Automated identification of diffuse radio emission in all-sky surveys with Radio U-Net

USC8 AI Workshop, 21th May 2025

Chiara Stuardi
INAF – IRA Istituto di Radioastronomia (Italy)

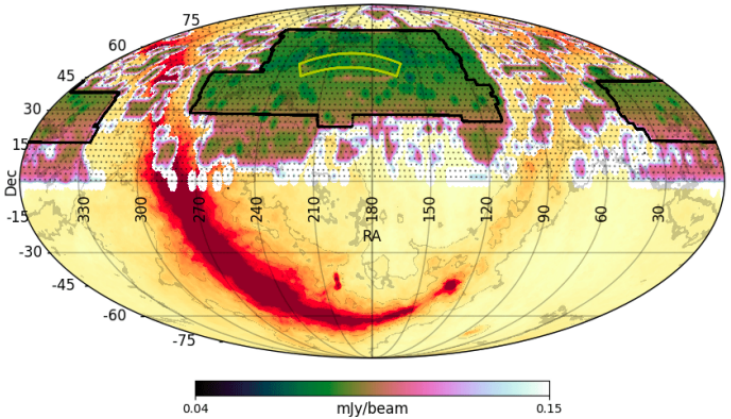
Collaborators: C. Gheller (INAF-IRA), N. Sanvitale (INAF-IRA), G. Di Gennaro (INAF-IRA), F. Braga (U. of Bologna), A. Botteon (INAF-IRA), F. Vazza (U. of Bologna)

SPOKE 3
Astrophysics & Cosmos Observations
ICSC Italian Research Center on High-Performance Computing, Big Data and Quantum Computing

Scientific rationale

Current radio surveys are challenging our detection and cataloging strategies

- large data size (PB/year)

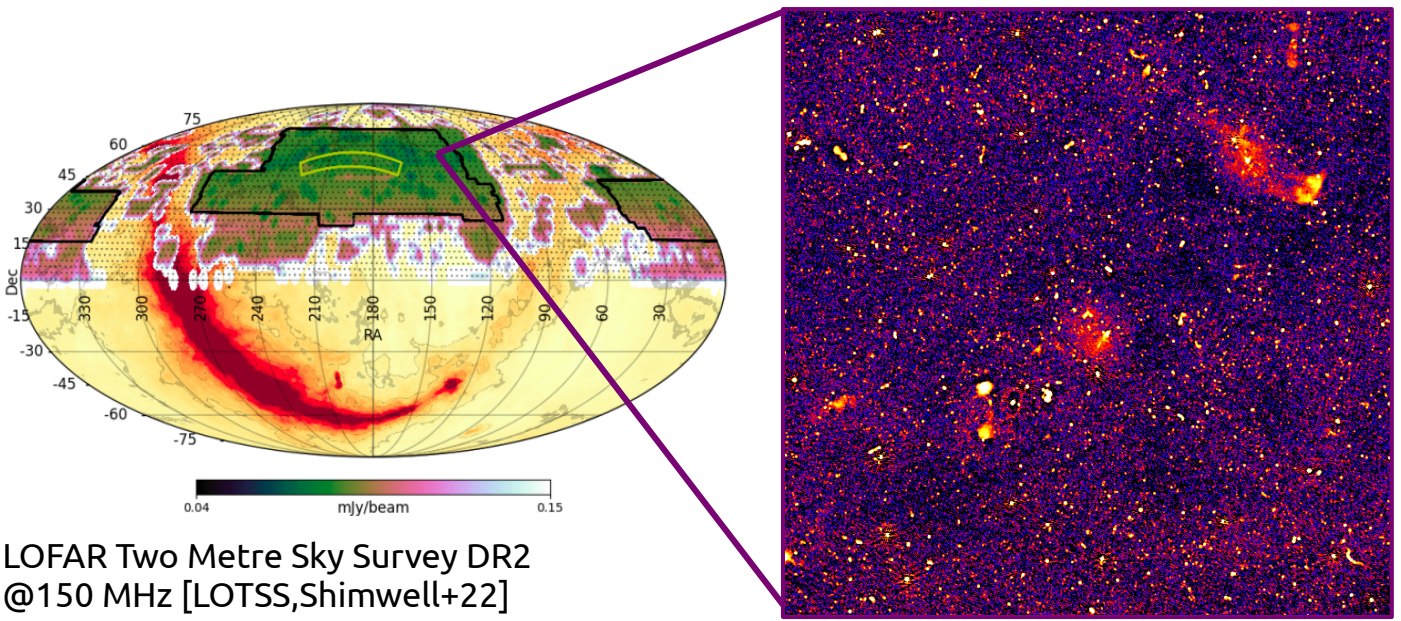


LOFAR Two Metre Sky Survey DR2
@150 MHz [LOTSS,Shimwell+22]

Scientific rationale

Current radio surveys are challenging our detection and cataloging strategies

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

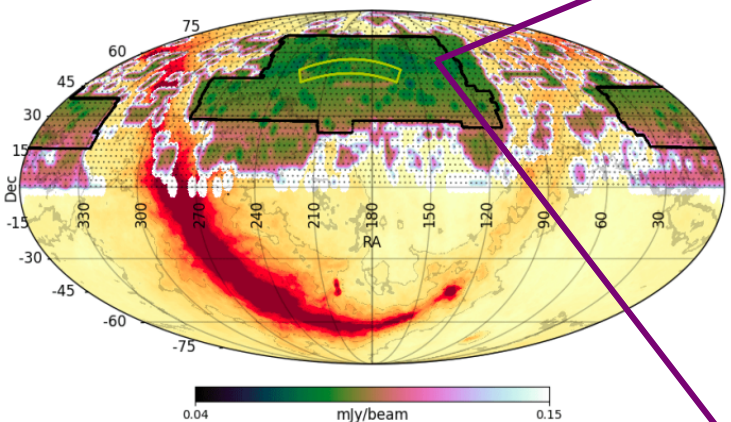


LOFAR Two Metre Sky Survey DR2
@150 MHz [LOTSS,Shimwell+22]

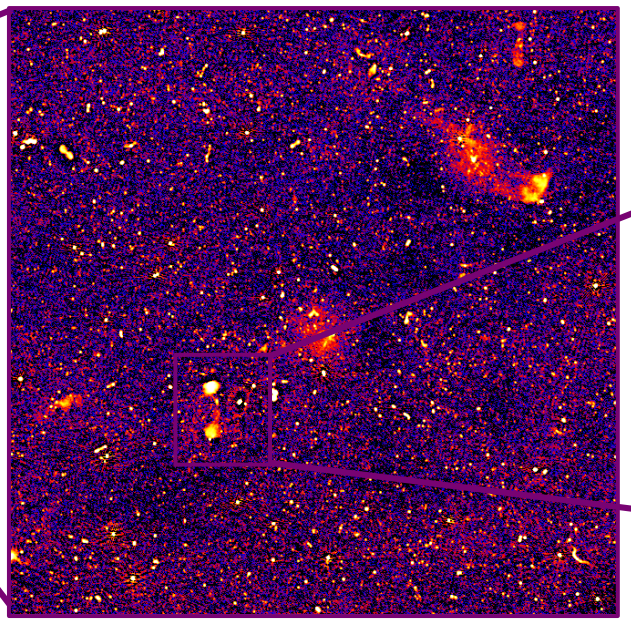
Scientific rationale

Current radio surveys are challenging our detection and cataloging strategies

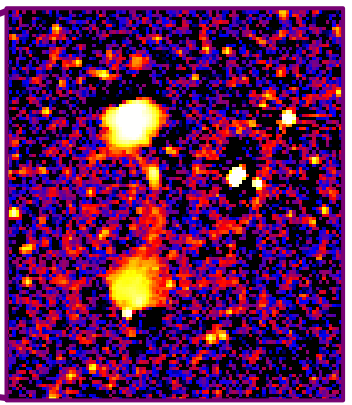
- large data size (PB/year)
- time-consuming and computationally expensive data reduction procedures
- non-Gaussian noise and imaging artifacts
- sources with complex and irregular morphology



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]

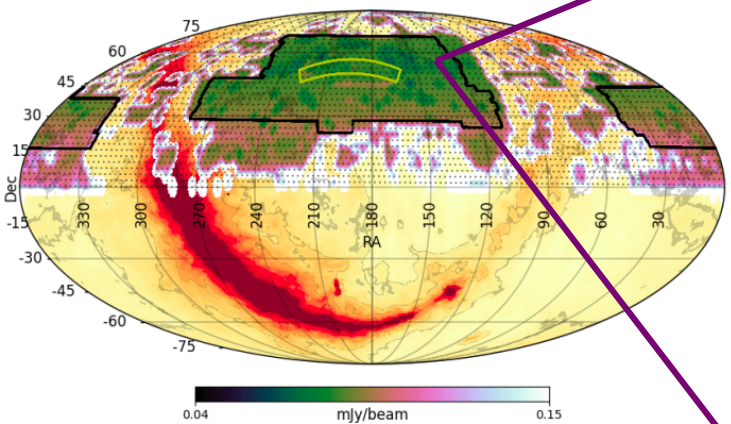


Scientific rationale

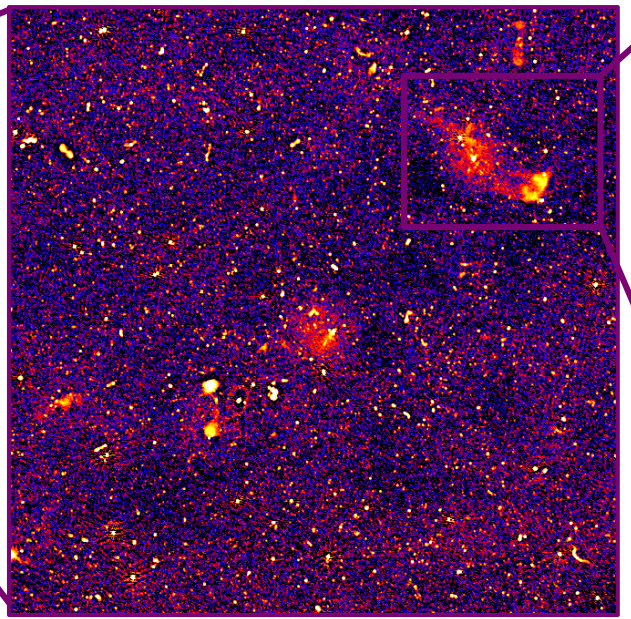
SPOKE 3 Astrophysics & Cosmos Observations

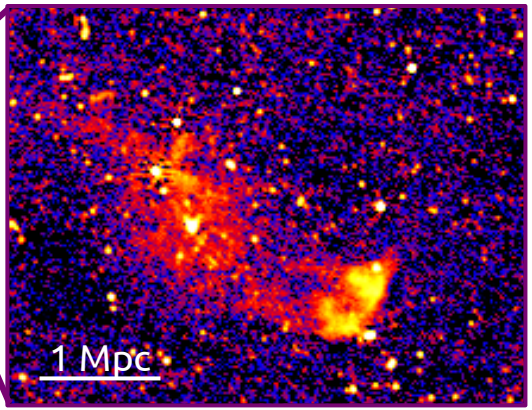
Current radio surveys are challenging our detection and cataloging strategies

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



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





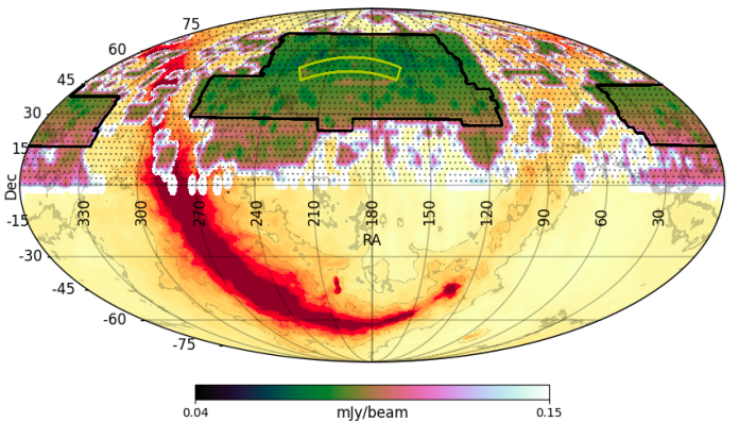
Diffuse radio sources in
galaxy clusters and
cosmic web filaments

Scientific rationale

Current radio surveys are challenging our detection and cataloging strategies

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

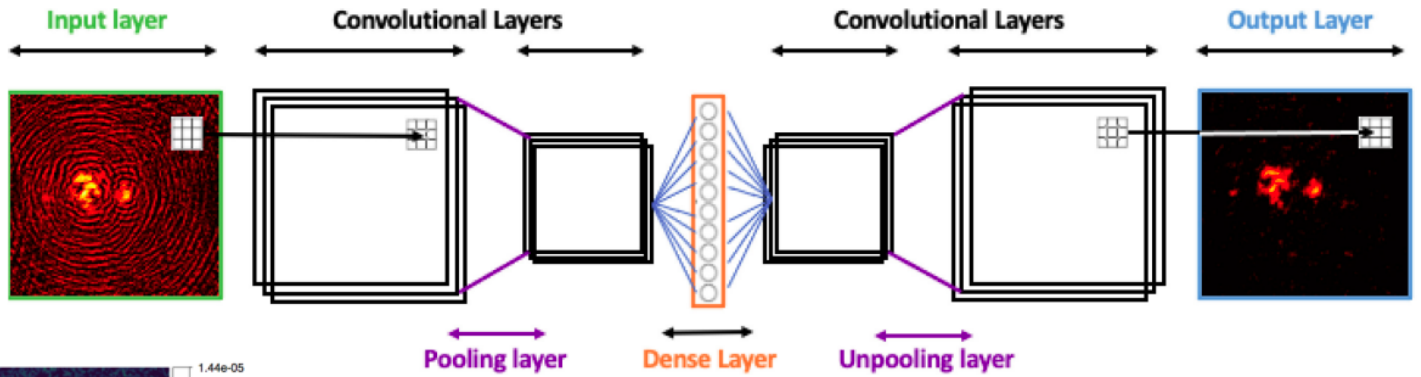
→ new strategies to minimize human intervention in data processing



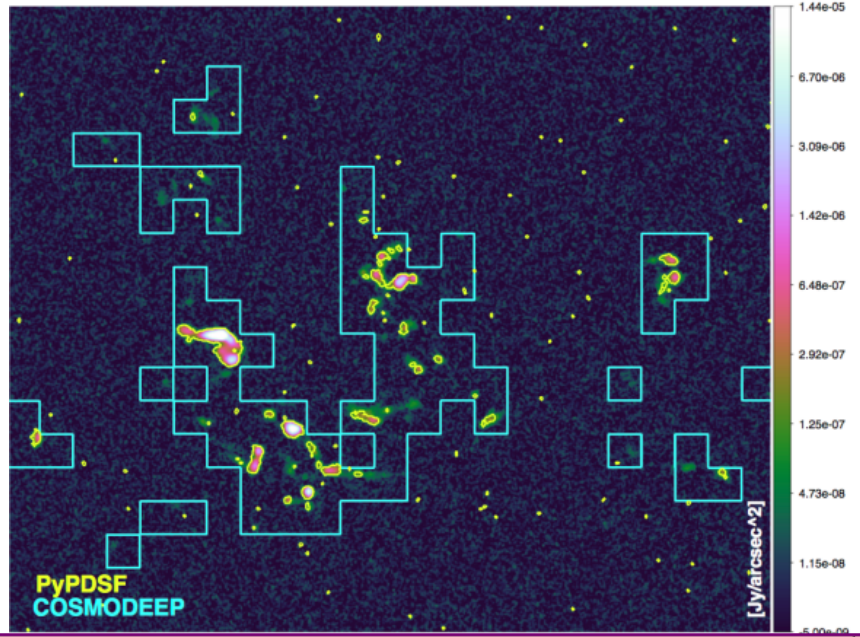
LOFAR Two Metre Sky Survey DR2
@150 MHz [LOTSS,Shimwell+22]

Square Kilometre Array, operational in 2030

Previous works

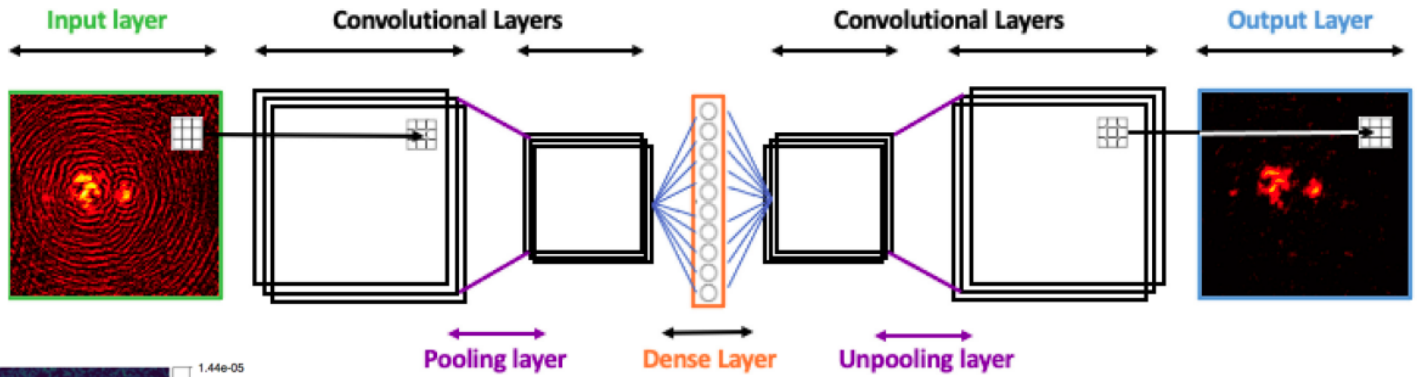


COSMODEEP
[Gheller,Vazza,Bonafede18]

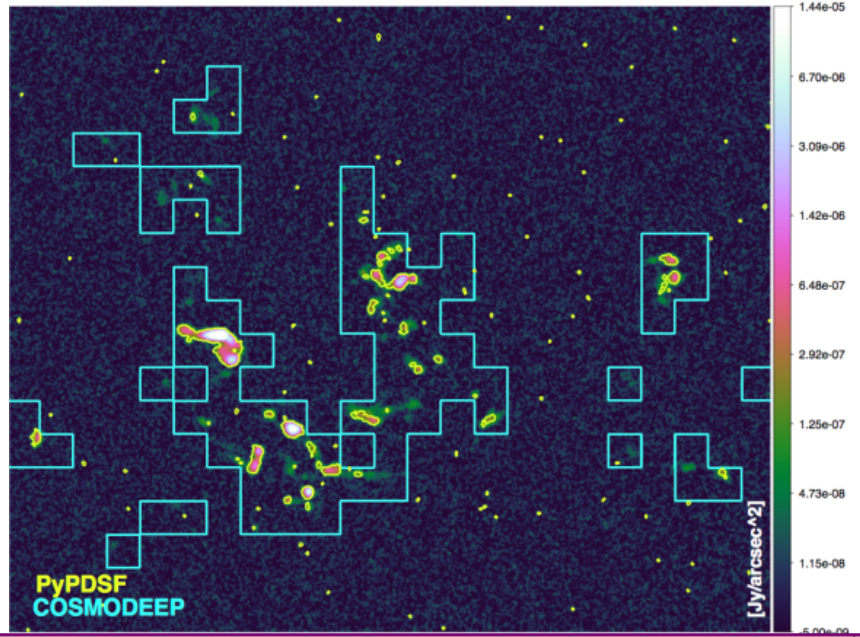


De-noising autoencoder
[Gheller&Vazza+22]

Previous works



COSMODEEP
[Gheller,Vazza,Bonafede18]



De-noising autoencoder
[Gheller&Vazza+22]

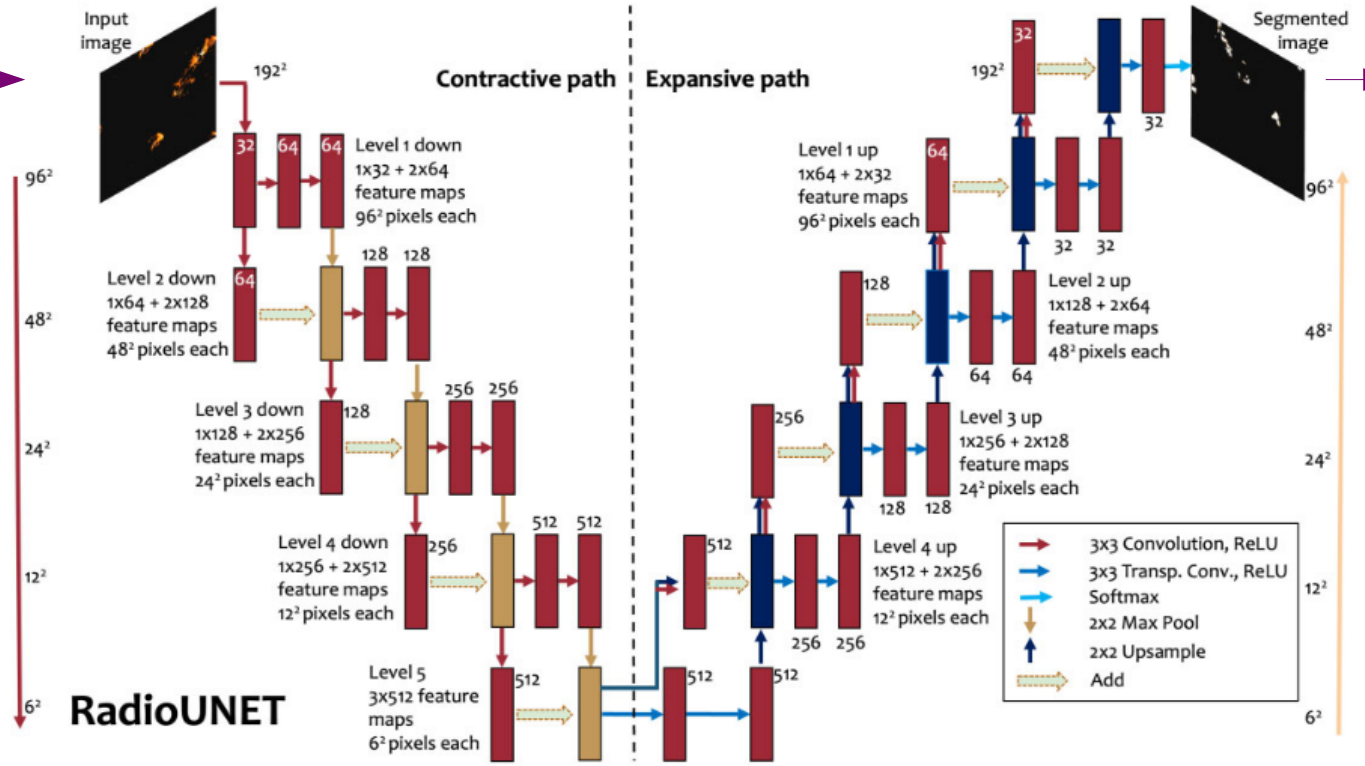
Never tested on real data!

Radio U-Net: architecture and training strategy

SPOKE 3 Astrophysics & Cosmos Observations

Convolutional neural network based on the U-net architecture [Ronneberger+15]
to perform the segmentation of diffuse radio emission in radio astronomical surveys

Tiling



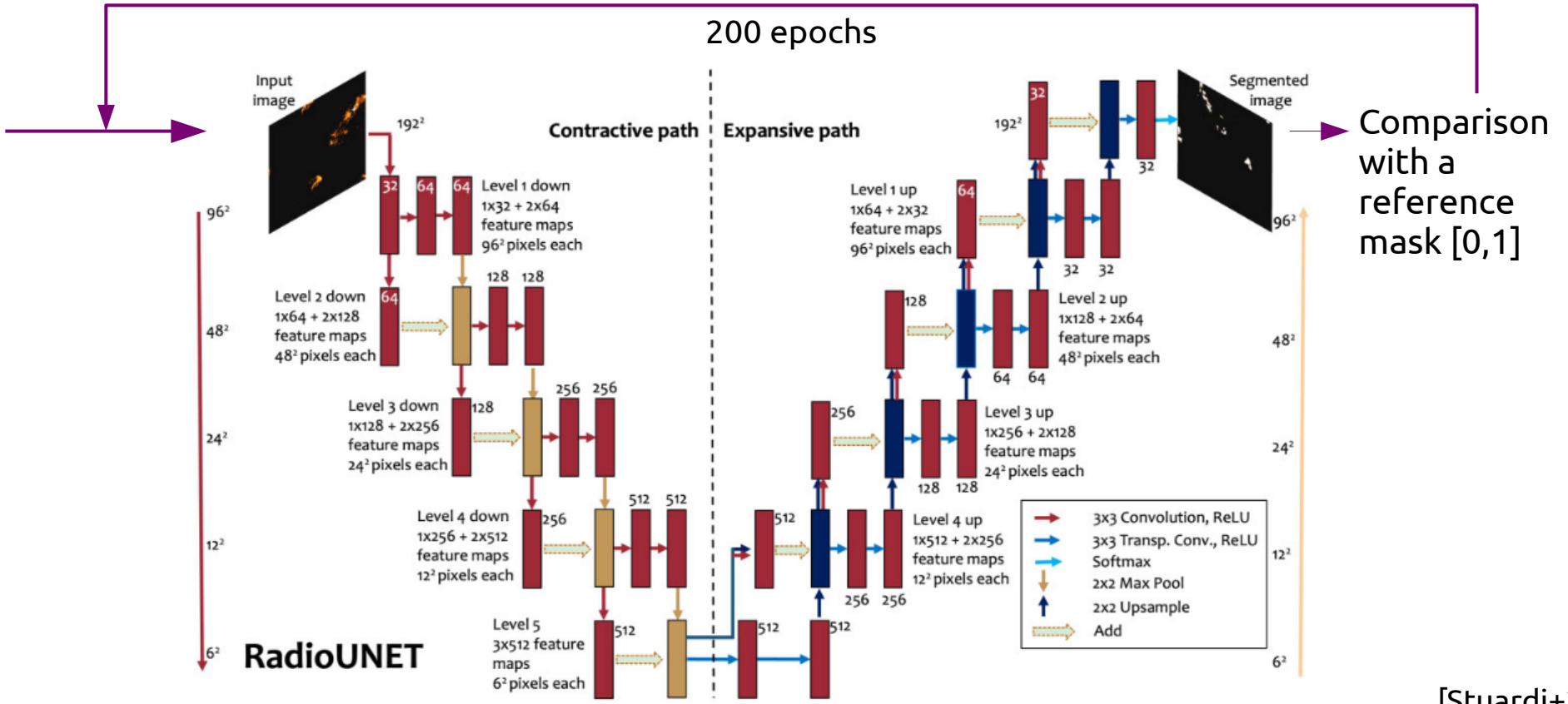
Comparison with a reference mask [0,1]

[Stuardi+24]

Radio U-Net: architecture and training strategy

SPOKE 3 Astrophysics & Cosmos Observations

Convolutional neural network based on the U-net architecture [Ronneberger+15]
to perform the segmentation of diffuse radio emission in radio astronomical surveys

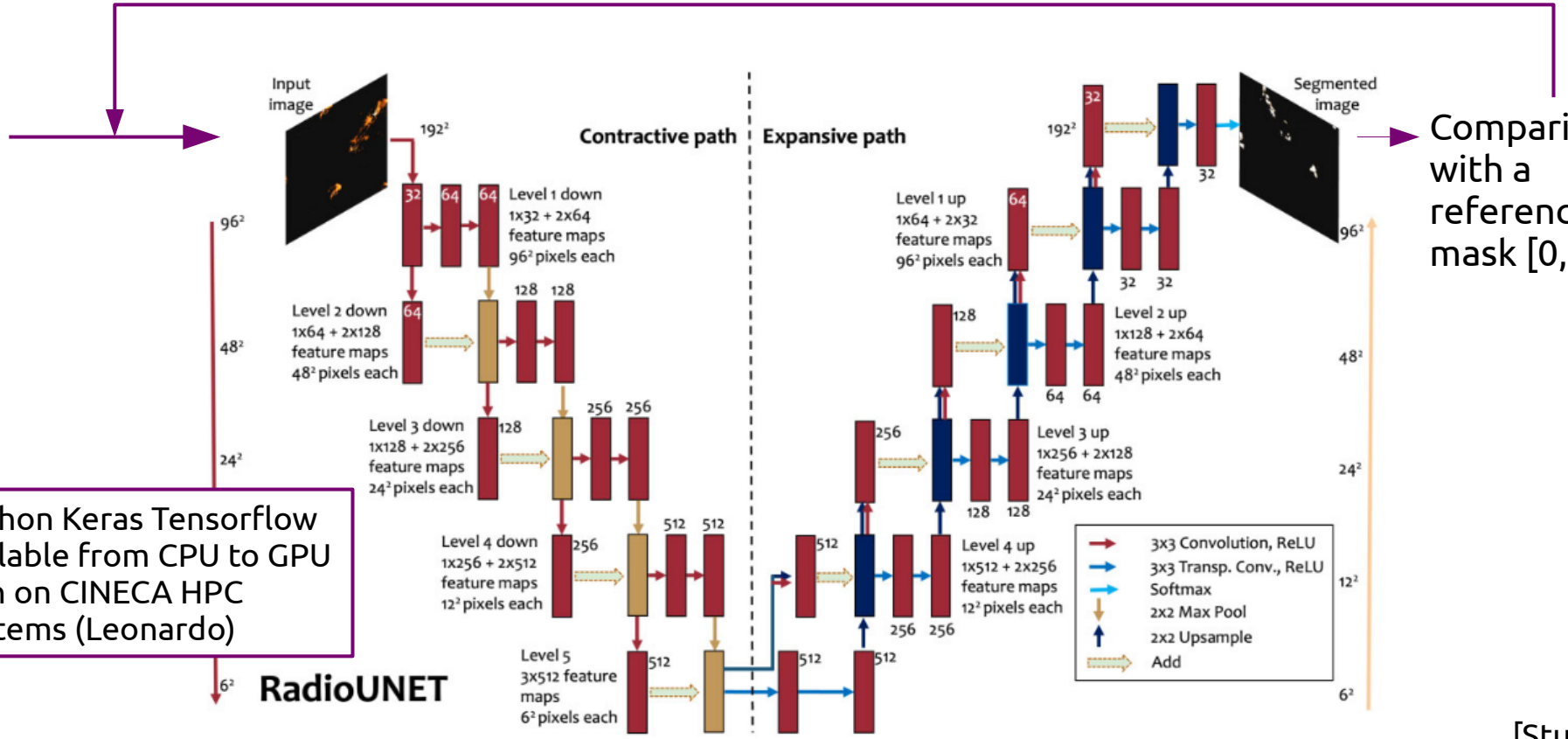


[Stuardi+24]

Radio U-Net: architecture and training strategy

SPOKE 3 Astrophysics & Cosmos Observations

Convolutional neural network based on the U-net architecture [Ronneberger+15]
to perform the segmentation of diffuse radio emission in radio astronomical surveys



- Python Keras Tensorflow
- Scalable from CPU to GPU
- Run on CINECA HPC systems (Leonardo)

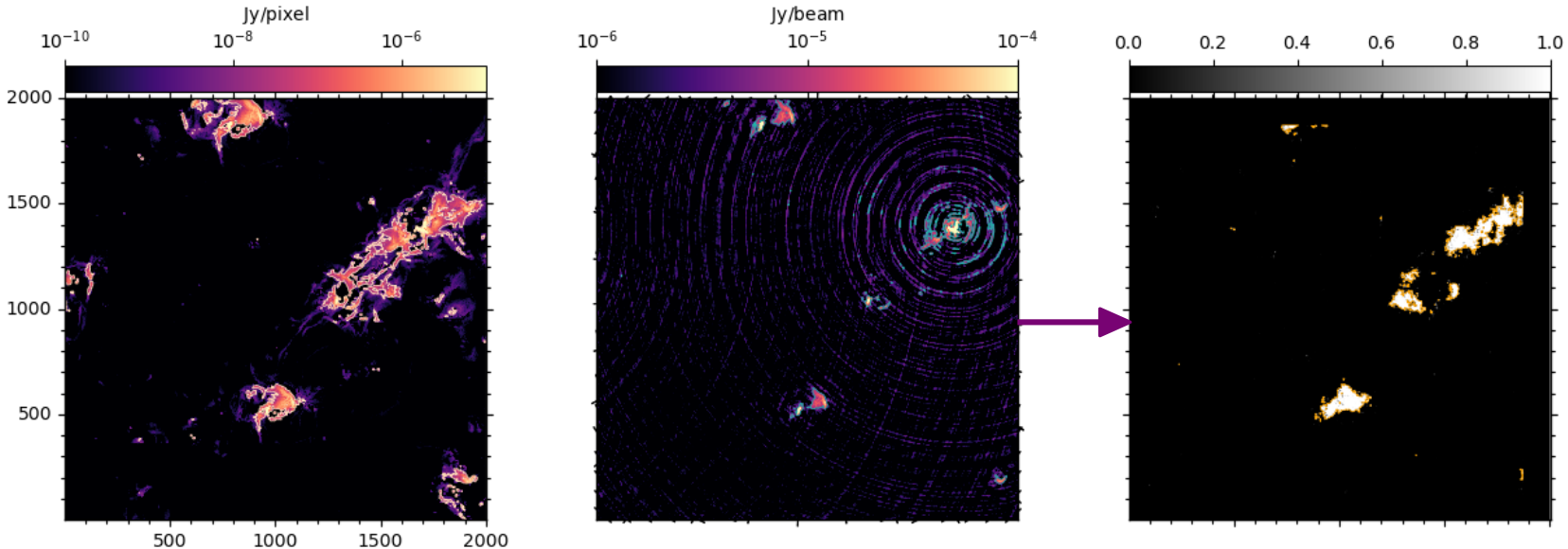
[Stuardi+24]

Radio U-Net: architecture and training strategy

SPOKE 3 Astrophysics & Cosmos Observations

Convolutional neural network based on the U-net architecture [Ronneberger+15]
to perform the segmentation of diffuse radio emission in radio astronomical surveys

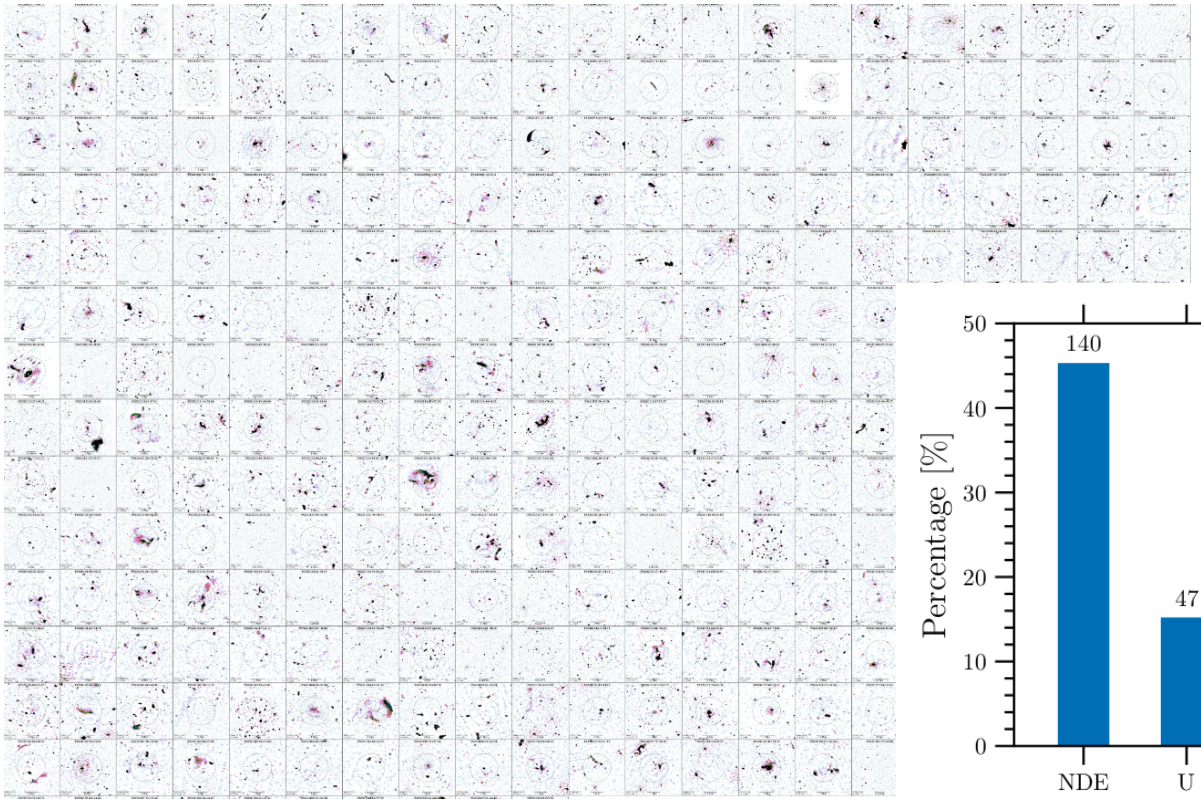
Training on synthetic LoTSS observations built on cosmological simulations [Gheller&Vazza22]
- residual imaging artifacts
- logarithmic normalization
- reference mask at 10^{-3}xnoise of the input image

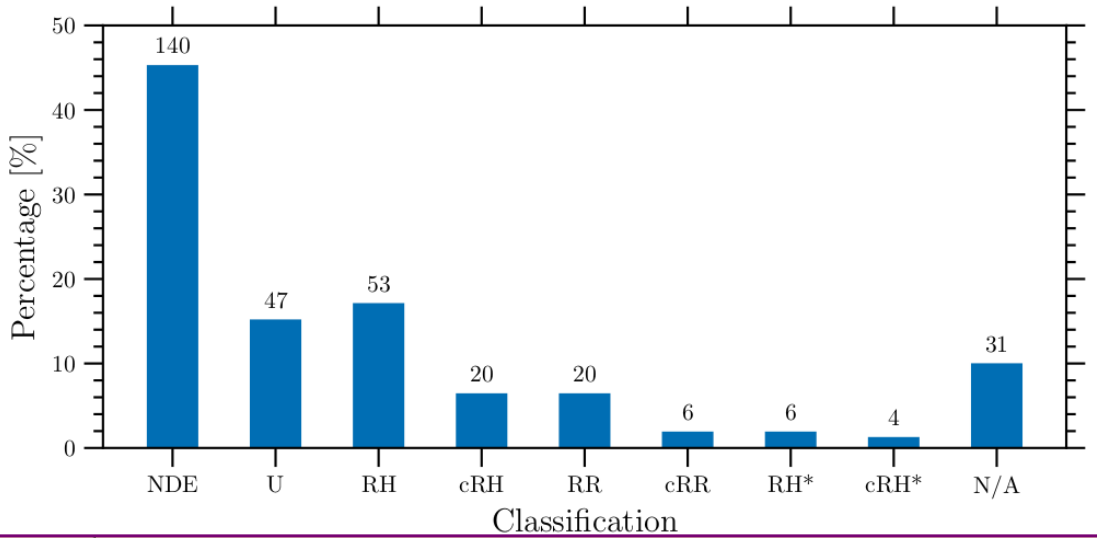


Sky image – Reference mask | Input image | Output segmented image [Stuardi+24]

Radio U-Net: performance verification on LoTSS data

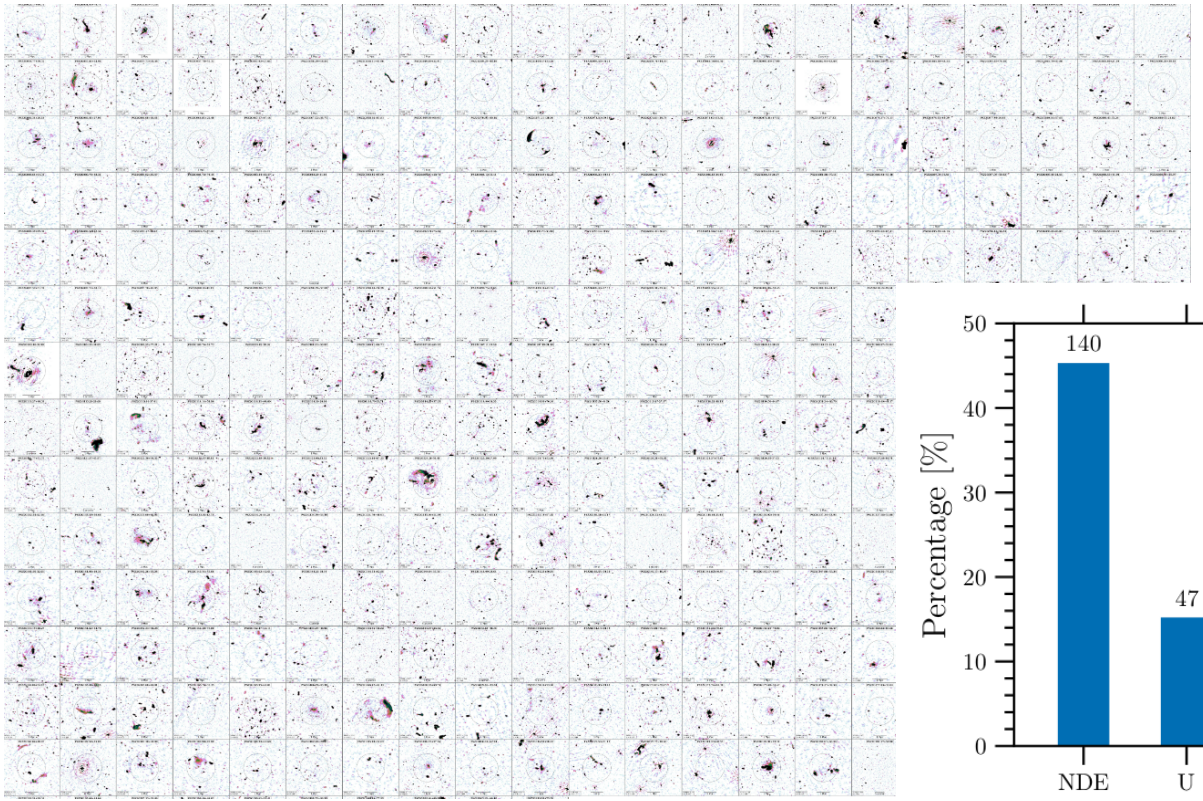
LoTSS dr2: **309** galaxy clusters with visual detection and classification [Botteon+22]
Images directly downloaded from the survey archive without any tailored processing



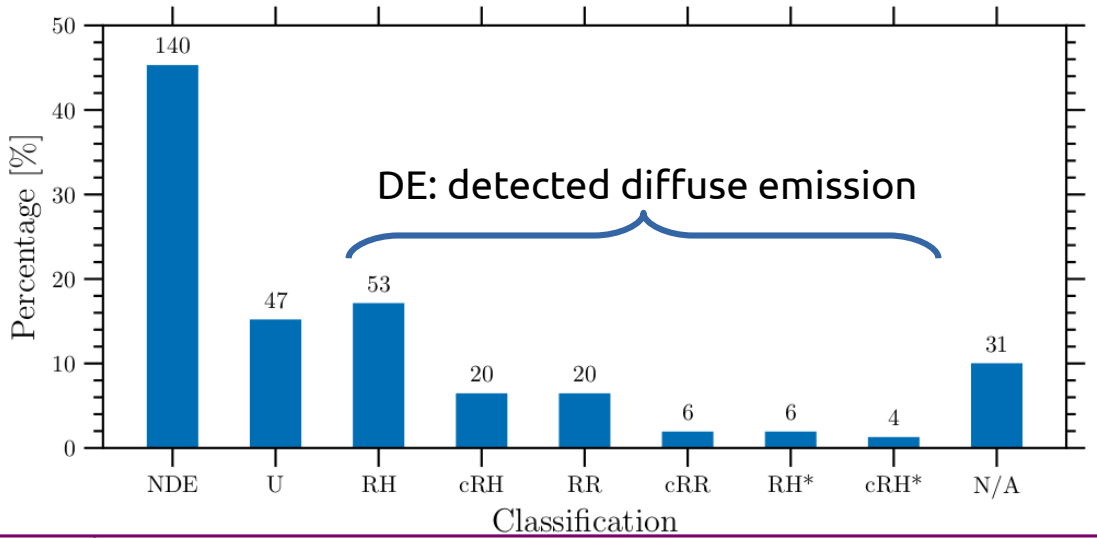


Radio U-Net: performance verification on LoTSS data

LoTSS dr2: **309** galaxy clusters with visual detection and classification [Botteon+22]
Images directly downloaded from the survey archive without any tailored processing

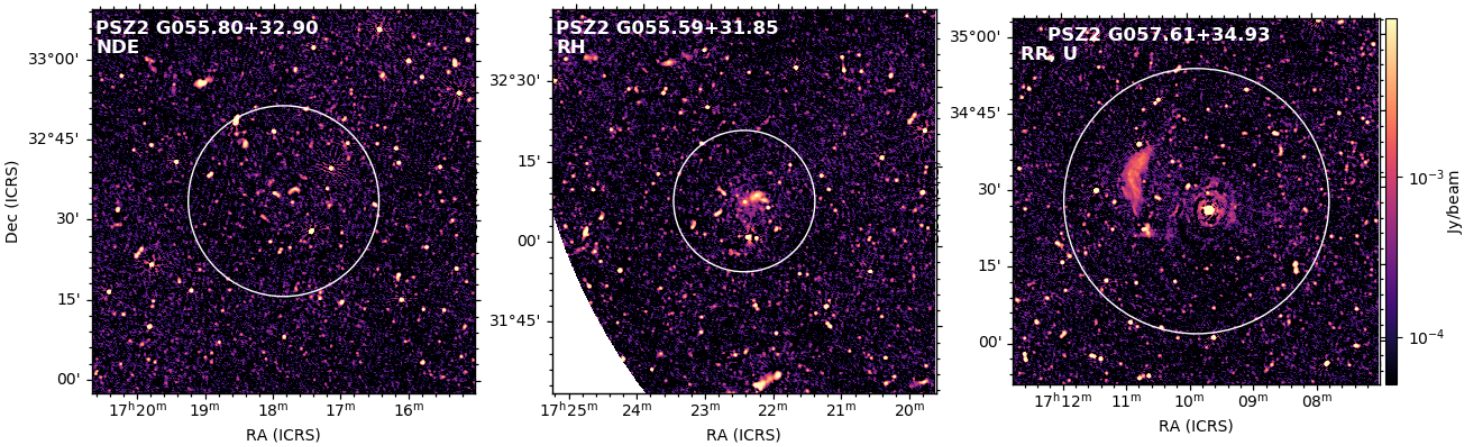


Applied to all 309 clusters
+
Test sample: 114 NDE, 85 DE, 47 DEU
- all with z and good data quality

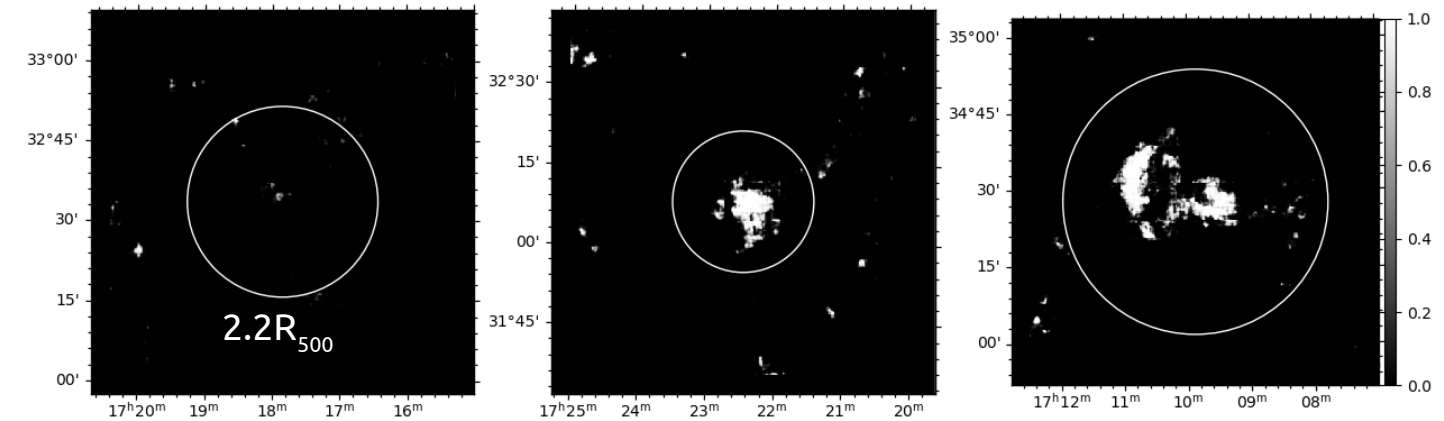


Radio U-Net: performance verification on LoTSS data

LoTSS data



Radio U-Net segmented

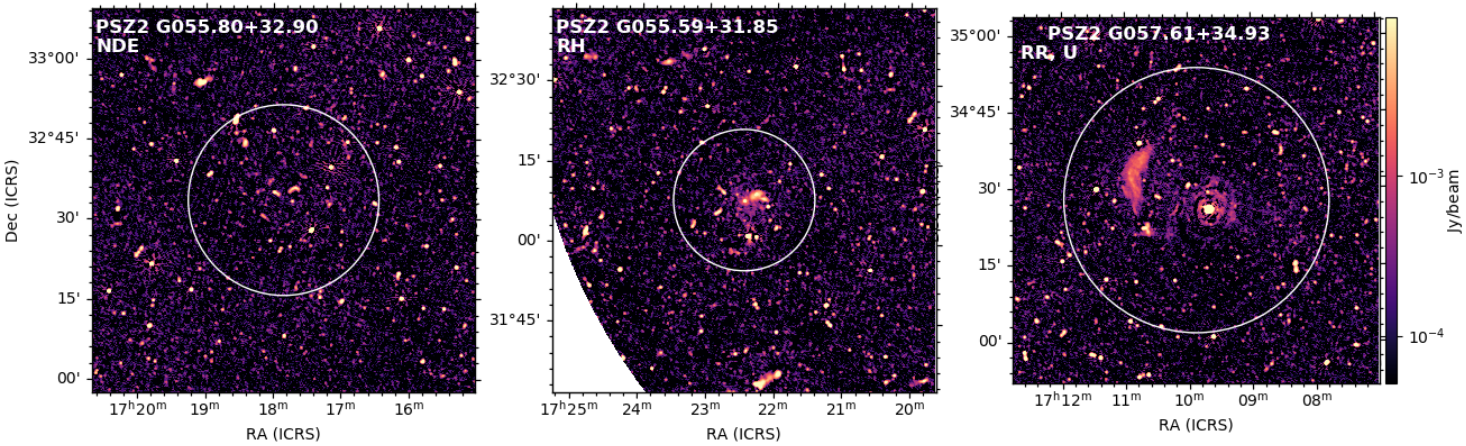


Blind search of diffuse radio emission

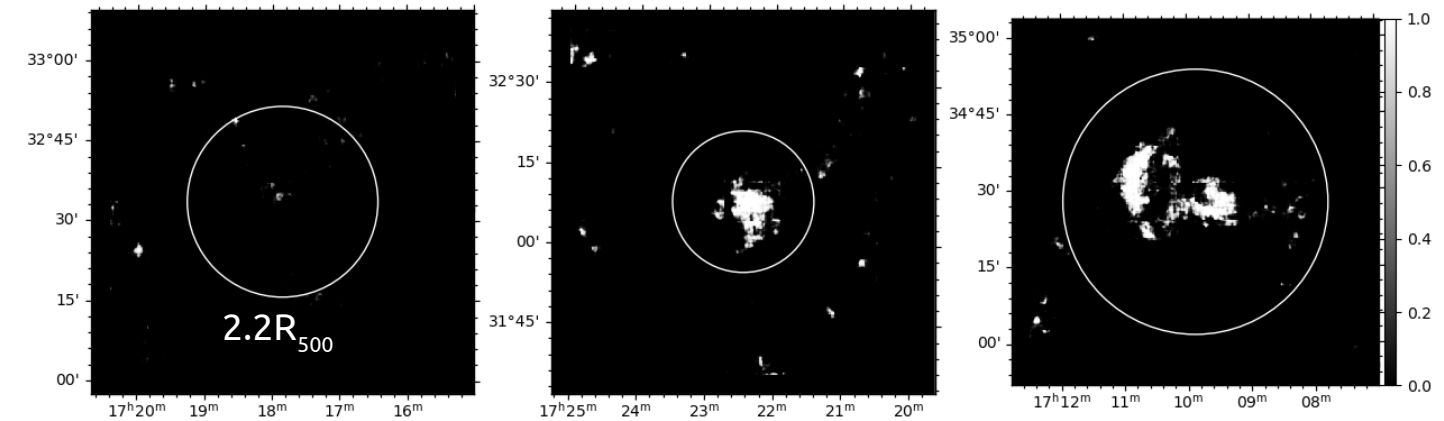
[Stuardi+24]

Radio U-Net: performance verification on LoTSS data

LoTSS data



Radio U-Net segmented



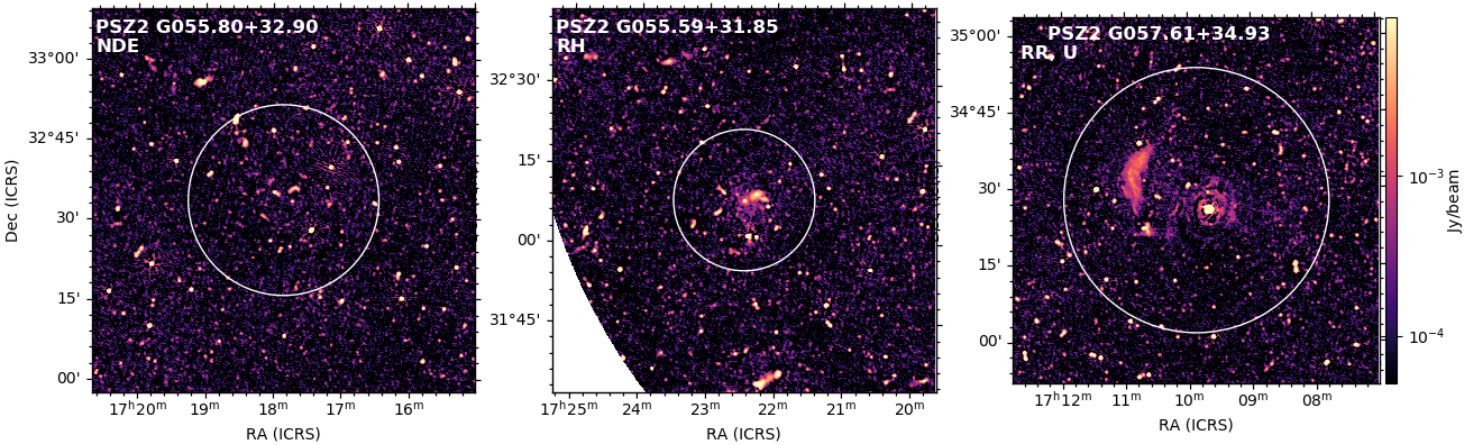
Detection ratio for test the detection with labeled galaxy clusters

R = (sum probability/number of pixels)$_{2.2R500}$

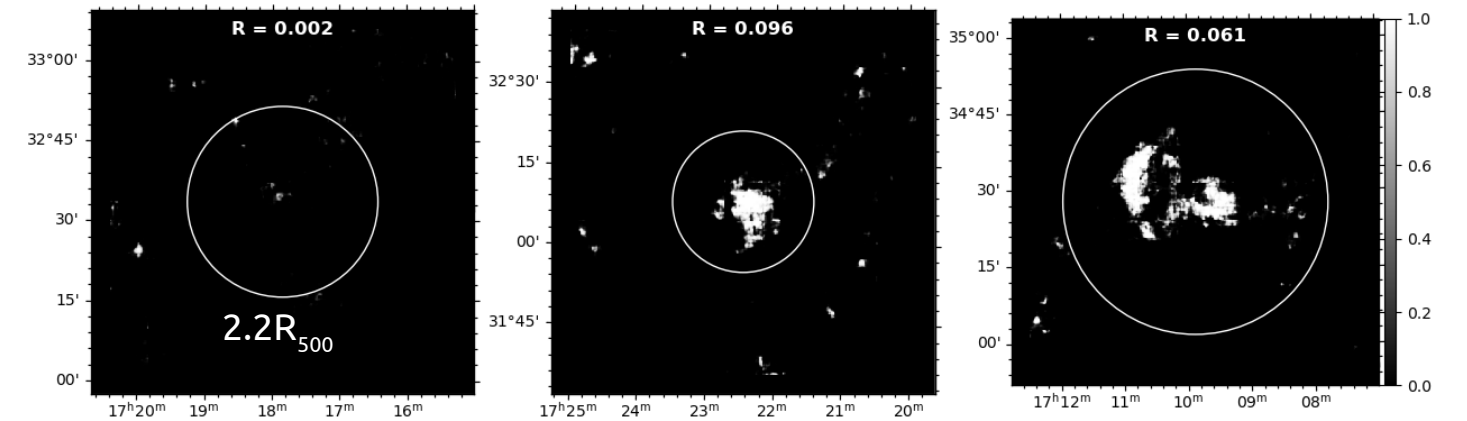
[Stuardi+24]

Radio U-Net: performance verification on LoTSS data

LoTSS data



Radio U-Net segmented

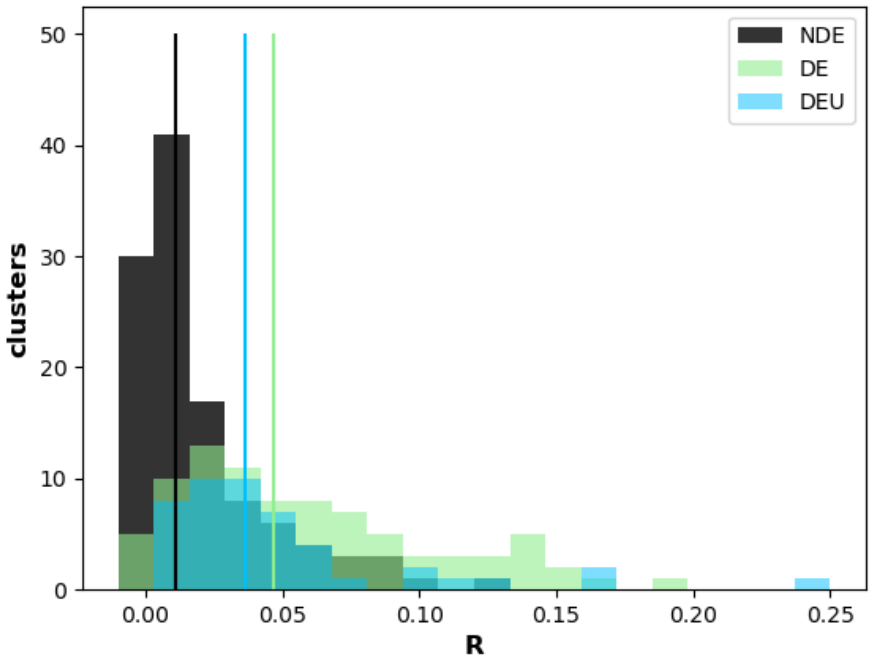


Detection ratio for test the detection with labeled galaxy clusters

R = (sum probability/number of pixels)$_{2.2R500}$

[Stuardi+24]

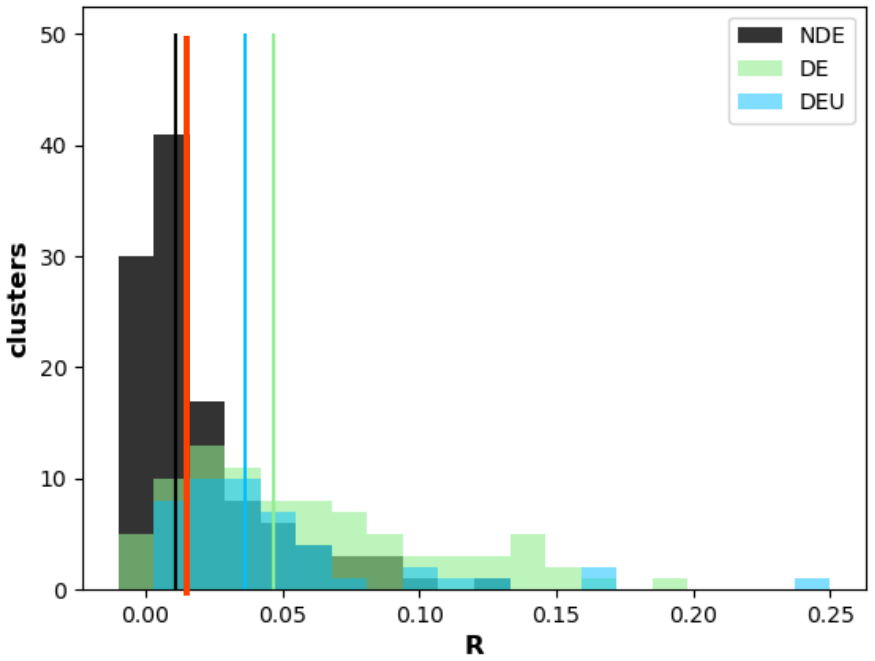
Radio U-Net: performance verification on LoTSS data

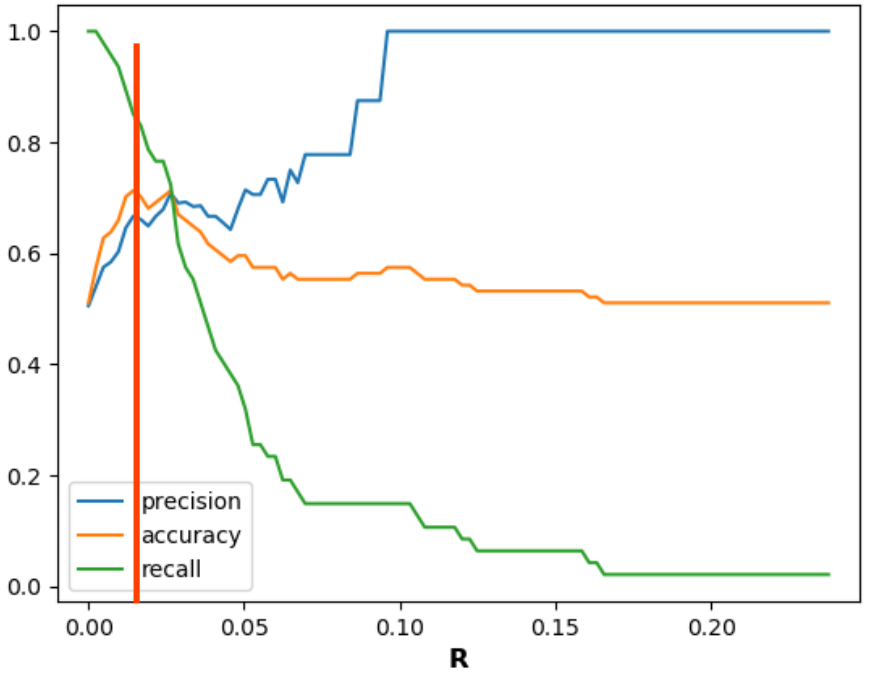


We can put a threshold on R to separate NDE from DE+DEU

[Stuardi+24]

Radio U-Net: performance verification on LoTSS data





Maximize the accuracy: 73% (@R=0.015)

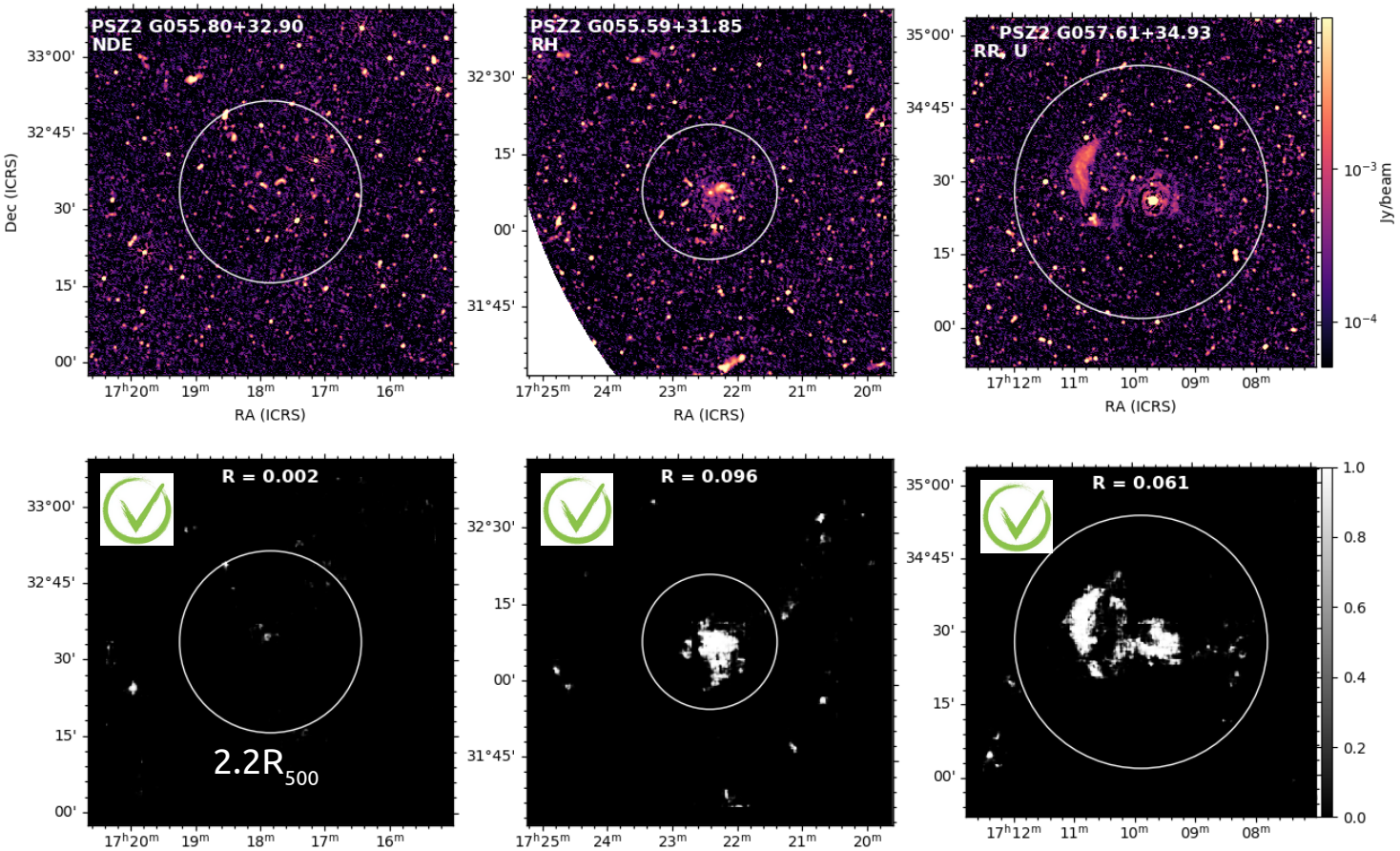
Precision: 72% → 43 out of 114 NDE are wrongly detected
Recall: 83% → 109 out of 132 galaxy clusters with diffuse emission were correctly detected

[Stuardi+24]

Radio U-Net: performance verification on LoTSS data

LoTSS data

Radio U-Net segmented

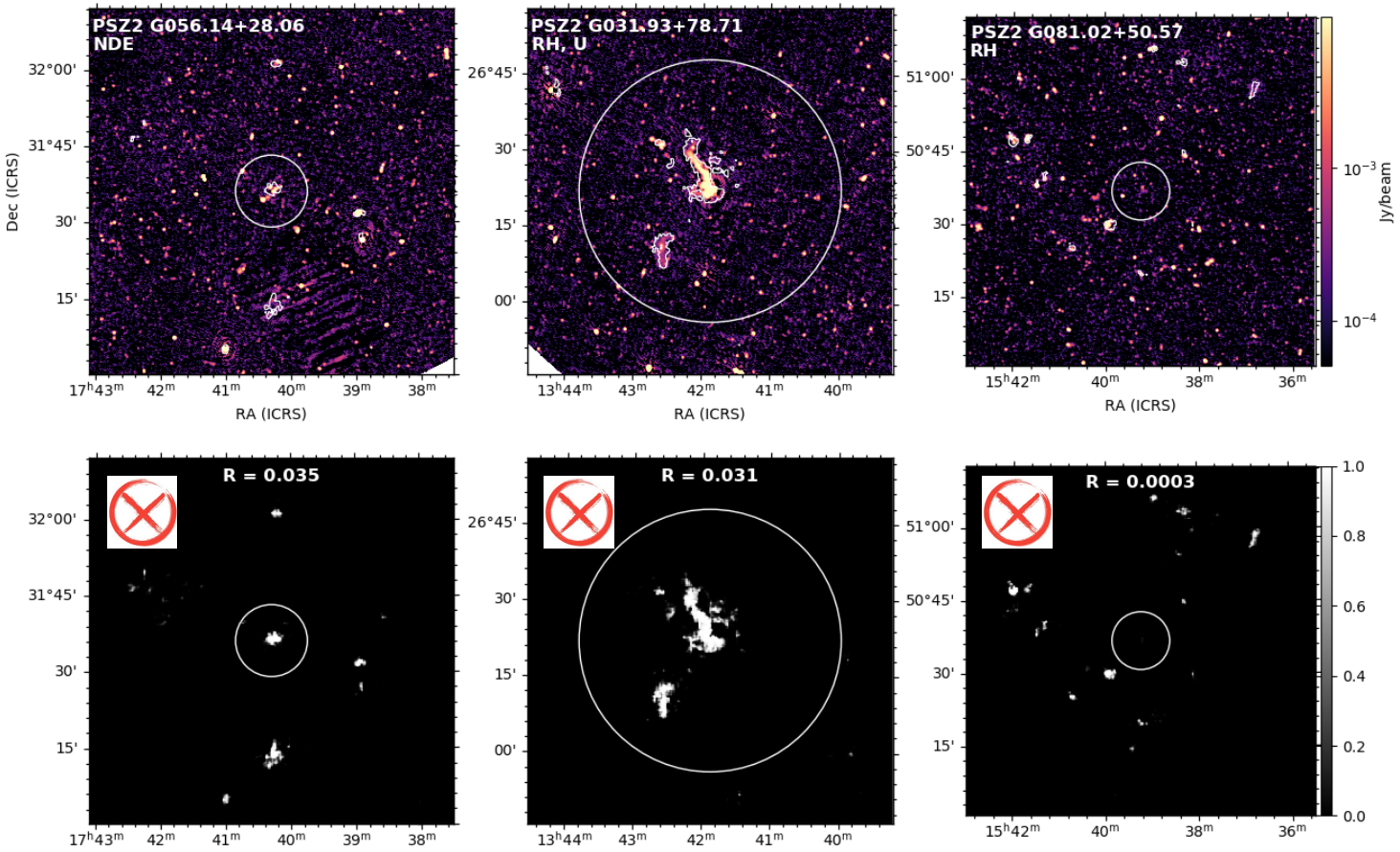


[Stuardi+24]

False positive and false negative

LoTSS data

Radio U-Net segmented

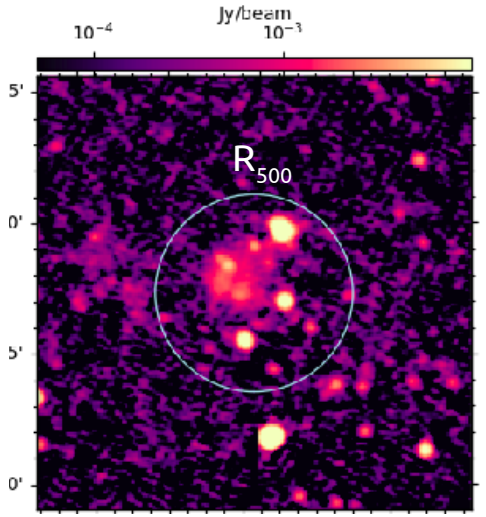


False positive mostly due to galaxy over-densities and/or extended radio galaxies

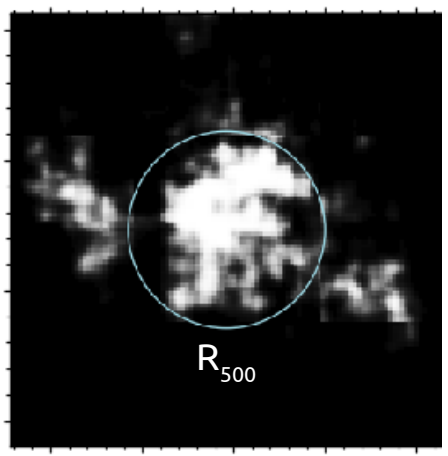
False negative for high redshift or low mass clusters

Detection beyond galaxy clusters

LoTSS data



Radio U-Net segmented

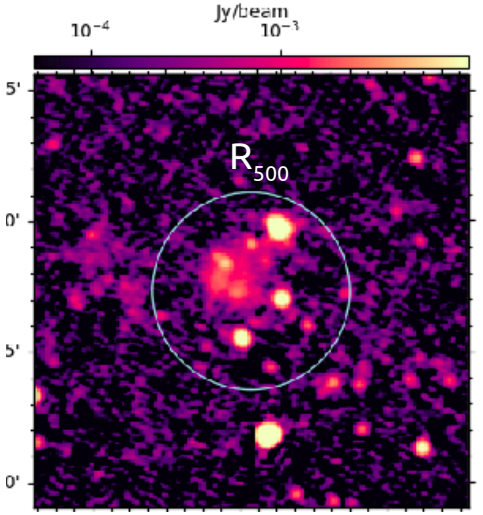


Detection of diffuse radio emission beyond galaxy clusters and below classical detection limits

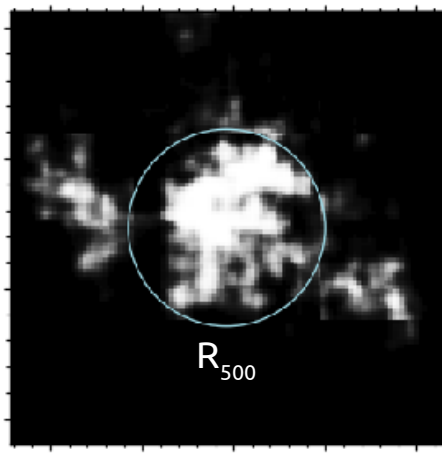
Detection beyond galaxy clusters

SPOKE 3 Astrophysics & Cosmos Observations

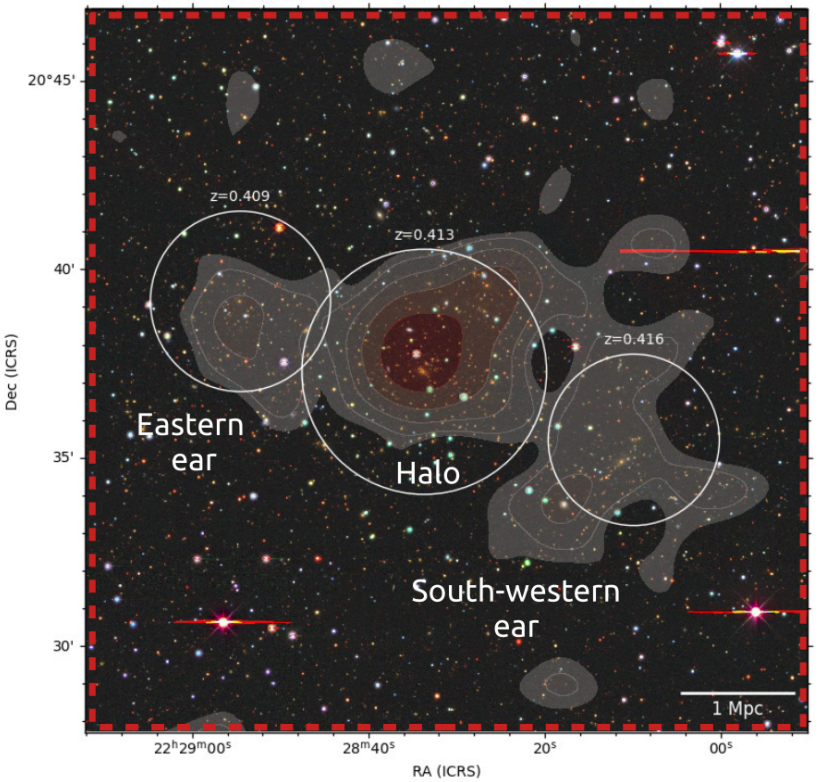
LoTSS data



Radio U-Net segmented

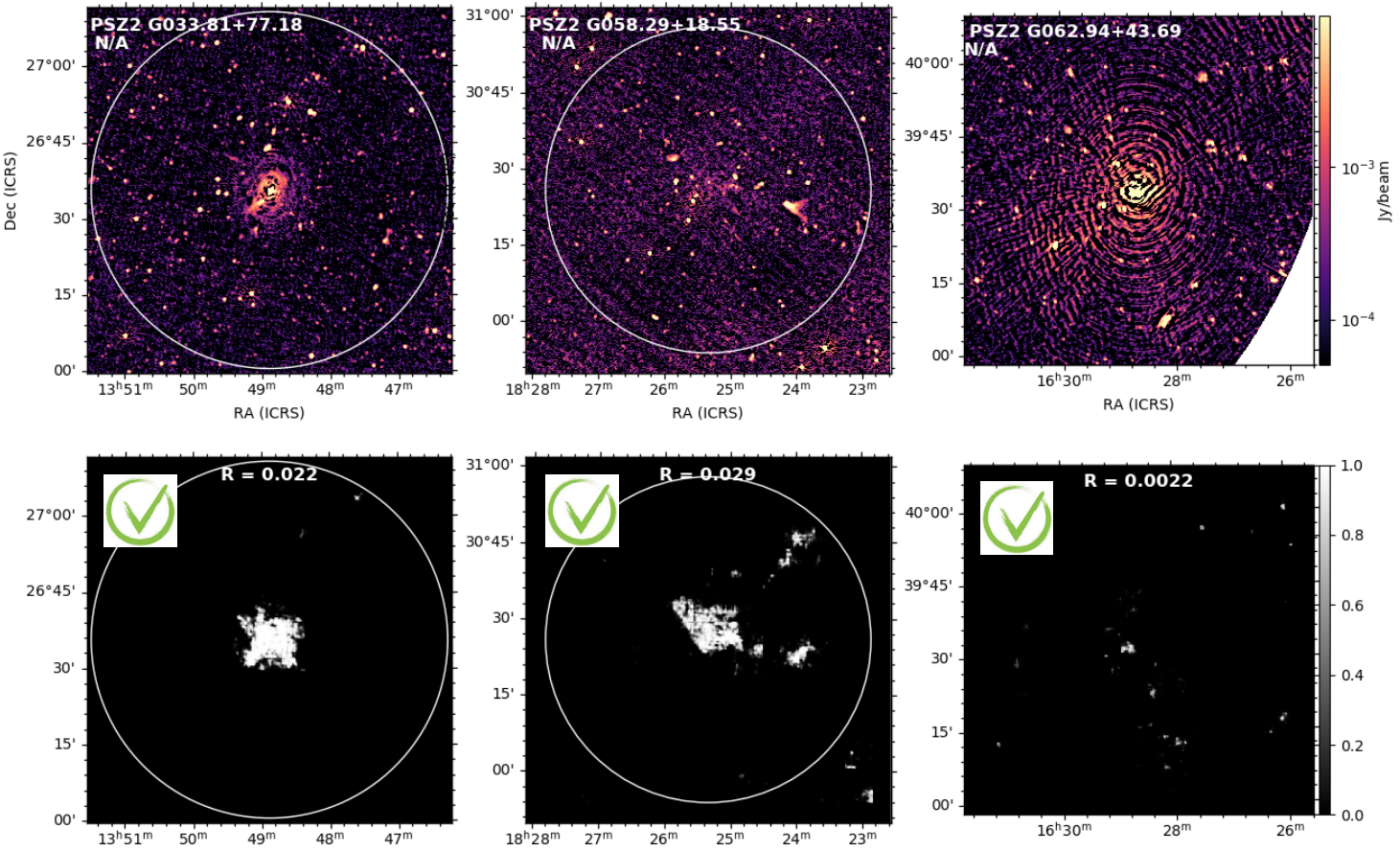


Detection of diffuse radio emission beyond galaxy clusters and below classical detection limits



[Stuardi+25]

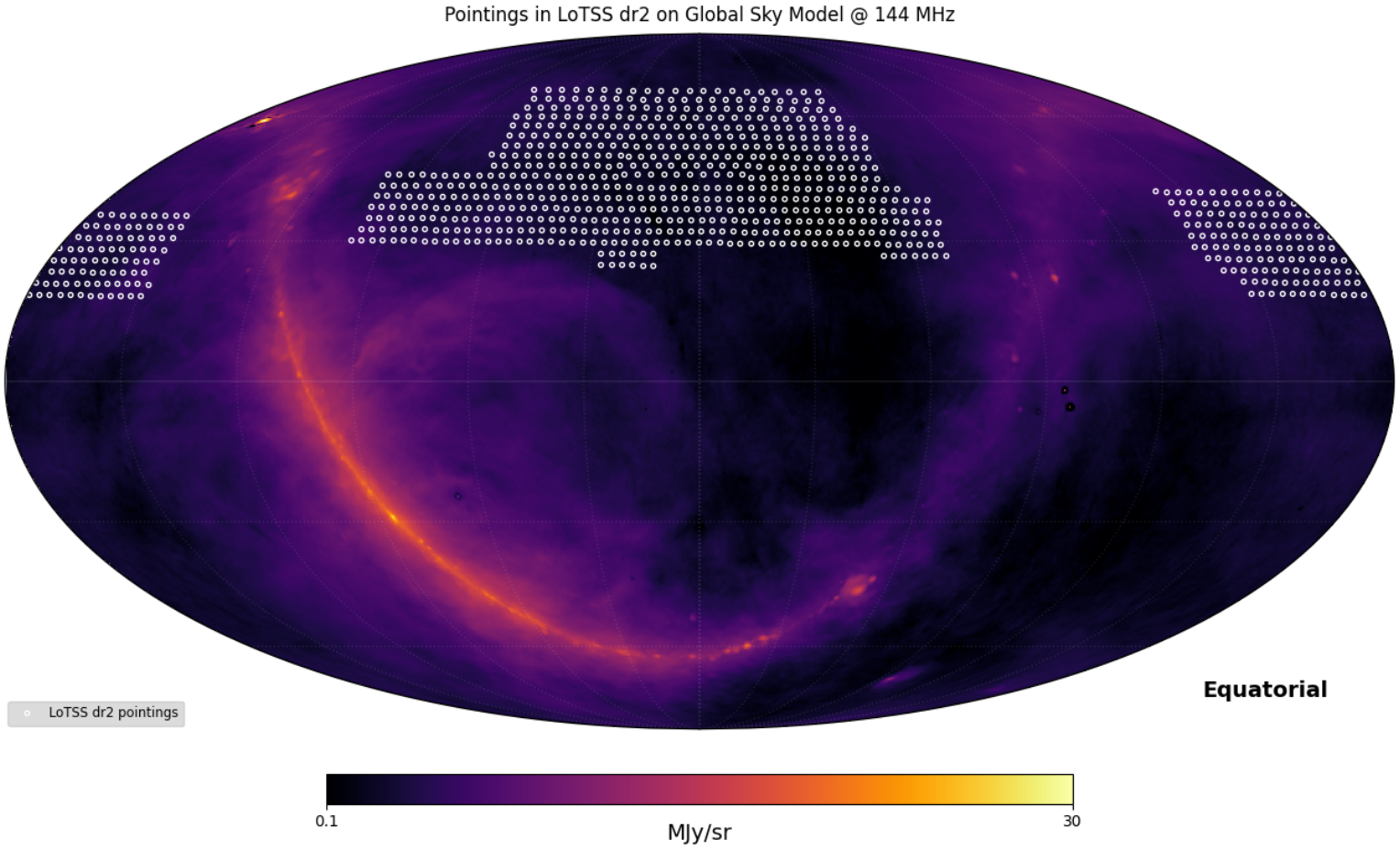
Low quality images



Correct segmentation of low-quality images

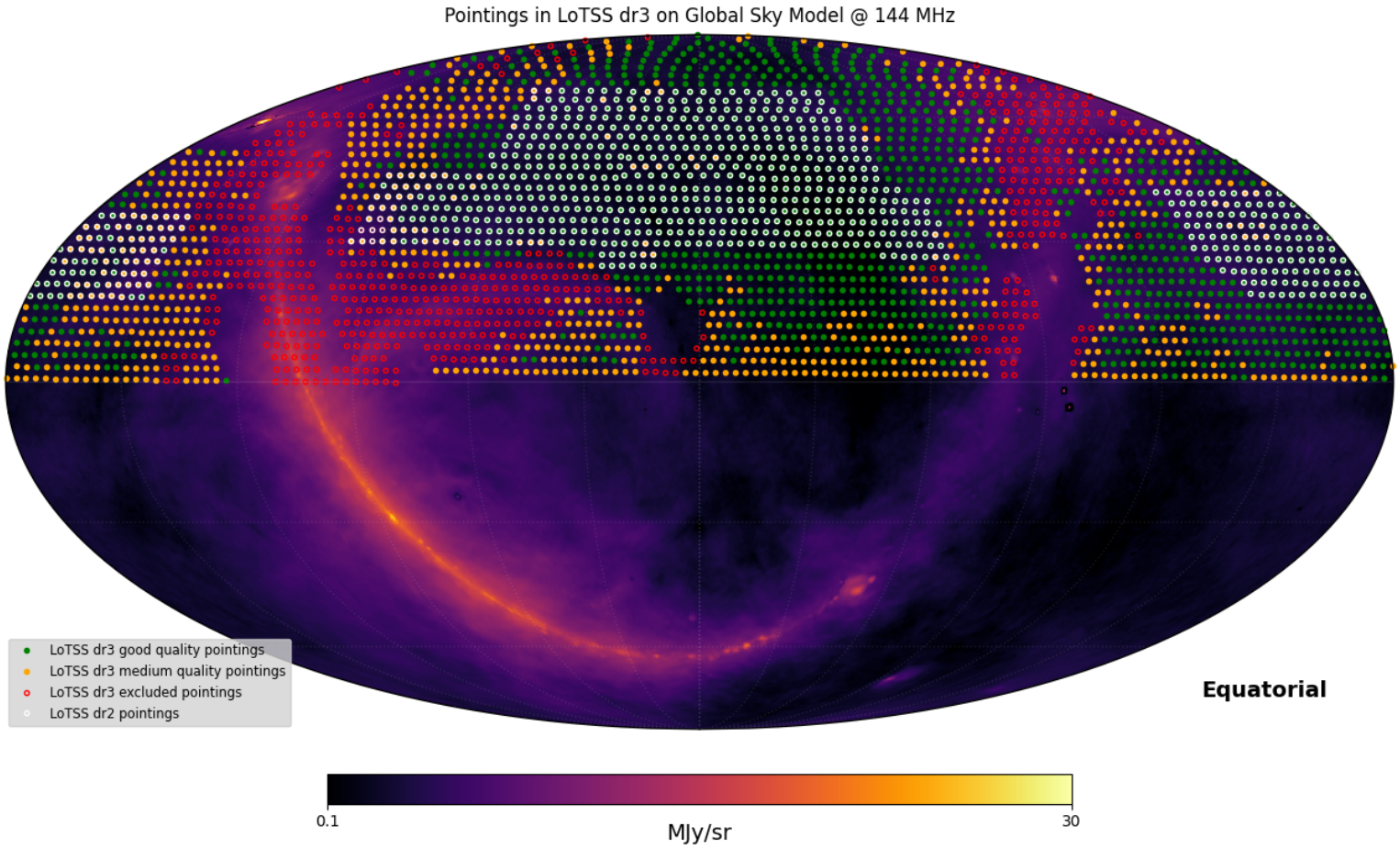
[Stuardi+24]

Radio U-Net: towards the LoTSS dr3



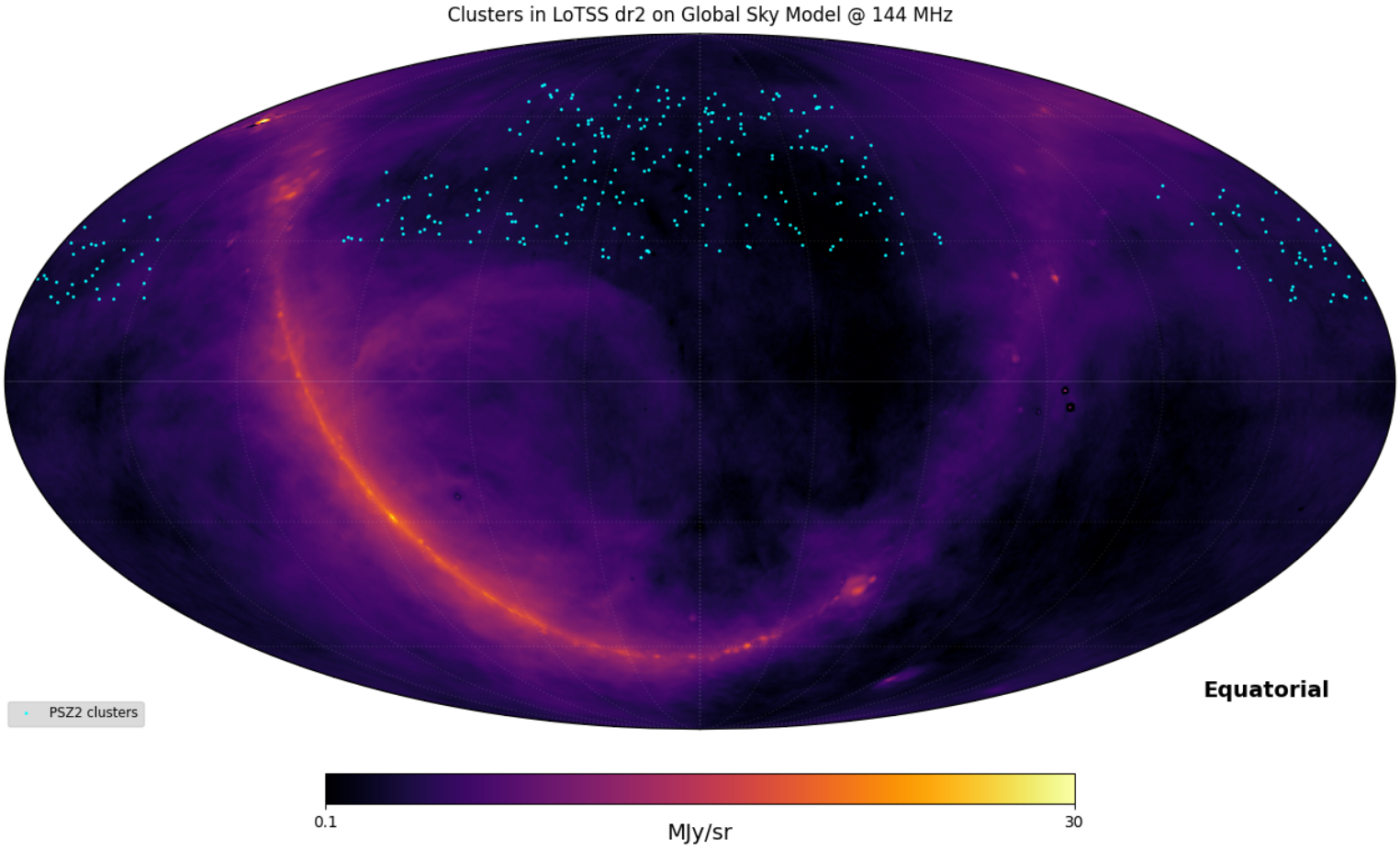
Radio U-Net: towards the LoTSS dr3

SPOKE 3
Astrophysics
& Cosmos Observations

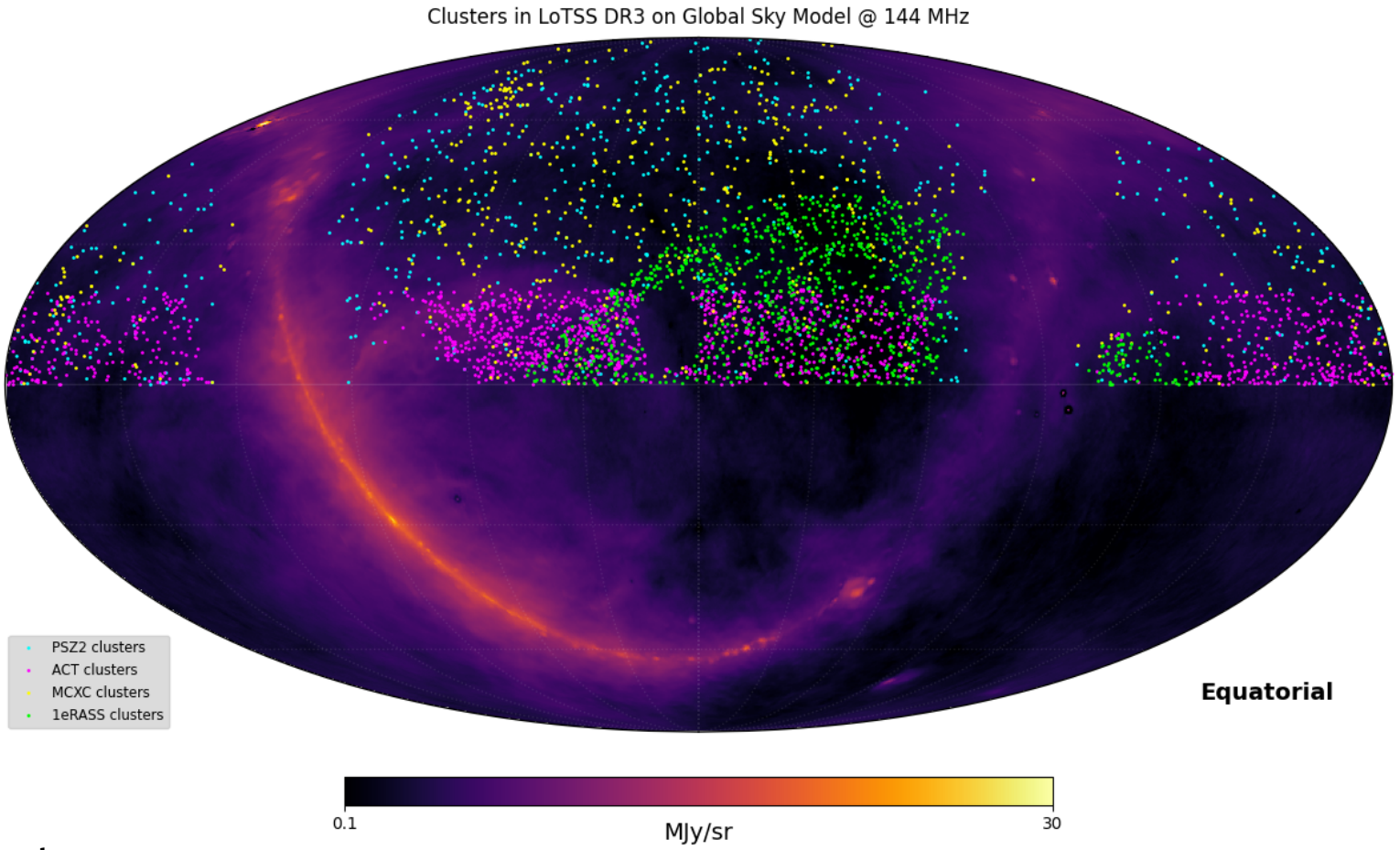


841 → 2913 pointings

Radio U-Net: towards the LoTSS dr3



Radio U-Net: towards the LoTSS dr3



309 → 3821 clusters

Radio U-Net: final remarks and next steps

SPOKE 3 Astrophysics & Cosmos Observations

- Automated and fast segmentation of diffuse radio sources in large surveys
- Used for classification: 73% accuracy on a balanced data-set, 83% recall
- Successful on low-quality images
  [Stuardi et al. 2024]

- Enabling new discoveries
  [Stuardi et al. 2025]

- Machine-learning-based LoTSS dr3 galaxy cluster catalog in preparation

Thank you for your attention!