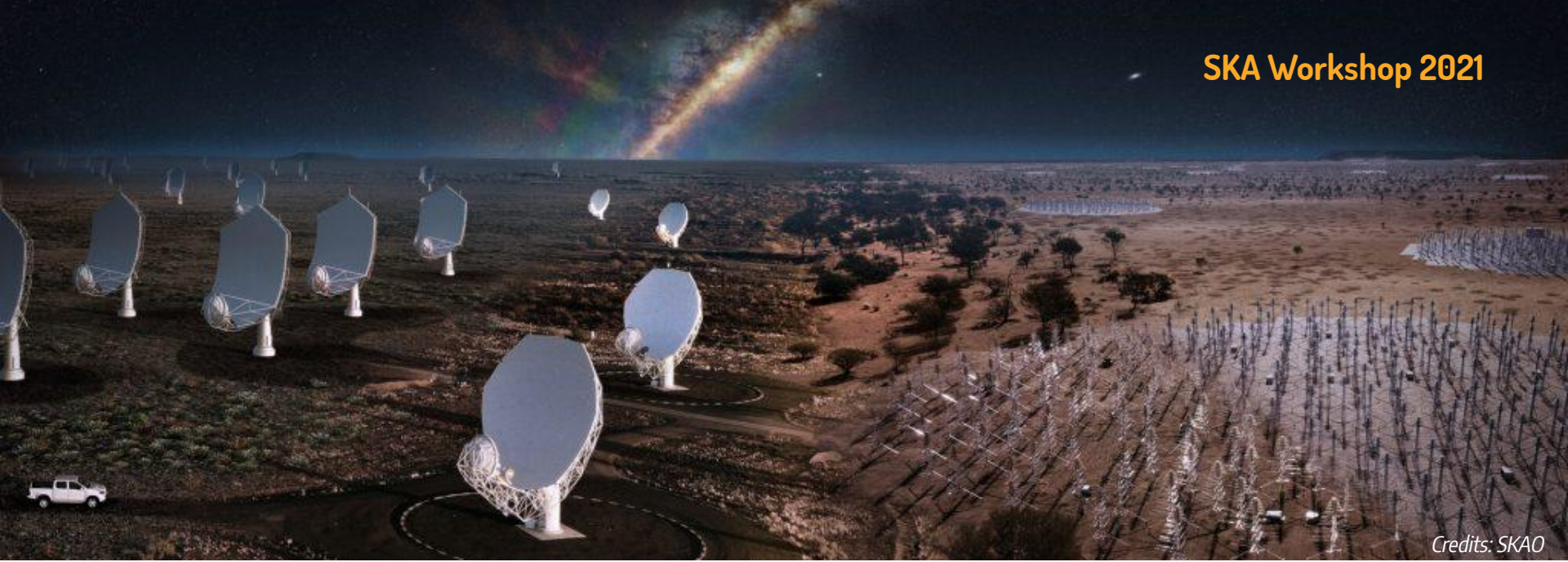


VISUALIZATION AND SOURCE ANALYSIS SERVICES FOR THE SRC

Simone Riggi
INAF-OACT

INAF
ISTITUTO NAZIONALE
DI ASTROFISICA

+ CIRASA, NEANIAS & ML4ASTRO collaborators

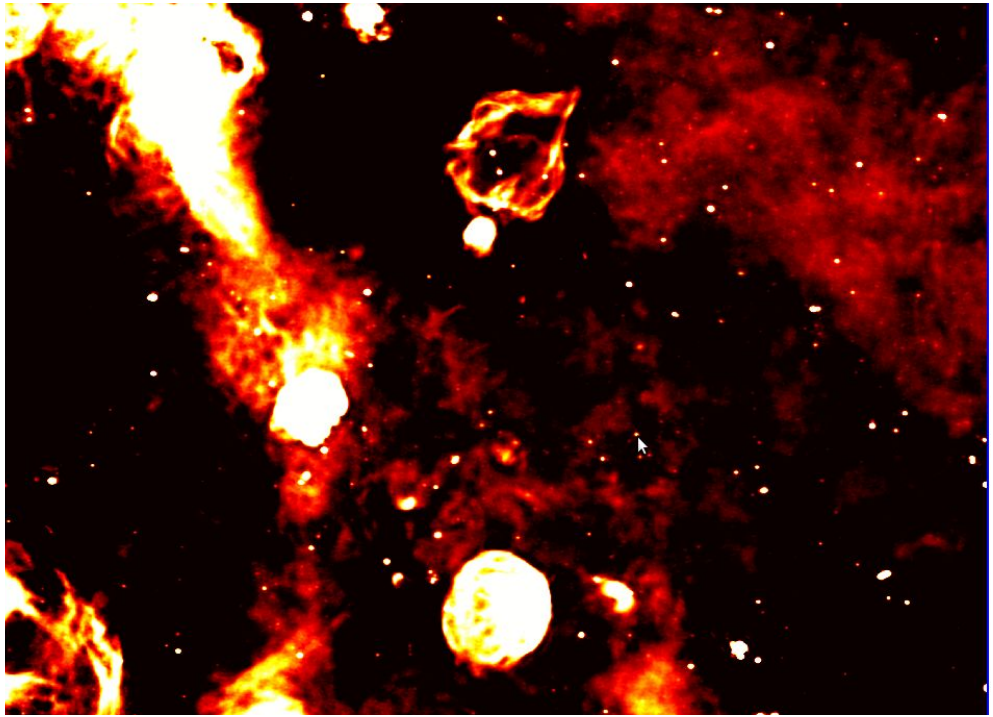
SRC SCIENCE USE CASES

- **SKAO & SKA Regional Centers (SRCs) responsible for *(see Possenti's presentation)*:**
  - **generation and lifecycle of Advanced Data Products (ADPs)** from Observatory or Project level data products (ODPs, PDPs)
  - **providing computing infrastructures and distributed services** to user community to deliver SKA science (following FAIR principles)

- **ADP generation and lifecycle**
  - joint responsibility of the science users and SRC Network
  - usually requiring interactive visualisation and comparison to other data
  - subjected to provenance and reproducibility policies
  - possibly complying with IVOA standards and accessed through VO services (TBD)

- **Several science cases being studied by the SRC WG3 & WG6 groups**
  - Our activities currently focusing on:
    - ✓ **SRC-WG3-TP2-USE-01:** Generation of Source Catalogues
    - ✓ **SRC-WG3-TP2-USE-04**: Postage Stamps & Stacked Postage Stamps

SCIENTIFIC DRIVERS



Patch of the Scorpio field observed with ASKAP @ 912 MHz

Galactic science mainly driving our technological developments

(see F. Cavallaro's talk)

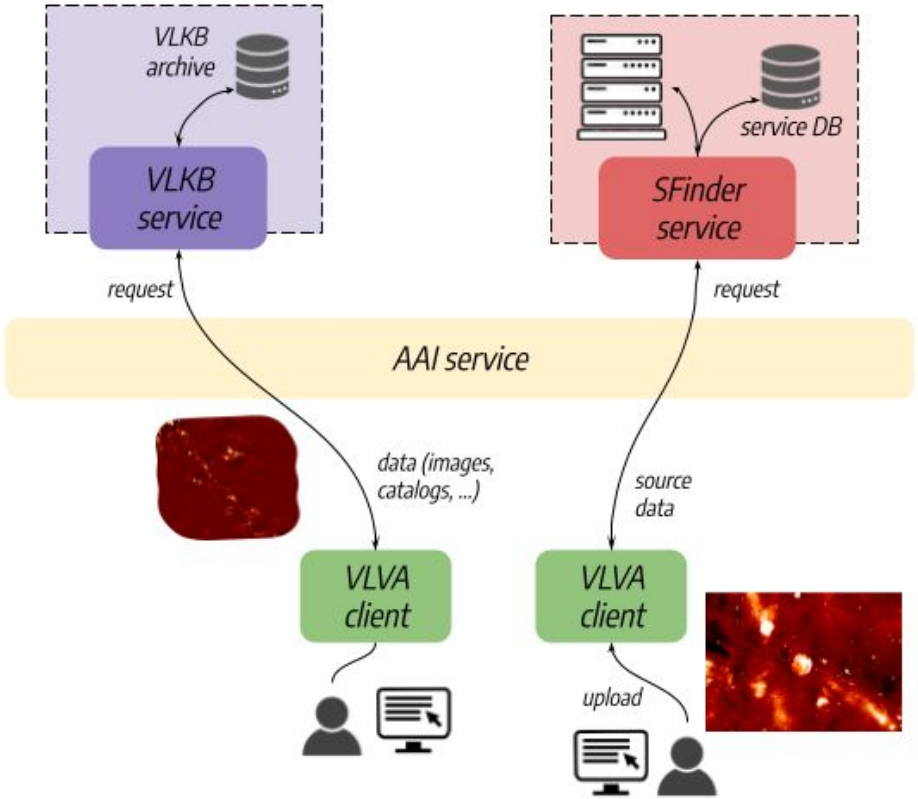
- **ASKAP EMU Survey**
  - Early Science observations of the SCORPIO field @ 912-1615 (2018-2019)
  - EMU pilot survey Phase 1 @ band 1 (2019)
  - EMU pilot survey Phase 2 @ band 1 (2021), including the SCORPIO field and other GP regions
- **MeerKAT Galactic Plane Survey**
  - Covering a large part of the GP @ 886 – 1678 MHz ($|b|<1.5°$, $2°<l<60°$, $252<l<358°$)
- **SKA Our Galaxy KSP**
  - Building on precursor experience, providing inputs to SKA WGs

⚠ | ADP GENERATION ISSUES

Limitations found when producing ADPs from SKA precursor data

- **Increased data product size (e.g. image, source density)**
  - Scalability issues in existing tools
  - Procured additional and dedicated resources for analysis and software development
- **Lack of tools to improve catalogue production automation and reproducibility**
  - Extended source extraction algorithms
  - Source classification (spurious vs real, galactic vs extragalactic, galactic classes) tools
  - Tools for unknown/anomaly detection, removal of duplicates,
- **Data visualization**
  - Starting to be challenging on local viewers
  - Limited source analysis functionalities in existing viewers
- **Knowledge gaps**
  - Astronomer community: best practises on computing, analysis automation and reproducibility
  - IT team: understanding of scientific use cases, science goal orientation

CIRASA PROJECT



CIRASA INAF PRIN TEC objectives

- Tackle selected SKA science cases from Galactic science perspective
- Speed-up the source cataloguing process for next-generation radio surveys
- Develop a visual analytic platform through integration of source finding, visualization, and knowledge base services
- Develop and integrate new ML source finders in the platform
- Add new interactive source visualization and analysis features
- Develop proto SKA Regional Center solutions, scalable to larger infrastructures

RELATED PROJECTS

NEANIAS

Horizon 2020
European Union Funding
for Research & Innovation

Novel EOSC services for Emerging Atmosphere, Underwater & Space Challenges

- Design innovative thematic services, following FAIR and Open Science principles
- Integrate services into the European Open Science Cloud (EOSC)
- Deliver services to end-user communities (Atmosphere, Underwater, Space)

ML4ASTRO Collaborations

- Development of ML techniques for astronomy
- MoU INAF-OACT-UniMalta: promoting scientific cooperation and mobility of researchers and PhD students
- Similar collaboration ongoing with the University of Catania and Milano Bicocca

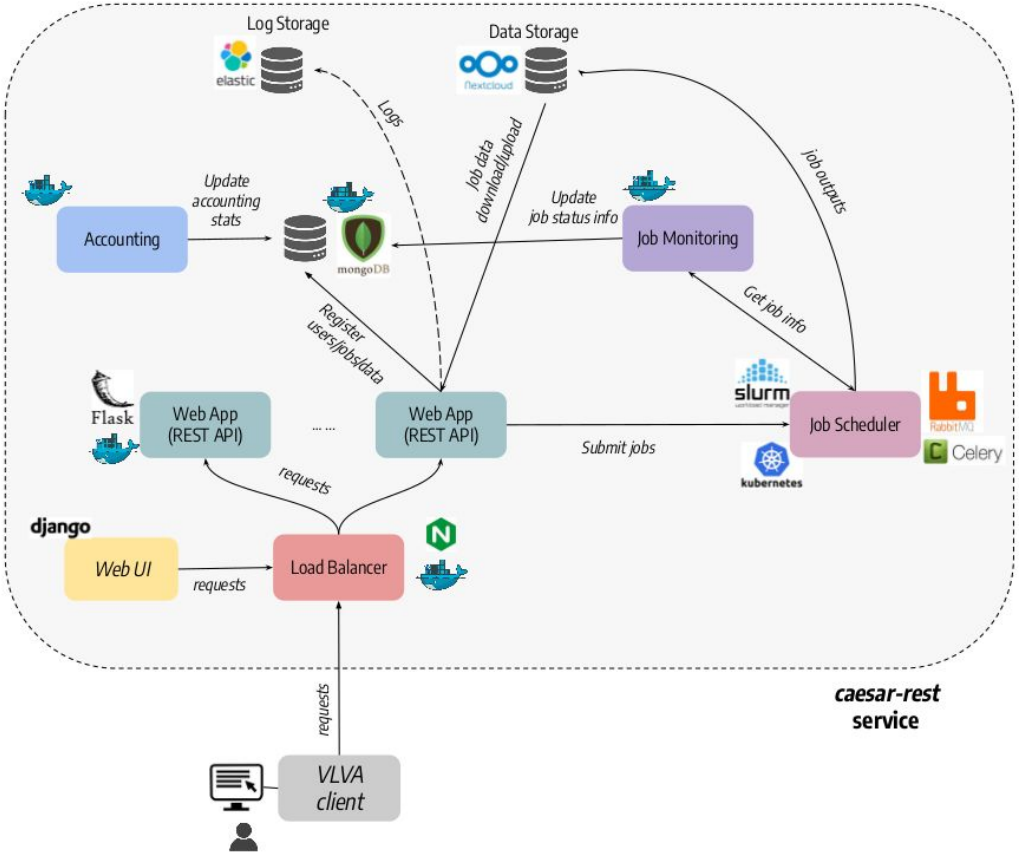
INAF ISTITUTO NAZIONALE DI ASTROFISICA

STVDIVM GENERALE SICILIAE 1434

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

L-Università ta' Malta
Institute of Space
Sciences & Astronomy

SFINDER SERVICES

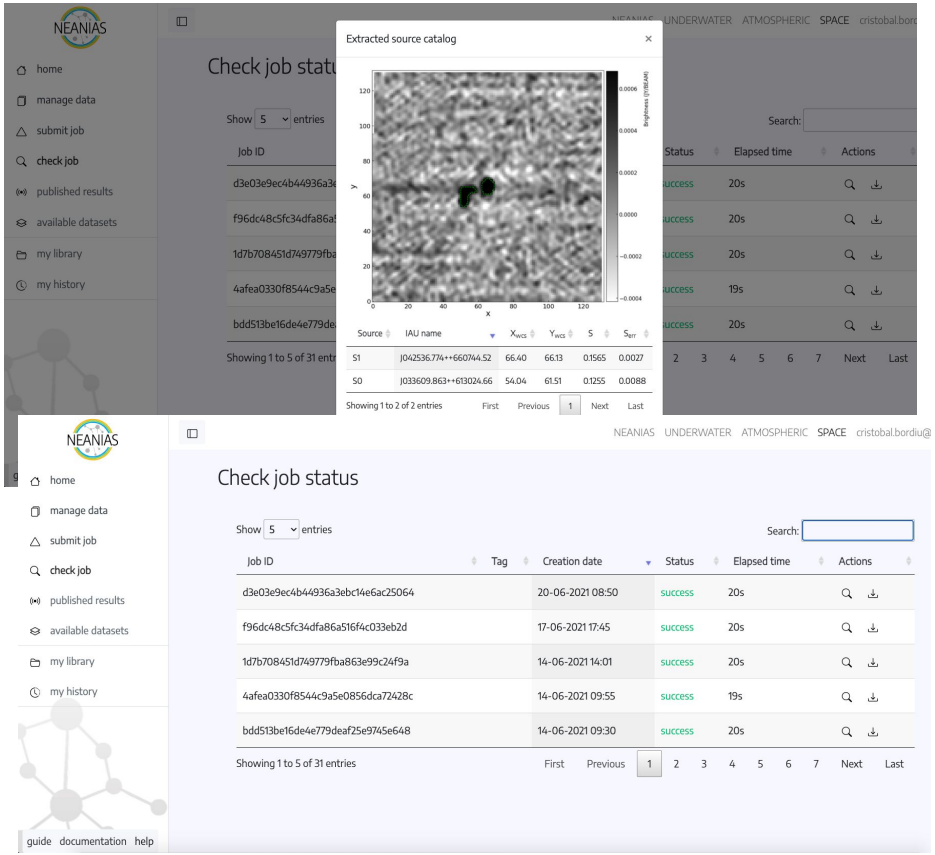


caesar-rest: *A REST-ful web service based on Flask framework for running CAESAR source finding jobs*

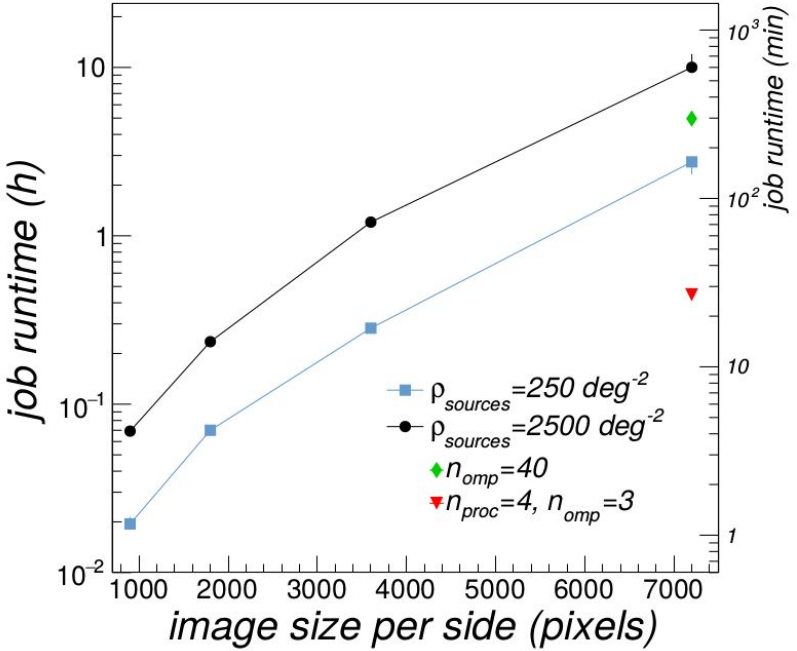
- Deployed and tested on GARR OpenStack Kubernetes cluster + CIRASA dedicated resources
- Multiple run strategies
  - Kubernetes Jobs (Docker)
  - Slurm Jobs (Singularity)
  - Celery async tasks
- Integrated with NEANIAS EOSC services (AAI, Logging, Accounting)
- Other source finders (ASGARD, CUTEX, Aegean) under integration
- Integration with ViaLactea visualization client ongoing

More details @ https://github.com/SKA-INAF/caesar-rest

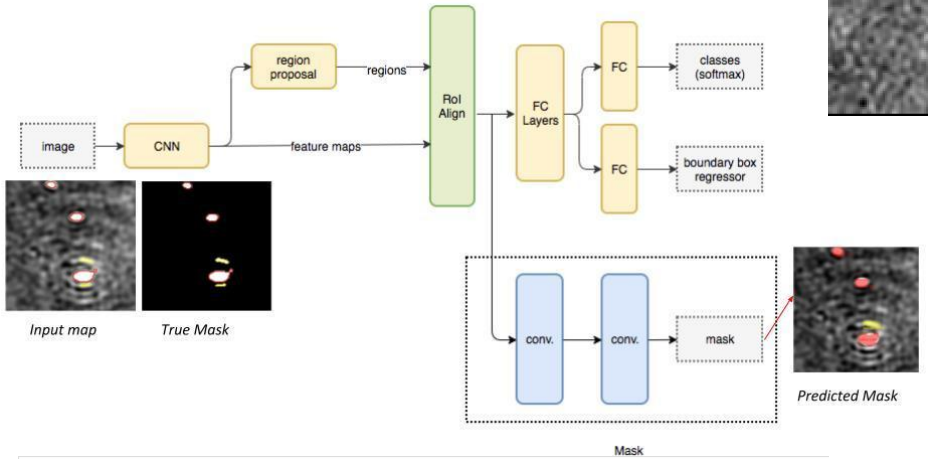
SFINDER SERVICES (contd)

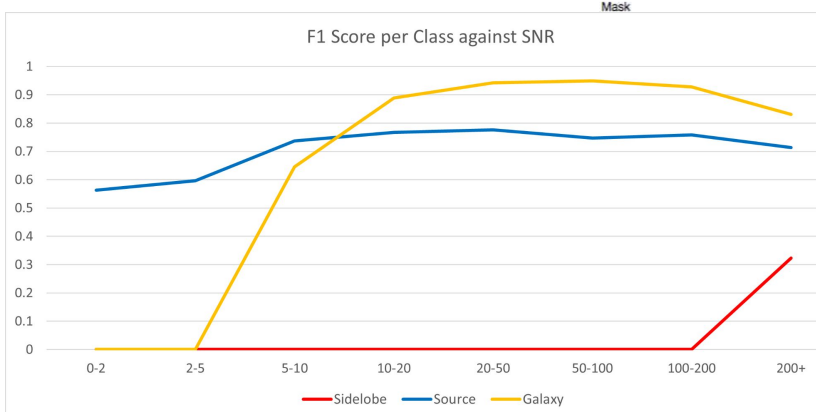


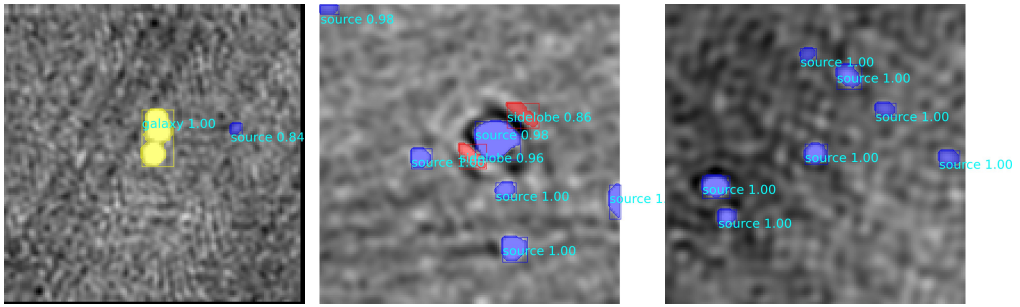
- Currently available on the EOSC marketplace, accessible from Django web UI
- Testing activities ongoing with ~30 users
- Searching for additional resources for larger scalability tests



ASGARD







Automated Source, Galaxy, and Artefact R-CNN Detector (ASGARD)

- Based on Mask R-CNN object detection framework
- Allowing detection of compact sources, sidelobes and extended radio galaxies
- Trained on ASKAP, RGZ FIRST, ATCA & MeerKAT data (galaxies, sidelobes, point-sources), ~15k images
- Used as Caesar classification step or as a new finder (detection + classification)
- MPI parallel version in alpha version
- Performance optimization studies ongoing

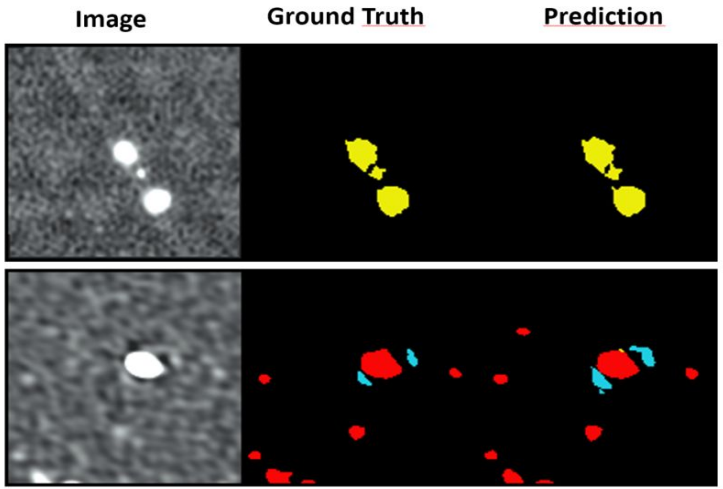
More details @ https://github.com/SKA-INAF/mrcnn

D. Magro & S. Riggi, PASA, 2021

NEW ML APPS

DE:TR

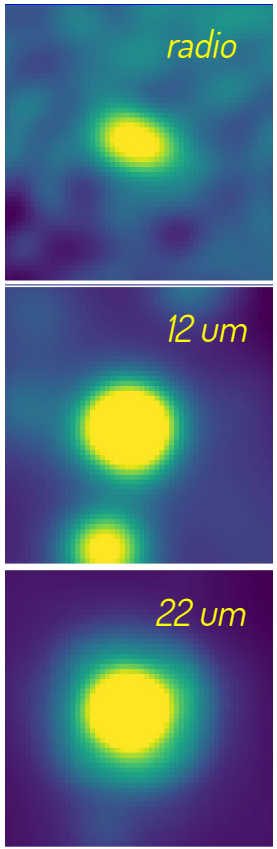
- Transformer model for object detection
- Removes the necessity for RPN, typical of R-CNN based models
- Heavier in terms of resources
- Preliminary results in images below



Semantic Segmentation with Tiramisu model

- Uses semantic segmentation, a different approach than object detection, to achieve the same goal of source detection
- Based on U-Net model
- Comparable results with ASGARD

SCUTOUT SERVICE



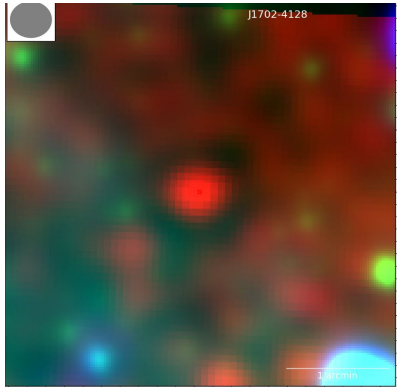
Source cutout service

- **Functionalities**
  - identify 2D survey data overlapping with a circle region and cutout on position
  - regridding outputs from multiple surveys to common WCS, resolution, map units
  - background subtraction
- **Technical details**
  - Mostly based on Montage
  - ~20 radio and infrared surveys available (~4 TB), no DB
- **Status**
  - Currently provided as a batch application for producing ML datasets
  - Service to be likely integrated in *caesar-rest*

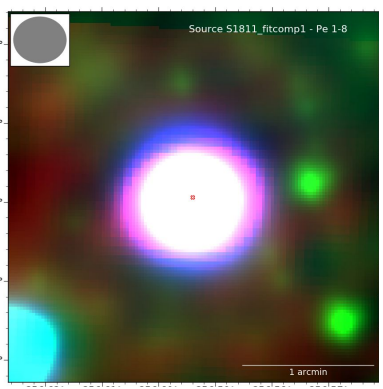
More details @ https://github.com/SKA-INAF/scutout

SCLASSIFIER SERVICE

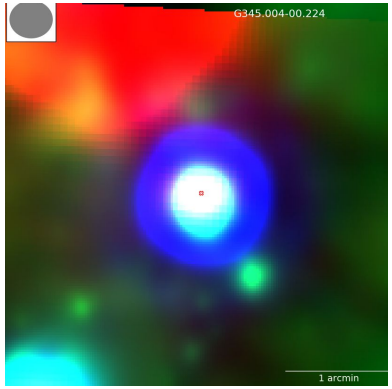
Known pulsar



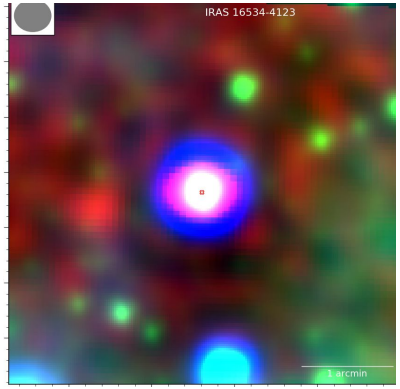
Known PN



Known HII region

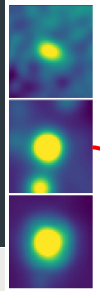


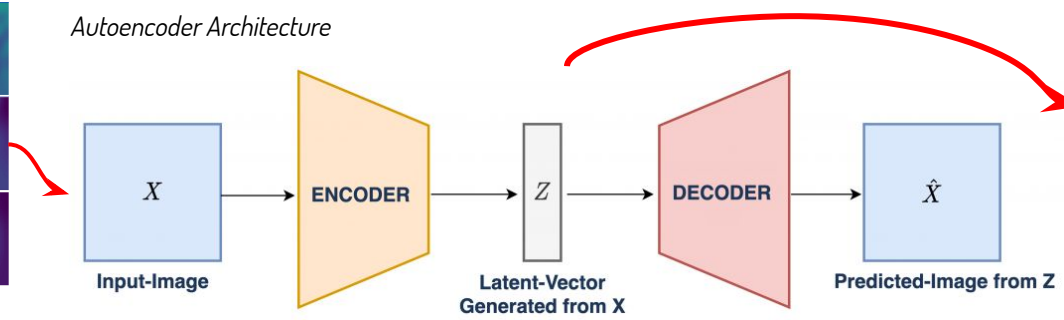
Known star

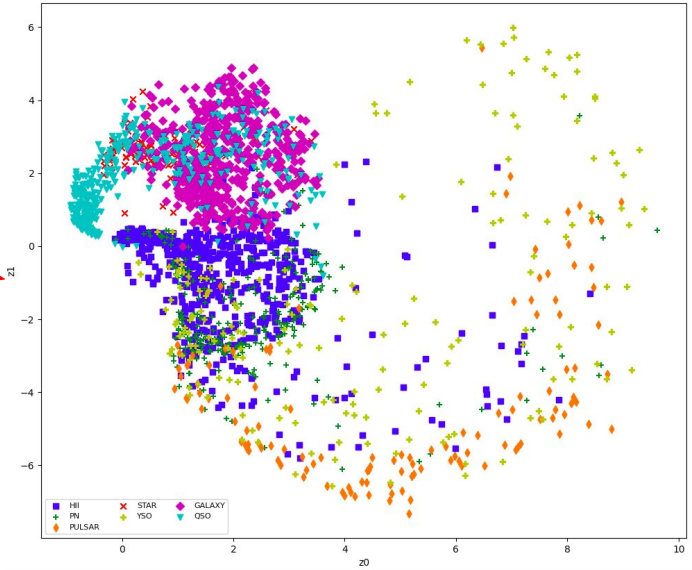


- **Goal: Developing classification/diagnostic tools using multiwavelength data:**
  - identify Galactic vs extragalactic objects, or discriminate individual Galactic classes
  - discover unexpected/anomalous objects, identifying the most similar known objects
- **Major issues**
  - Lack of homogeneous radio labelled datasets, possibly at multiple frequencies
  - Limited number of training data for some Galactic classes (~hundreds)
  - No catalogued extragalactic objects (e.g. RG, QSO) covering the GP, using extragalactic surveys
  - No full-sky coverage for some discriminant surveys (e.g. HiGAL @ 70 micron)

SCLASSIFIER SERVICE (contd)

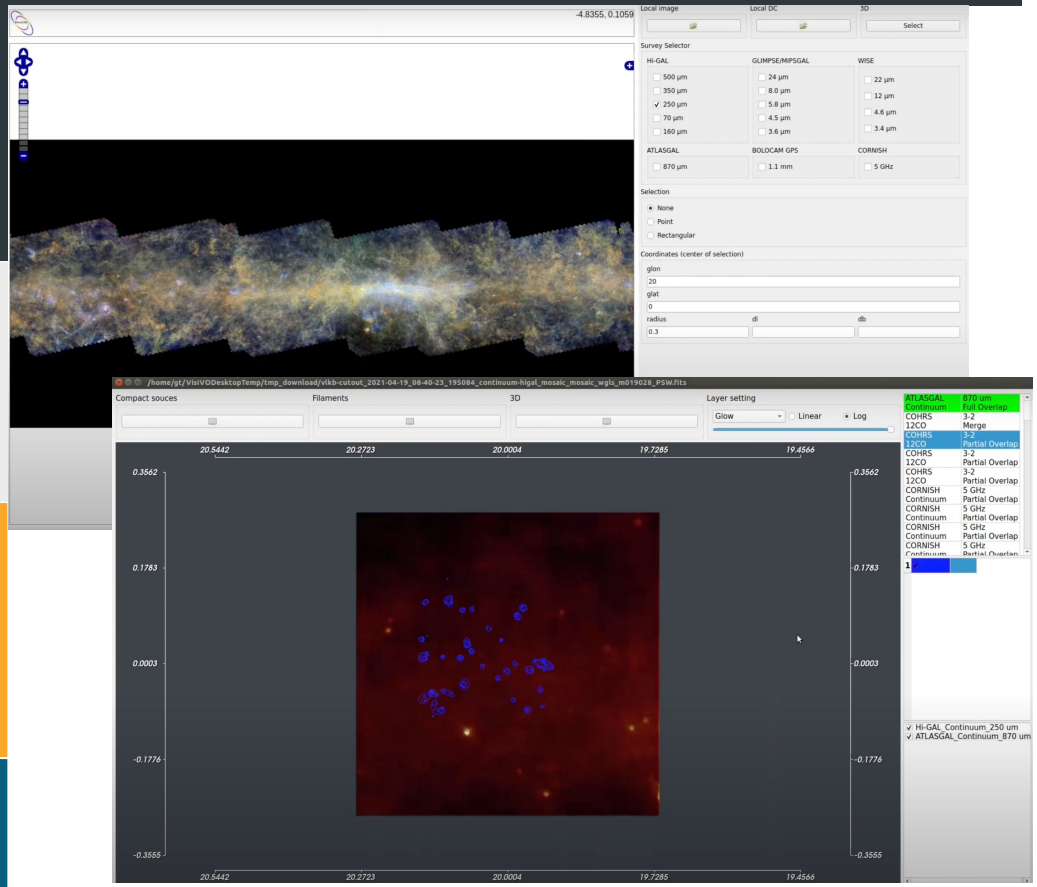






- **Dataset preparation ongoing**
  - MeerKAT GPS survey providing high-resolution data for a large portion of the GP, at multiple frequencies
  - Data cleaning requires a considerable effort
  - ~15 K objects (⅓ of the available data) currently available
- **Different unsupervised methods being considered**
  - First runs performed with different tools (autoencoders, UMAP) on a small (not fully cleaned) dataset
  - Some methods highly dependent on hyperparameters or network architecture
  - Tests to be repeated with improved dataset
- **UNIMIB collaborators developing a dedicated service (including different algorithms) for EOSC deployment**

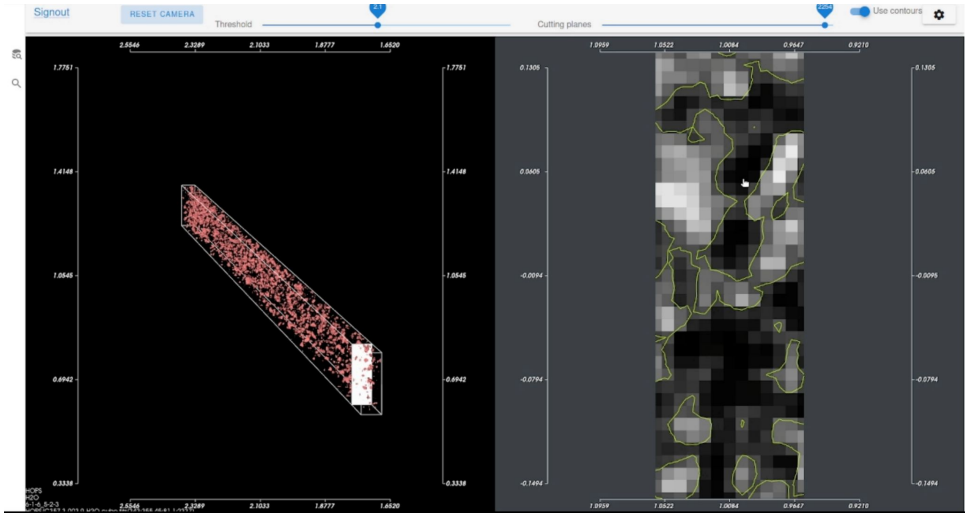
ViaLactea Visual Analytics



ViaLactea Visual Analytic (VLVA)

- Provides access to GP radio & infrared surveys archived in the Knowledge Base (VLKB)
- Supporting visualization of 2D images and 3D velocity datacubes (vol. renderings, slices), loaded locally or from VLKB
- Enabling visualization of compact sources and filaments from VLKB
- Currently being integrated with source finding services
- Available as desktop application for Mac OS and Ubuntu and also as Docker container.
- Open Source at GitHub: https://github.com/NEANIAS-Space/ViaLacteaVisualAnalytics
- See user manual at: https://vlva.readthedocs.io

ViaLactea Web

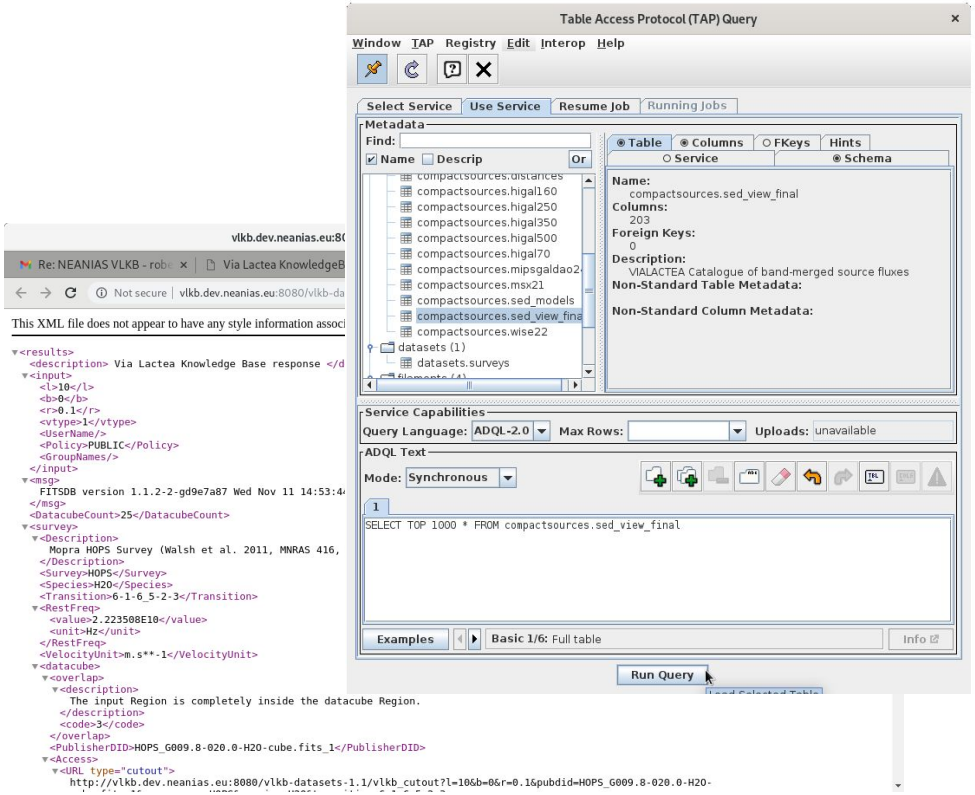


ViaLactea Web application (VLW)

- Work-in-progress simplified web version of the VISUAL ANALYTICS TOOL, developed in collaboration with University of Portsmouth (UK)
- Currently supporting only 3D velocity datacubes visualization
- Multi-user support within web environment
- Provides efficient offscreen visualisation (GPU and CPU rendering) on remote server

- Web browser solution with desktop/mobile/tablet support that is flexible for future functional extension
- Working on full containerization for optimized deployment on cloud and distributed infrastructures

ViaLactea Knowledge Base



ViaLactea Knowledge Base (VLKB)

- Provides discovery services and access to data collections and catalogues of the galactic plane, initially made available by the EU FP7 VIALACTEA Project.
- Accessible through a Virtual Observatory enabled infrastructure (by the TAP protocol).
- Data access available through REST-API services: *search*, *cutout* and *merge*.
- Also provides information about compact sources, filament structures and numerical SED models
- Fully exploited by ViaLactea Visual Analytic client tools (desktop and web).
- Secured under Authentication and Authorization Infrastructure (AAI)
- Handles user privacy roles regarding access to specific surveys

Summary

- **Technological solutions being developed to support SRC and precursor use cases**
  - CIRASA: development and integration of existing and new services for source analysis, visualization and knowledge base
  - NEANIAS: development, integration and deployment with external EOSC service and infrastructure

- **What we have done so far?**
  - Source finding services developed. integrated and deployed on EOSC
  - VLKB services integrated and deployed on EOSC, integrated with VLVA
  - New ML source finders and source cutout tools developed

- **What we are doing now?**
  - Integration of other existing applications (e.g. Aegean, CUTEX, SoFIA) in source finding service
  - Integration of source finding services with VLVA
  - Setting up of larger computing infrastructure (Kubernetes) for scalability tests
  - Dataset preparation for new source classifiers
  - Developing new features in source cutout application and service integration
  - Developing source visualization and tagging features in VLVA client

THANKS TO ALL CONTRIBUTORS

F. Vitello (IRA), G. Tudisco (INAF), E. Sciacca (INAF), C. Bordiu (INAF), C. Pino (INAF), D. Magro (UoM/INAF), R. Sortino (UNICT/INAF), A. De Marco (UoM), C. Spampinato (UNICT), F. Bufano (INAF), M. Molinaro (INAF), S. Molinari (INAF), M. Benedettini (INAF), K. Zarb Adami (UoM), T. Cecconello (UNIMIB), G. Vizzari (UNIMIB), G. Fiameni (CINECA)

+ the INAF-OACT radio & IT group