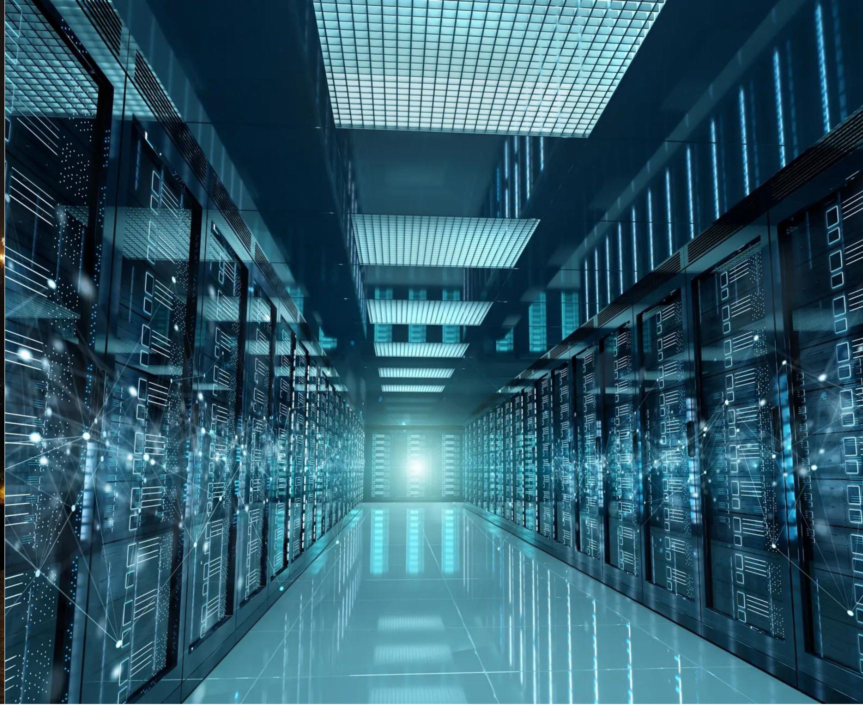


Infrastrutture di calcolo e sviluppo di software scientifico @ INAF-OACT

Simone Riggi & Eva Sciacca

INAF-OACT

+ other contributors

INAF
ISTITUTO NAZIONALE DI ASTROFISICA

CONTENT OUTLINE

- ## Computing Infrastructures @ OACT
  - CHIPP and PLEIADI resources
  - Next-generation computing devices for astronomical data processing

- ## Software for astronomical data simulation, processing and visualization
  - **Visual analytics & virtual reality applications**
  - **Cloud services**
  - **HPC applications**
    - ✓ GAIA astrometric core solver
    - ✓ LOFAR & MeerKAT radio data processing pipelines
    - ✓ Radio source finding
    - ✓ Relativistic jet-ISM modelling
  - **Machine learning applications**
    - ✓ ML for the analysis of radio data
    - ✓ ML for the analysis of solar data

COMPUTING HARDWARE @ OACT

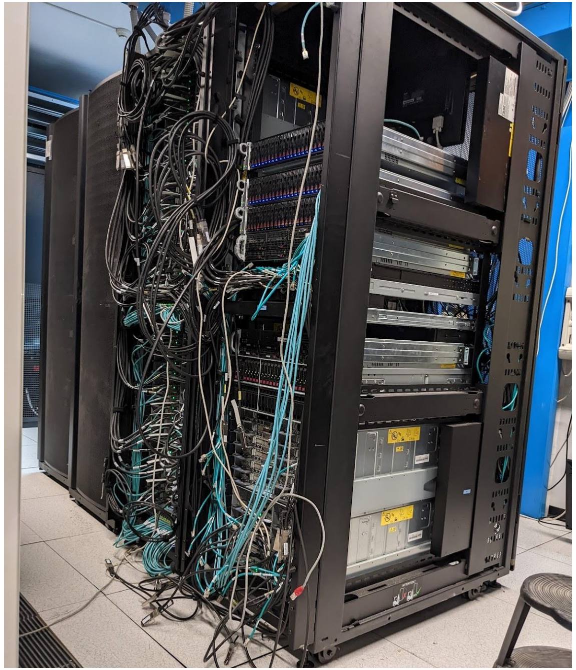
FROM CHIPP TO PLEIADI

- **Experience in the design, setup and maintenance of HPC/HTC computing systems**
  - **PI2S2:** Creation of the first Sicilian GRID infrastructure for scientific and industrial application, integrated with the italian and European GRIDs;
  - **CHIPP** project funded by the Italian Institute for Astrophysics (INAF) and promoted by the ICT office of INAF. The main purpose of the CHIPP project is to coordinate the use of, and access to, **already existing** high throughput computing and high-performance computing and data processing resources (for small/medium size programs) for the INAF community.
  - **Pleiadi** CHIPP successor's. **Re-engineering** of Galileo Cineca Cluster to meet the computing needs of the INAF community

- **User Support for:**
  - Use of the resources with one to one call, seminars, documentation
  - Porting of application, drive the user to obtain the best performance from their applications using the available resources

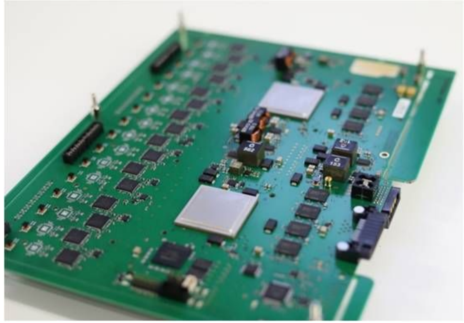
PLEIADI OACT INFRA

- **12 nodes 256 GB RAM**
- **60 nodes 128 GB RAM**
- **6 nodes with GPU (128 GB RAM)**
  - 2 Tesla V100 PCIe 16 GB
  - 4 Tesla K40m 12 GB
- **Storage: 170 TB**



- **Nodes:** 2 x Intel Intel(R) Xeon(R) CPU E5-2697 v4 @ 2.30GHz in total 36 cores with 128 or 256 GB RAM
- **Network:** Omni-Path HFI Silicon 100 Series, 100Gbits interconnect

NEXT-GENERATION DEVICES



- **ITPM 1.6 :**
  - Designed by INAF/Uni Oxford /Uni Malta/ Sanitas EG (PI Schillirò)
  - Is the LFAA Tile Processing Module ready for mass production

- **Versal ACAP VCK190**
  - Used for AI commercial and RESEARCH Use Case (MOSAICO Project , PI Schillirò)

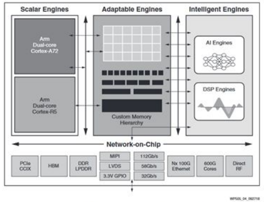


Figure 4: **Xilinx Versal ACAP Functional Diagram**

Table 2: **Versal ACAPs and Targeted Markets**

Market	Benchmark	vs. CPU	vs. GPU	vs. FPGA	Notes
Data Center	Image Recognition (Inference) - Latency insensitive	43X	2X	5X	GoogLeNet v1 (unlimited Batch size)
	Image Recognition (Inference) - 2ms latency	n/a	8X	5X	GoogLeNet v1 (< 2ms) CPU latency floor 5ms
	Risk Analysis	89X	n/a	>1X	Value at Risk (VaR) for Interest Rate Swaps Maxeler result
	Genomics	90X	n/a	>1X	Human Gene Analysis Edico Genome result
	Elastic Search	91X	n/a	>1X	91X lower latency BlackLynx result on 1TB data
Wireless 5G	16x16 5G Remote Radio	n/a	n/a	>5X	>5X more radio bandwidth for 5G Remote Radio
	Beam Forming	n/a	n/a	>5X	>5X more compute
A&D Radar	DSP TMACs	n/a	n/a	>5X	Over 27 TMACs
	Algorithm iteration time	n/a	n/a	>100X	Software programmable intelligent engines compile in minutes

- **Ultrascale Zynq RFSoC**
  - A possible candidate for Meerkat Band5 Digital Processor Module.

BIG DATA ANALYSIS/VISUALIZATION
(VisIVO Lab)

VisIVO Lab

→ See Poster + Light Talk
G. Tudisco

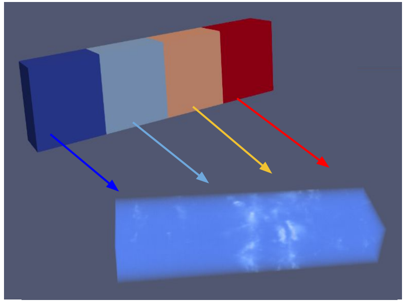
VisIVO LAB is a competence center for the development of Big Data Analysis techniques based on Visualization.

About 20 years of experience in the Scientific visualization field, with the development of the **VisIVO Framework**, a suite of software tools for creating customized views of 3D renderings from astrophysical data tables.

Latest development includes:

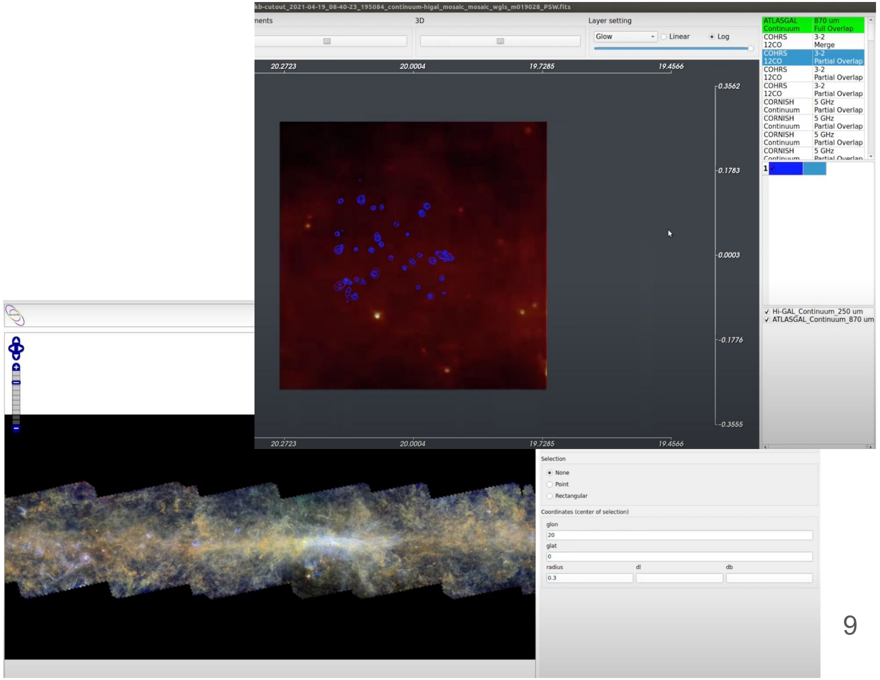
- **VisIVO for Pluto**
  - Creating a visualization environment for Pluto Code (http://plutocode.ph.unito.it), including wrapper to create Visualisation with Python scripting
- **ViaLactea Visual Analytics (VLVA)**
  - Open-source desktop application that offers a visual analytics environment to conduct research activities on two-dimensional regions of space and three-dimensional datacubes.
  - Official client of the VLKB which includes 2D and 3D surveys, numerical model outputs, point-like and diffuse object catalogues.
- **Virtual Reality (VR)**
  - Collaboration with Inter-University Institute for Data Intensive Astronomy (IDIA, SA) to develop iDaVIE a VR software to allow data cube investigation (Meerkat HI Cubes).

VISUALIZATION FOR SKA



- **INAF leading "*Orange*" team within Science Regional Centres (SRCs) WGs**
  - SAFe® agile team with INAF Scrum Master & Product Owner roles + team members
  - Working on Prototype 4 to **deliver interoperable visualization services** for SKA SRCs

- **VLVA being upgraded to address SKA visualization and data analysis challenges**
  - Some new features from CIRASA & ECOGAL projects
    - ✓ Multiple local image file & ds9 region importers
    - ✓ source finding and SED fitting service interface
    - ✓ source catalogue visualization & filtering
    - ✓ VLKB multi-cutout service interface
  - VLVA upgrade for SKA
    - ✓ Remote visualization with scalable infrastructures
    - ✓ Parallel and distributed visualization pipelines



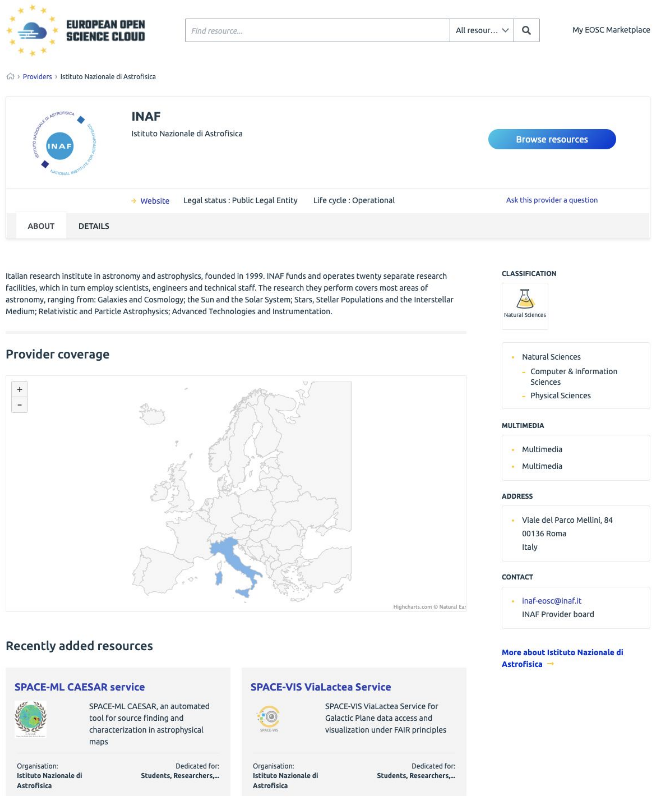
→ More details @ https://visivo.readthedocs.io/en/latest/

CLOUD SERVICES

CLOUD SERVICES for ASTROPHYSICS

- **Porting of tools and services to the Cloud**
  - ViaLactea, available a cloud based knowledge base and a simplified web version of the visual analytic client.
  - CAESAR, available via REST-APIs and developed a fully web base UI
  - ASTRO-ML & Tiramisu @ AI Gateway on JupyterHub
  - VisIVO @ Visualization Gateway on JupyterHub
- **Deployed on the GARR Cloud**
  - OpenStack based Cloud Platform
  - Kubernetes based Container Platform
- **Available on the European Open Science Cloud**
  - Access from the EOSC Marketplace
  - INAF as resource Provider
  - Onboarded ViaLactea and CAESAR services
  - Available on the NEANIAS service catalogue and on the SPACE Thematic Portal.

→ See Poster E. Sciacca



HPC SOFTWARE APPLICATIONS

GAIA Solver on pre-Exascale systems

The Gaia AVU-GSR (Astrometric Verification Unit - Global Sphere Reconstruction) solver

- **Objective**: find positions and proper motions of ~100M stars in the Milky Way with a micro-arcsec resolution, by solving a system of linear equations
- Based on a parallel (MPI+OpenMP) implementation of the LSQR algorithm for large and sparse matrices
- In production for the ESA Gaia mission since 2014 under a INAF-CINECA agreement, with the ASI support
- GPU porting with CUDA:
  - **Performance gain factor of ~14 wrt the CPU production version**
  - Tested on the Marconi100 CINECA platform (4 NVIDIA V100 16 GB GPUs per node)
  - System solution consistent with CPU version (numerical stability)
  - Future porting on the Leonardo CINECA infra (4 A100 GPUs per node)
- Further extensions planned for PNRR Centro Nazionale (e.g. towards Green Computing)

→ See Poster + Light talk
V. Cesare

gaia

The Marconi100 cluster (CINECA)

Streaming Multiprocessor of V100

MeerKAT/LOFAR pipelines

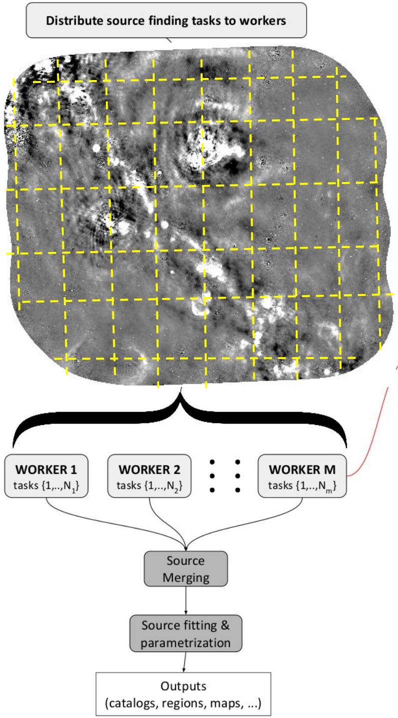
- **Reducing MeerKAT data on PLEIADI infrastructure**
  - **Objective: to re-process Galactic plane survey data (and possibly others) to fulfil scientific goals**
  - Adapting the general-purpose pipeline developed by IDIA for the Ilifu infrastructure
  - The pipeline was tuned to meet PLEIADI specifications (#CPU, available RAM per node)
  - CASA is provided by a singularity container and the MPI-CASA wrapper
  - Tests on a real MeerKAT data set are ongoing

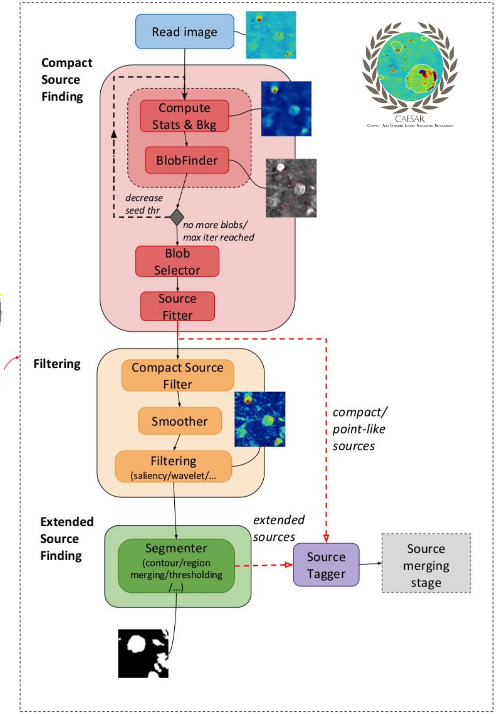
- **Exploring Multi-Node Parallel and Accelerated FFTs Implementations for Imaging Pipelines**
  - **Objective: Speeding-up radio imaging pipeline using parallel & accelerated FFT implementations (multi-node + GPU)**
  - Fast Fourier Transforms affecting the performance of the imaging phase
  - Analyse potential performance improvements with FFTW library (multi-node) and new NVIDIA implementations for multi-GPU
  - Future tests on a real LOFAR imaging pipeline proposed by Gheller C. et al.

Radio source finding

CAESAR: Compact And Extended Source Automated Recognition

- **Objective**: **automated source detection for Galactic plane radio surveys on HPC systems**
- Providing specialized algorithms for compact and extended radio source detection
- Developed and used within **ASKAP EMU and MeerKAT GPS surveys** for catalogue/dataset production
- Parallel implementation (MPI+OpenMP) for large-map processing, tested on OACT HPC infra & GARR OpenStack cloud
- REST-API service deployed in EOSC prototype and integrated with ViaLactea visualization (VLVA) client
- Further activities proposed for the *PNRR Computing National Centre*





→ See Poster S. Riggi

Relativistic jet-ISM modelling

- **A Python *glue* interface between spatially separated codes**
  (V. A.-D., Cielo, Janiuk, 2020-2022)

Motivation: There exist many *monolithic* CFD codes to model accretion and relativistic jets. However, even those exploiting sophisticated *Adaptive Mesh Refinement* (AMR) parallel domain decomposition are *in practice* not suited to deal with the large data structures required to model the formation and propagation of Supermassive Black Holes relativistic jets, even on the largest available HPC systems (memory req. blow up).

Glue **together 2+ different codes**

- FLASH: (Special) relativistic AMR MFD code, customised to deal with jet-cold gas interactions on galactic scales;
- HARMPI: (General) Relativistic AMR MHD code, for SMBH's accretion/jet production.

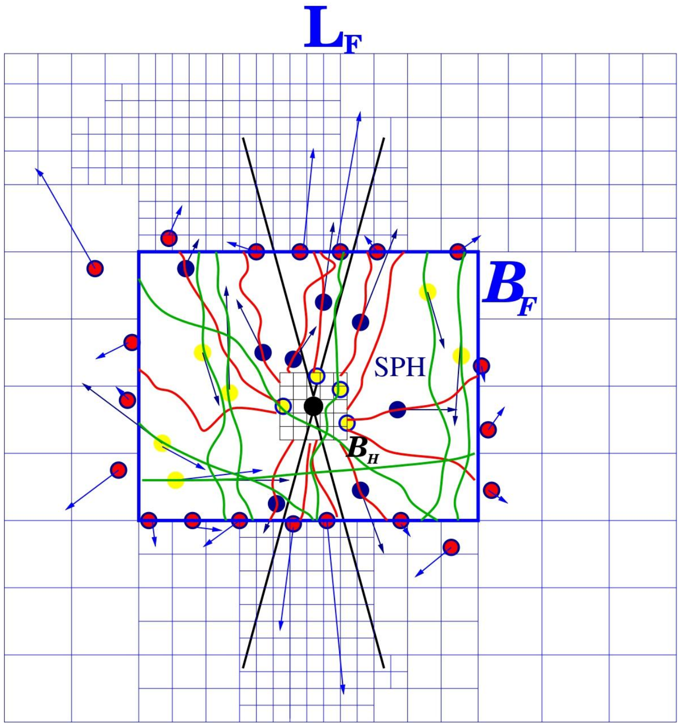
Relativistic jet-ISM modelling

General scheme: The ***space-time*** domain of the SMBH accretion region (dealt with by ***HARMPI***) is *embedded* in that of its host galaxy (dealt with by ***FLASH***)

- A Python SPH code (***PySPH***) takes care of the propagation of the information in the hollow region external to the *HARMPI* box and internal to the *FLASH* boundary B_F (*glue*).

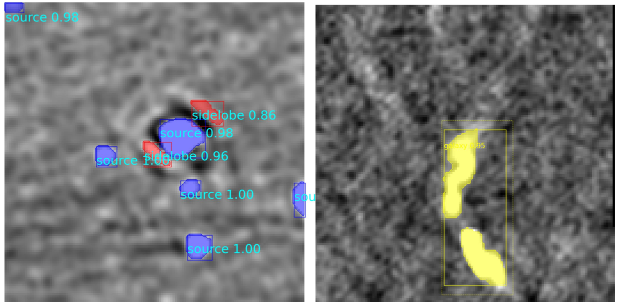
- *Synchronization* is realised using the same SPH particles used to propagate the fields

- *Resolution* fixed at the start → variable amount of SPH particles in the intermediate region.



MACHINE LEARNING APPLICATIONS

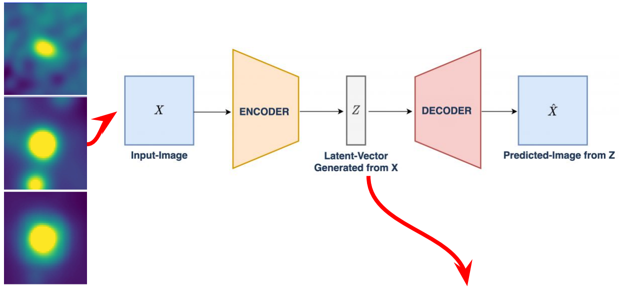
ML FOR RADIO DATA ANALYSIS



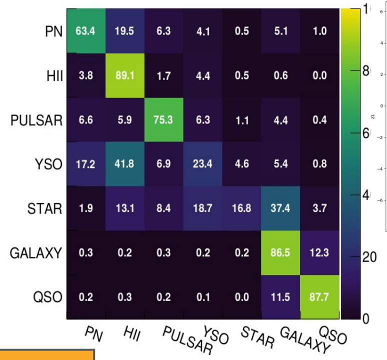
Exploiting deep learning methodologies for radio source analysis

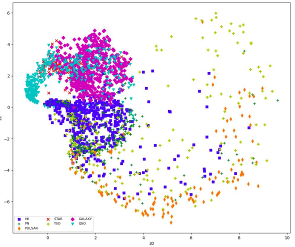
→ Activities proposed for the "PNRR Computing National Centre"

- **Radio source and artefact detection with object detection frameworks**
  - Models (*Mask R-CNN*, *Detectron2*, *DETR*) trained on different GPUs (K40, RTX 6000, CINECA DGX A100, V100)
  - Dataset includes ~20k images from different surveys (FIRST, ATCA, ASKAP, MeerKAT GPS)
  - Mask R-CNN detection step parallelized with MPI for large-map processing



- **Radio compact source classification and anomaly detection**
  - Pipeline using different ML methods (autoencoders, LightGBM, iForest)
  - Dataset including ~15k images from radio (FIRST, ASKAP/MeerKAT GPS) and IR (WISE, Hi-GAL) surveys
  - Parallel processing with MPI for large catalogues

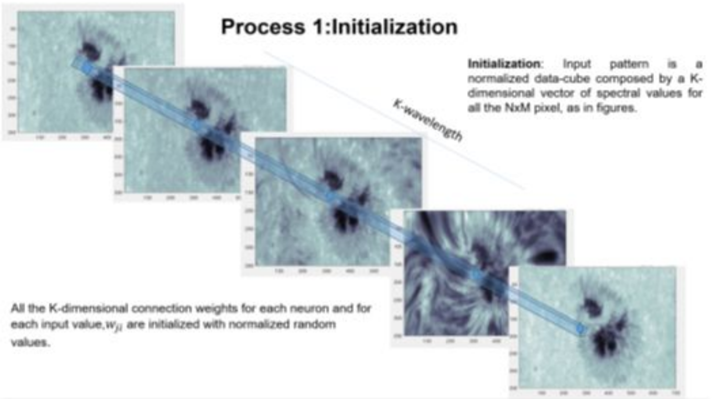




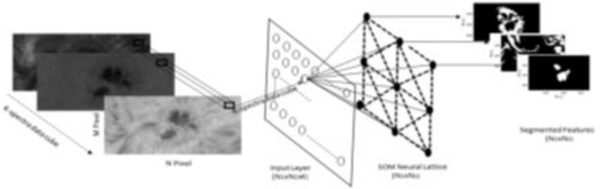
- **Other ML-based tools under development**
  - SNR unsupervised classification (autoencoders+UMAP+density-based clustering)
  - Radio source morphological classification (autoencoders, SOM)
  - Filament detection with deep learning networks

→ See Poster S. Riggi

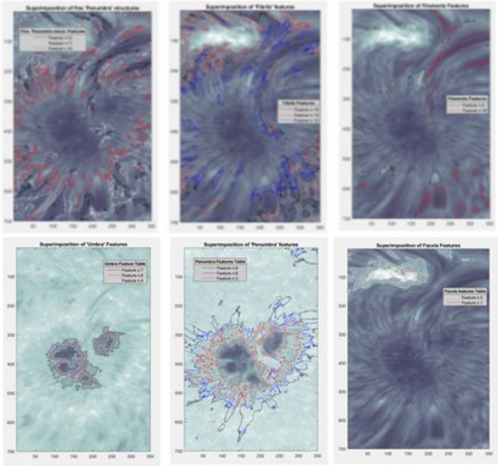
ML for SOLAR DATA ANALYSIS

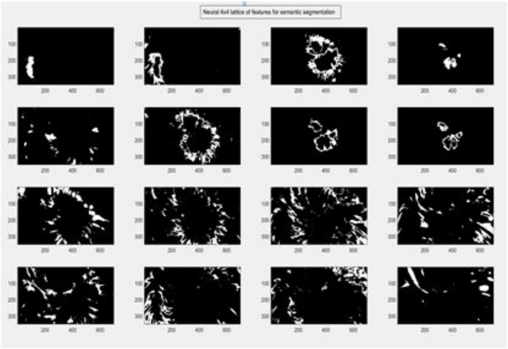


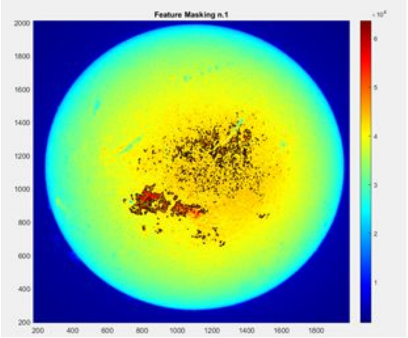
The Segmentation Algorithm

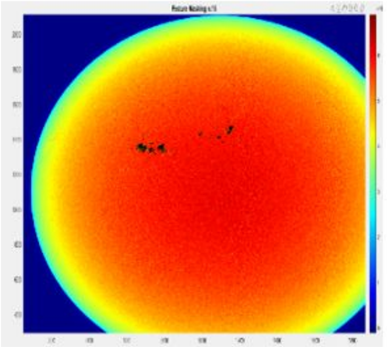


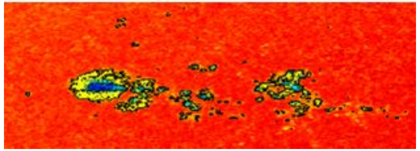
- Feature produced are complementary and mutually exclusive;
- Computationally very fast;
- Parallelizable and portable on different platform;
- Very good candidate as ‘near real time’ monitoring and processing tools





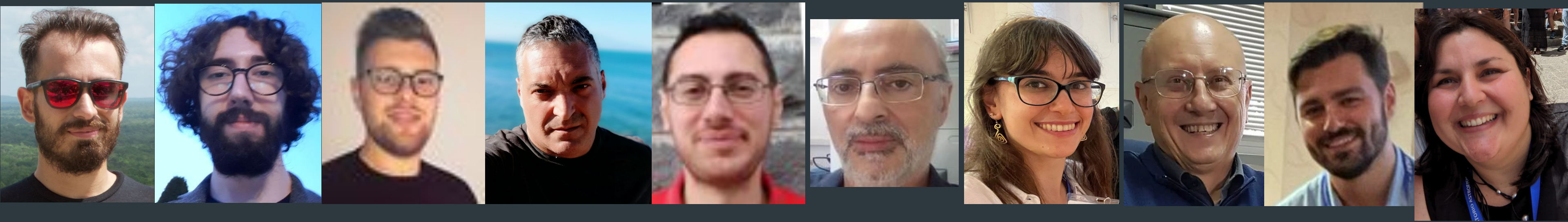






THANKS TO ALL CONTRIBUTORS

F. Vitello, G. Tudisco, M. Raciti, F. Schillirò, A. Ingallinera, V. Antonuccio, V. Cesare, U. Becciani, C. Bordiu, F. Bufano



Ongoing projects:

MOSAICO - Metodologie Open Source per l'Automazione Industriale e delle procedure di CalcOlo in astrofisica

MeerKAT

LOFAR

ECOGAL

NEANIAS

CIRASA

SKA SQUARE KILOMETRE ARRAY

ASKAP EMU Evolutionary Map of the Universe

gaia