and filaments), from 3D point-cloud data that result from cosmological simulations simulation is a time-consuming task. This lengthy process significantly slows down the pace of scientific research.

Various tools like SUBFIND, Rockstar, VIDE e DisPerSE, ZOBOV, Popcorn, and SCMS are currently used to **identify different components of the Cosmic Web**. However, each tool specializes in identifying only one type of component and the analysis process remains time-consuming.

DeepCosmoNet aims to develop specialized **Machine Learning (ML) models** for each Cosmic Web structure component that are significantly faster than current state-of-the-art tools. Additionally, the project will create an **integrated pipeline** capable of identifying multiple types of components simultaneously.

To achieve optimal performance, **Neural Architecture Search (NAS) techniques** will be investigated and implemented to optimize these ML models.

Technical Objectives and Methodologies

Advanced Segmentation: development of machine learning models specifically designed for the automatic segmentation of complex 3D point-cloud data. These models will identify Cosmic Web structures such as halos, filaments, and voids within cosmological numerical simulations. Training will occur using datasets labelled with existing catalogues, with the primary goal of significantly reducing the time required for identifying cosmic structures, allowing researchers to explore more models in less time.

Cosmological Segmentation Pipeline: creation of a comprehensive cosmological segmentation pipeline that leverages the advanced segmentation models developed previously. This pipeline will support the simultaneous detection of various types of cosmic structures, eliminating the need for separate tools dedicated to individual components. This approach will further decrease the time researchers spend on non-scientific activities.

Optimize Analysis: use of Neural Architecture Search (NAS) techniques to enhance the efficiency of identifying the optimal neural networks for the machine learning models, including 3D Convolutional Neural Networks (3D CNNs) and Graph Neural Networks (GNNs). This optimisation will ensure that the models are effective and adaptable to varying requirements in cosmic structure analysis.

Computing Resources

Advanced Segmentation (DL training + NAS) & Cosmological Pipeline (inference + automation):
Proceed with trial and error, testing several approaches and neural network architectures on smaller regions and then training the final models on a larger dataset.

Expected:
1700 hours spread across different machines
from 16 to 192 vCPUs, from 64 to 2048 GiB Memory,
with NVIDIA A10G Tensor Core GPU and NVIDIA H100 Tensor Core (from 35 to 536 TFLOPs)
Cost on AWS ~11k $

Involved Staff and new recruitments

Spoke Scientific Advisor: Carmelita Carbone, INAF - IASF Milano

Koexai Team size: 2.24 FTE = 3852 hours

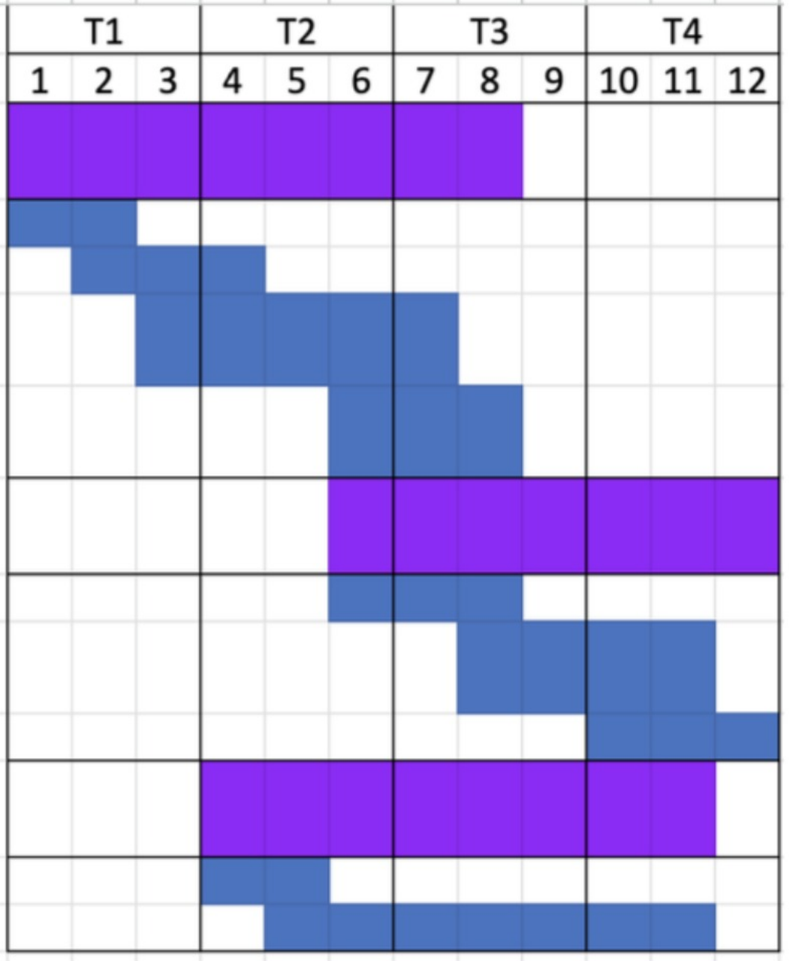
1. **Luca Naso:** Project Lead, Senior Data Scientist, Ph.D. in Astrophysics, and CEO of Koexai
2. **Marco Cataldo:** Project Manager and Project Administrative Manager
3. **Vincenzo Del Zoppo:** Senior Data Scientist and AI Engineer
4. **Giuseppe Puglisi:** Data Scientist, Ph.D. in Machine Learning
5. **Fabio Spampinato:** Data Scientist

Scientific Consultant:

- **Fondazione C. Fillietroz:** the Foundation manages the Astronomical Observatory of the Autonomous Region of Valle d'Aosta and serves as the main scientific institution in the region for fundamental research, technology development, and public engagement in astronomy. It provides expertise in cosmology and connects it with machine learning to ensure that research aligns with astrophysics standards. The Foundation regularly collaborates with Italian and international research centres on advanced projects in contemporary astronomy. Over time, researchers at the centre have built a strong base of specific knowledge, enabling the Foundation to engage in technology transfer since 2011.

Gantt

OR	AR	Titolo
1		*Studio e definizione di un sistema di Machine Learning avanzato di segmentazione per ciascuna componente della rete cosmica*
	1.1	Studio dello stato dell'arte nello specifico contesto cosmologico
	1.2	Ricerca e preparazione dei cataloghi per ciascuna categoria di struttura cosmica
	1.3	Studio e definizione del sistema di ML avanzato di segmentazione per ciascuna categoria
	1.4	Evoluzione e determinazione del modello di segmentazione tramite confronto con i cataloghi esistenti
2		*Studio e definizione di una pipeline cosmologica di segmentazione*
	2.1	Studio e definizione di un sistema integrato di segmentazione cosmologica
	2.2	Evoluzione e determinazione della pipeline cosmologica di segmentazione tramite confronto con i cataloghi esistenti
	2.3	Realizzazione di cataloghi e condivisione dei risultati
3		*Studio e definizione di tecniche NAS per l'ottimizzazione dei modelli di Machine Learning per la segmentazione cosmologica*
	3.1	Studio di letteraturea sulle tecniche NAS
	3.2	Studio e creazione di tecniche NAS esaustive per il sistema di segmentazione



Timescale, Milestones, SAL

Project start: Sep/24

Milestone 1 - Nov/24
- Literature Study
- Dataset creation
- AR1.1

Milestone 2 - Feb/25
- Automatic Segmentation
- AR1.2
- AR3.1

Milestone 3 - May/25
- Cosmological Segmentation Pipeline
- AR1.3, AR1.4
- AR2.1

Milestone 4 - Aug/25
- End of project: all deliverables
- AR2.2, AR2.3
- AR3.2