



Finanziato  
dall'Unione europea  
NextGenerationEU



Ministero  
dell'Università  
e della Ricerca



Italiadomani

PIANO NAZIONALE  
DI RIPRESA E RESILIENZA



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

# *Machine learning to detect diffuse sources in radio surveys: Radio U-Net and new approaches*

*Chiara Stuardi  
INAF-IRA*

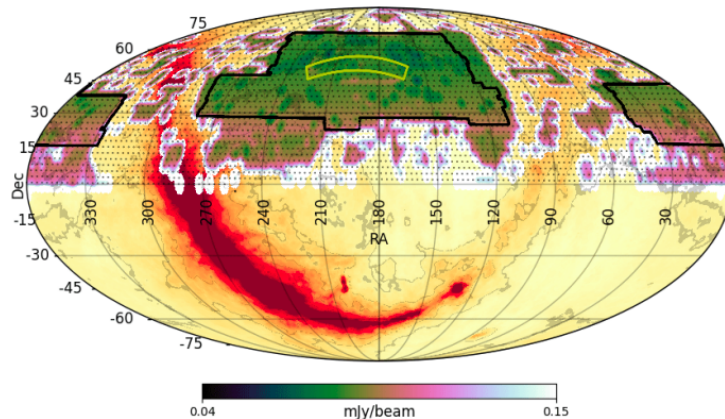
*Claudio Gheller (INAF-IRA), F. Vazza (UniBo), A. Botteon (INAF-IRA), N. Sanvitale (INAF-IRA)*

**Spoke 3 II Technical Workshop, Bologna Dec 17 -19, 2024**

# Scientific Rationale

Current radio surveys are challenging our detection and cataloging strategies

- large data size (PB/year)

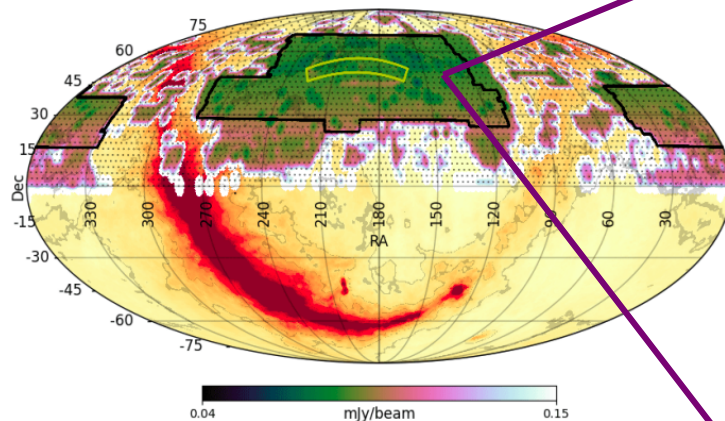


LOFAR Two Metre Sky Survey DR2  
@150 MHz [LOTSS,Shimwell+22]

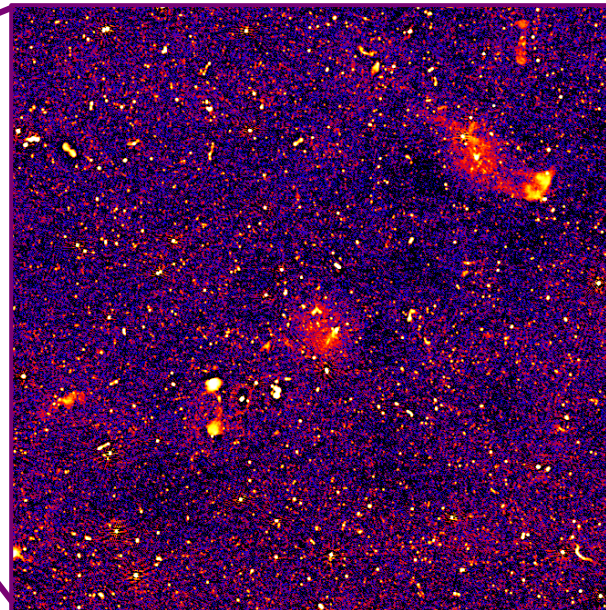
# Scientific Rationale

Current radio surveys are challenging our detection and cataloging strategies

- large data size (PB/year)
- time-consuming and computationally expensive data reduction and imaging procedures
- non-Gaussian noise and imaging artifacts



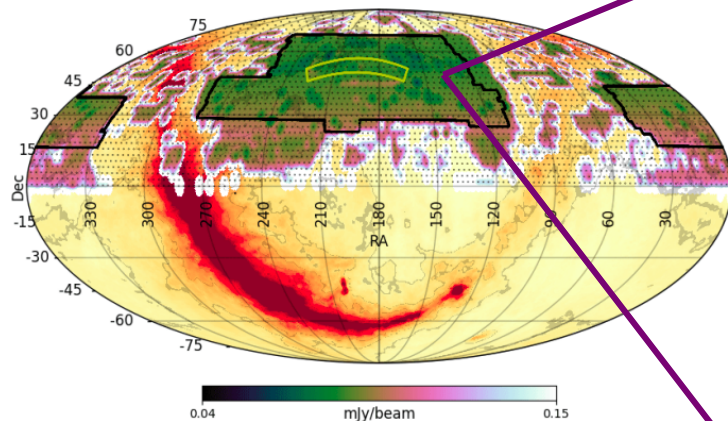
LOFAR Two Metre Sky Survey DR2  
@150 MHz [LOTSS,Shimwell+22]



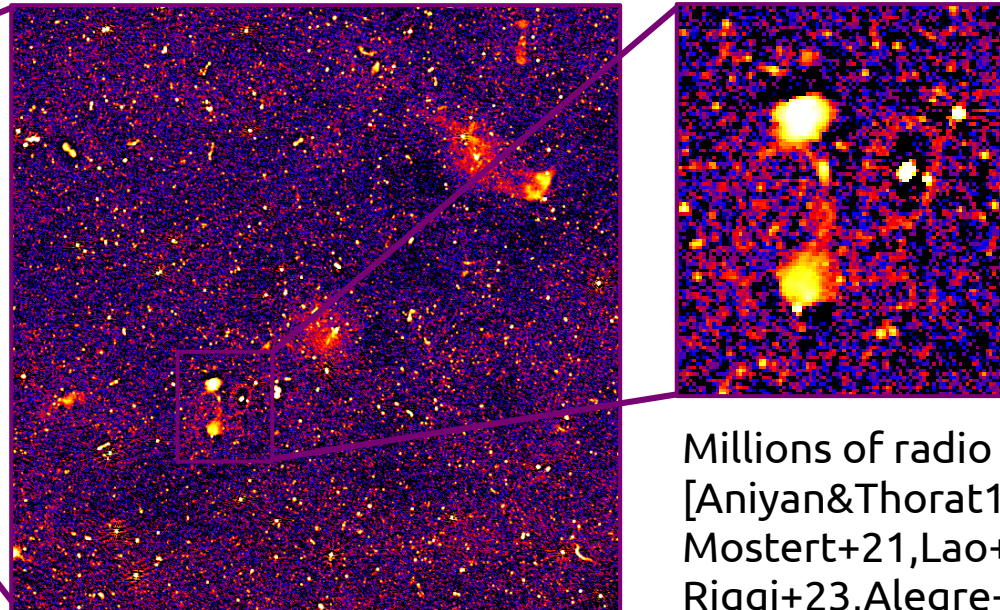
# Scientific Rationale

Current radio surveys are challenging our detection and cataloging strategies

- large data size (PB/year)
- time-consuming and computationally expensive data reduction and imaging procedures
- non-Gaussian noise and imaging artifacts



LOFAR Two Metre Sky Survey DR2  
@150 MHz [LOTSS,Shimwell+22]

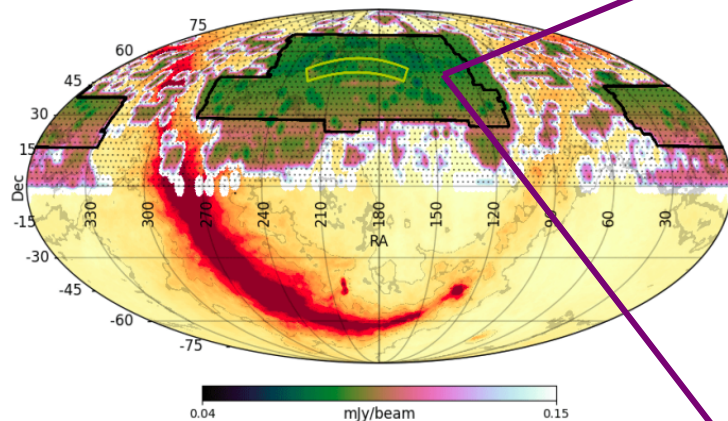


Millions of radio galaxies  
[Aniyan&Thorat17,Lukic+18,  
Mostert+21,Lao+23,  
Riggi+23,Alegre+24,Gupta+24  
Riggi+24,Slijepcevic+24]

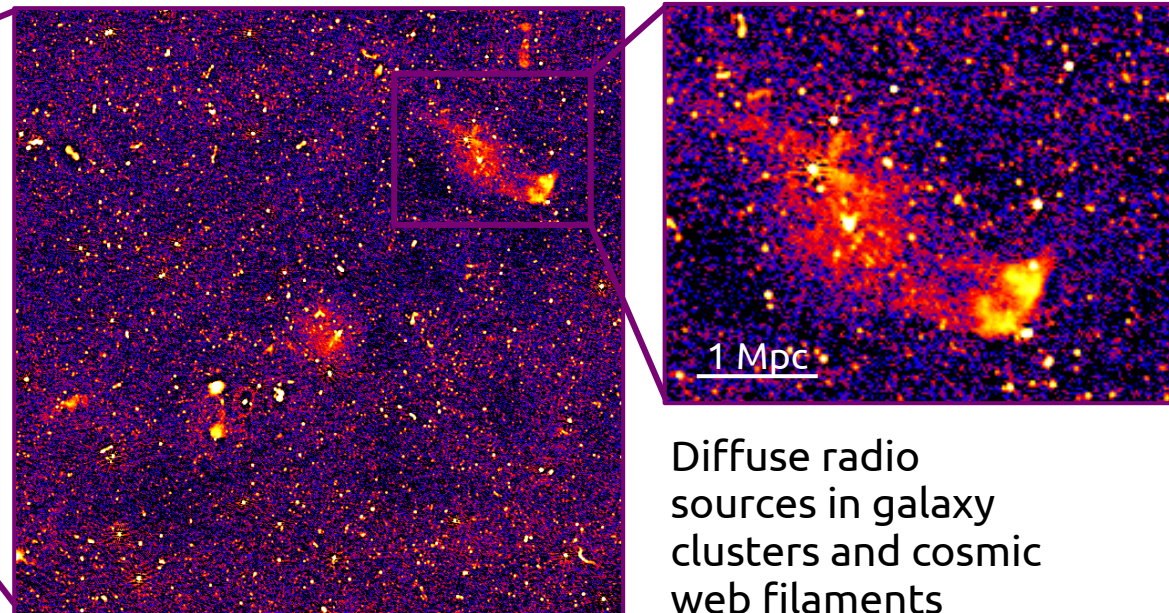
# Scientific Rationale

Current radio surveys are challenging our detection and cataloging strategies

- large data size (PB/year)
- time-consuming and computationally expensive data reduction and imaging procedures
- non-Gaussian noise and imaging artifacts
- rare sources with complex and irregular morphology



LOFAR Two Metre Sky Survey DR2  
@150 MHz [LOTSS, Shimwell+22]

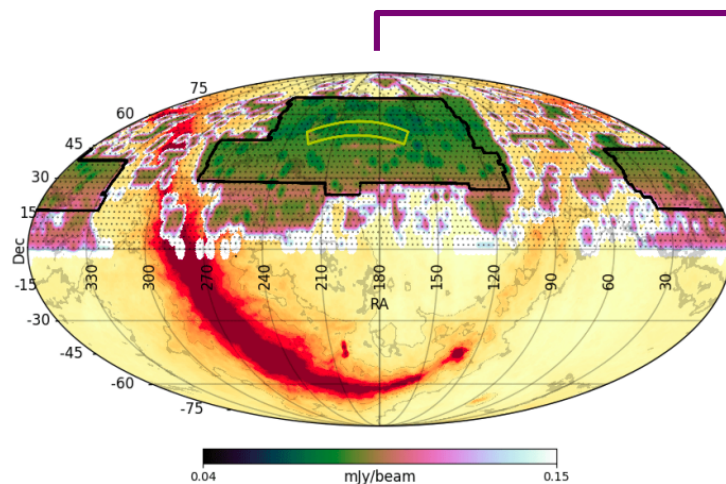


Diffuse radio sources in galaxy clusters and cosmic web filaments

# Scientific Rationale

Current radio surveys are challenging our detection and cataloging strategies

- large data size (PB/year)
  - time-consuming and computationally expensive data reduction and imaging procedures
  - non-Gaussian noise and imaging artifacts
  - rare sources with complex and irregular morphology
- new strategies to minimize human intervention in data processing



LOFAR Two Metre Sky Survey DR2  
@150 MHz [LOTSS,Shimwell+22]

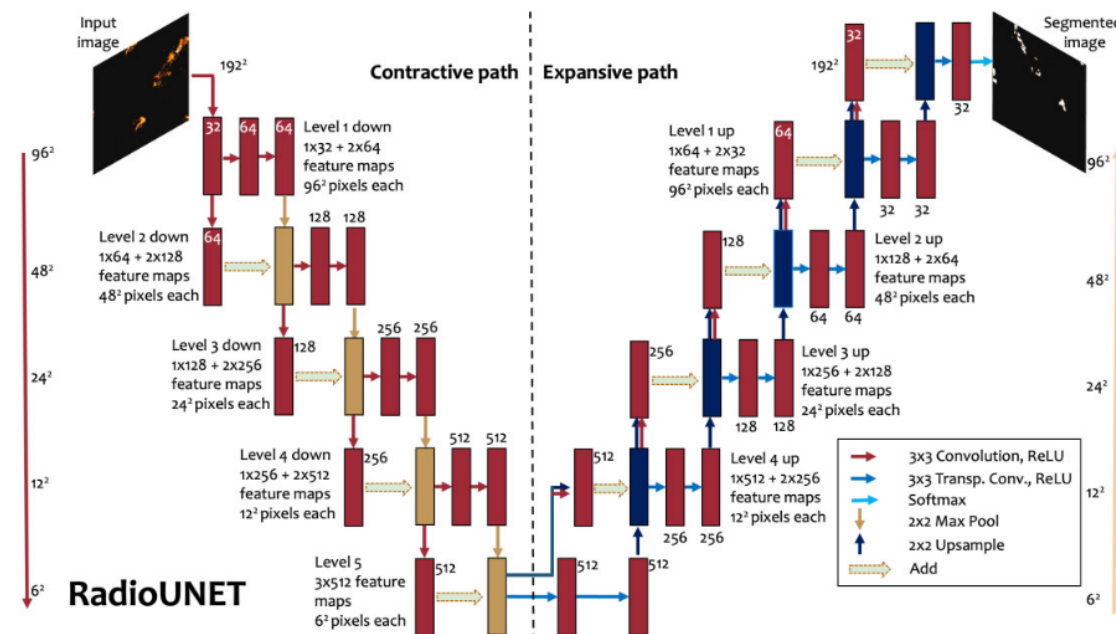
Square Kilometre Array, operational in 2030



# Technical Objectives, Methodologies and Solutions

**Objective:** Development and distribution of a machine learning tool for detection of diffuse emission in radio surveys

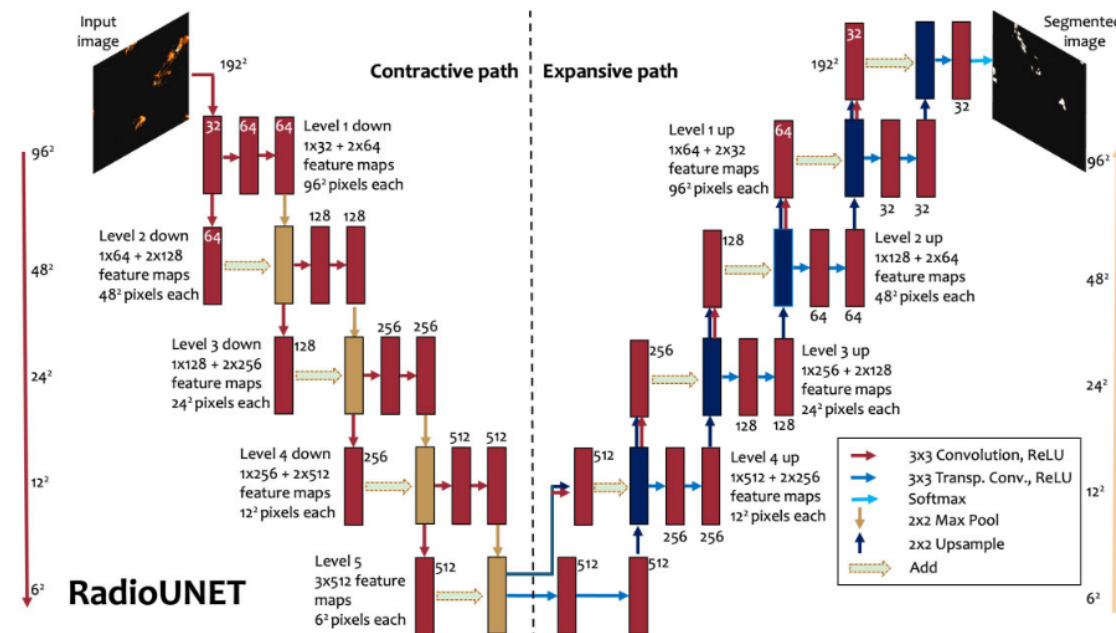
- U-Net architecture adopted perform fast segmentation of large datasets



# Technical Objectives, Methodologies and Solutions

**Objective:** Development and distribution of a machine learning tool for detection of diffuse emission in radio surveys

- U-Net architecture adopted perform fast segmentation of large datasets
- Scalability from CPU to GPU, run on Leonardo @CINECA

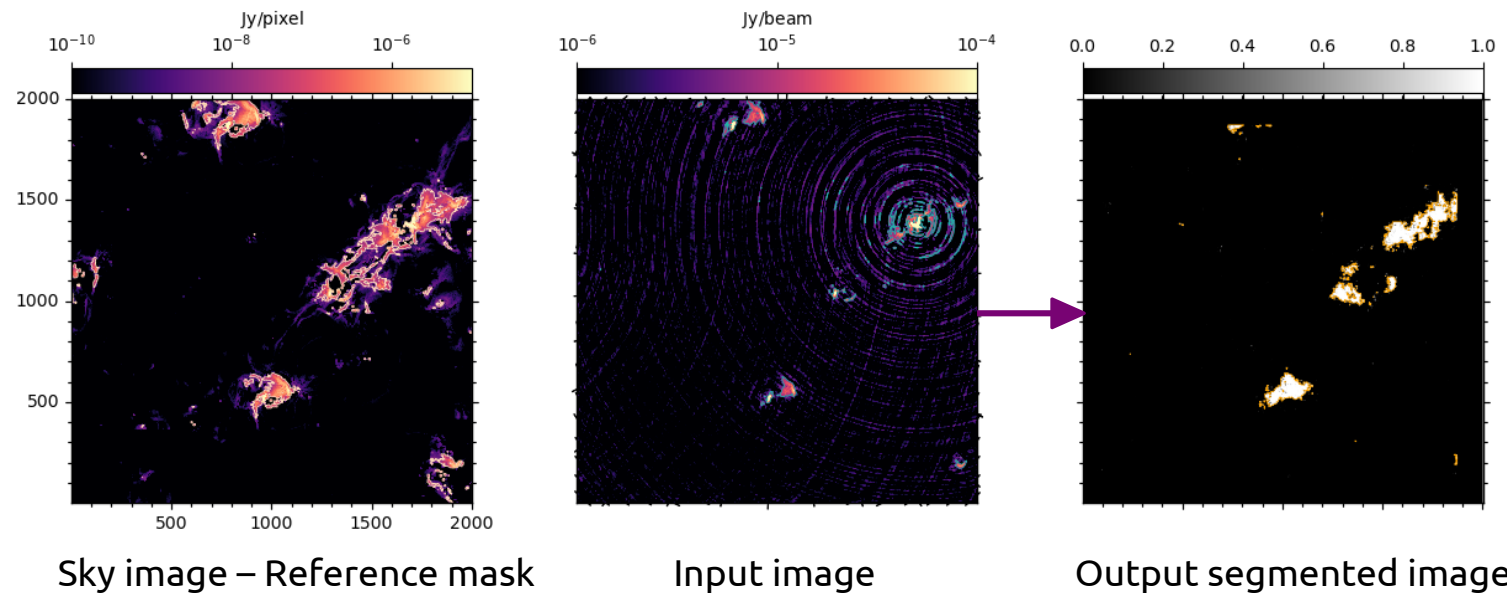




# Technical Objectives, Methodologies and Solutions

**Objective:** Development and distribution of a machine learning tool for detection of diffuse emission in radio surveys

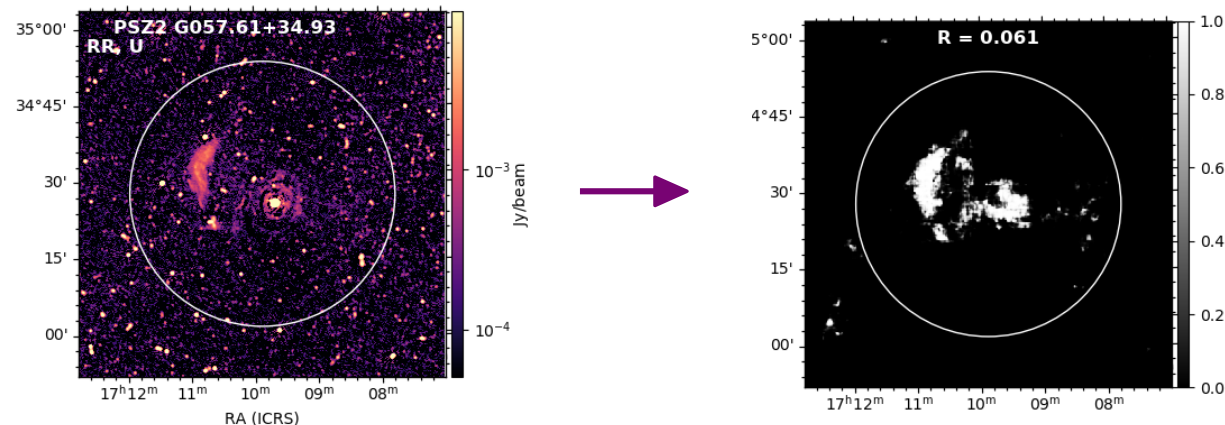
- U-Net architecture adopted perform fast segmentation of large datasets
- Scalability from CPU to GPU, run on Leonardo @CINECA
- Self-supervised learning with training on cosmological simulations [Gheller&Vazza22]



# Technical Objectives, Methodologies and Solutions

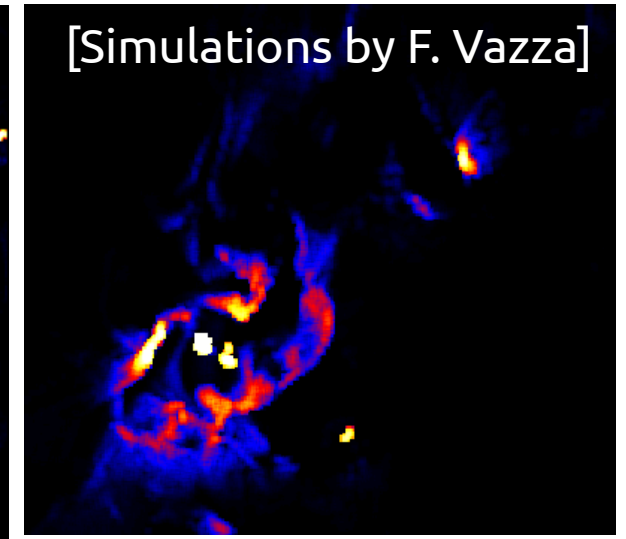
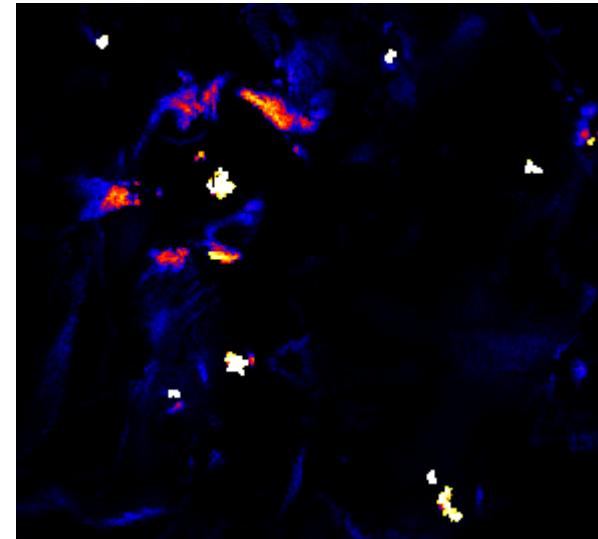
**Objective:** Development and distribution of a machine learning tool for detection of diffuse emission in radio surveys

- U-Net architecture adopted perform fast segmentation of large datasets
- Scalability from CPU to GPU, run on Leonardo @CINECA
- Self-supervised learning with training on cosmological simulations
- Application and performance verification on LOFAR Two Metre Sky Survey data



# Technical Objectives, Methodologies and Solutions

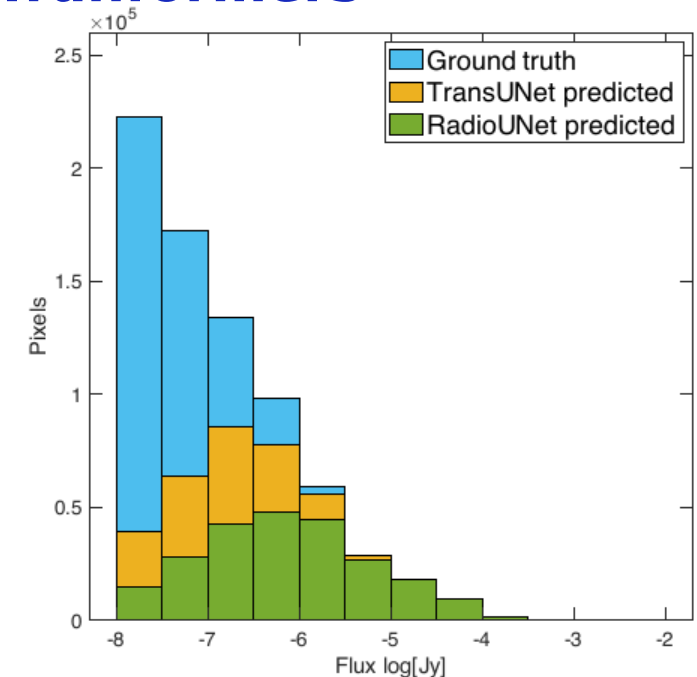
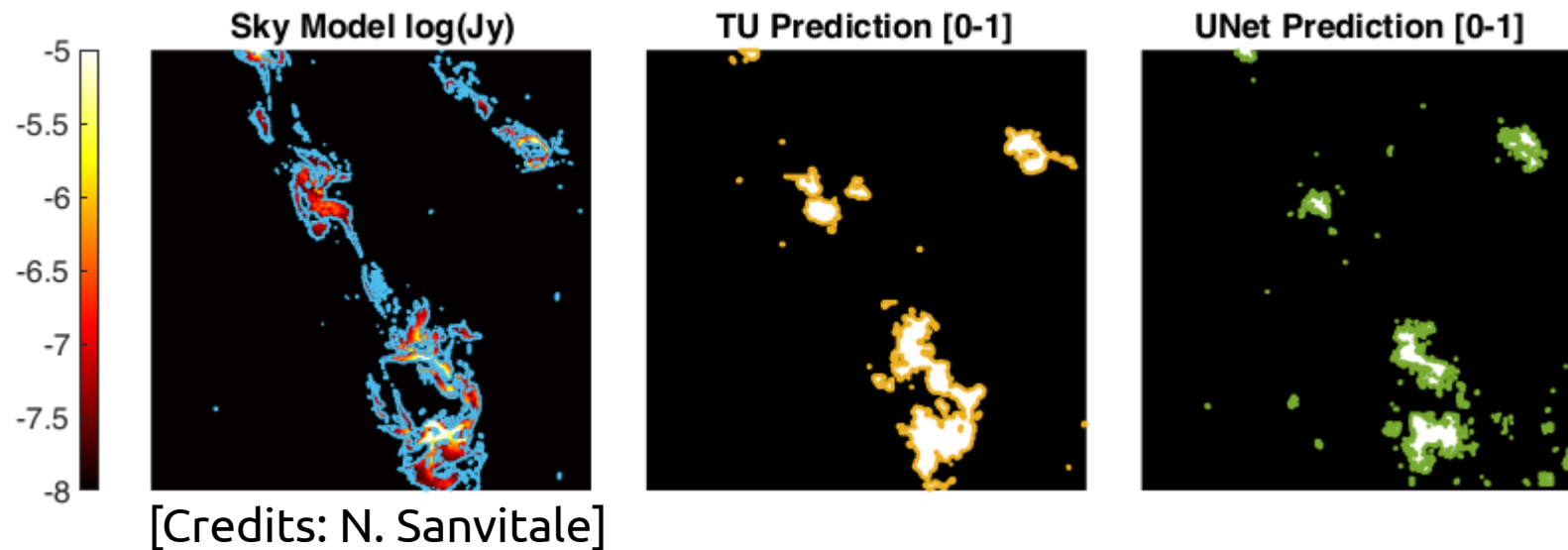
**Criticality: confusion with extended radio galaxies which are not present in the simulation used for training**  
→ possible solution: new simulations



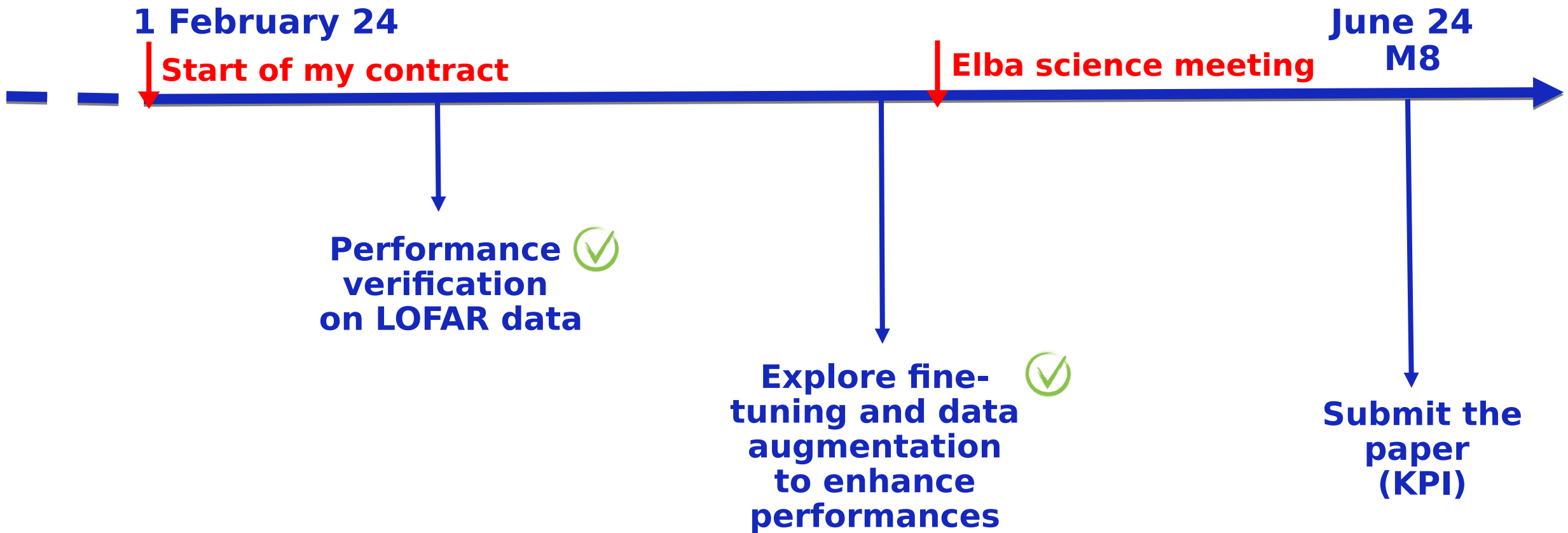
# Technical Objectives, Methodologies and Solutions

**Criticality:** confusion with extended radio galaxies which are not present in the simulation used for training  
→ possible solution: new simulations

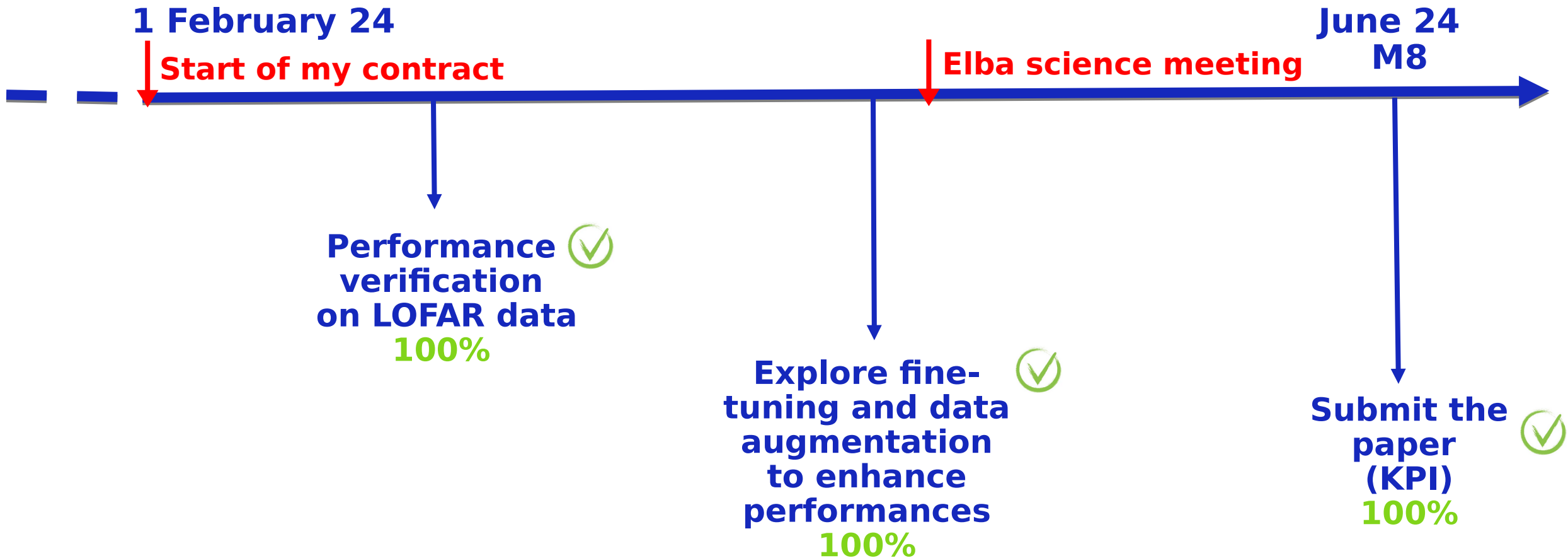
**Addition:** test of a modification of the U-Net including Transformers  
→ better performances but still to be explored on data



# Main Results



# Main Results



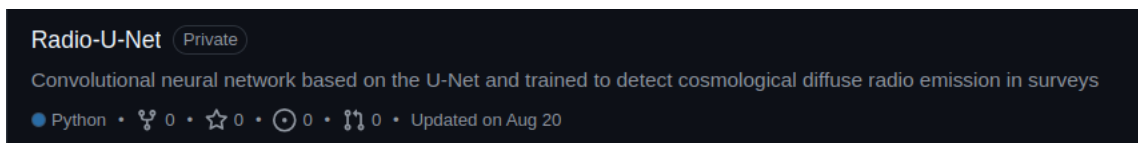
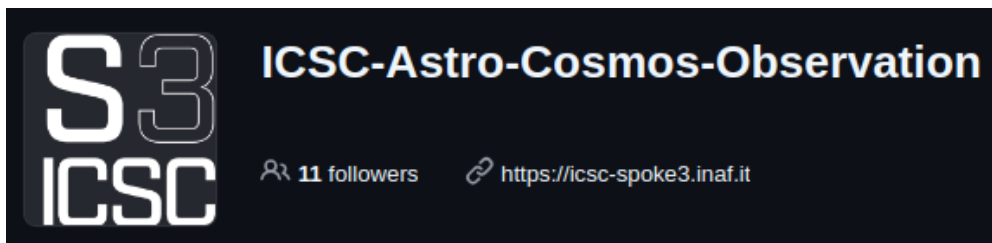
# Main Results

October 24  
M9

March 25

August 25  
M10

Paper published on MNRAS (KPI)  
+ follow-up A&A letter, under review  
Code on GitHub but private



# Main Results

October 24  
M9

Today

March 25

August 25  
M10

Paper published on MNRAS (KPI) ✓  
+ follow-up A&A letter, under review  
Code on GitHub but private 80%



# Final Steps

October 24  
M9

Today

March 25

August 25  
M10

Paper published on MNRAS (KPI) ✓  
+ follow-up A&A letter, under review  
Code on GitHub but private 80%

Creation of a new simulation set,  
training and test  
(intermediate KPI)

# Final Steps

October 24  
M9

Today

March 25

August 25  
M10

Paper published on MNRAS (KPI) ✓  
+ follow-up A&A letter, under review  
Code on GitHub but private 80%

Creation of a new simulation set, 50%  
training and test 20%  
(intermediate KPI)

# Final Steps

October 24  
M9

Today

March 25

August 25  
M10

Paper published on MNRAS (KPI) ✓  
+ follow-up A&A letter, under review  
Code on GitHub but private 80%

Creation of a new simulation set, 50%  
training and test 20%  
(intermediate KPI)

Application of the best network  
to the LoTSS DR3 and submit paper  
(KPI)



# Final Steps

October 24  
M9

Today

March 25

August 25  
M10

Paper published on MNRAS (KPI) ✓  
+ follow-up A&A letter, under review  
Code on GitHub but private 80%

Creation of a new simulation set, 50%  
training and test 20%  
(intermediate KPI)

Application of the best network  
to the LoTSS DR3 and submit paper  
(KPI) 10%



# Final Steps

**December 25  
M11(?)**



**Make the code publicly available.**

**TOTAL: 55%**

# Final Steps

December 25  
M11(?)



Make the code publicly available.

**TOTAL: 55%**

Thank you  
for your attention!