Probing the Universe with Radio-loud AGN

Matthew Alger

Supervisors:
Julie Banfield, Christian Wolf,
Cheng Soon Ong (Data61/ANU), Ivy Wong (UWA/ICRAR)

Slides: http://www.mso.anu.edu.au/~alger/thesis-proposal

Radio AGN

Radio Active Galactic Nuclei

- AGN are actively-accreting supermassive black holes in the middle of galaxies.
- AGN are part of the lifecycle of a galaxy and understanding them is important for understanding galaxy evolution.
- Radio AGN affect the host galaxy and neighbouring environment through AGN feedback processes.

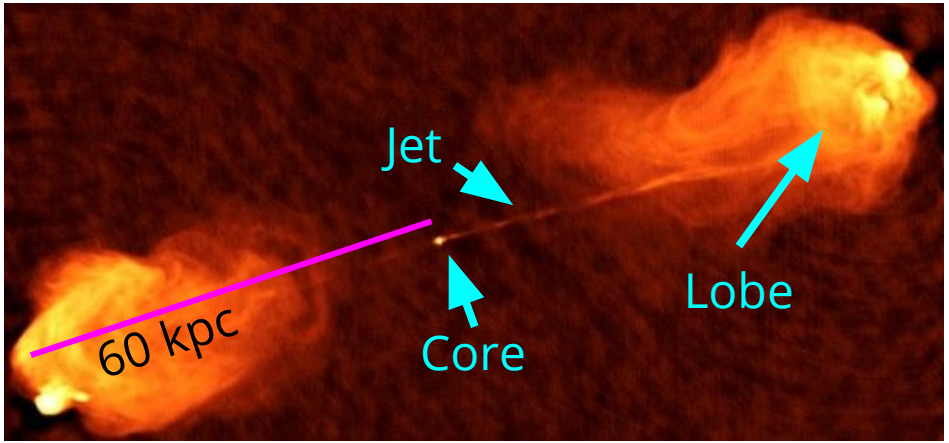
Centaurus A, a nearby radio AGN.

Image: ESO/WFI (Optical); MPIfR/ESO/APEX/A.Weiss et al. (Submillimetre); NASA/CXC/CfA/R.Kraft et al. (X-ray)

Radio Active Galactic Nuclei

Nearby(ish) radio galaxy Cygnus A:

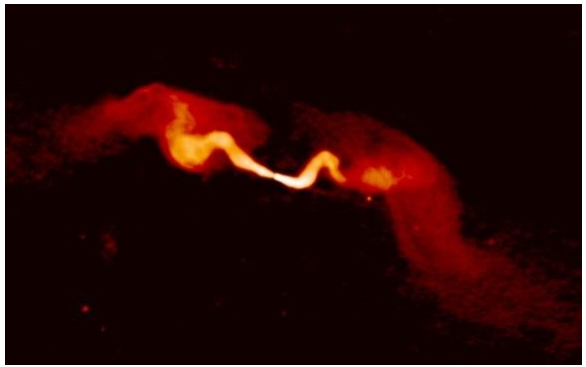
- ~120 kpc across
- z = 0.056
- $P_{1.4\ GHz} = 6 \times 10^{27}$ W Hz^{-1}



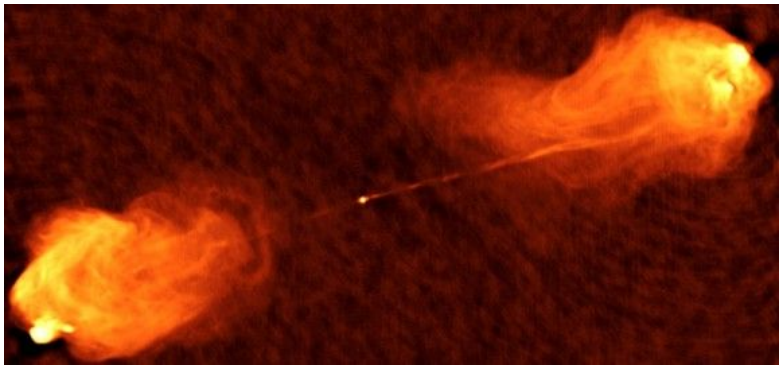
Cygnus A. *Image: NRAO/AUI*

Perley, Dreher & Cowan, 1984

Radio-loud Galaxy Classifications



3C31: Fanaroff-Riley type I.
Image: NRAO/AUI

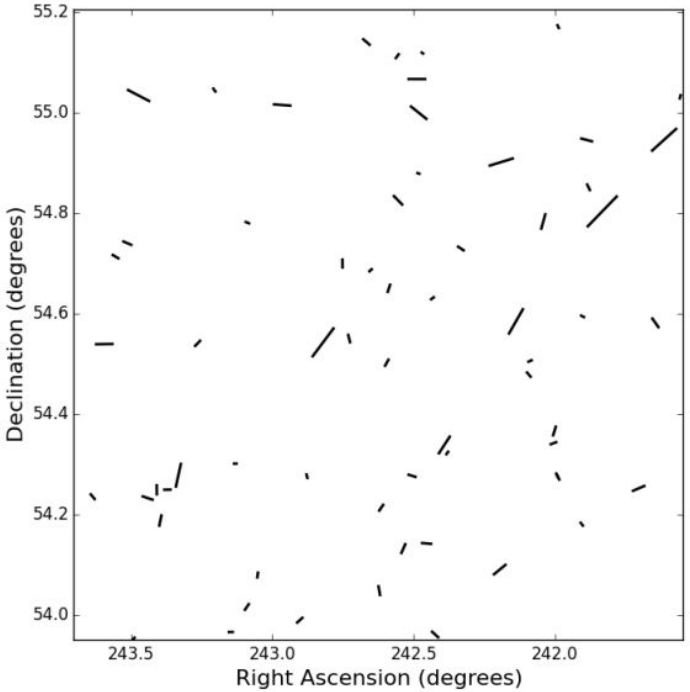


Cygnus A: Fanaroff-Riley type II.
Image: NRAO/AUI

+ Radio quasars, BL Lac, …

Radio Alignment

- Radio galaxies align at scales of 20–40 Mpc
  (e.g. Taylor & Jagannathan 2016, Contigiani et al. 2017)
- Bent galaxies trace clusters
  (e.g. Banfield et al. 2016, Johnston-Hollitt et al. 2014)



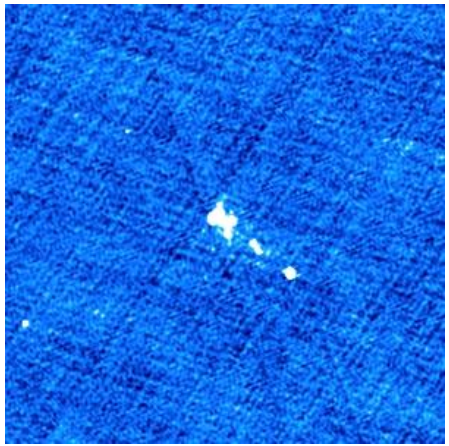
Angle and size of radio jets in the ELAIS-N1 field.
Image: Taylor & Jagannathan (2016).

Complementary Problems

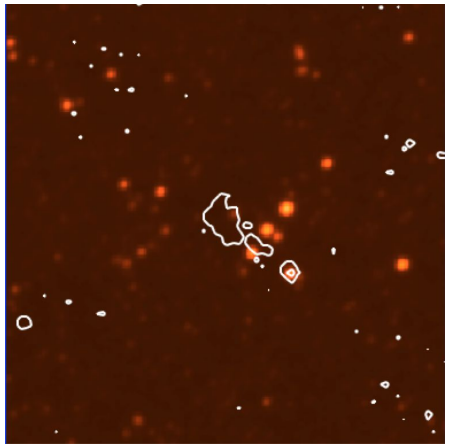
- Host galaxy cross-identification
- Radio morphology identification

Host Galaxy Cross-identification

- Match radio emission to its host galaxy.
- Hard in radio: Radio emission can be diffuse and extended.
  - Radio emission can be extended at scales of tens of arcminutes and multiple megaparsecs.
  - Often no clear relationship between radio emission and host galaxy.
  - High-resolution surveys like FIRST make this easier.



FIRSTJ023838.0+023450 at 1.4 GHz.
Image: FIRST

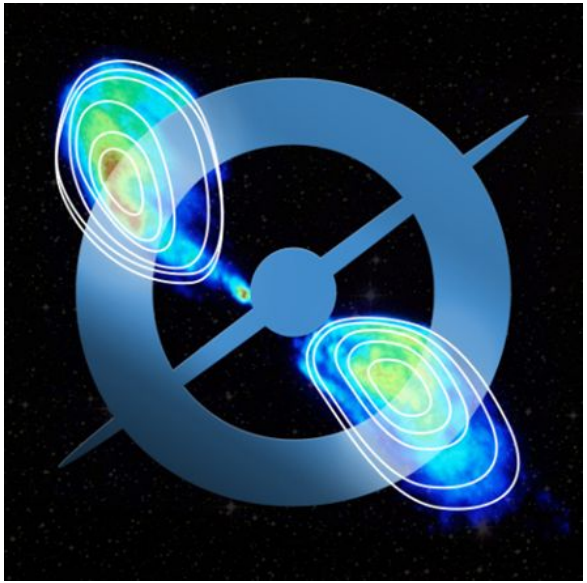


FIRSTJ023838.0+023450 in optical/infrared.
Image: WISE (Infrared), SDSS (optical)

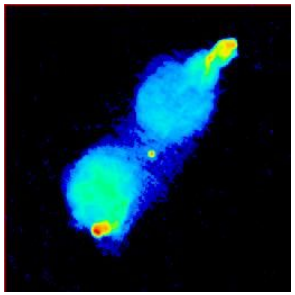
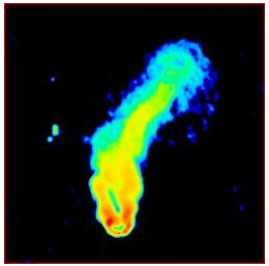
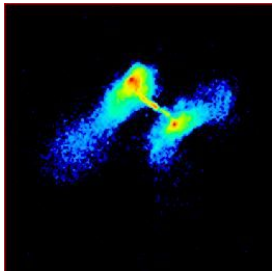
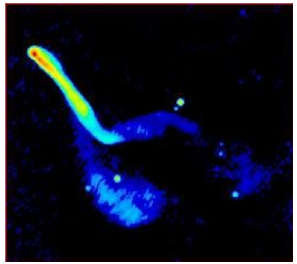
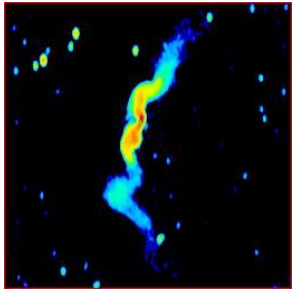
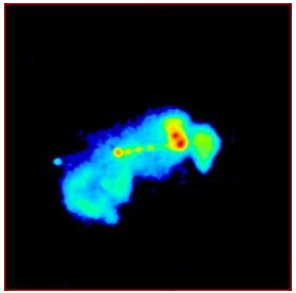
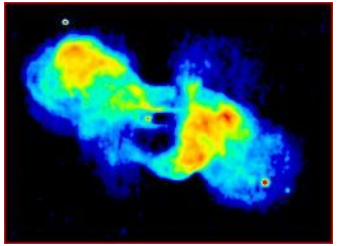
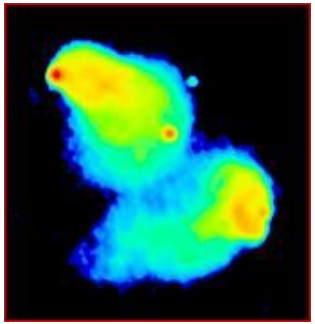
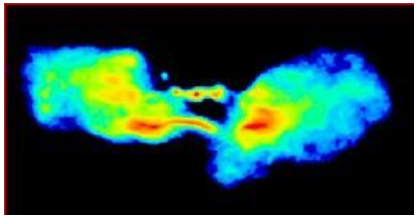
Host Galaxy Cross-identification

Current approaches:

- Manual
  - CDFS — Norris et al. (2006)
  - ELAIS-S1 — Middelberg et al. (2008)
  - CoNFIG — Gendre & Wall (2008)
- Crowdsourced
  - Radio Galaxy Zoo — Banfield et al. (2015)
- Automated
  - Nearest neighbour — Norris et al. (2006)
  - Bayesian hypothesis testing — Fan et al. (2015)
  - Likelihood ratio — Weston et al. (in prep)



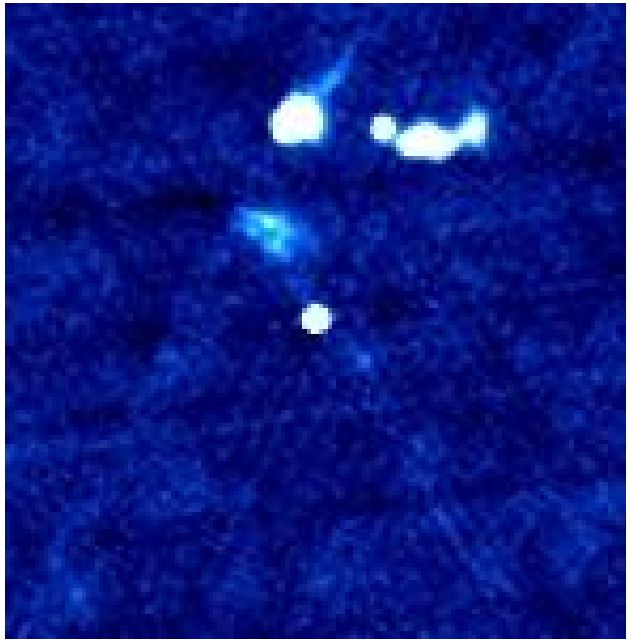
Radio Morphology Identification

http://www.jb.man.ac.uk/atlas

Radio Morphology Identification

- Complicated by the complex structure often present in radio galaxies.
- Often unclear which radio components are part of the same radio source.
- Two subproblems:
  - Classification
  - Source identification



Radio image centred on
10h 48m 01.177s +15° 14’ 38.40”.
Image: FIRST

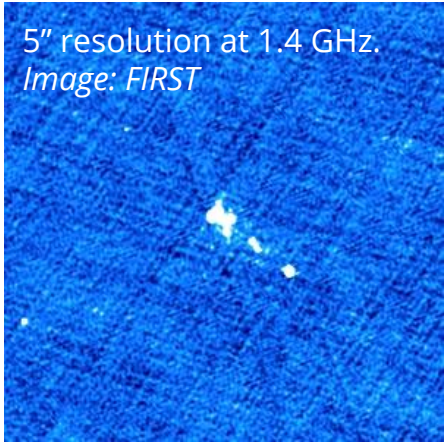
Radio Morphology Identification

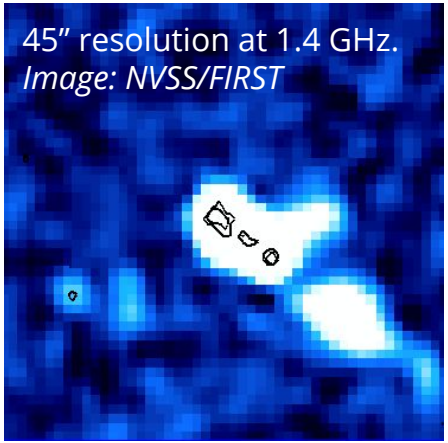
Current approaches:

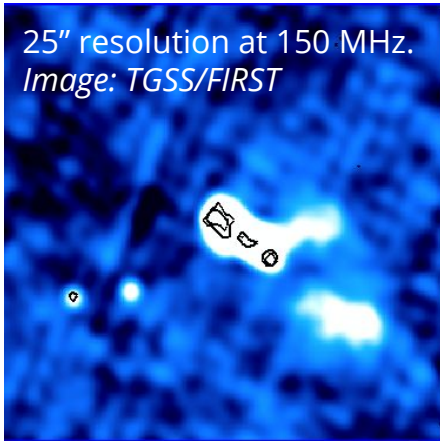
- Manual
  - CDFS — Norris et al. (2006)
  - ELAIS-S1 — Middelberg et al. (2008)
  - CoNFIG — Gendre & Wall (2008)
- Crowdsourced
  - Radio Galaxy Zoo — Banfield et al. (2015)
- Automated
  - Mixture models — Kirshner et al. (2003)
  - Bayesian hypothesis testing — Fan et al. (2015)
  - Neural networks — Aniyan & Thorat (2017)

Radio is Tricky

- Different wavelengths can show very different structure of radio galaxies.
- High resolution imaging is necessary, but diffuse radio emission is invisible at higher resolutions.







Astronomical Data

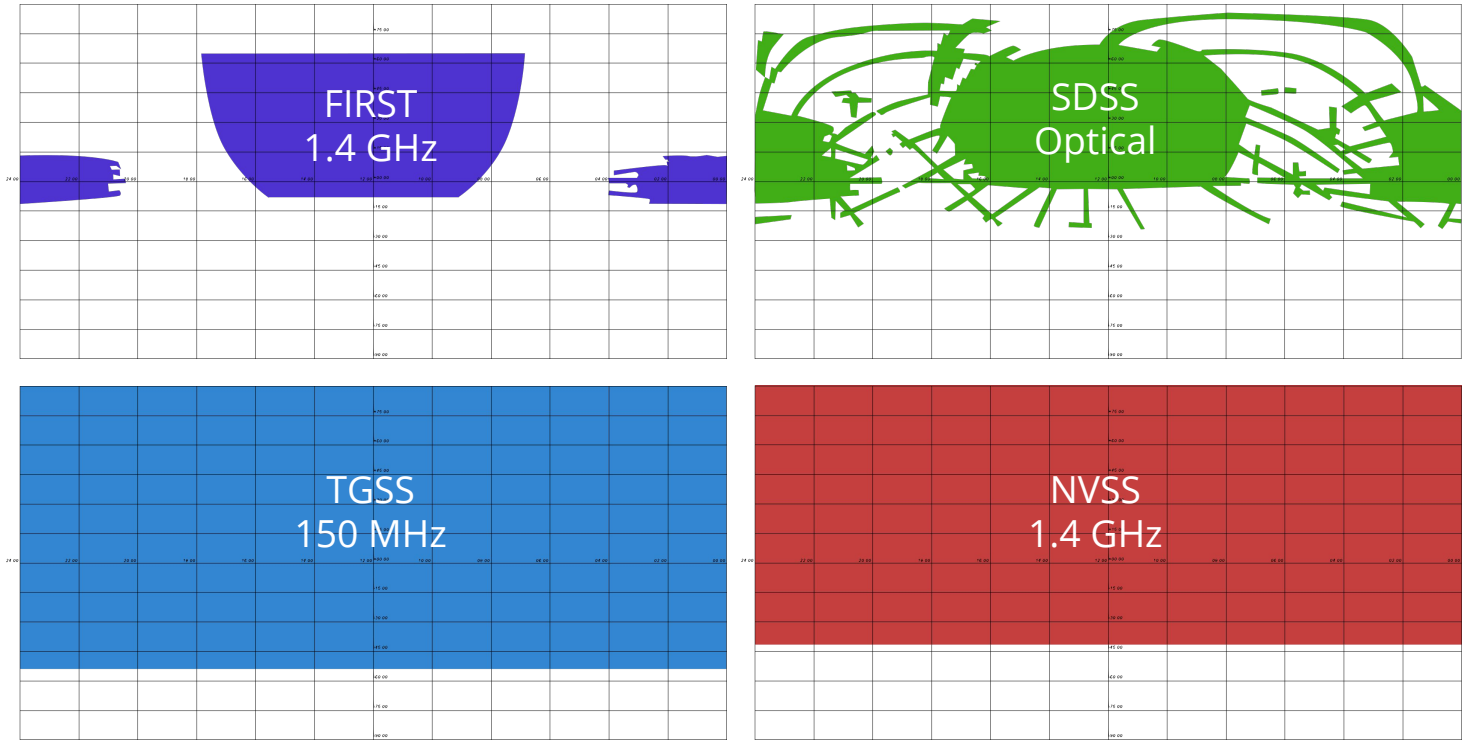
- Wide-area surveys
- Machine learning

Wide-area Surveys

- Radio
  - NVSS (1.4 GHz, 45” resolution, 1 800 000 sources)
  - FIRST (1.4 GHz, 5” resolution, 946 000 sources)
  - TGSS ADR1 (150 MHz, 25” resolution, 630 000 sources)
- Infrared
  - AllWISE
- Optical
  - SDSS

Image: NVSS

Wide-area Surveys



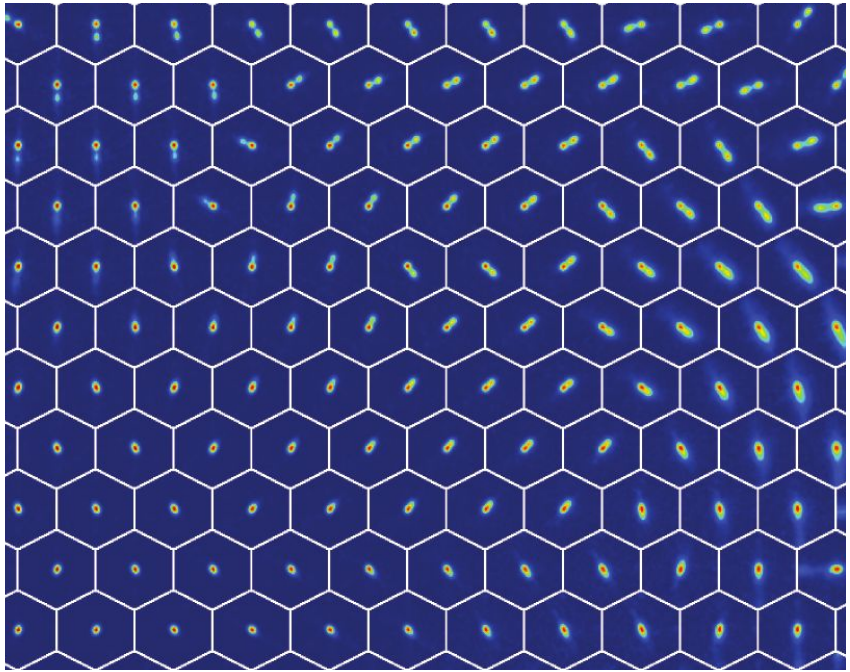
Data Mining in Radio

- Upcoming radio surveys will be *big*:
  - EMU/WODAN will find ~100,000,000 new radio sources.
  - If constructed, SKA could generate data up to 10 PB/year.
  - Need automated methods to process this data.

Australian SKA Pathfinder.
Image: CSIRO

Machine Learning in Astronomy

- Unsupervised
  - Dimensionality reduction
  - Clustering
  - Self-organising maps
    (e.g. Polsterer et al. 2015)
- Supervised
  - Regression (e.g. photo-z)
  - Classification
    (e.g. Aniyan & Thorat, 2017)
  - Object localisation (e.g. host galaxy cross-identification)



Part of a self-organising map of FIRST radio sources.
Image: Polsterer et al. (2015)

Proposal

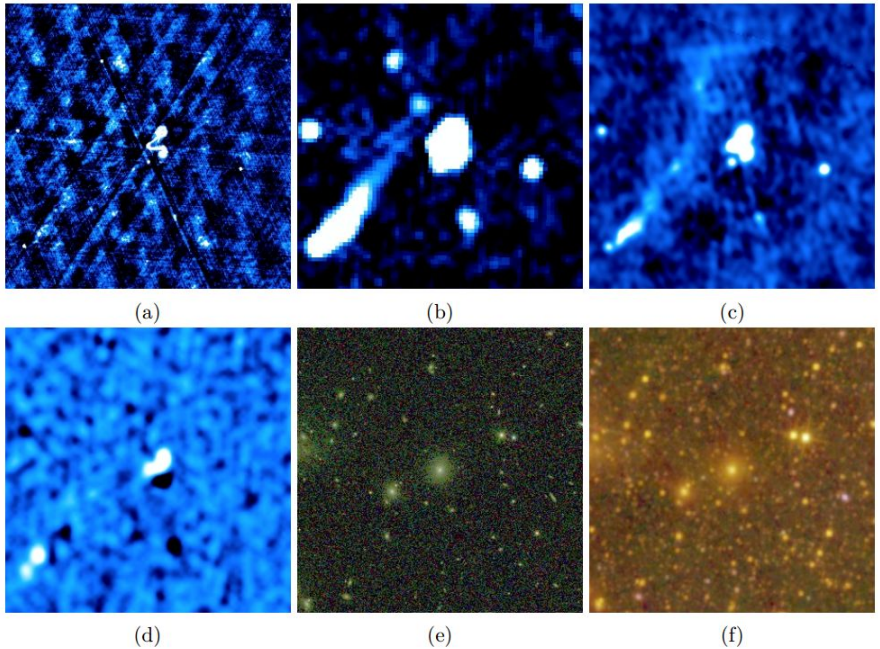
- Outline
- Current work
- Key problems
- Projected outcomes

Proposed Project

- Develop machine learning methods for use in wide-area surveys, incorporating multi-wavelength and multi-resolution imagery.
- Data-mine FIRST, NVSS, and TGSS for radio AGN (including morphologies and cross-identifications).
- Estimate proportions of FR I/FR II, distributions of geometric properties, polarisation properties, and host galaxy properties.
- Use radio alignment to map large-scale structure.

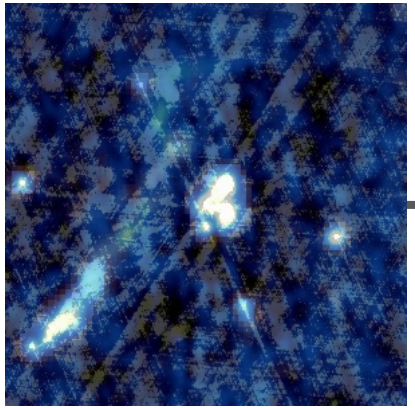
Multi-wavelength, Multi-resolution ML

- Methods I will develop will incorporate multiple surveys.
- Necessary to get the full picture of radio objects.
- Sensor fusion is an active research area in machine learning (self-driving cars, geophysics).

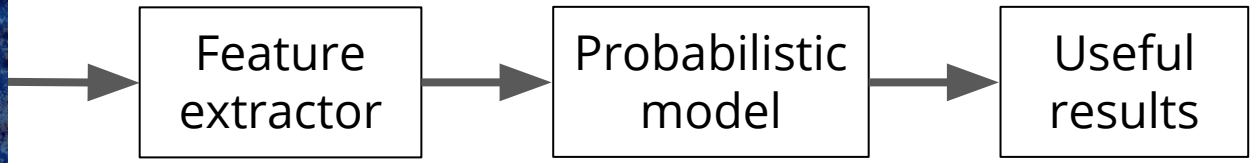


IC 708 in (a) FIRST, (b) NVSS intensity, (c) TGSS, (d) NVSS polarisation, (e) SDSS, and (f) WISE.

Generic Approach

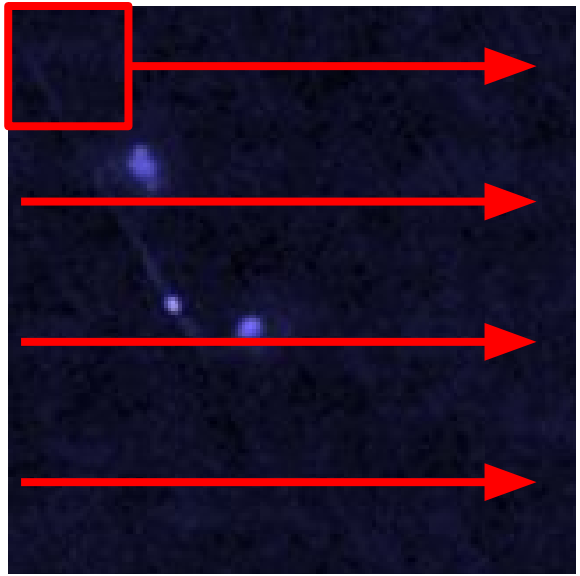


Multi-wavelength, multi-resolution image



Automating Cross-identification

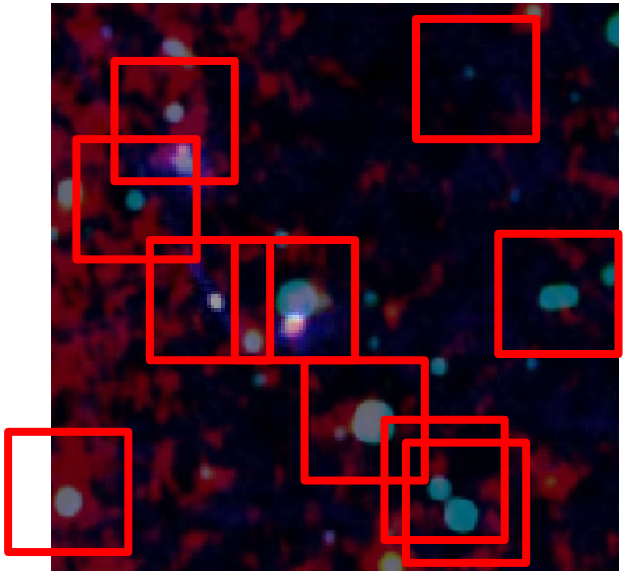
- First attempt:
  - Given an image of radio emission, check each square patch to see if the AGN is located there
  - Not terribly efficient



Scanning to find the AGN.
Image: FIRST

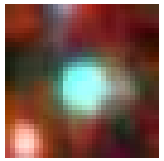
Automating Cross-identification

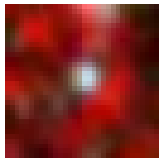
- First attempt:
  - Given an image of radio emission, check each square patch to see if the AGN is located there
  - Not terribly efficient
- Second attempt:
  - Given an image of radio emission, check each *galaxy* in that image to see if it looks like it is the host galaxy
  - Much more efficient!



Candidate host galaxies.
Image: FIRST/WISE

Automating Cross-identification

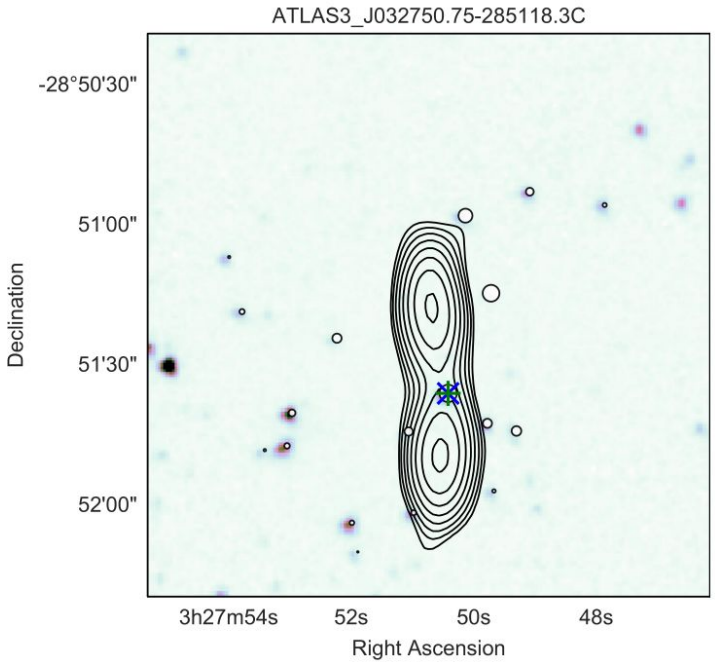
( , ) ⟶ 1

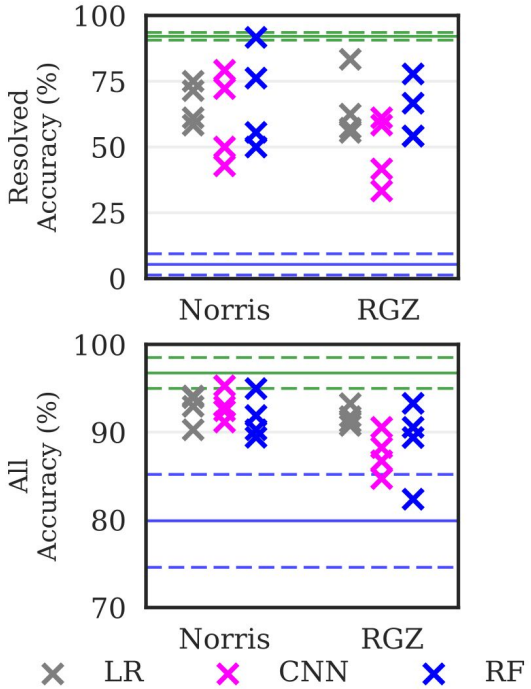
( , ) ⟶ 0

Representation of galaxy

Whether galaxy has an AGN

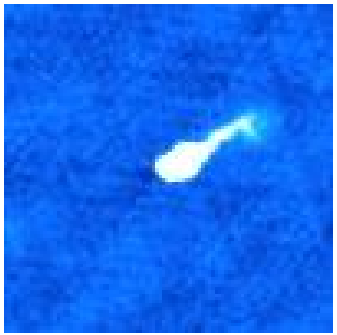
Automating Cross-identification

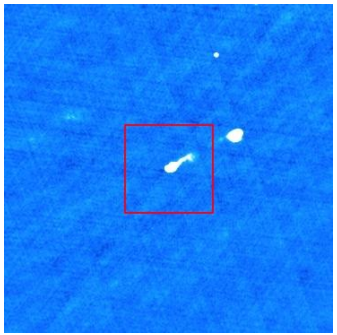




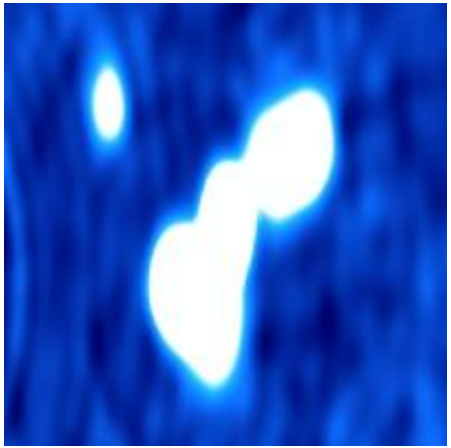
Automating Cross-identification

- Morphology is a big problem!
  - There may be multiple AGN in one image (especially for deeper surveys like EMU).
  - The host galaxy may not be anywhere near the radio emission.





The host galaxy of FIRSTJ151227.2+454026 is several arcminutes away.
Image: FIRST



ATLAS3 J033402.87-282405.8C breaks the assumption that there is only one source in the image.
Image: ATLAS

Key Problems

- Host galaxy cross-identification
  - Casting the problem as an ML task
  - Finding relevant radio data requires morphology
- Radio morphology identification
  - Classification tasks in ML expect clear, well-defined classes (which we don’t have)
  - No good way to encode “other” classes
  - Finding morphology made easier with cross-identifications
- Source detection
  - Radio components comprising a source may be disconnected
  - Components can be arbitrarily far apart on the sky
  - Difficult to identify objects with unknown size

Other Problems

- Feature extraction
  - Much better understood in optical than radio
  - Limited research, e.g. Proctor (2006), Polsterer et al. (2015), Aniyan & Thorat (2017), Lukic et al. (in prep)
  - Standard image feature extraction techniques may not work
- Latent angle of radio objects
  - Could be considered as latent variables in probabilistic model
  - Leads to potentially difficult EM algorithm
  - Links to AGN unification

Outcomes

- Methods for getting science out of wide-area radio surveys at scale.
- Methods for combining multi-wavelength and multi-resolution surveys at scale.
- Bounds on radio AGN class proportions.
- Estimated distributions of polarisation and geometry.
- More complete redshift–class distributions.

Where to now?

- Finish writing paper on host cross-identification.
- Develop a basic feature extractor.
- Combine FIRST and TGSS for morphology classification and visualisation.
- Work on Radio Galaxy Zoo DR1.

Feature Extraction Approaches