

Detection and characterization of galaxies with deep-learning in radio continuum surveys, preparation to SKAO

The upcoming Square Kilometer Array Observatory (SKAO) will revolutionize the field of radio astronomy, particularly in the study of the Epoch of Reionization. However, its projected data flow (exceeding 700 PB/year) demands a shift toward exascale-ready pipelines. In addition, current automatic detection methods are beginning to be outpaced by the ever-increasing complexity of the data. In this Big Data era, Machine Learning (ML) stands out for efficient astronomical data processing.

During this talk, I will present our work on developing a robust and fast pipeline for radio-galaxy detection and characterization in 2D continuum images for SKAO, using precursor instruments (such as LOFAR, ASKAP, MeerKAT...) for methodological developments.

Building on the MINERVA team's success with the YOLO-inspired (You Only Look Once) regression-based method during SKA Data Challenges (SDC) 1 and 2, I will show how we generalized the method to train a new source detector model on our custom simulation pipeline. The latter was developed with the aim of simulating any deconvolved radio interferometer response for full-field observations, including point sources, but also resolved Active Galactic Nuclei (AGN) jets, star-forming galaxies, ionospheric artifacts... I will then show how we deploy our network models trained on simulations to the LOFAR Two-meter Sky Survey containing 13 million sources identified with automatic detection methods, and I will discuss how it compares to our detection performances, especially for "hard detections" like artifacts around bright sources, AGN with complex morphologies, low-SNR sources, and also in terms of computation time.

Finally, I will present the exploration of alternative self-supervised strategies, pre-training networks on target data, in order to improve recall and characterization accuracy for the faintest sources.