

Radio U-Net: a fully convolutional neural network to detect diffuse sources in radio surveys

Machine Learning for Astrophysics, 8th July 2024

Chiara Stuardi

INAF – IRA Istituto di Radioastronomia (Italy)

Collaborators: C. Gheller (INAF-IRA), F. Vazza (U. of Bologna), A. Botteon (INAF-IRA)

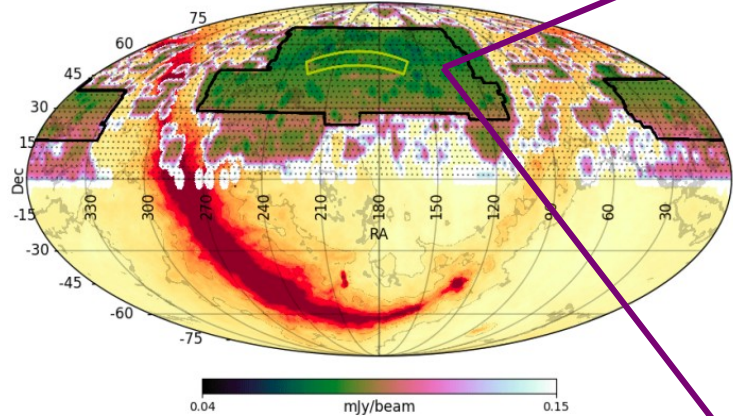


ICSC Italian Research Center on High-Performance
Computing, Big Data and Quantum Computing

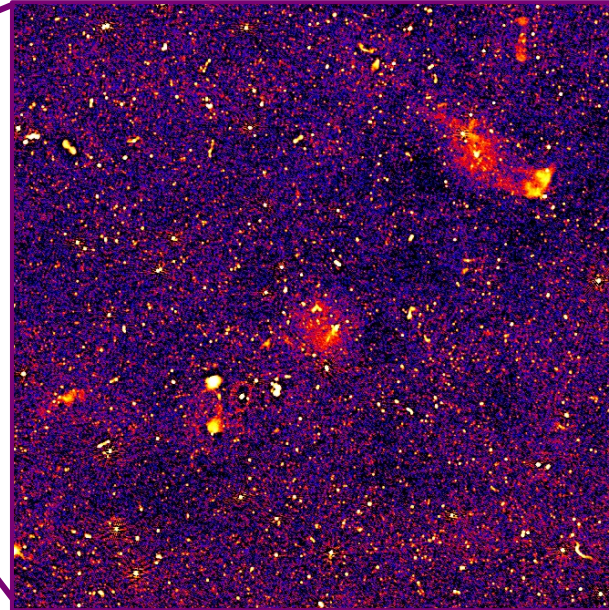
Scientific rationale

Current radio surveys are challenging our detection and cataloging strategies

- large data size (PB/year)
 - 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



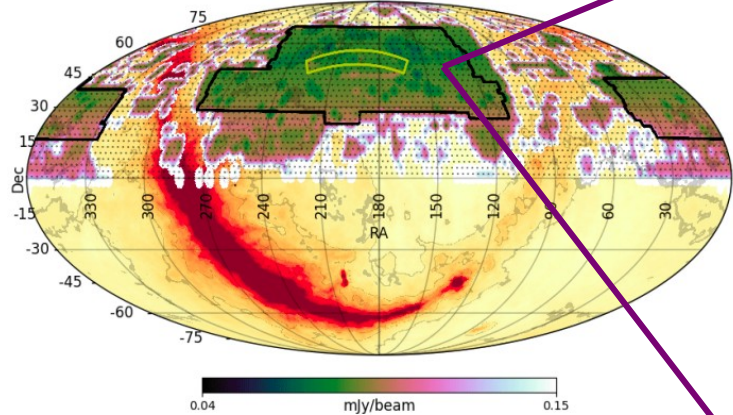
LOFAR Two Metre Sky Survey DR2
@150 MHz [LOTSS, Shimwell+22]



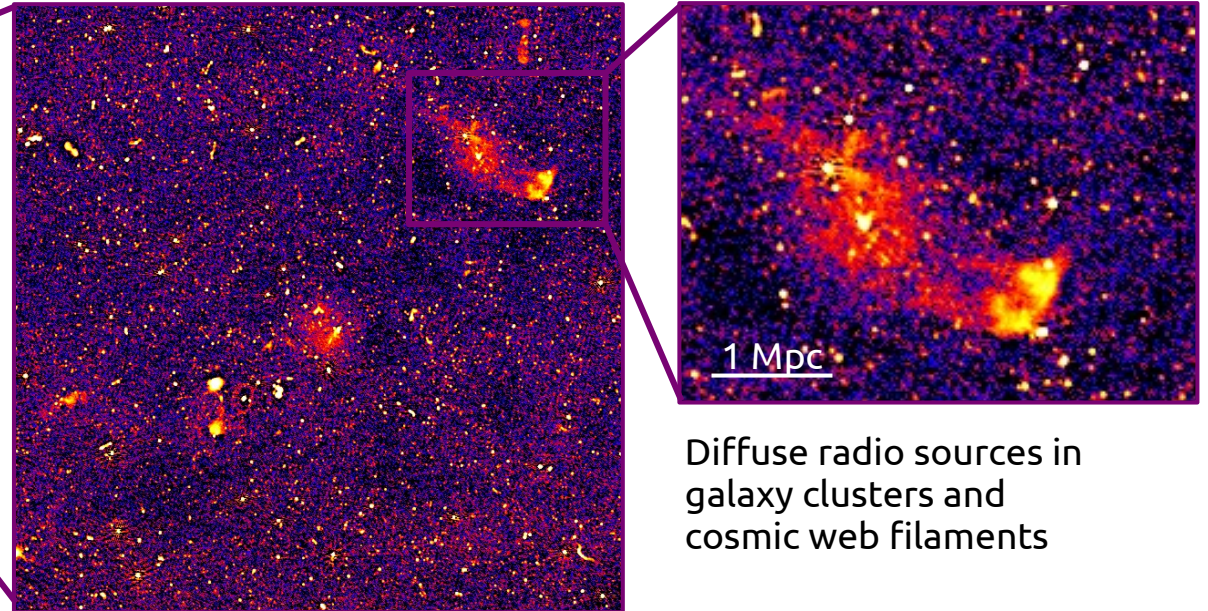
Scientific rationale

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]

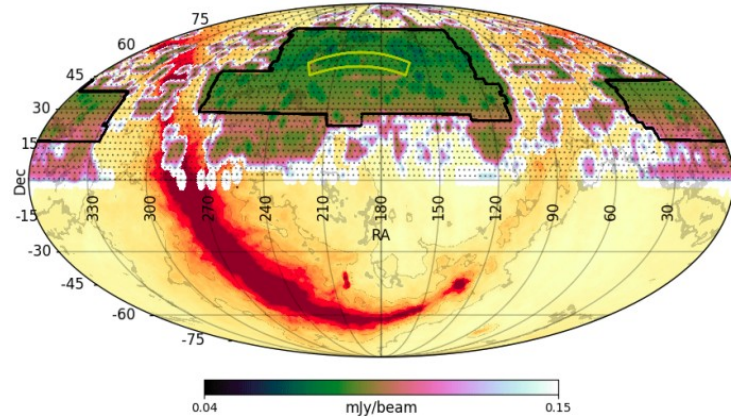


Diffuse radio sources in
galaxy clusters and
cosmic web filaments

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

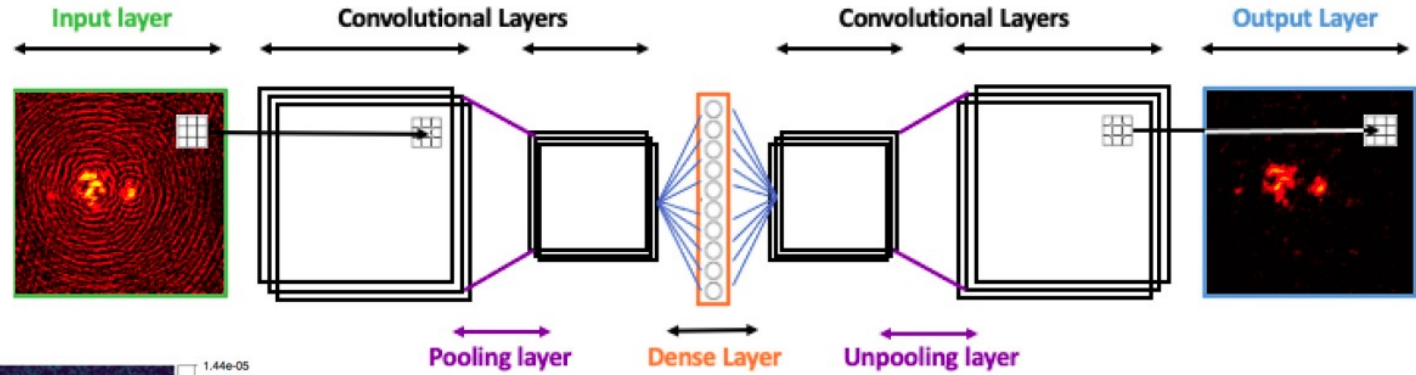
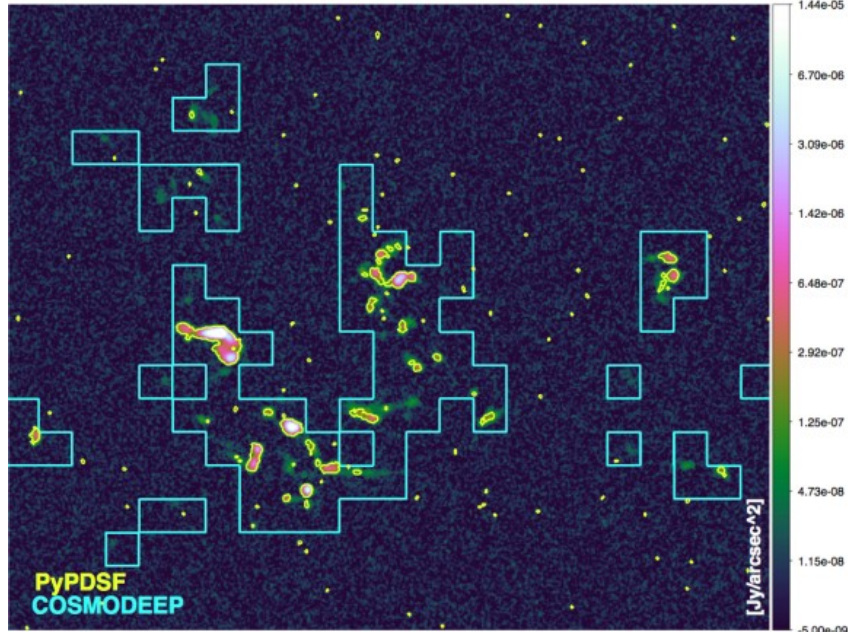
[credits: SKAO]



Previous works

COSMODEEP

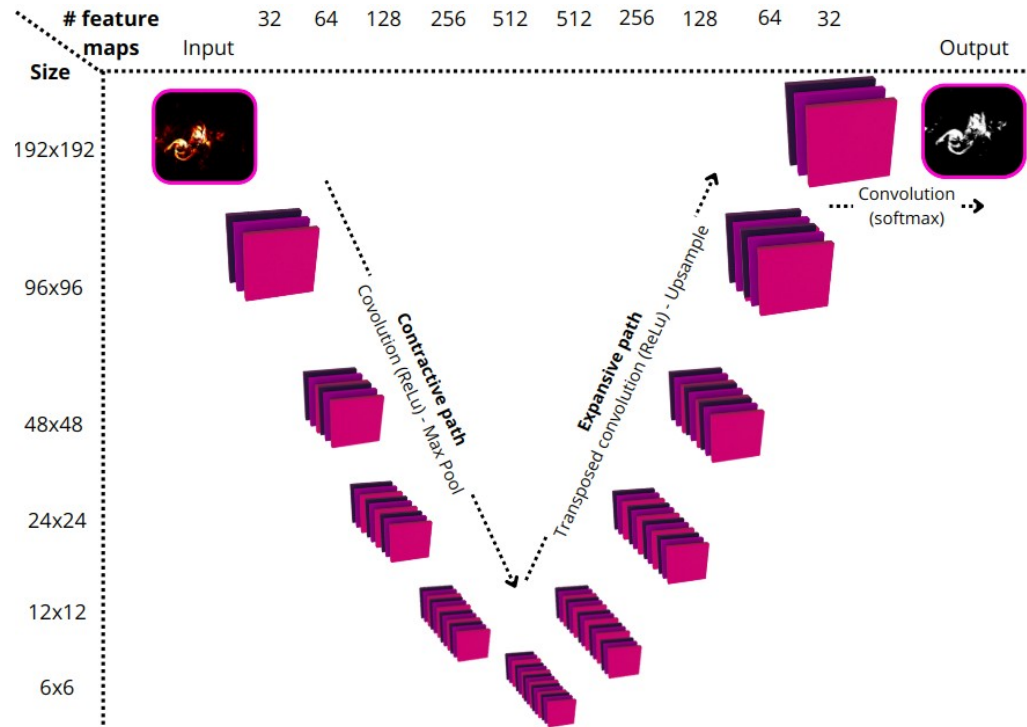
[Gheller,Vazza,Bonafede18]



De-noising autoencoder
[Gheller&Vazza+22]

Radio U-Net: architecture and training strategy

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

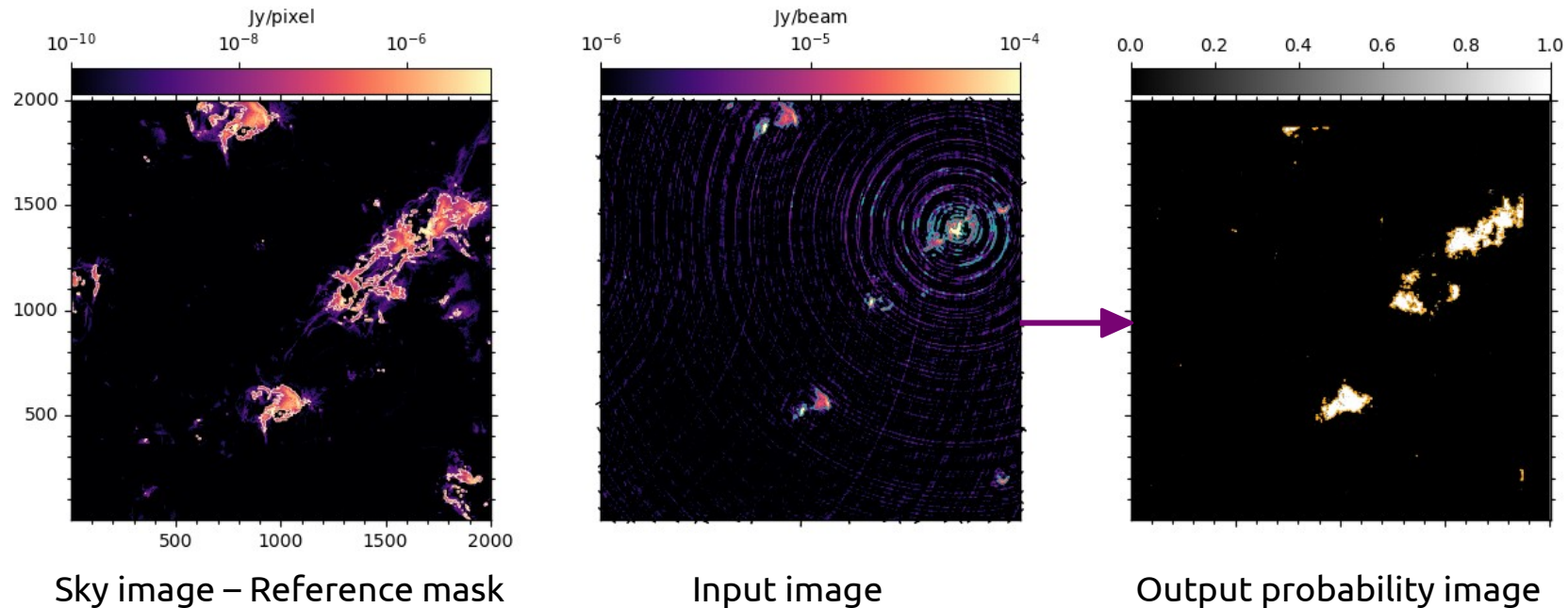


- Python Keras Tensorflow
- Scalable from CPU to GPU
- Run on CINECA HPC systems (Leonardo)

Radio U-Net: architecture and training strategy

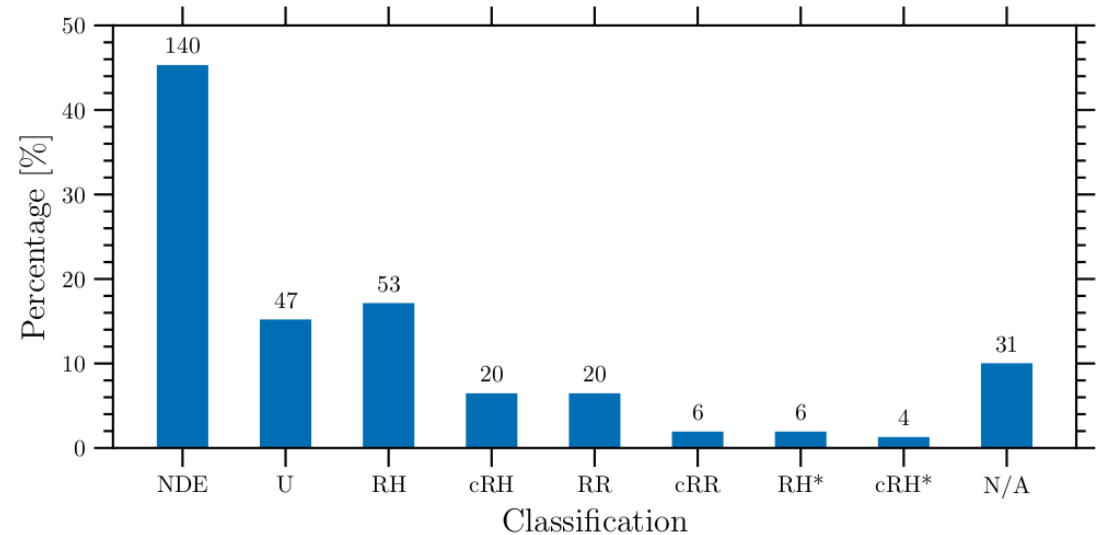
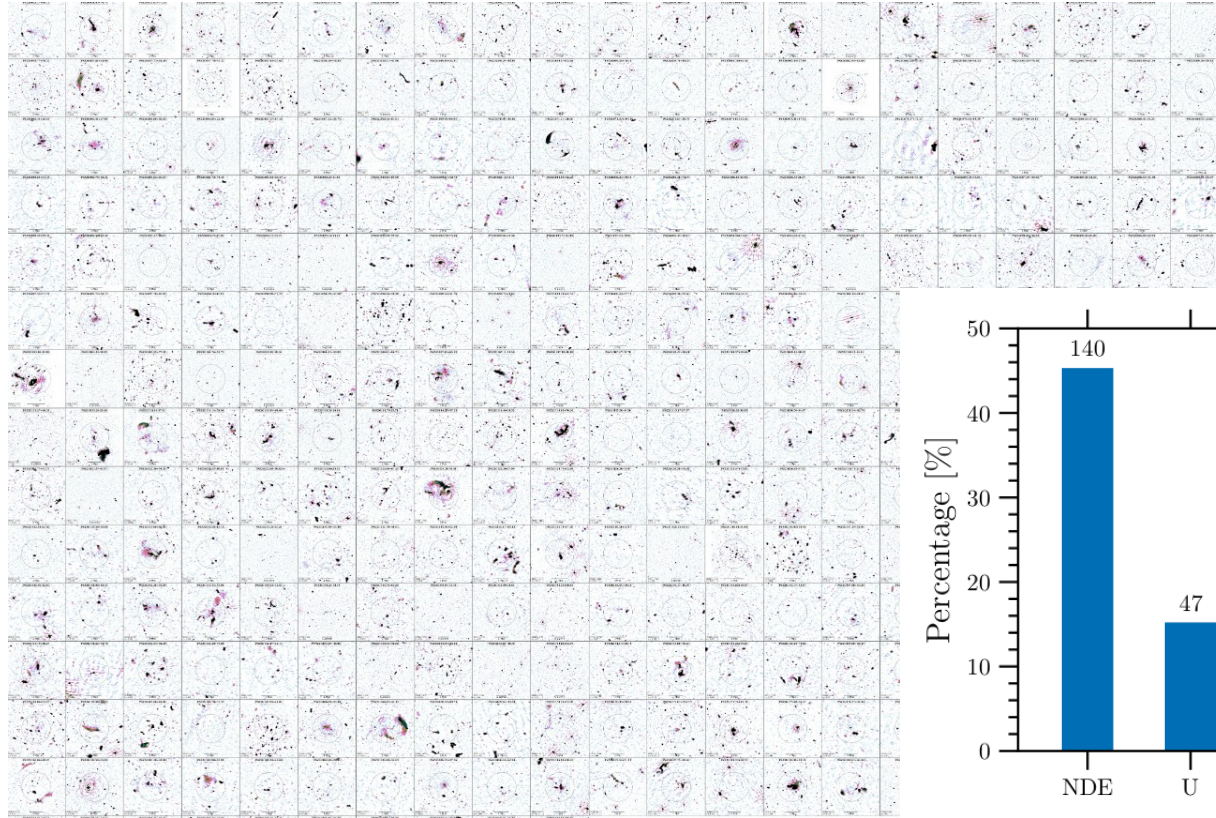
Convolutional neural network based on the U-net architecture [Ronneberger+15] to perform the segmentation of diffuse radio emission in radio astronomical surveys

Training on synthetic observations built on cosmological simulations [Gheller&Vazza22]



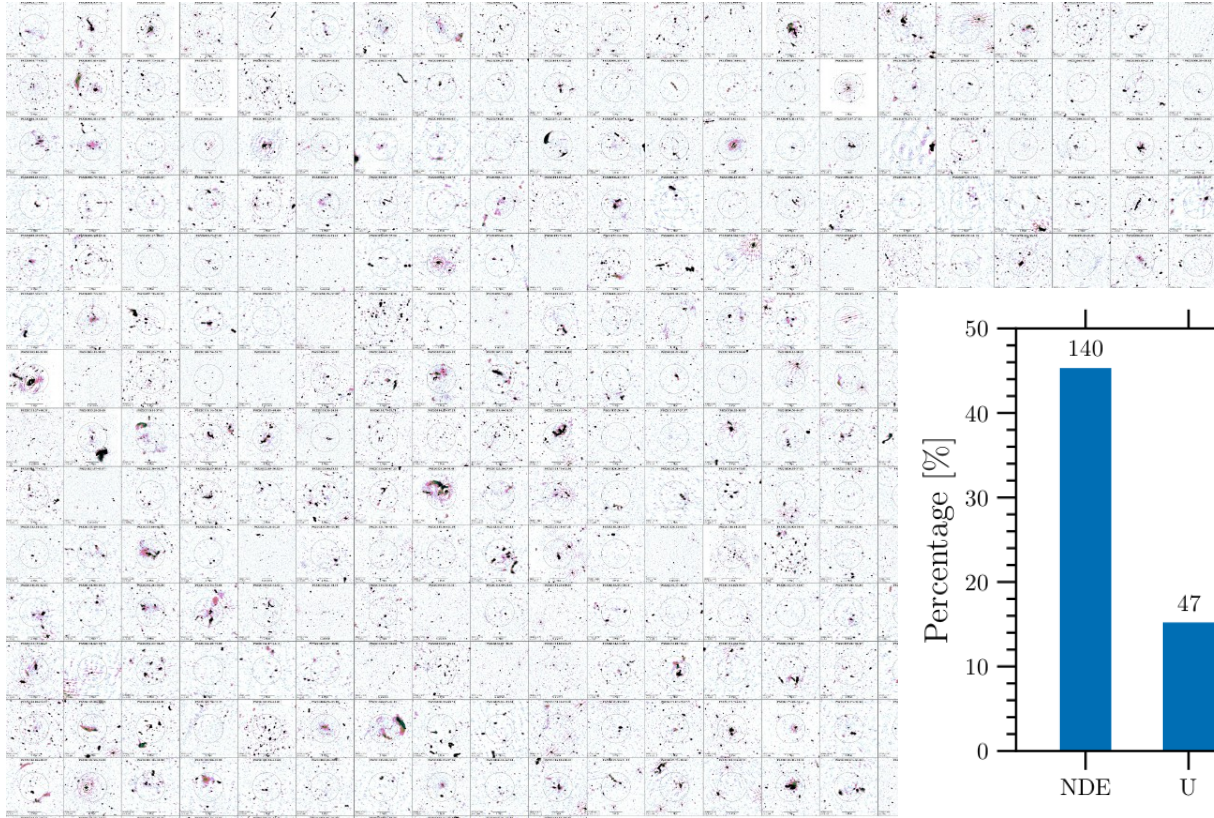
Radio U-Net: performance verification on LoTSS data

LoTSSDR2/PSZ2: 309 galaxy clusters with visual detection and classification [Botteon+22]
Images directly downloaded from the survey archive without any tailored processing

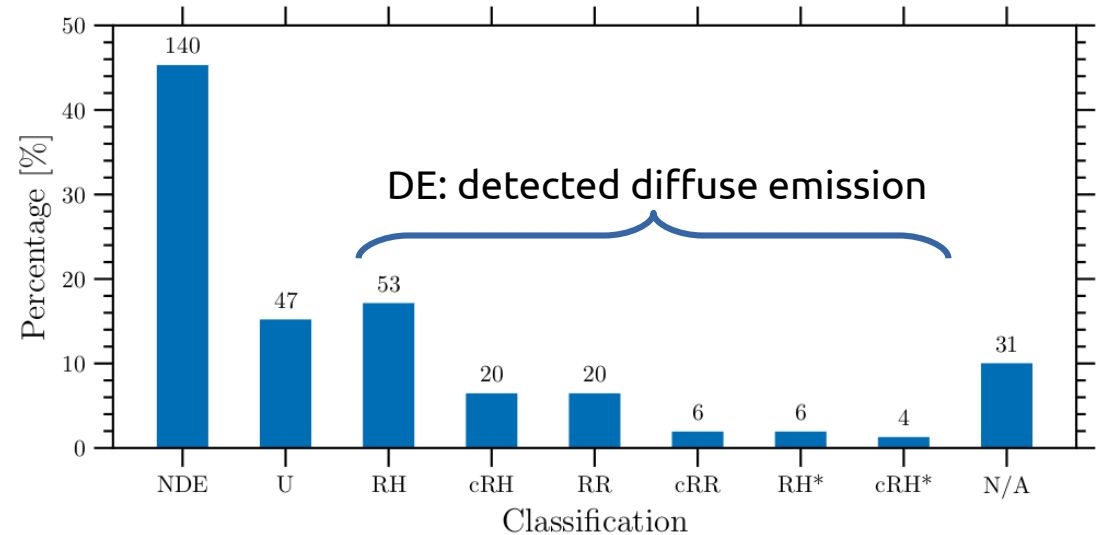


Radio U-Net: performance verification on LoTSS data

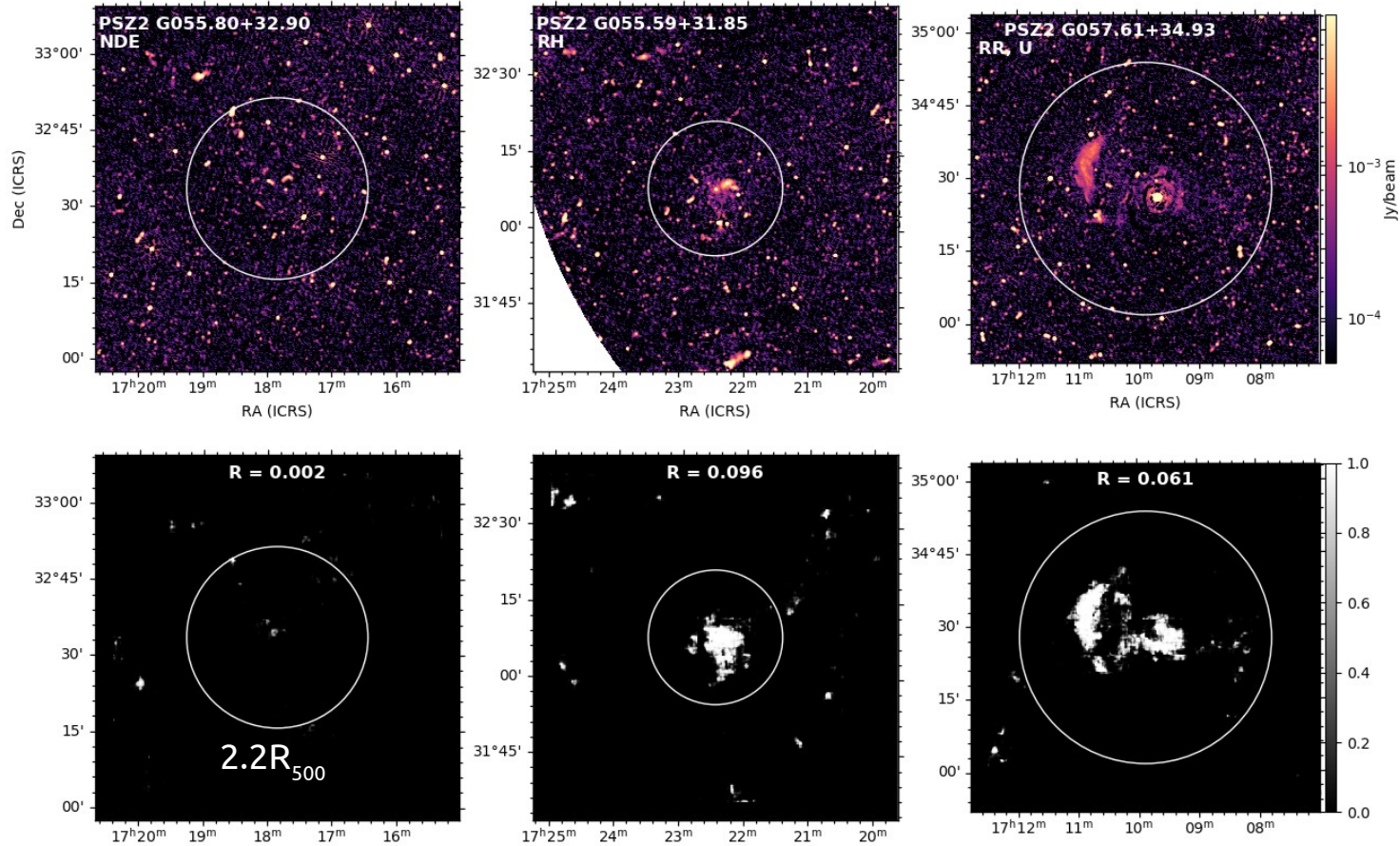
LoTSSDR2/PSZ2: 309 galaxy clusters with visual detection and classification [Botteon+22]
Images directly downloaded from the survey archive without any tailored processing



Test sample: 114 NDE, 85 DE, 47 DEU



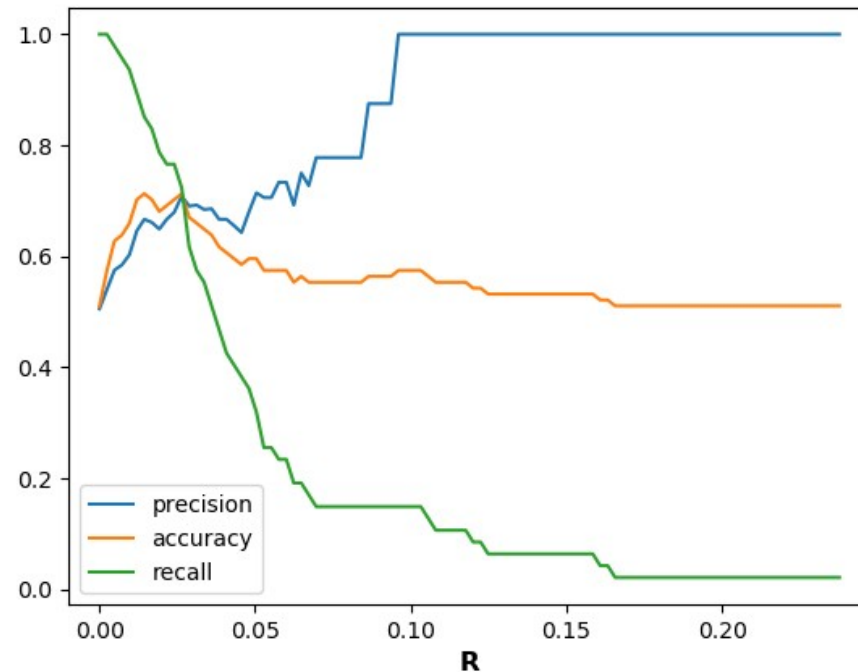
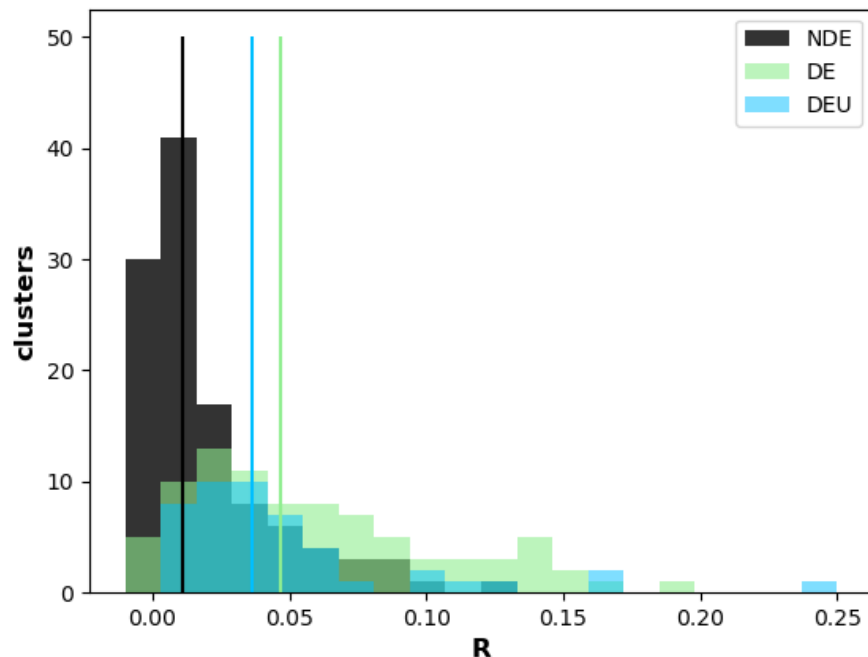
Radio U-Net: performance verification on LoTSS data



Detection ratio:

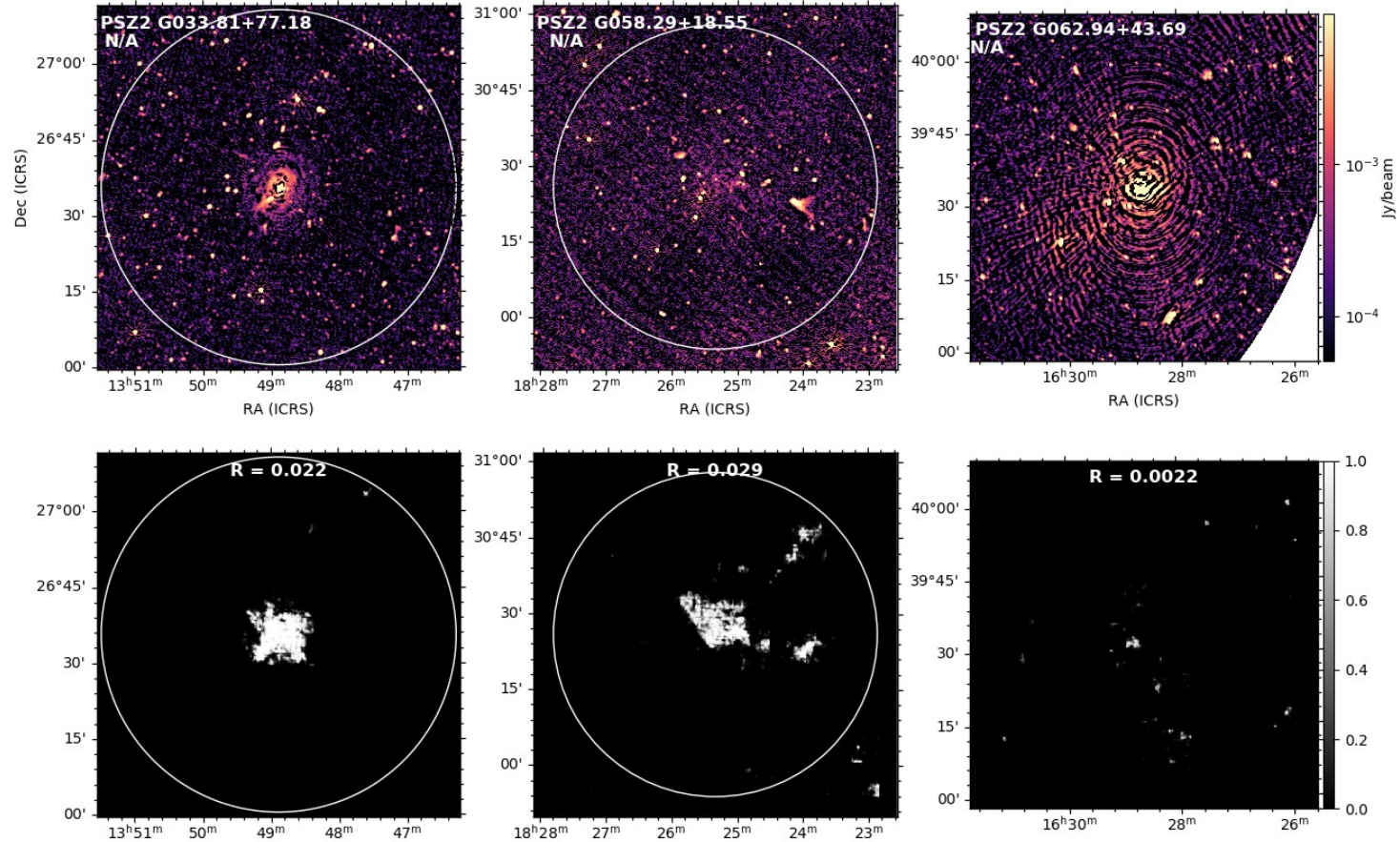
$$R = (\text{sum probability/number of pixels})_{2.2R_{500}}$$

Radio U-Net: performance verification on LoTSS data



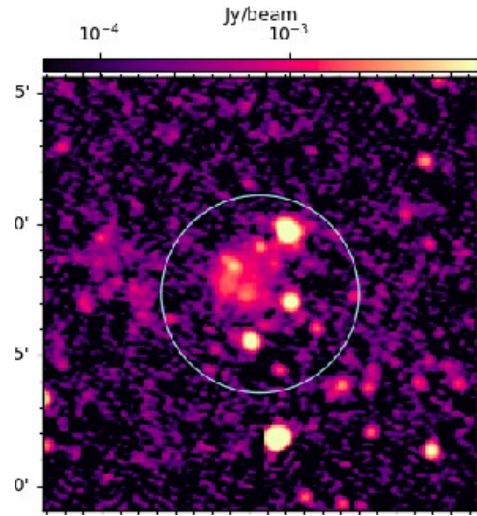
	Non-detected	Detected	Uncertain	tot
	NDE	DE	DEU	
Initial test set	114	85	47	246
True _R =0.015	71 (62%)	70 (82%)	39 (83%)	180 (73%)

Radio U-Net: performance verification on LoTSS data

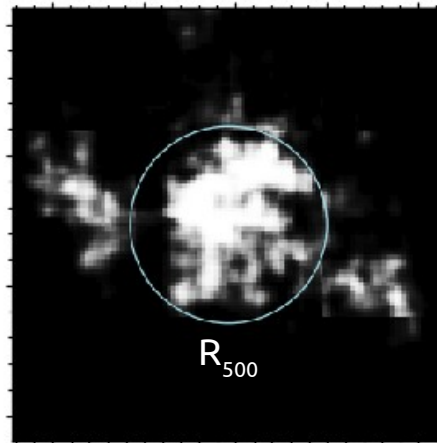


Correct segmentation of low-quality images

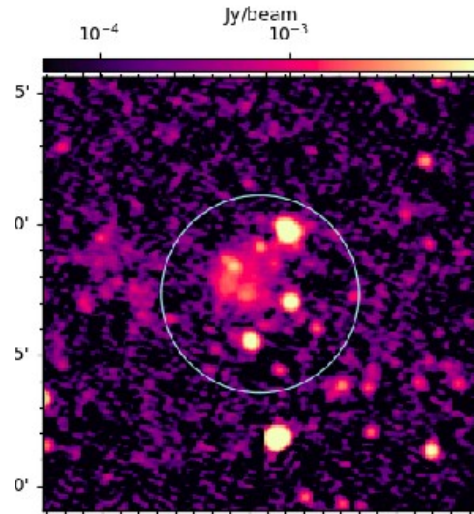
Radio U-Net: performance verification on LoTSS data



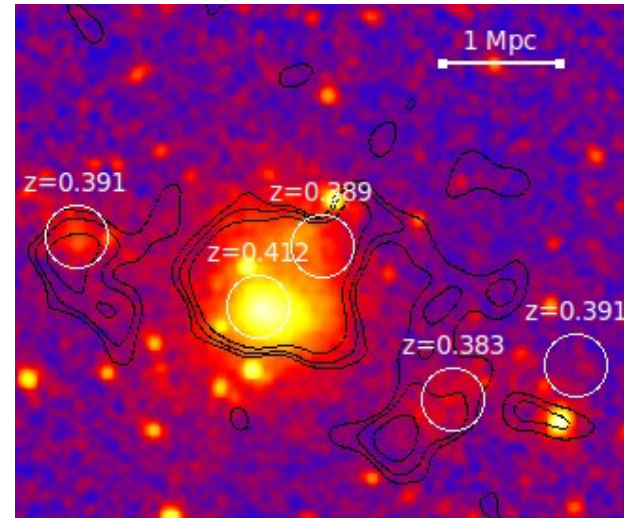
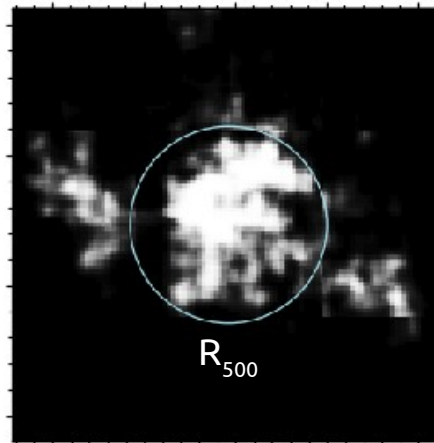
Detection of diffuse radio emission
beyond galaxy clusters



Radio U-Net: performance verification on LoTSS data



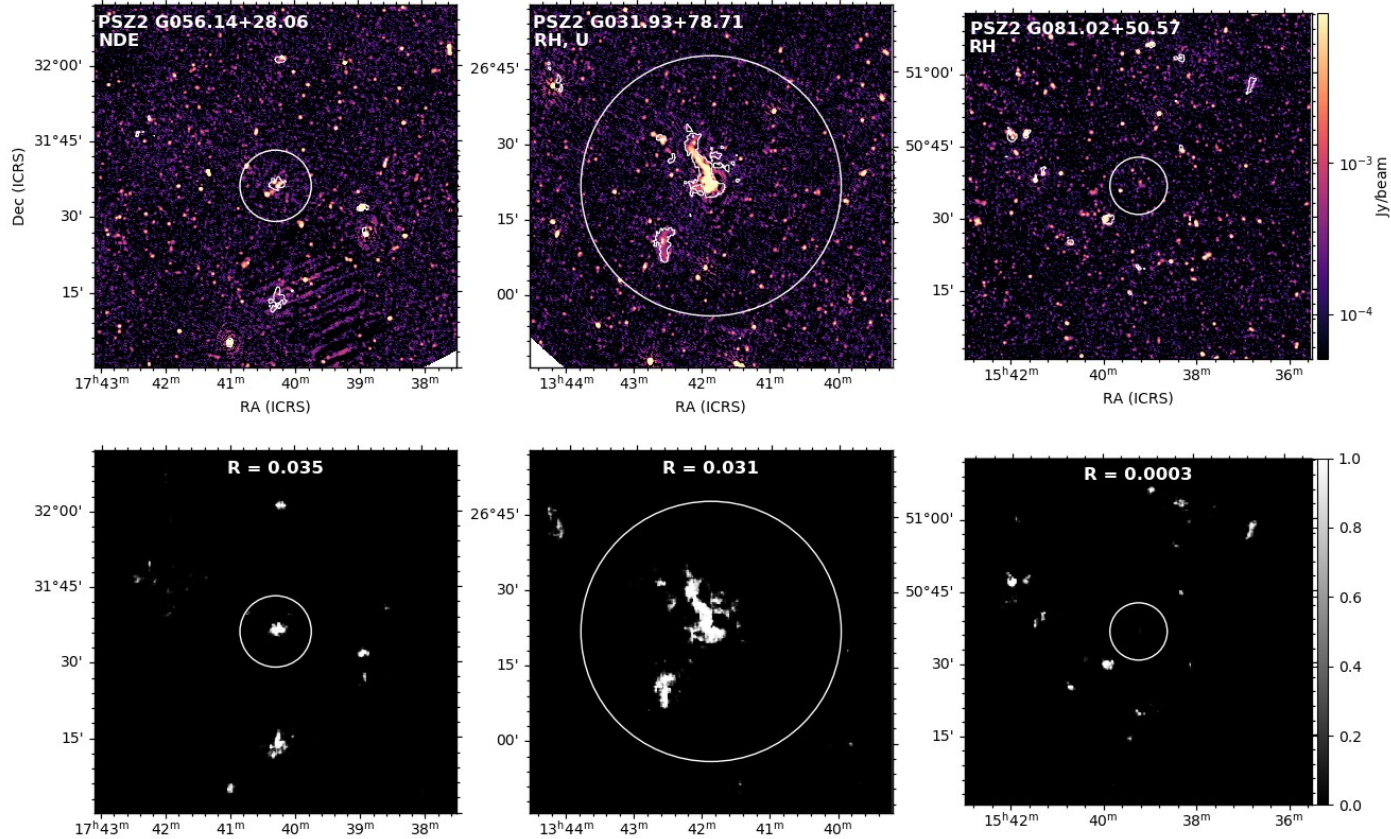
Detection of diffuse radio emission
beyond galaxy clusters



X-ray (XMM-Newton)
+radio contours
(LOFAR HBA) from
sources subtracted
image
+NED galaxy clusters

[Stuardi+ in prep.]

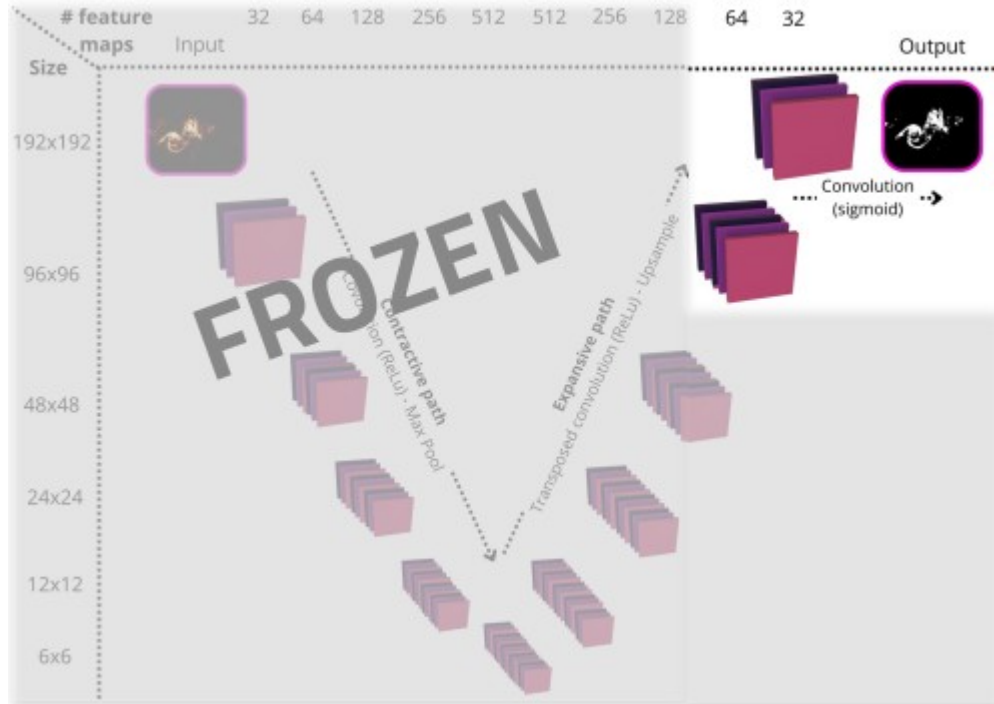
False positive and false negative



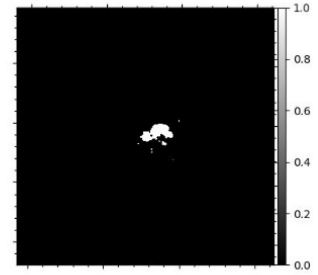
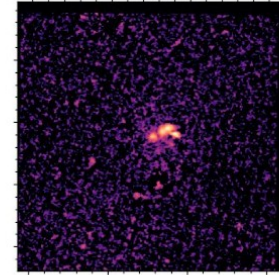
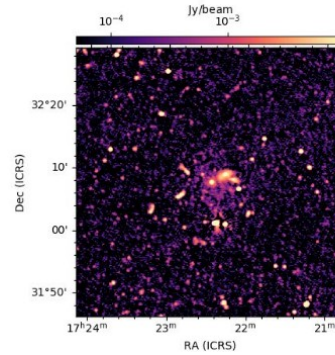
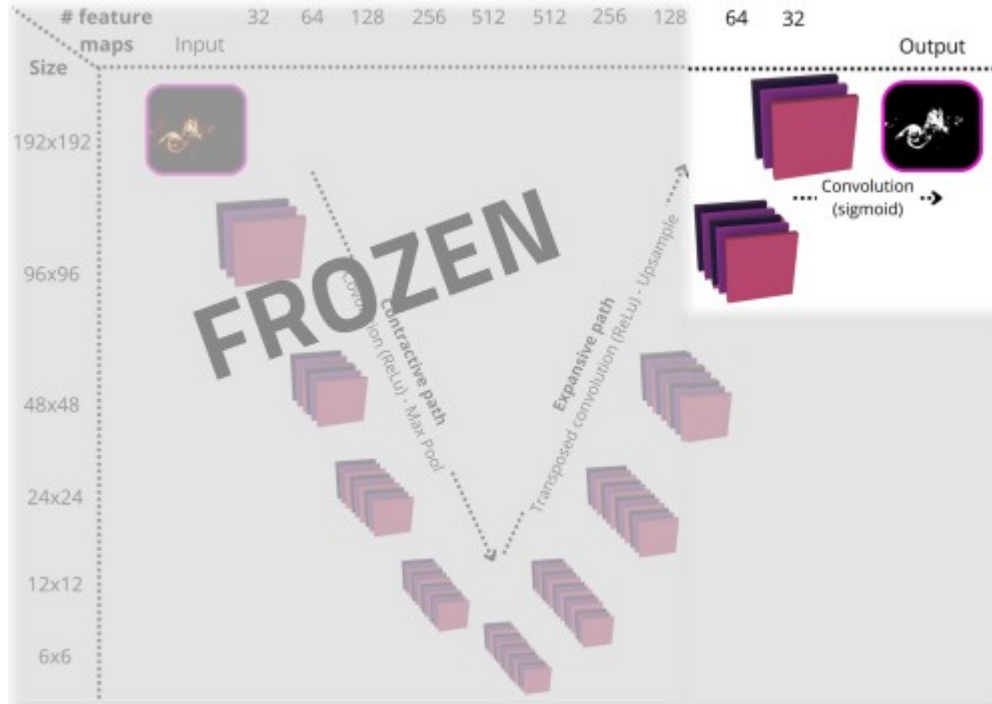
False positive mostly due to galaxy over-density and/or extended radio galaxies

False negative for high redshift or compact sources

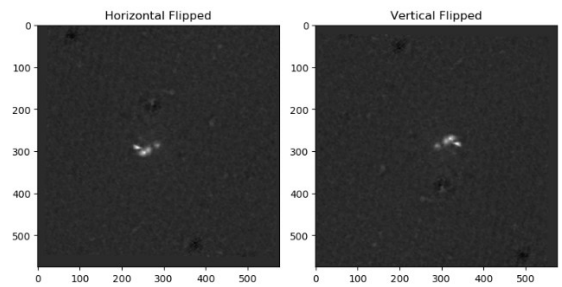
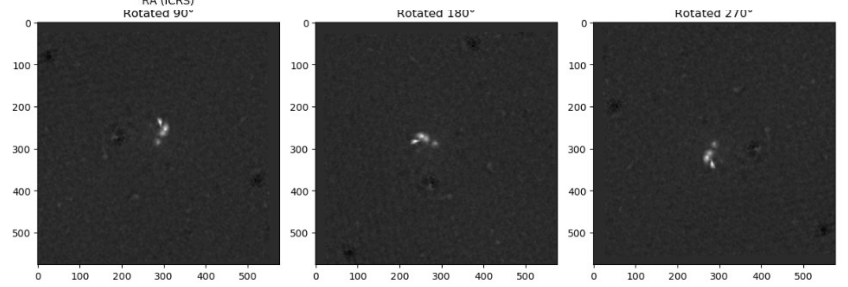
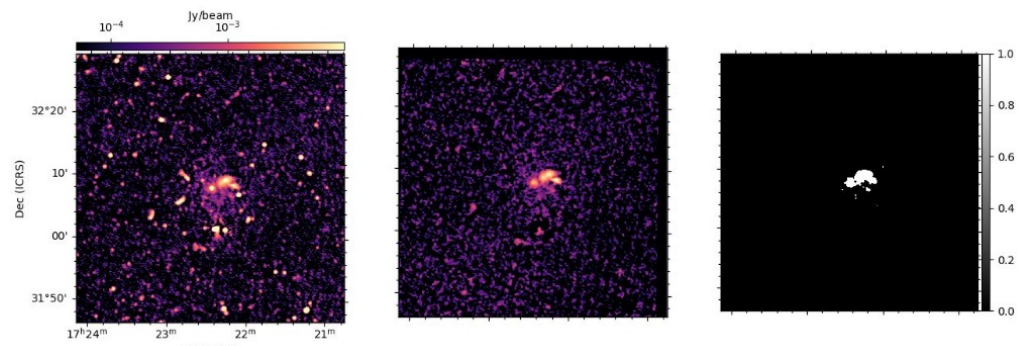
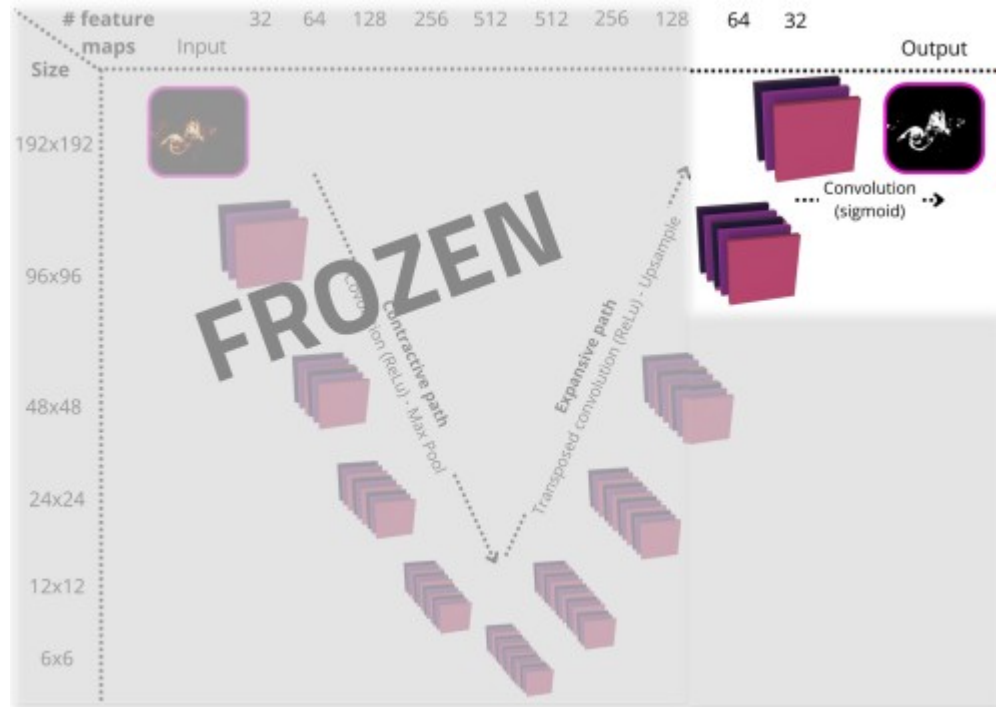
Data augmentation and fine tuning



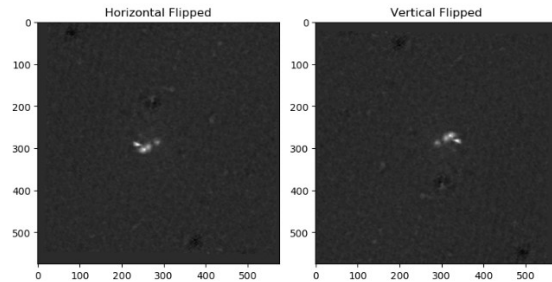
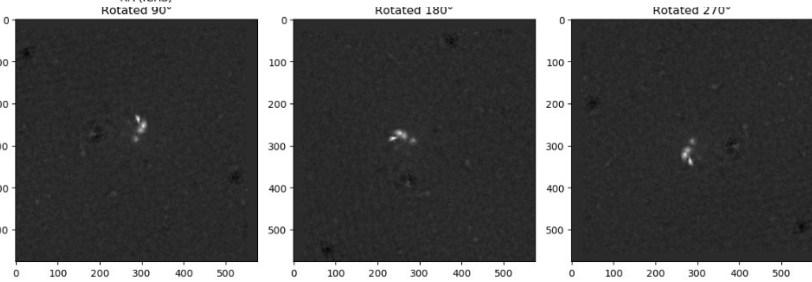
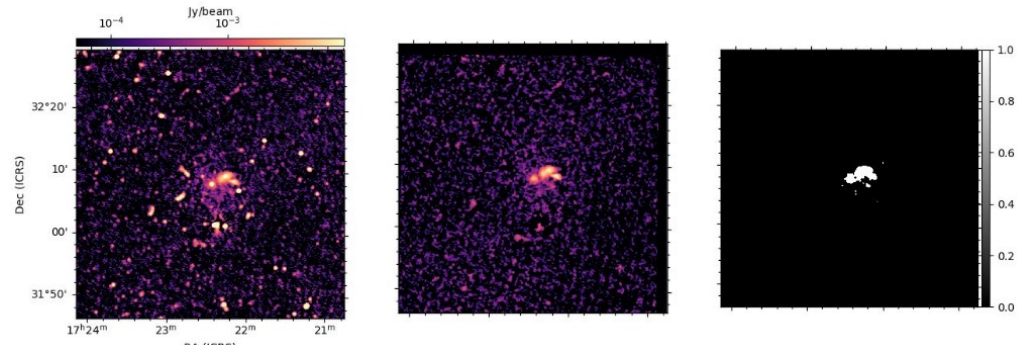
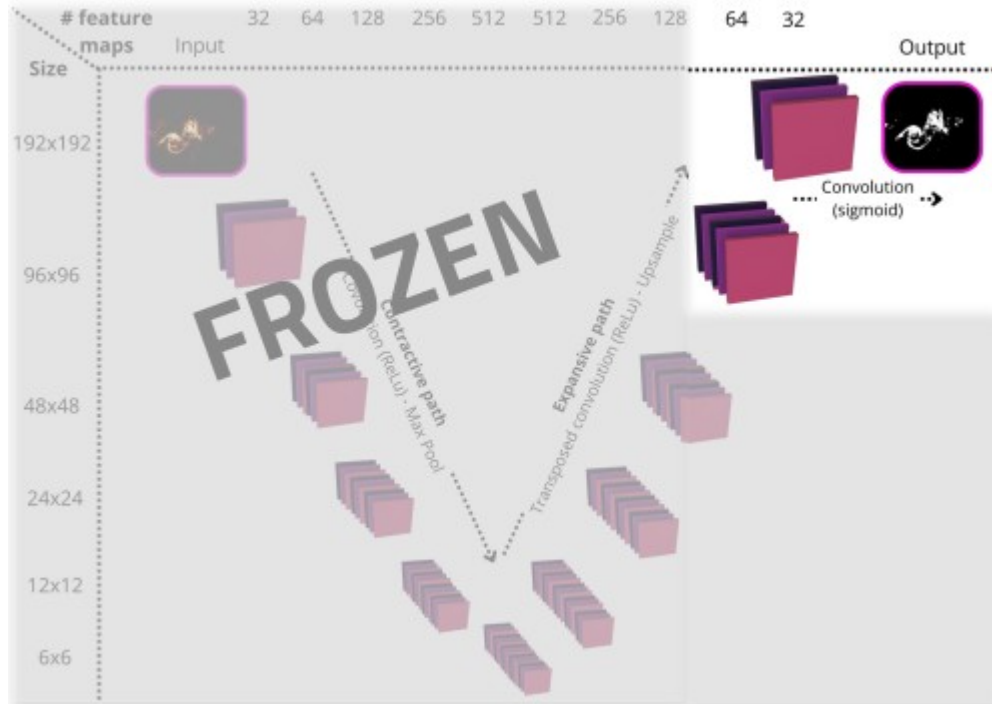
Data augmentation and fine tuning



Data augmentation and fine tuning



Data augmentation and fine tuning



Result: higher recall (87%)
but similar accuracy (72%)
With fine tuning on small subset
(78 DE galaxy clusters)

Radio U-Net: final remarks and next steps

- Automated and fast segmentation of diffuse radio sources in large surveys
- 73% accuracy on a balanced data-set, 83% recall
- Successful on low-quality images
[Stuardi et al. MNRAS, under minor revisions]
- Follow-up paper with new detection of galaxy cluster chain
- Use for the LoTSS DR3 under discussion

Radio U-Net: final remarks and next steps

- Automated and fast segmentation of diffuse radio sources in large surveys
- 73% accuracy on a balanced data-set, 83% recall
- Successful on low-quality images
[Stuardi et al. MNRAS, under minor revisions]
- Follow-up paper with new detection of galaxy cluster chain
- Use for the LoTSS DR3 under discussion

Thank you for your attention!