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

V. Del Zoppo, F. Spampinato, G. Puglisi,
C. Carbone, M. Calabrese, L. Naso

Spoke 3 III Technical Workshop, Perugia 26-29 Maggio, 2025

Team Koexai

https://koexai.com

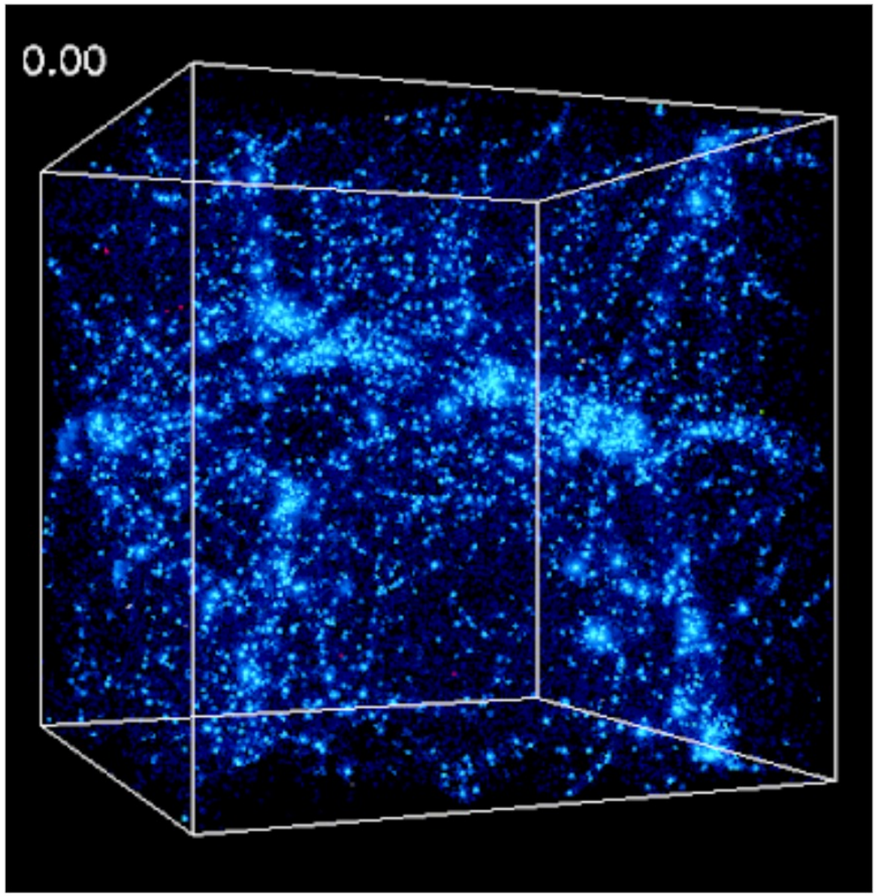
Scientific Rationale

Nbody simulations are a fundamental instrument and vastly exploited in cosmological research.

Modern Nbody simulations with billions of particles require high amount of computational resourced to be analysed.

Current algorithms like FoF, SubFind and ZOBOV are sequential and take weeks to analize a single snapshot on HPC resources.

The aim of DeepCosmoNet is to reduce the amount of time and resources required to detect and segment the cosmological structures to speed up their analysis.

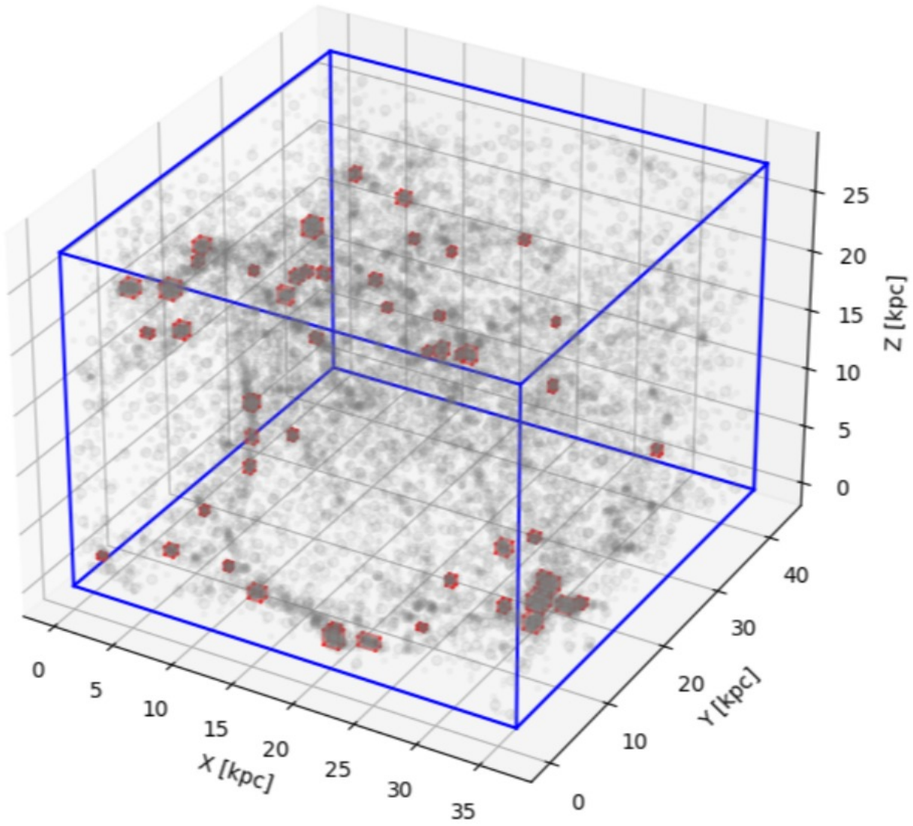


Technical Objectives, Methodologies and Solutions

Two datasets with 10^9 and 10^{10} particles have been provided with labels of halos, sub-halos and voids to test the scalability of our solution and the accuracy of the results.

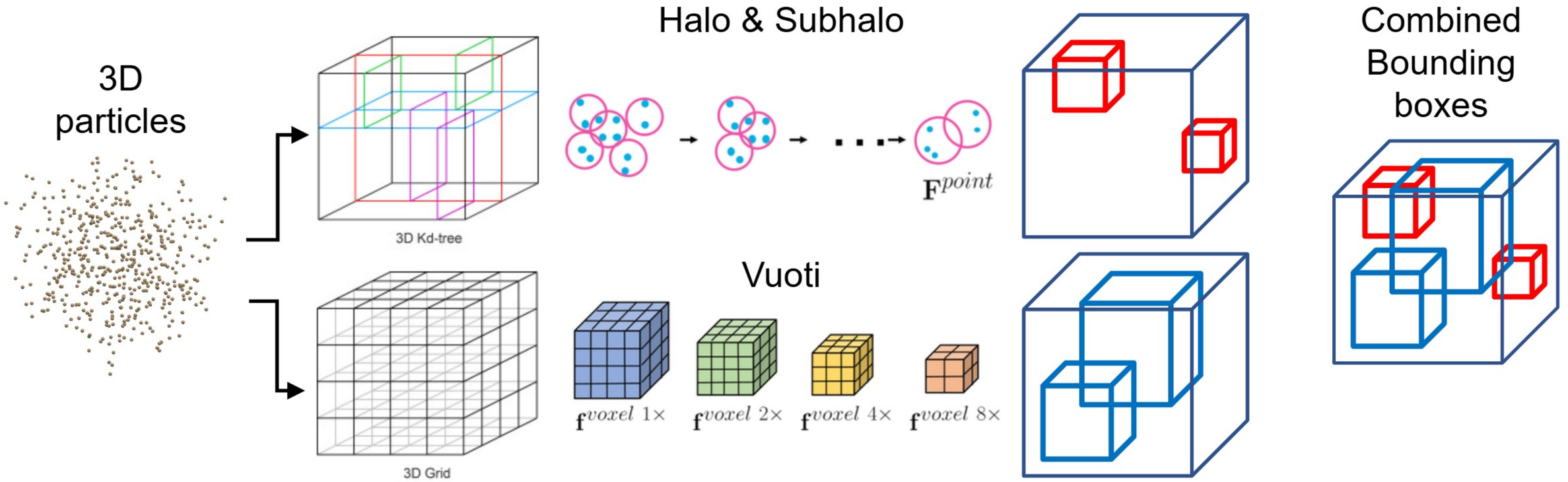
We have build an hyerarchical indexing or "paging" of the particles in smaller chunks to reduce the size of the problem since in any case the complexity of KNN can't go lower than O(n log(n)).

This partitioning, made the problem linear and scalable at the expense of a negligible drop in accuracy due to border effects at the edges of each partition.



Technical Objectives, Methodologies and Solutions

The full pipeline branches in two different partition strategies and models that have been optimsed to different structure types accounting for scales and expected accuracy of the solution.



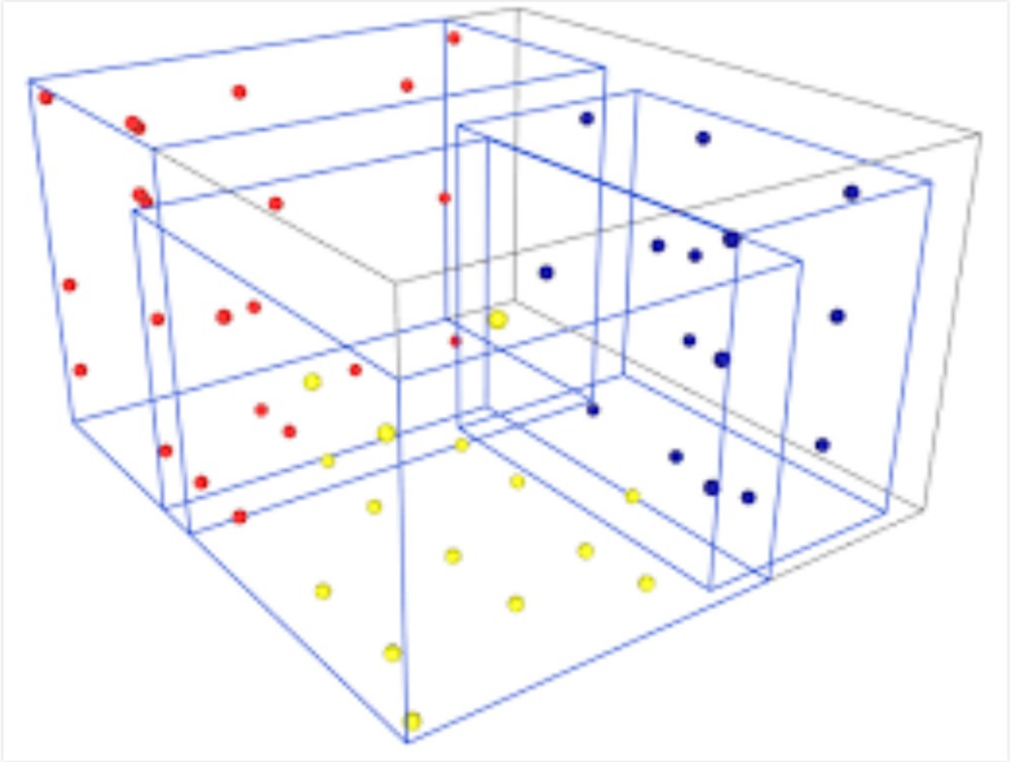
Technical Objectives, Methodologies and Solutions

For the halos and sub-halos the particles have been sorted using a KDtree.

Features have been extracted using optimized KNN queries on neighbors.

An MLP has been employed to perform a first binary semantic segmentation of the particles. The particles selected as part of high density structure are then clustered by HDBSCAN.

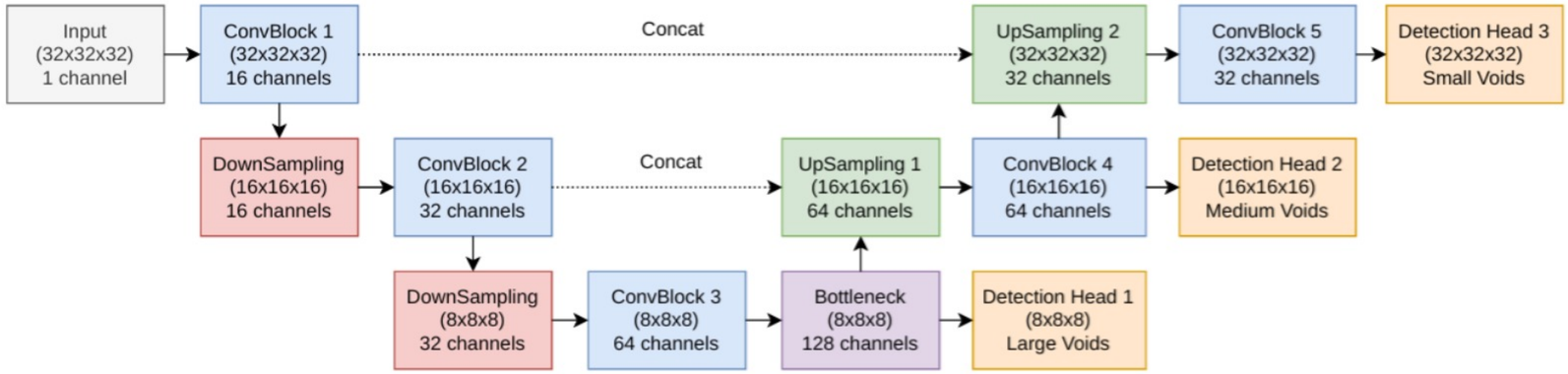
The code has been profiled and for memory and CPU usage. Each of the algorithms employed doesn't exceed O(n log(n)) complexity.

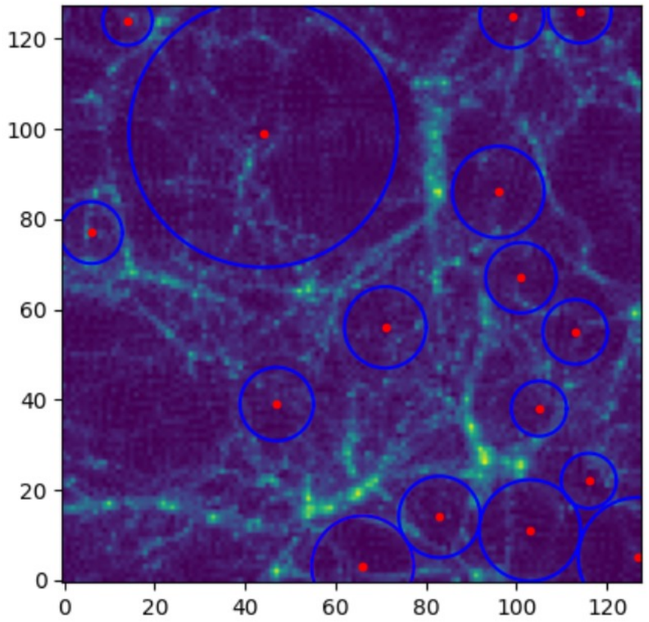


Technical Objectives, Methodologies and Solutions

For the voids detection the particles have been sorted in a 3D grid and passed to a 3DCNN. Here we share the schema of the 3D FPN used to extract voids at different scales.

It mimics the behavior of a YoloV3 model where the standard Convolutions have been replaced by 3D spatial separable Conv blocks, while heads have been modified for 3D bounding box detection.



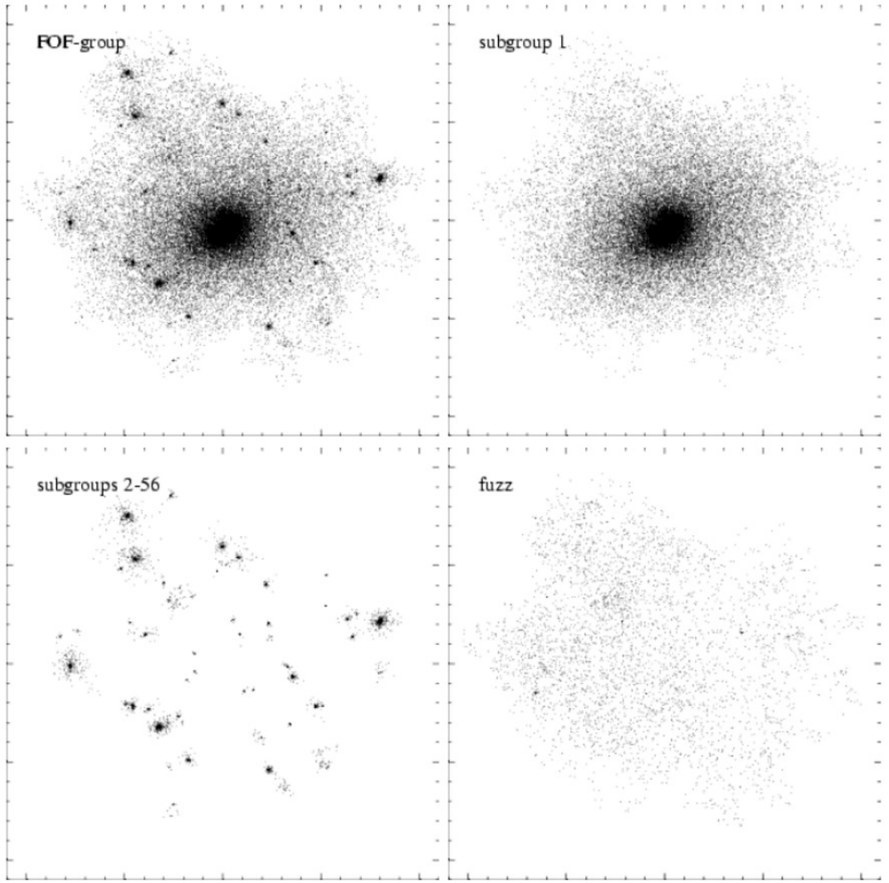


Main Results

The full segmentation trains and runs in minutes on the whole snapshot and achieves a 95% segmentation accuracy, an ARI score above 90% and 89% IoU.

The particles mislabeled by the algorithm are the ones belonging to the fuzz at the boundaries of the structures while the mislabeled structures are not total blunders but mostly clusters that have been merged or splitted compared to the ones given as ground truth.

The sizes of the structures found, follows the same power law distribution and there is no immediate evidence of artifacts in the output

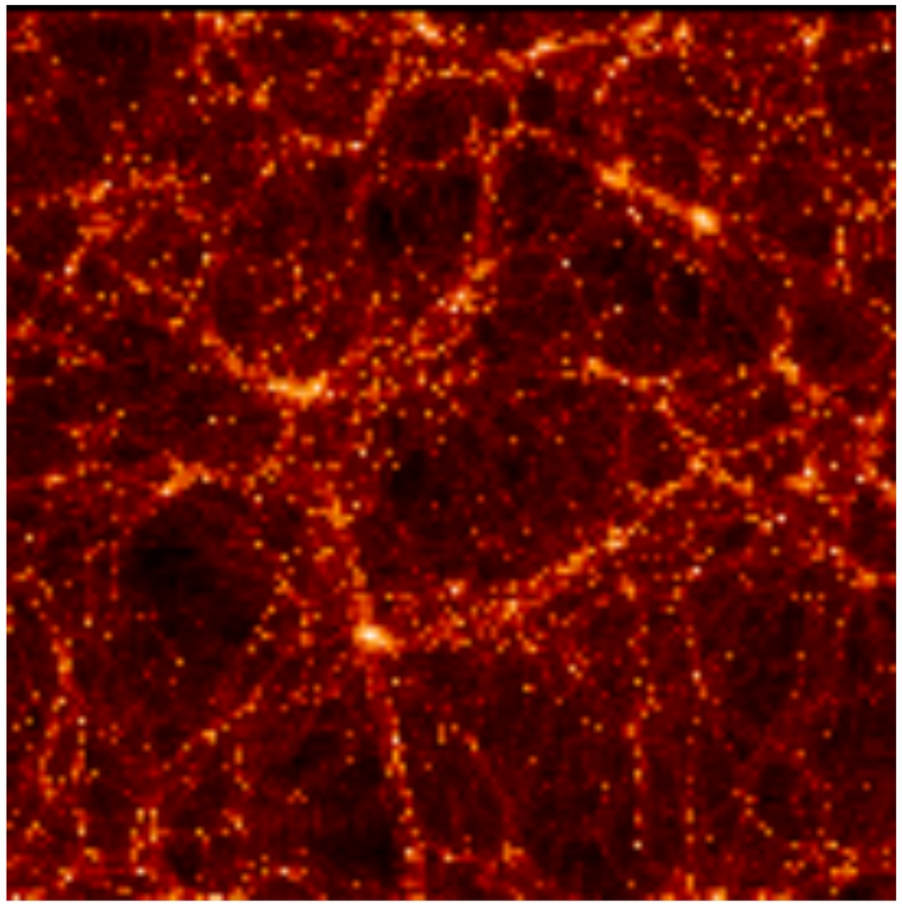


Final Steps

First results on the data available are encouraging to develop the research further in this field.

We aim to refine the feature extracted for each structure type in order to improve the quality of the output and make it robust and accurate at different mass resolutions.

The code will be published and made available for the scientific community for free usage.



Thank you :)

vincenzo.delzoppo@koexai.com