Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

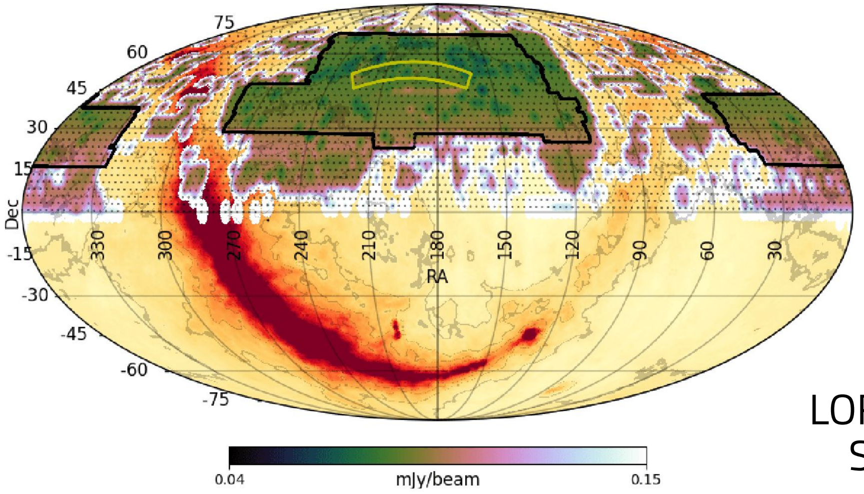
Radio U-Net: *the convolutional neural network for diffuse radio sources detection*

Chiara Stuardi
(IRA Bologna)
Claudio Gheller, Franco Vazza, Andrea Botteon

Spoke 3 General Meeting, Elba 5-9 / 05, 2024

Scientific Rationale

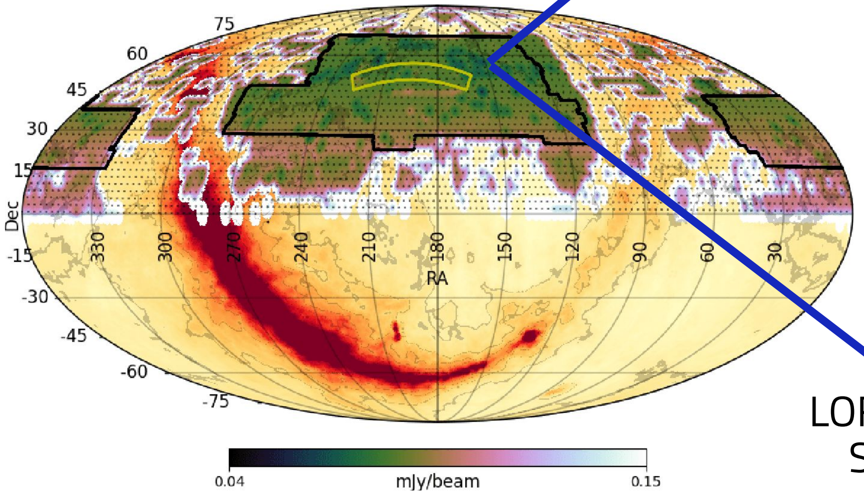
Current radio surveys are challenging our detection and cataloging strategies
-> new strategies to minimize human intervention in the data processing

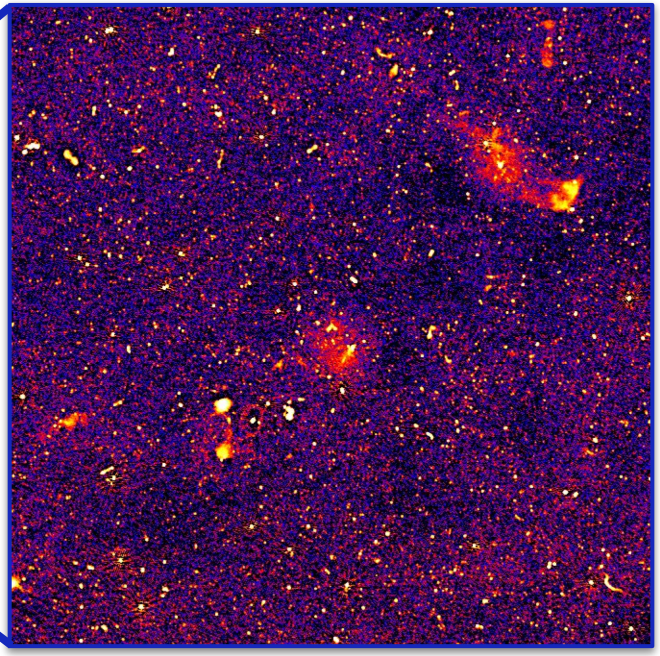


LOFAR Two Metre Sky Survey DR2 (LOTSS, Shimwell+22)

Scientific Rationale

Current radio surveys are challenging our detection and cataloging strategies
-> new strategies to minimize human intervention in the data processing

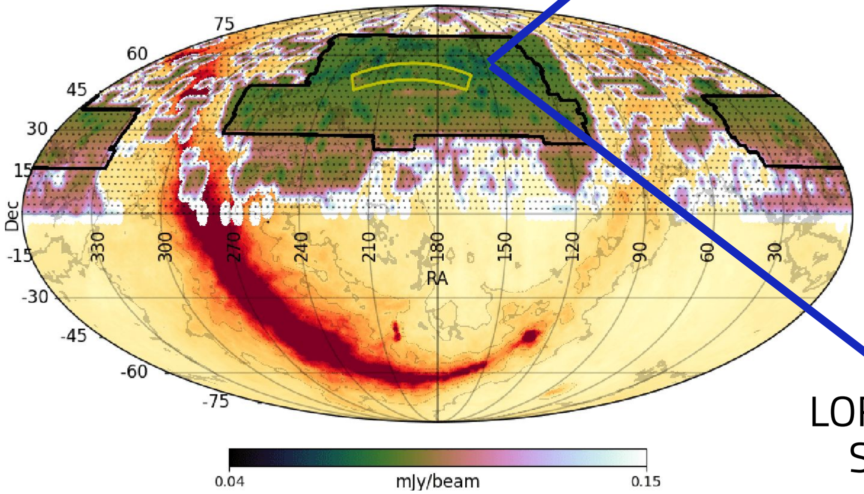


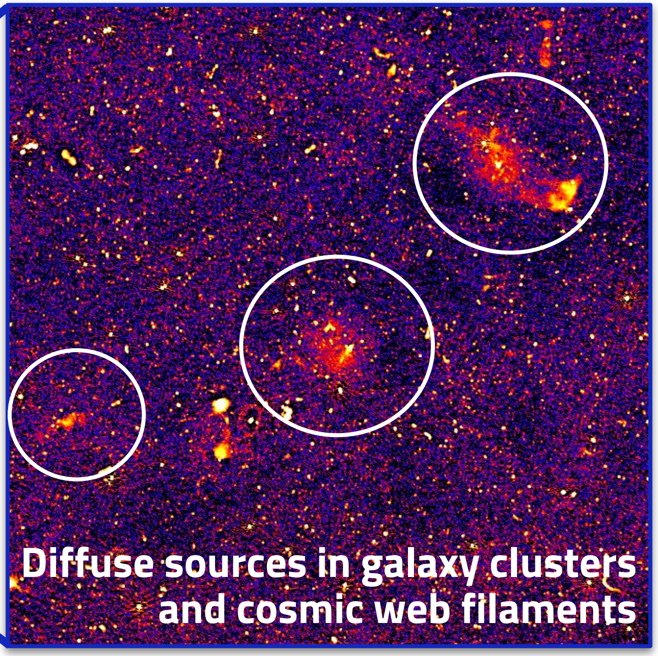


LOFAR Two Metre Sky Survey DR2 (LOTSS, Shimwell+22)

Scientific Rationale

Current radio surveys are challenging our detection and cataloging strategies
-> new strategies to minimize human intervention in the data processing

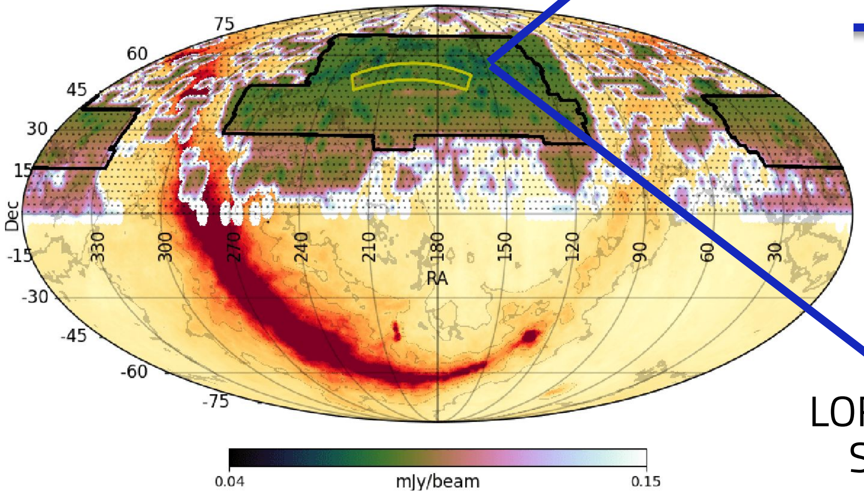


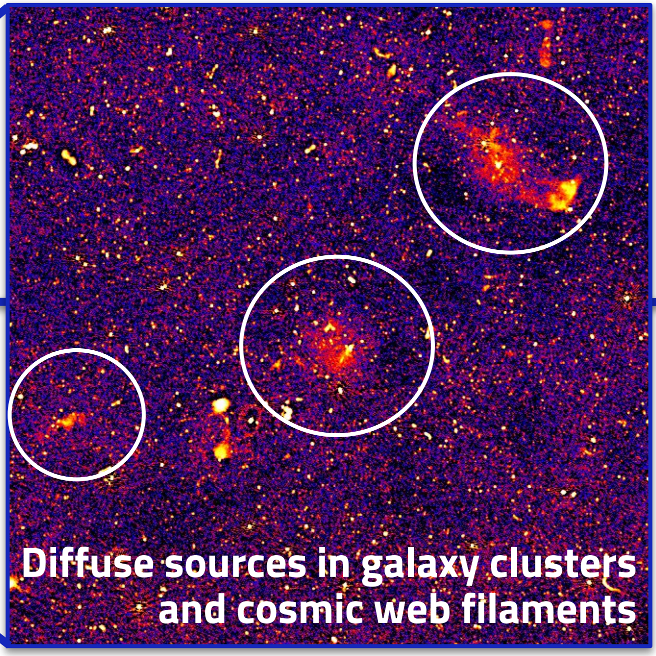


LOFAR Two Metre Sky Survey DR2 (LOTSS, Shimwell+22)

Scientific Rationale

Current radio surveys are challenging our detection and cataloging strategies
-> new strategies to minimize human intervention in the data processing

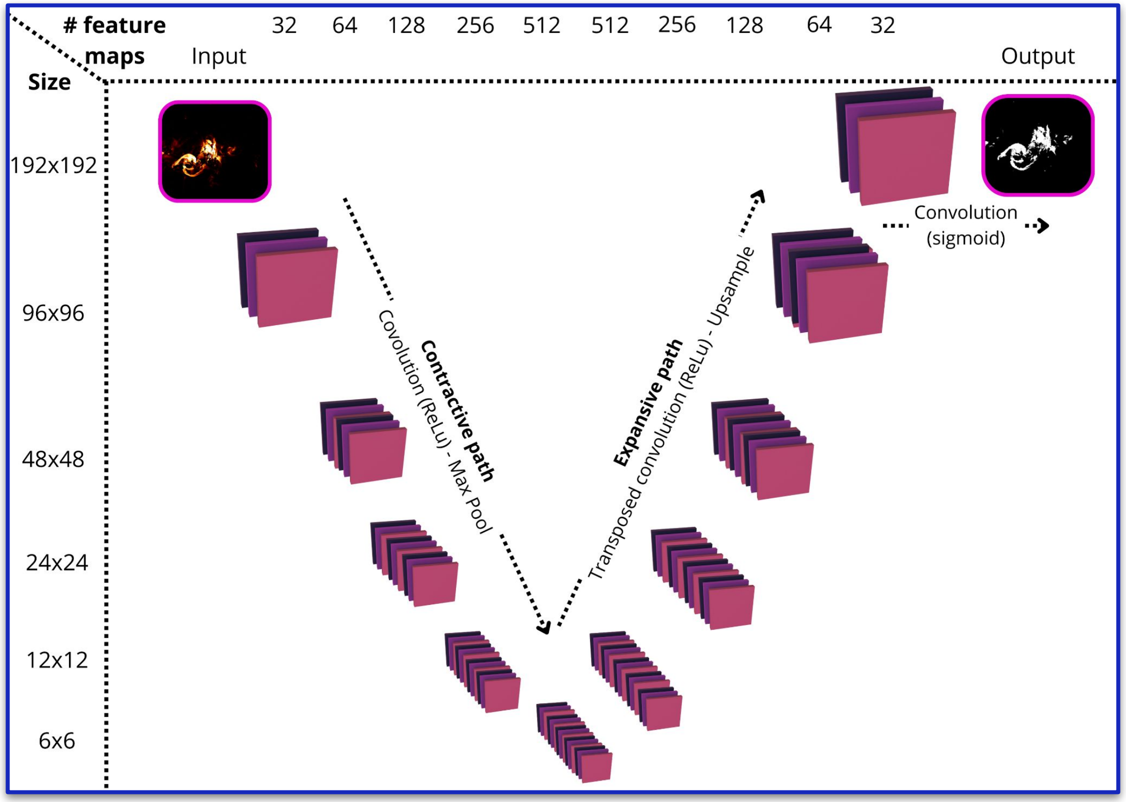




LOFAR Two Metre Sky Survey DR2 (LOTSS, Shimwell+22)

Technical Objectives, Methodologies and Solutions

Development of a Convolutional Neural Network (CNN) to perform the automated segmentation of diffuse radio emission in radio astronomical surveys: Radio U-Net, based on the U-Net architecture.

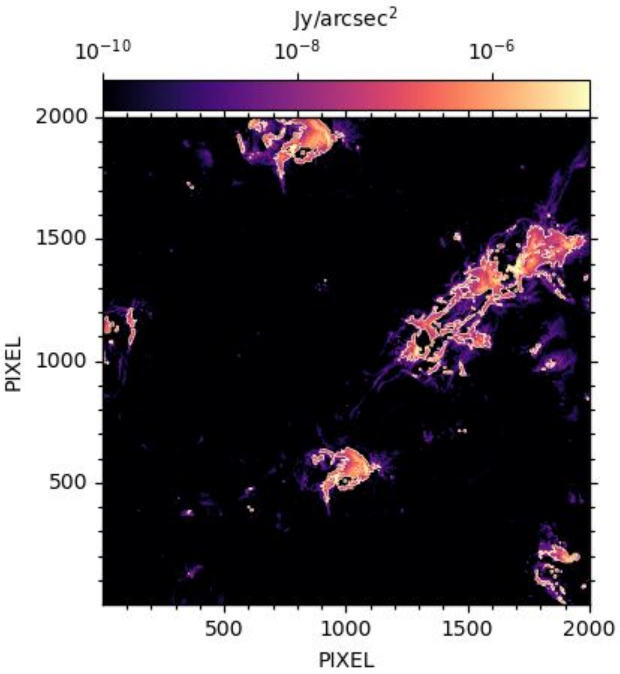


Python code with Keras Tensorflow

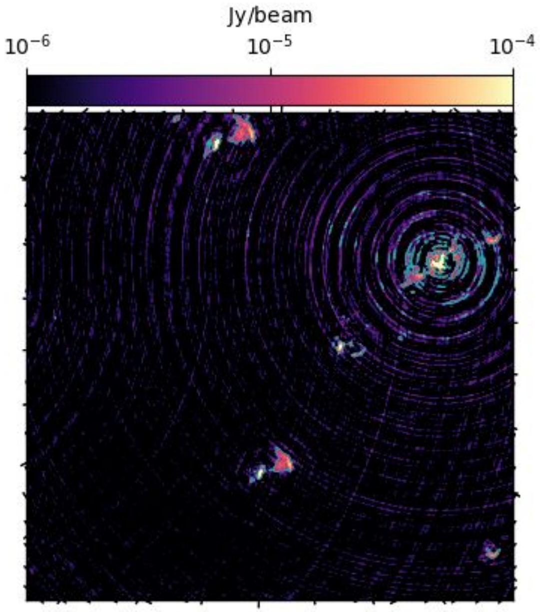
Scalable from CPU to GPU
Run on CINECA HPC systems (Leonardo)

Technical Objectives, Methodologies and Solutions

Problem: small number of reference images for the training
Solution: train the network on synthetic observations built on cosmological simulations
(Gheller&Vazza 2022)



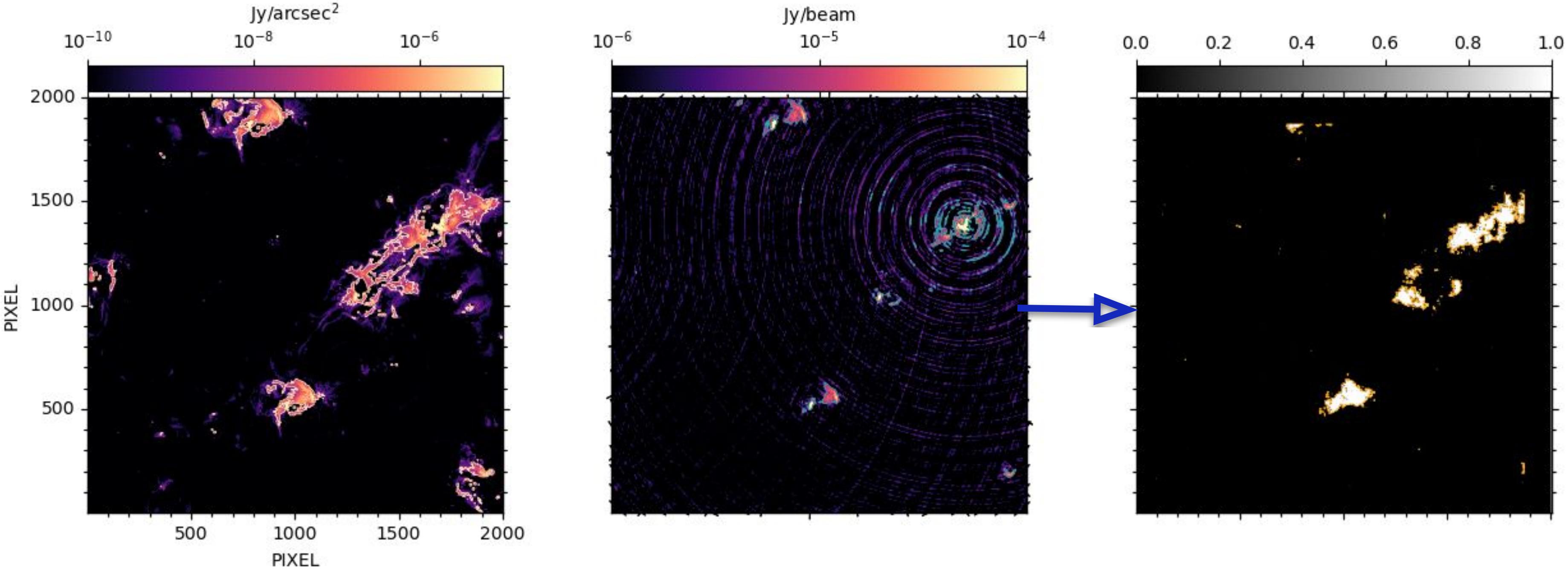
Sky image - Reference mask



Clean image - Input

Technical Objectives, Methodologies and Solutions

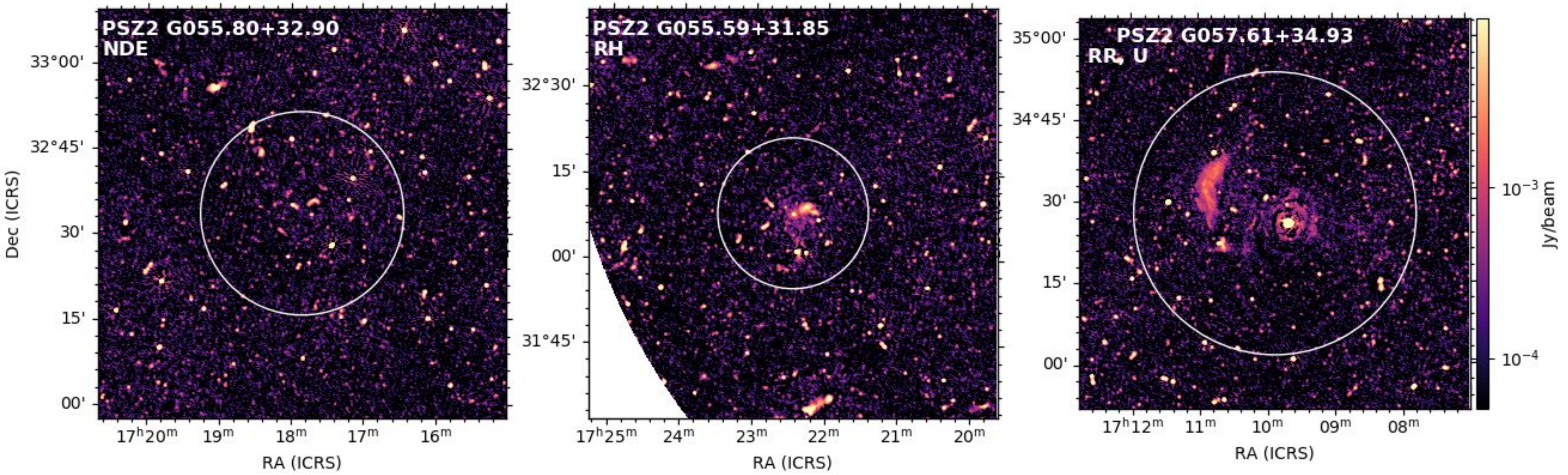
Problem: small number of reference images for the training
Solution: train the network on synthetic observations built on cosmological simulations
(Gheller&Vazza 2022)



Sky image - Reference mask **Clean image - Input** **Output probability**

Technical Objectives, Methodologies and Solutions

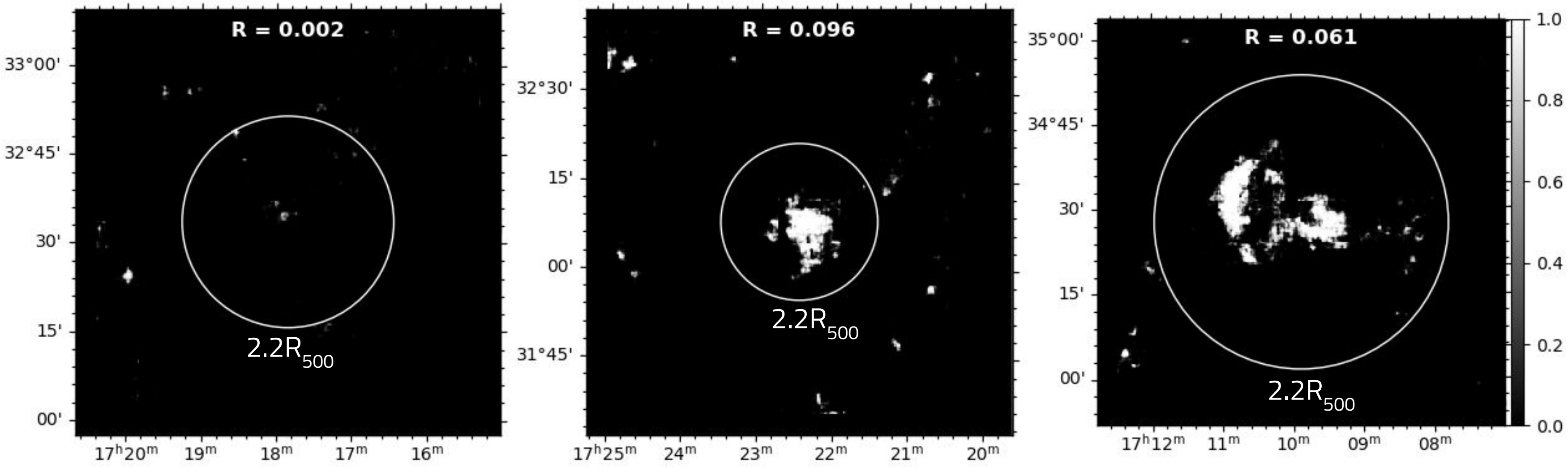
Application: apply Radio U-Net on LOFAR Two Metre Sky Survey Data
-> PSZ2-LoTSS2 sample of 309 galaxy clusters (Botteon+ 2023)



Images directly downloaded from the survey archive, without any tailored processing

Technical Objectives, Methodologies and Solutions

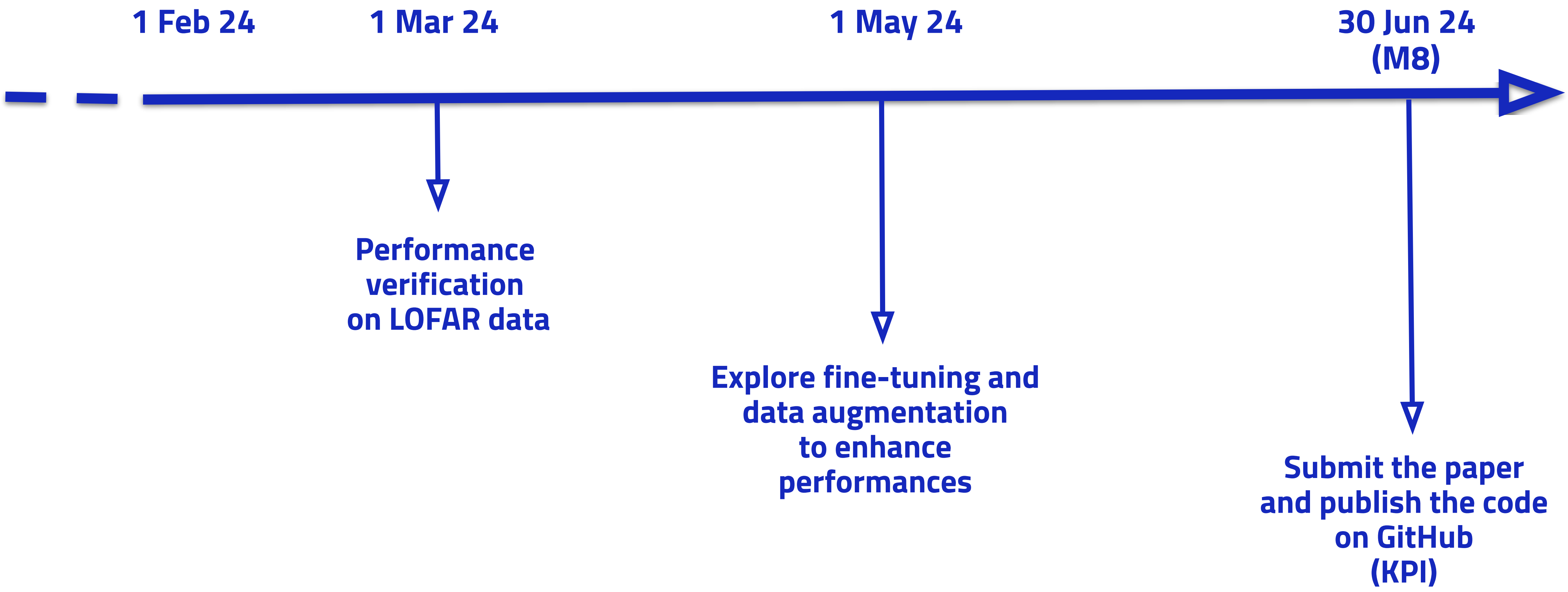
Application: apply Radio U-Net on LOFAR Two Metre Sky Survey Data -> PSZ2-LoTSS2 sample of 309 galaxy clusters (Botteon+ 2023)



Detection ratio

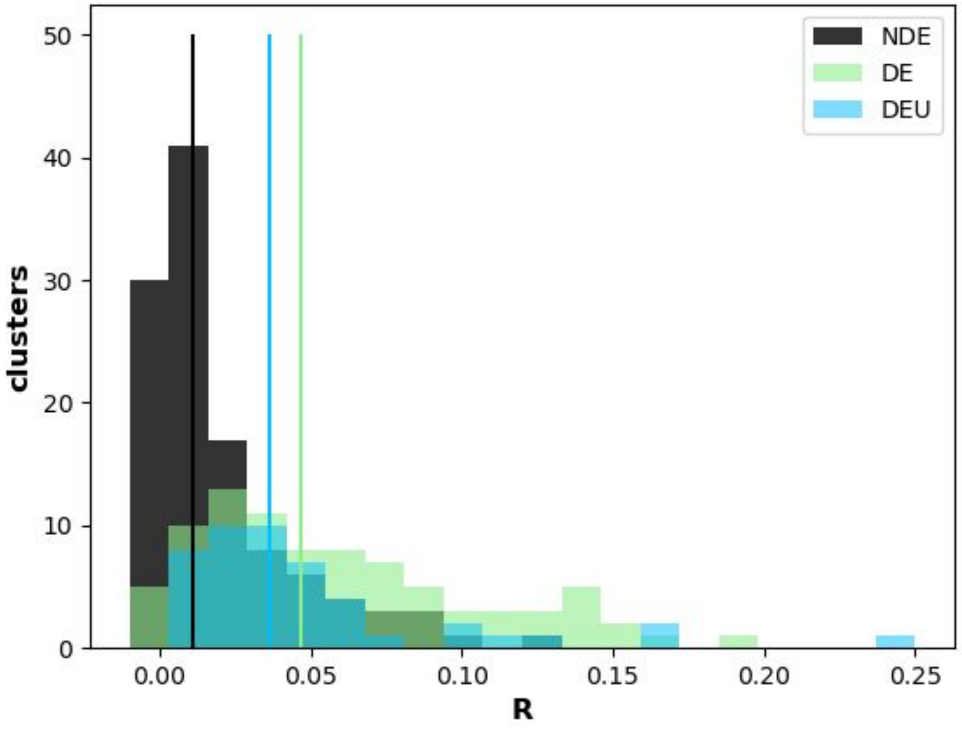
$R = (\text{sum probability}/\text{number of pixels})_{2.2R500}$

Timescale, Milestones and KPIs



Accomplished Work, Results

Performance verification on LOFAR data

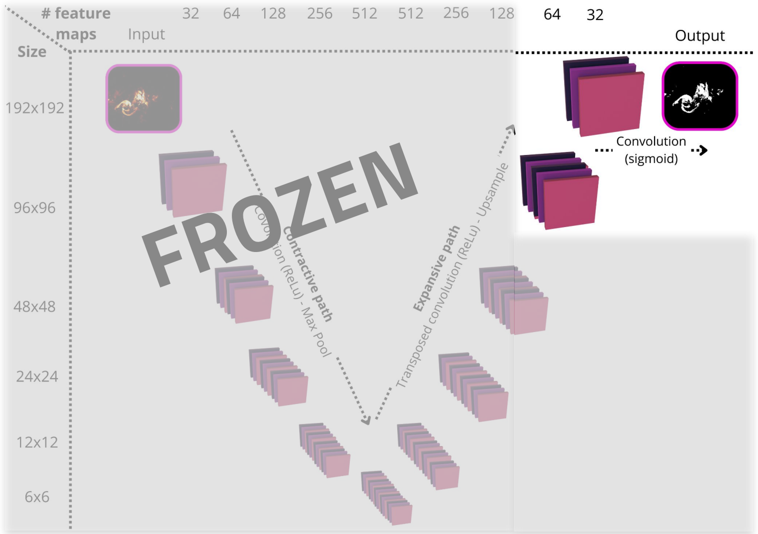


	non-detected NDE	detected DE	uncertain DEU	tot
Initial test set	114	85	47	246
$\mathrm{True}_{R=0.015}$	71 (62%)	70 (82%)	39 (83%)	180 (73%)

Result: detection accuracy 73% (how many are correctly classified?)
precision 72% (how many detections are true?)
false negative 17% (how many sources are missed?)

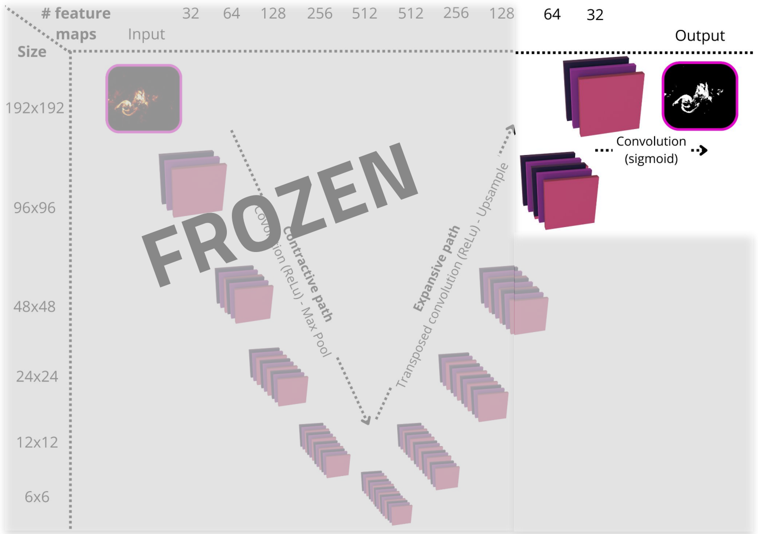
Accomplished Work, Results

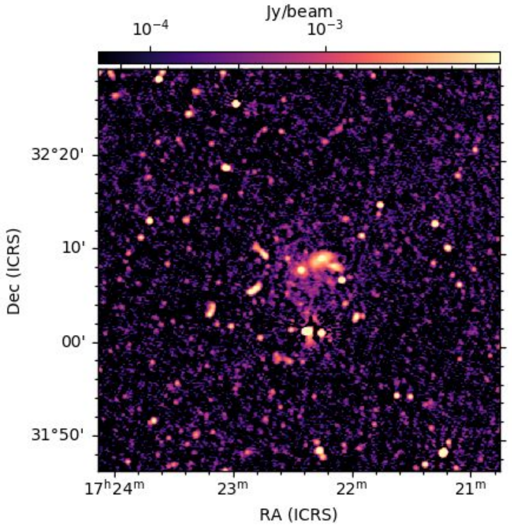
Explore fine-tuning using LOFAR data



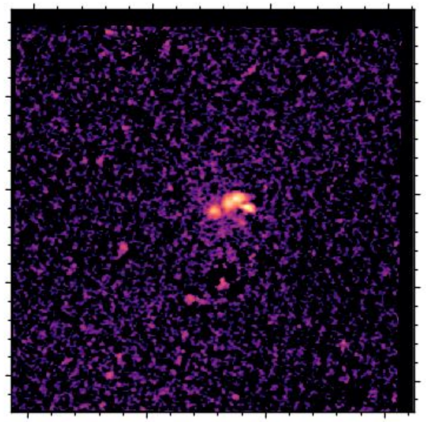
Accomplished Work, Results

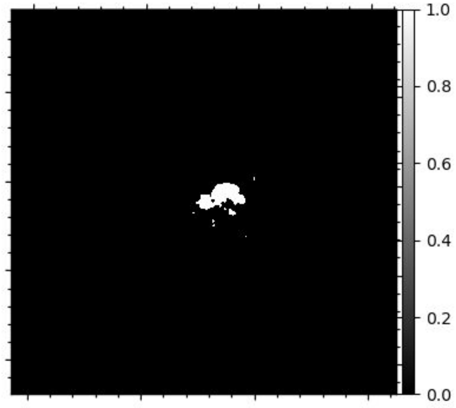
Explore fine-tuning using LOFAR data





Input image (DE)

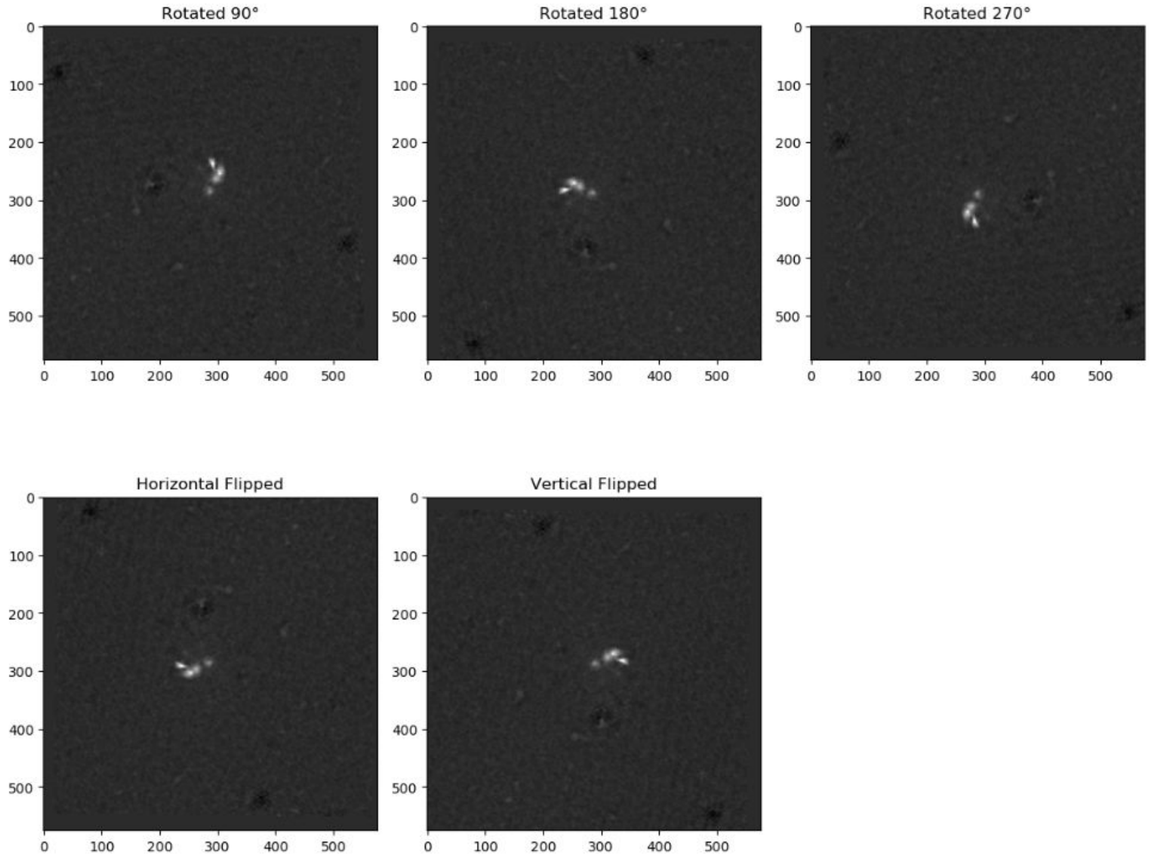




Reference mask

Accomplished Work, Results

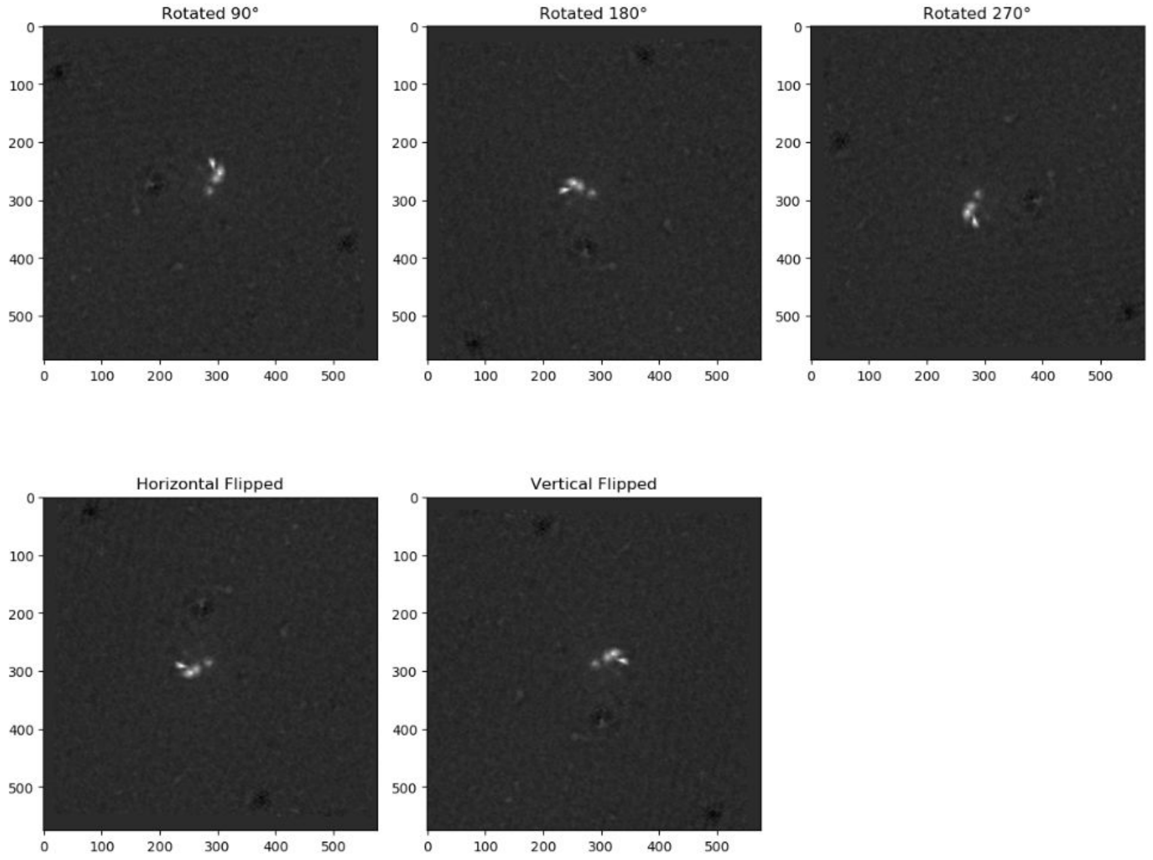
Explore fine-tuning using LOFAR data and data augmentation.



Test of several hyperparameters

Accomplished Work, Results

Explore fine-tuning using LOFAR data and data augmentation.



Test of several hyperparameters

	NDE	DEU	tot
Test set	54	47	108
$\text{True}_{\text{original}} \text{R}=0.026$	40 (74%)	34 (72%)	71%
$\text{True}_{\text{ft}} \text{R}=0.00084$	31 (57%)	42 (89%)	71%
$\text{True}_{\text{ft-da}} \text{R}=0.00081$	33 (61%)	41 (87%)	72%

Result: fine-tuning reduces false negative rate but does not increase the overall accuracy

Next Steps and Expected Results

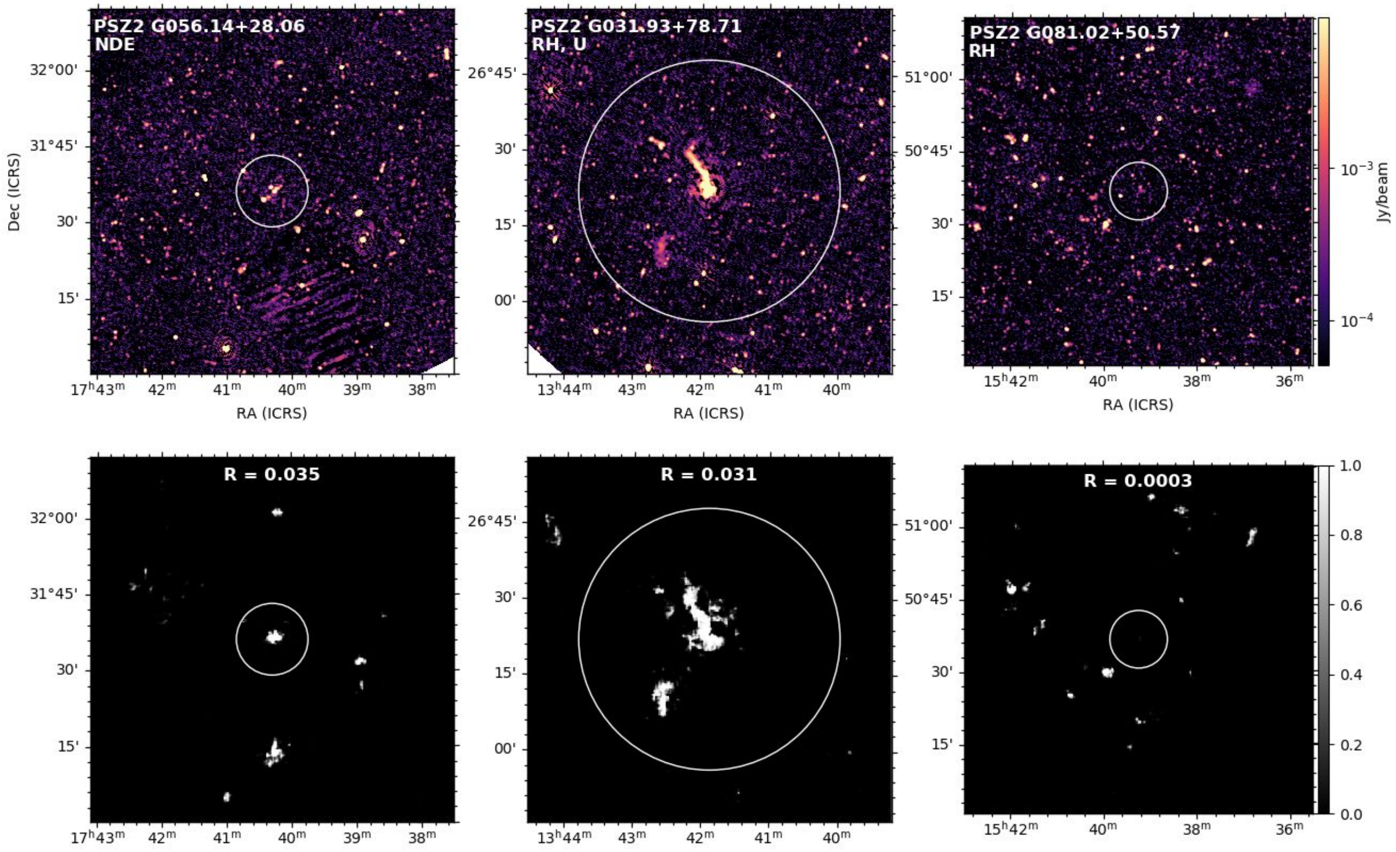
30 Jun 24 (M8) — M9 — M10

Finalize and submit the paper
Make the code publicly available

Further developments
- Include radio galaxies in synthetic observations for the training
- Apply to other radio surveys
- Add a classification layer to the network

Accomplished Work, Results

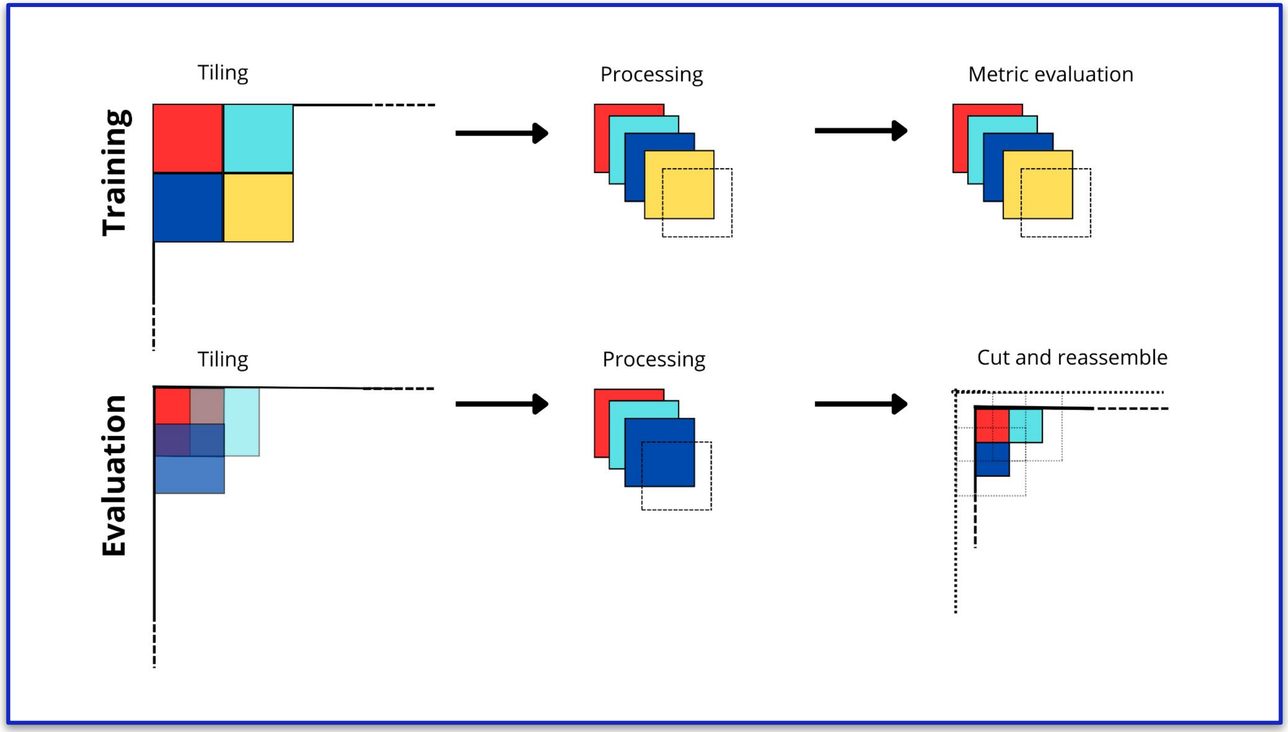
Performance verification on LOFAR data



False positive mainly caused by extended and/or grouped radio galaxies which are not present in simulations

Technical Objectives, Methodologies and Solutions

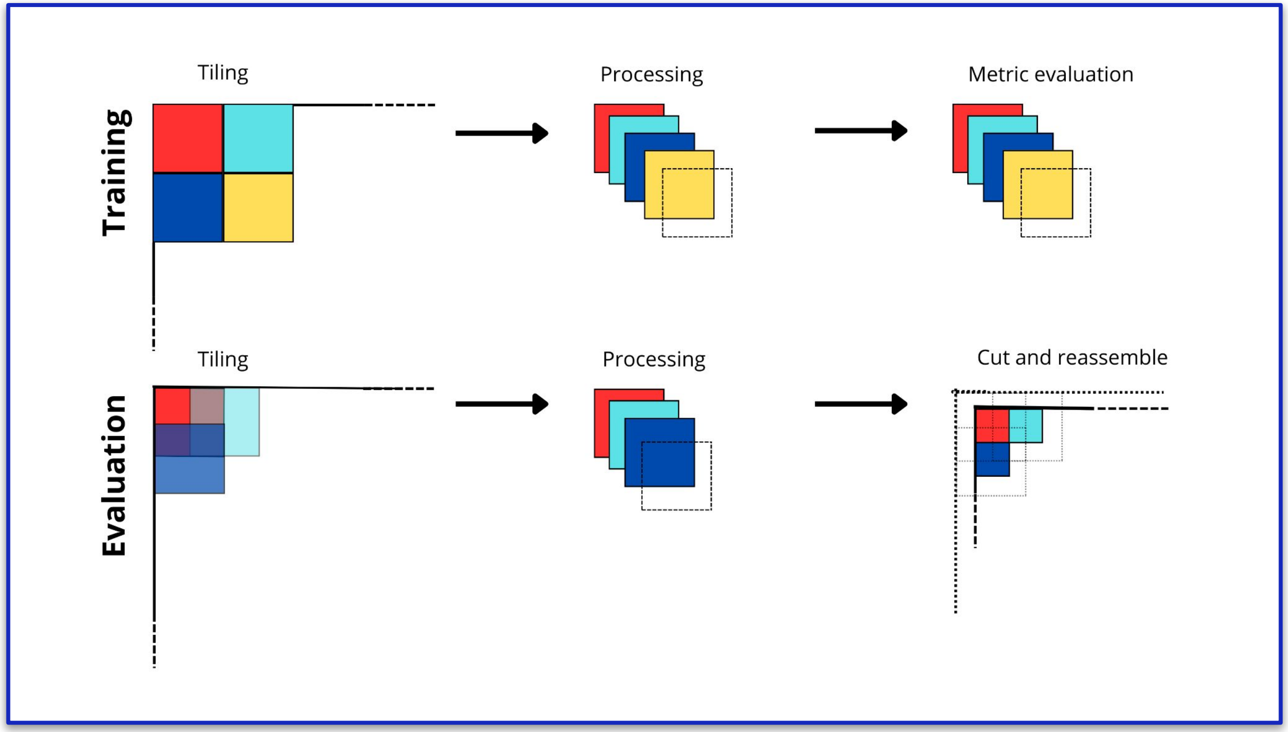
Problem: small number of labelled images for the training
Solution: train the network on synthetic observations built on cosmological simulations
+ tiling strategy: increase the number of individual images and exploit parallel processing



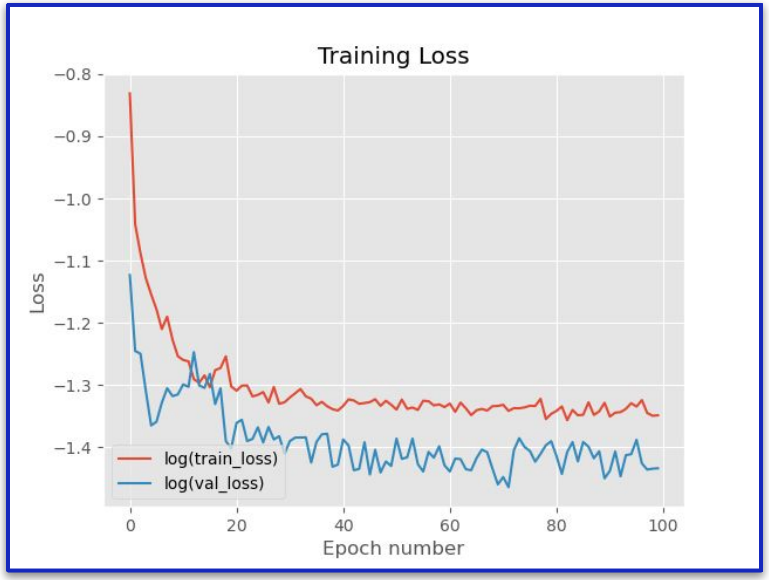
1000 images 2000x2000
100·000 tiles 192x192

Technical Objectives, Methodologies and Solutions

Problem: small number of labelled images for the training
Solution: train the network on synthetic observations built on cosmological simulations
+ tiling strategy: increase the number of individual images and exploit parallel processing

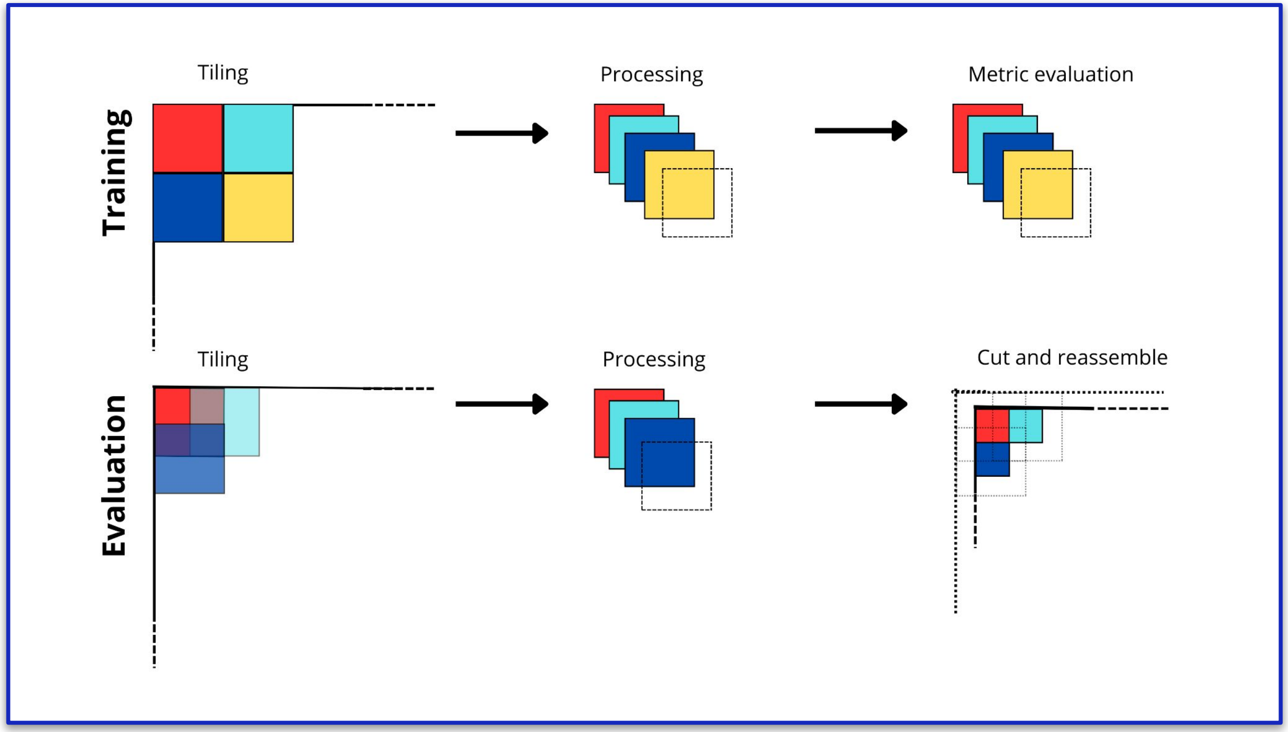


1000 images 2000x2000
100·000 tiles 192x192



Technical Objectives, Methodologies and Solutions

Problem: small number of labelled images for the training
Solution: train the network on synthetic observations built on cosmological simulations
+ tiling strategy: increase the number of individual images and exploit parallel processing



1000 images 2000x2000
100·000 tiles 192x192

Solving boundary effects by cropping overlapped tiles when re-assembling the output mask