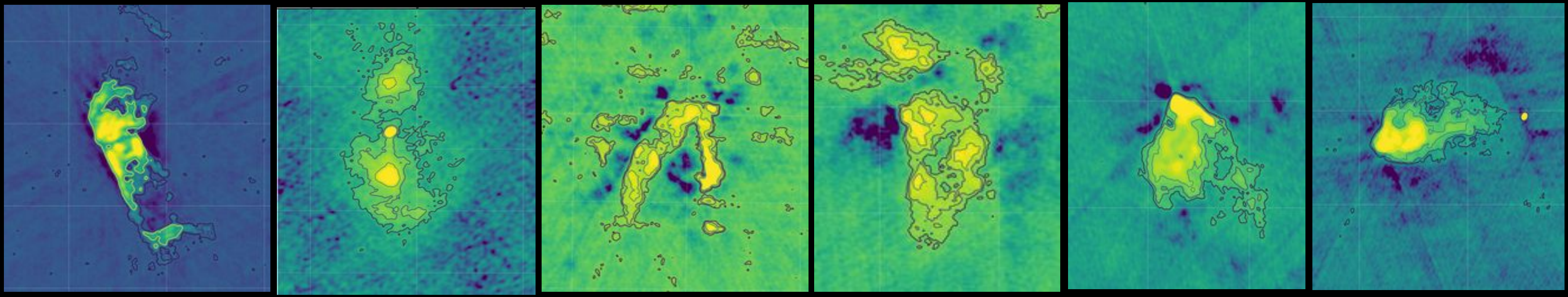
Anomaly Detection in the VLA Sky Survey (VLASS)



Viral Parekh

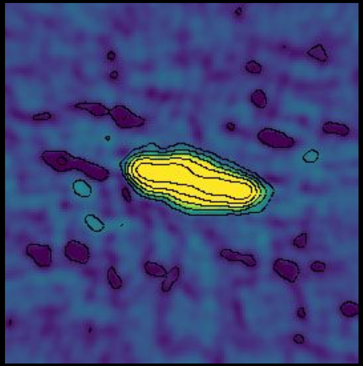
National Radio Astronomy Observatory
Socorro, NM

Acknowledgement to the VLASS team

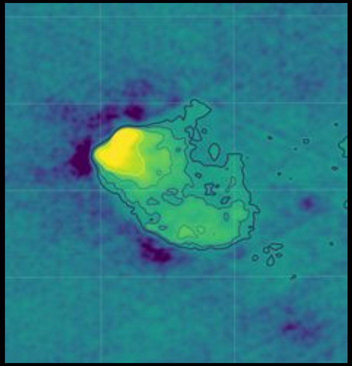
2nd Annual Cosmic Horizons Conference, July 15, 2026

VLA Sky Survey (VLASS)

- VLA S-band (2-4 GHz) wide-area sky survey (δ > -40deg, 33,885 deg^2 of sky)
- VLA B-configuration provides ~ 2” resolution (~1’ Largest angular scale)
- Multi-epoch observations (2017-2026)
- Single epoch rms ~ 120 μJy/beam; combined images rms ~ 70 μJy/beam
- Fully automated calibration and imaging pipeline
- Data products: Quick look, SE (self-cal) images, alpha (spx) maps, coarse cubes, catalogues, image combination, and image cutouts

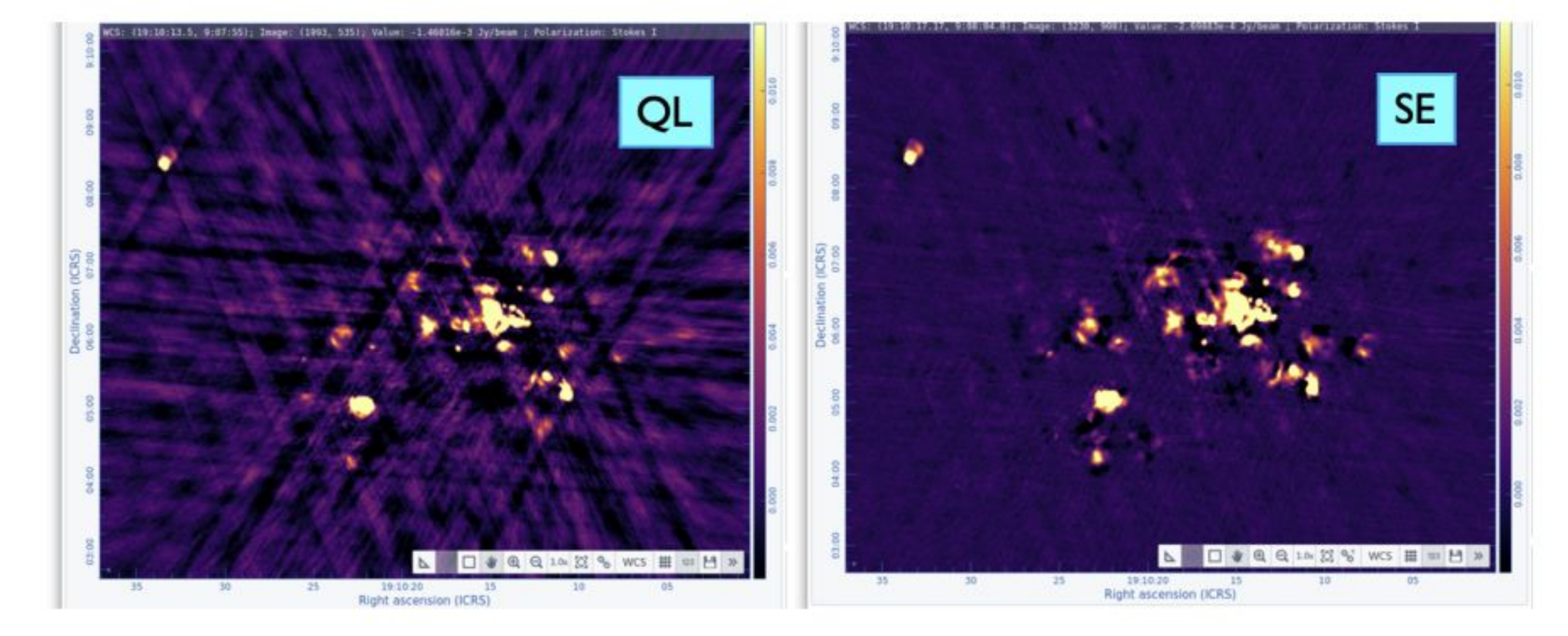


TGSS



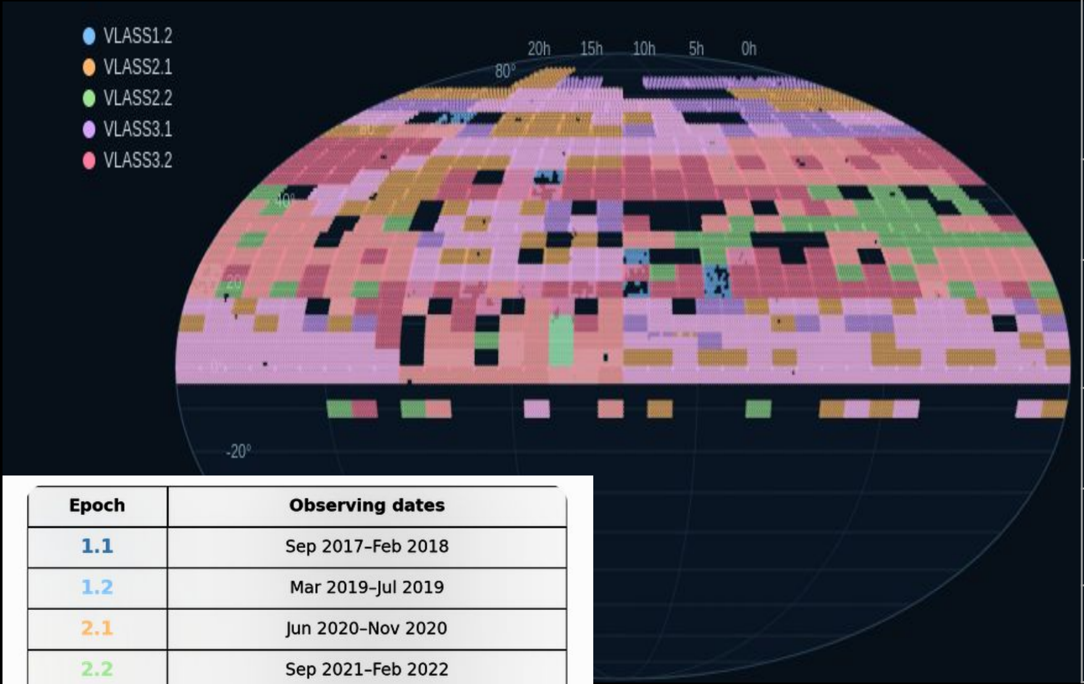
VLASS

QL vs SECI maps



(Amy Kimball et al. 2026)

VLA Sky Survey (VLASS) SECI Images



Epoch	Observing dates
1.1	Sep 2017–Feb 2018
1.2	Mar 2019–Jul 2019
2.1	Jun 2020–Nov 2020
2.2	Sep 2021–Feb 2022
3.1	Jan 2023–Jun 2023
3.2	Apr 2024–Oct 2024
4.1	Aug 2025–Feb 2026

VLASS Epoch	No. of images	Area (deg^2)	Area (sr)	% of survey area
1.2	3560	2795	0.85	8.24
2.1	9106	7151	2.17	21.10
2.2	6213	4879	1.48	14.39
3.1	7905	6208	1.89	18.31
3.2	8219	6455	1.96	19.04

Extended Radio Sources in the VLASS

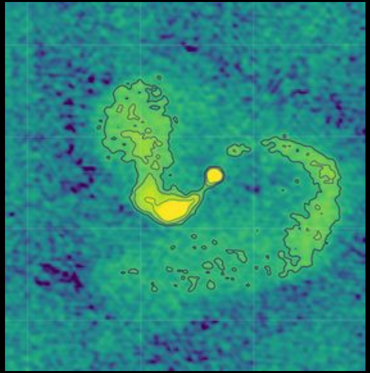
- Visual inspections of early QL maps (Hernandez and Andernach) of VLASS 1.1
  - Discovered 9 new giant radio galaxies of size > 1 Mpc

- DRAGNs in VLASS (Gordon et al. 2023): candidate double radio sources identified using the QL component catalogue
  - > 17,000 DRAGNs, 13,000 host galaxies, 31 new GRGs

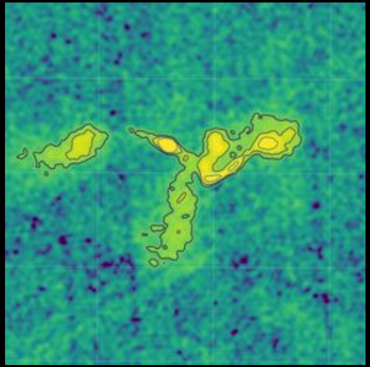
- Morphology clustering of VLASS QL sources using Self-Organizing Maps (SOMs) by the CIRADA team

What is an anomaly?

- Non-Gaussian, elongated, complex, or irregular radio morphology
- Flexible across multiple source morphologies rather than targeting only one source class
- **Goal:** prioritize extended and morphologically unusual radio sources while suppressing compact sources, artifacts, and sidelobe structures
- **Benefit:** reduces the volume of data requiring expert visual inspection while preserving candidates with high scientific discovery potential



VLASSJ0615+5158

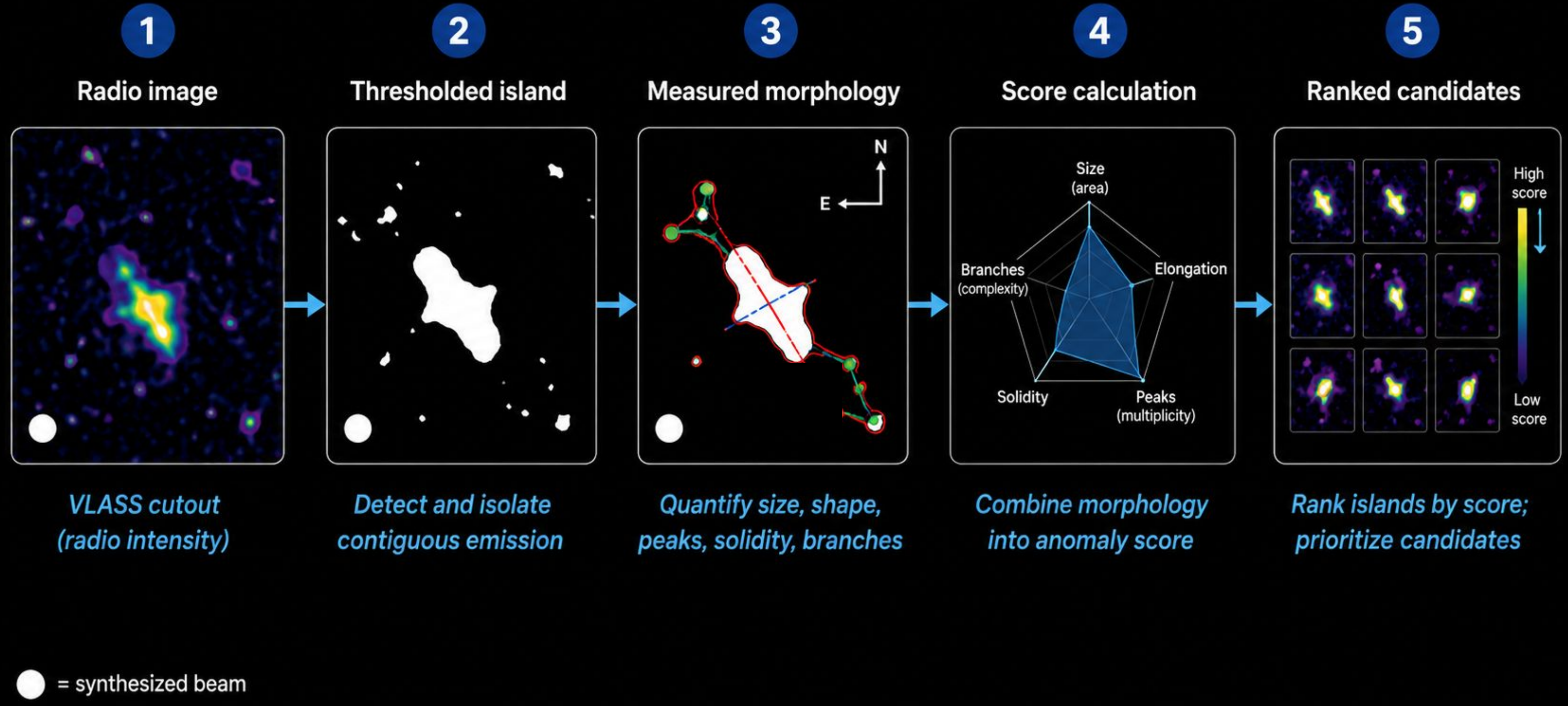


VLASSJ0707+3503

Anomaly detection in VLASS data

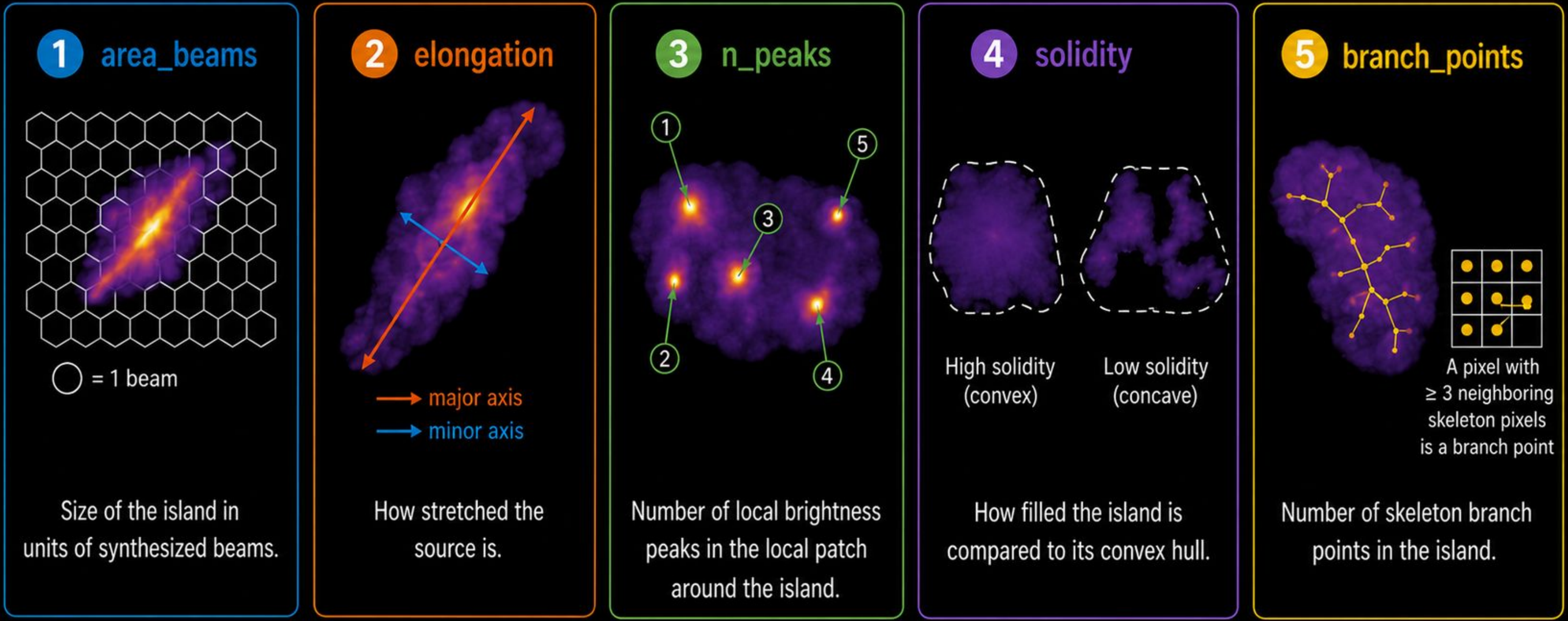
- Applied to VLASS 2.1 and 2.2 SECI images (9106 and 6213, respectively)
- Radio-source morphology-based anomaly detection
- Detect radio-source islands above a defined threshold
- Measure source morphology and structural complexity
- Assign an anomaly score to every detected candidate, including real sources and artifacts
- Use local background diagnostics to identify sidelobes and artifacts
- Build a sample of extended and unusual radio sources for future ML classification

Method



1
Radio image
2
Thresholded island
3
Measured morphology
N
E
4
Score calculation
Size
(area)
Elongation
Peaks
(multiplicity)
Solidity
Branches
(complexity)
5
Ranked candidates
High
score
Low
score
VLASS cutout
(radio intensity)
Detect and isolate
contiguous emission
Quantify size, shape,
peaks, solidity, branches
Combine morphology
into anomaly score
Rank islands by score;
prioritize candidates
= synthesized beam

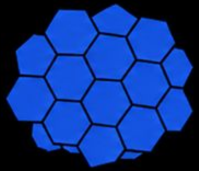
Morphology parameters



How the score is calculated

$$\text{score} = S_{\text{area}} + S_{\text{elongation}} + S_{\text{peaks}} + S_{\text{solidity}} + S_{\text{branches}}$$

size

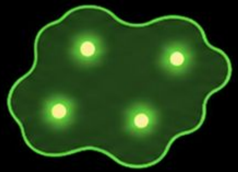


How large the island is

shape

How stretched or elongated the island is

multiple peaks



How many local brightness peaks it contains

irregularity

How irregular or non-convex the island shape is

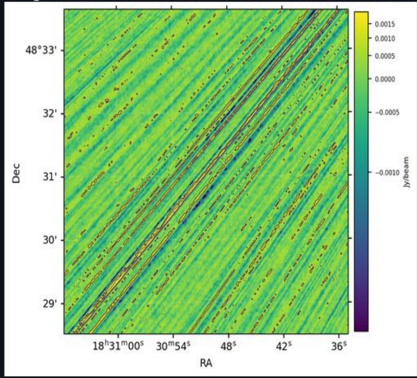
complexity

How complex the skeleton structure is

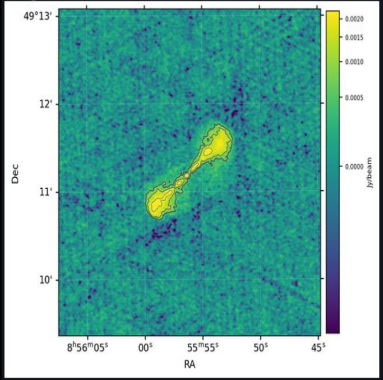
Weights set importance;
caps prevent one parameter from dominating.

Rejecting artifacts and sidelobes

VLASSJ1830+4831
RA,Dec: 277.709717, 48.517760 (18:30:50.33 +48:31:03.9)
Peak RA,Dec: 277.739458, 48.496685 (18:30:57.47 +48:29:48.1)
SIMBAD @ object: No match **Type**: NA **Sep**: nan arcsec **z**: nan **Radio flux**: nan
NED @ object: No match **Type**: NA **Sep**: nan arcsec **z**: nan **Radio flux**: nan
SIMBAD @ peak: No match **Type**: NA **Sep**: nan arcsec **z**: nan **Radio flux**: nan
NED @ peak: WISEA J183057.89+482951.1 **Type**: UvS **Sep**: 5.25 arcsec **z**: nan **Radio flux**: 2.84e-08 Jy
Flux: **Int**=0.3281 ± 0.0166 Jy **Peak**=0.003229 Jy/beam **Min**=-0.001908 **Std**=0.0004545 **RMS**=0.0004545
Score: 6411.34 (orig 6411.34) **peakSNR**: 21.3 **area(beams)**: 228.46 **elong**: 146.38 **peaks**: 1585 **solidity**: 0.66 **branches**: 57
bgAlign: 1.00 **bgAniso**: 0.92 **negSide**: 0.620 **ridgeStraight**: 1.00
Image: VLASS2.1.se.T23t19.J183252+483000.06.2048.v1.I.iter3.image.pbcor.tt0.subim.fits



VLASSJ0855+4911
RA,Dec: 133.983713, 49.187000 (08:55:56.09 +49:11:13.2)
Peak RA,Dec: 133.984886, 49.186395 (08:55:56.37 +49:11:11.0)
SIMBAD @ object: NVSS J085556+491112 **Type**: Radio **Sep**: 0.73 arcsec **z**: nan **Radio flux**: nan
NED @ object: No match **Type**: NA **Sep**: nan arcsec **z**: nan **Radio flux**: nan
SIMBAD @ peak: SDSS J085556.36+491110.6 **Type**: RadioG **Sep**: 0.32 arcsec **z**: nan **Radio flux**: nan
NED @ peak: No match **Type**: NA **Sep**: nan arcsec **z**: nan **Radio flux**: nan
Flux: **Int**=0.2304 ± 0.0116 Jy **Peak**=0.005139 Jy/beam **Min**=-0.0004731 **Std**=0.0001184 **RMS**=0.0001184
Score: 89.21 (orig 151.92) **peakSNR**: 46.5 **area(beams)**: 169.18 **elong**: 5.53 **peaks**: 22 **solidity**: 0.59 **branches**: 38
bgAlign: 0.87 **bgAniso**: 0.20 **negSide**: 0.002 **ridgeStraight**: 0.49
Image: VLASS2.2.se.T23t09.J085726+493000.06.2048.v1.I.iter3.image.pbcor.tt0.subim.fits

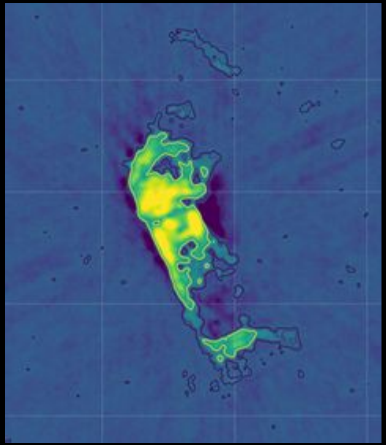


Example candidate-retention cuts: neg_sidelobe < 0.20; elongation < 14; n_peaks < 50

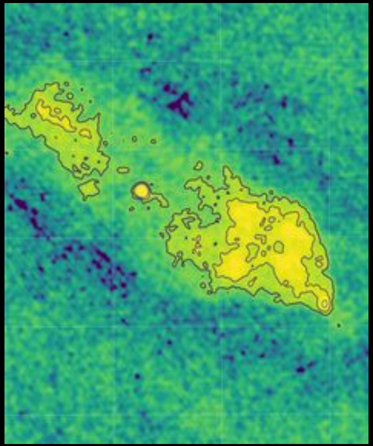
Anomaly data volume

- VLASS 2.1: 26,158 initial cutouts → 8,413 candidates after filtering, corresponding to a **67.8% reduction** in visual-inspection volume.
- VLASS 2.2: 22,115 initial cutouts → 6,875 candidates after filtering, corresponding to a **68.9% reduction** in visual-inspection volume.

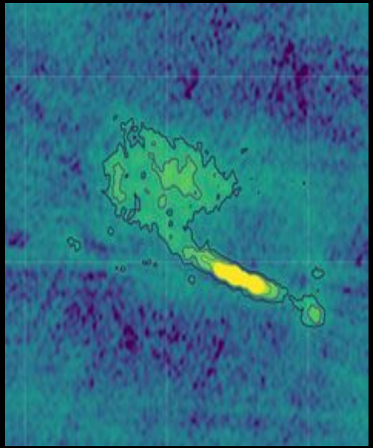
Results: Top-ranked complex radio sources



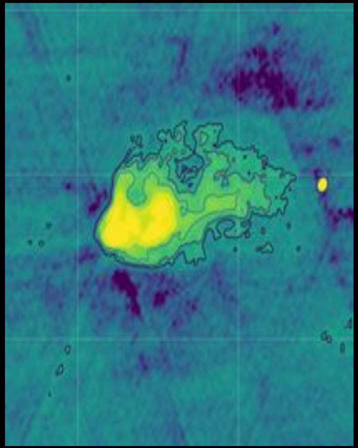
VLASSJ2027+3722
Score 424



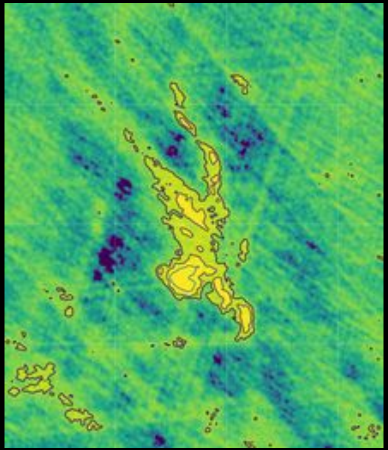
VLASSJ0207+2930
Score 419



VLASSJ1724-0242
Score 404



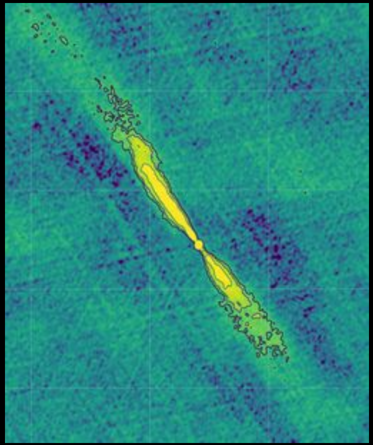
VLASSJ1602+0157
Score 395



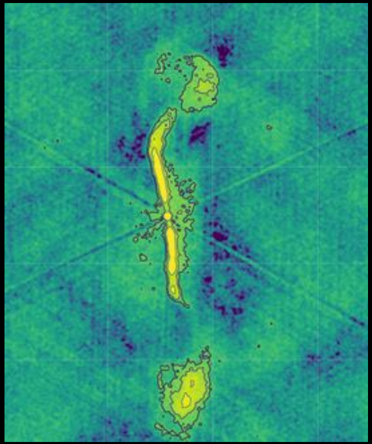
VLASSJ2032+4016
Score 390

Results: Top-ranked extended radio sources

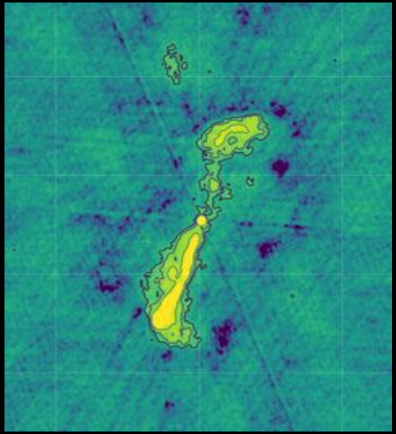
Each candidate is cross-matched with the SIMBAD and NED databases



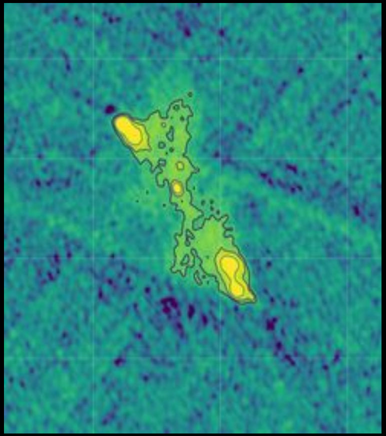
VLASSJ1416+1048
Score 232



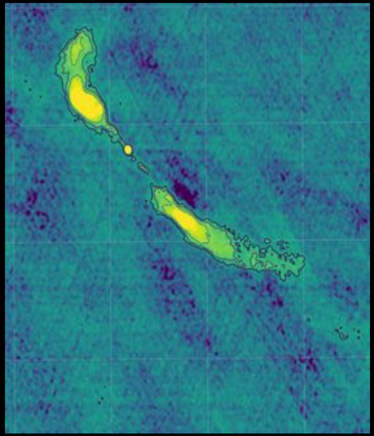
VLASSJ2231+3921
Score 167



VLASSJ0057-0123
Score 162



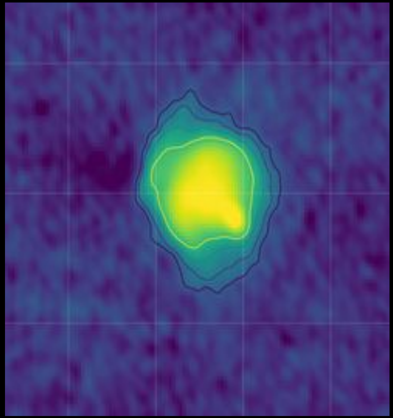
VLASSJ0849+0949
Score 152



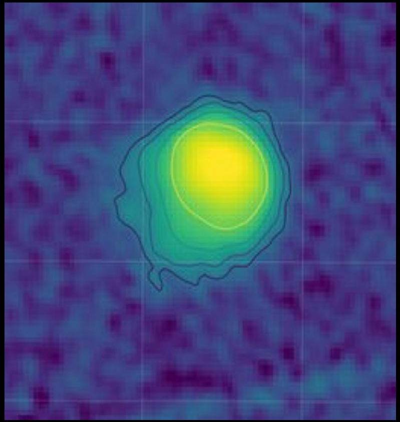
VLASSJ0452+5204
Score 142

Results: Top-ranked roundish radio sources

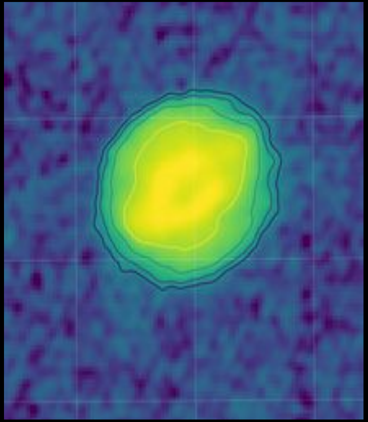
<u>Science interest</u>: Searching for radio halos in galaxy clusters, Odd Radio Circles (ORCs), and high-redshift objects.



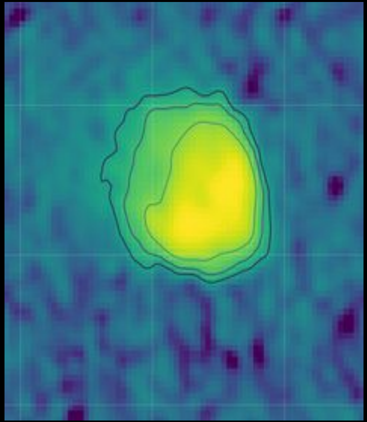
VLASSJ0643-0108
Score 190



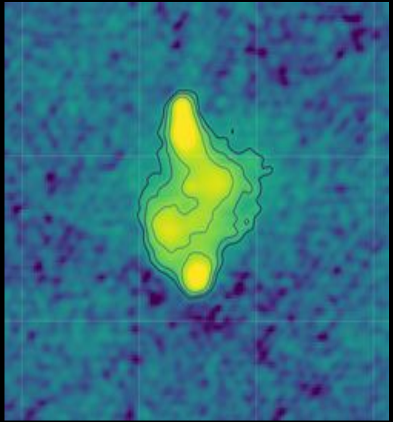
VLASSJ0436+5112
Score 180



VLASSJ1944+5031
Score 170



VLASSJ1901+0504
Score 160



VLASSJ0752+3252
Score 120

https://github.com/viralp/VLASS-anomaly

Future ML-based classification of radio sources

Use the current anomaly catalogue and cutouts as a seed dataset for image-similarity search, self-supervised learning, and active learning.

1 Current outputs

candidates.csv
FITS/PNG cutouts
Morphology scores

2 Embedding search

Convert each cutout into an image vector “find more like this”

FiftyOne + FAISS

3 Self-supervised ML

Train on unlabeled VLASS cutouts so model learns radio morphology

PyTorch + Lightly/BYOL

4 Active learning

Human labels only the most useful examples: real / artifact / type

modAL + scikit-learn

5 ML catalogue

Class labels
probabilities
review notes +
exports

Immediate next step

Load a small VLASS anomaly subset into FiftyOne → compute pretrained embeddings → select seed examples → retrieve nearest neighbours → export a ranked review set.

Then train BYOL/Lightly on clean radio cutouts and use active learning to improve labels with minimal human review.