Radio luminosity functions
with machine learning and Radio Galaxy Zoo

Matthew Alger (ANU/Data61)
O. Ivy Wong (ICRAR/UWA)
Cheng Soon Ong (Data61/ANU)
Naomi McClure-Griffiths (ANU)

Slides: http://www.mso.anu.edu.au/~alger/asa-19

Australian National University

Radio luminosity functions
with machine learning and Radio Galaxy Zoo

Comoving density of radio sources
as a function of radio luminosity

Radio luminosity functions

with machine learning and Radio Galaxy Zoo

Comoving density of radio sources
as a function of radio luminosity

Radio luminosity functions

with **machine learning** and Radio Galaxy Zoo

Approximating functions
based on existing data

Comoving density of radio sources as a function of radio luminosity

Radio luminosity functions

with **machine learning** and **Radio Galaxy Zoo**

Approximating functions based on existing data

A citizen science project for matching radio emission to infrared galaxies

Comoving density of radio sources as a function of radio luminosity

Radio luminosity functions

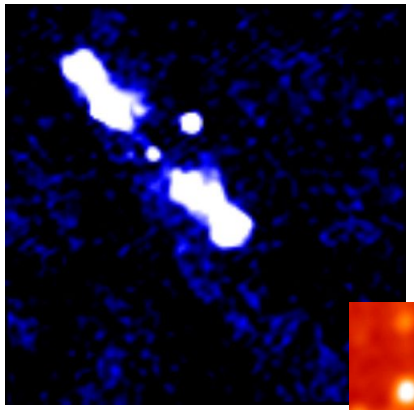
with **machine learning** and **Radio Galaxy Zoo**

Approximating functions based on existing data

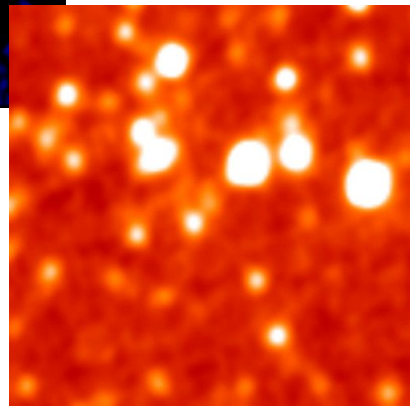
A citizen science project for matching radio emission to infrared galaxies

*Use **machine learning** to approximate how **citizen scientists** cross-identify, then cross-identify everything and make a **luminosity function** with a huge sample size*

Radio Galaxy Zoo

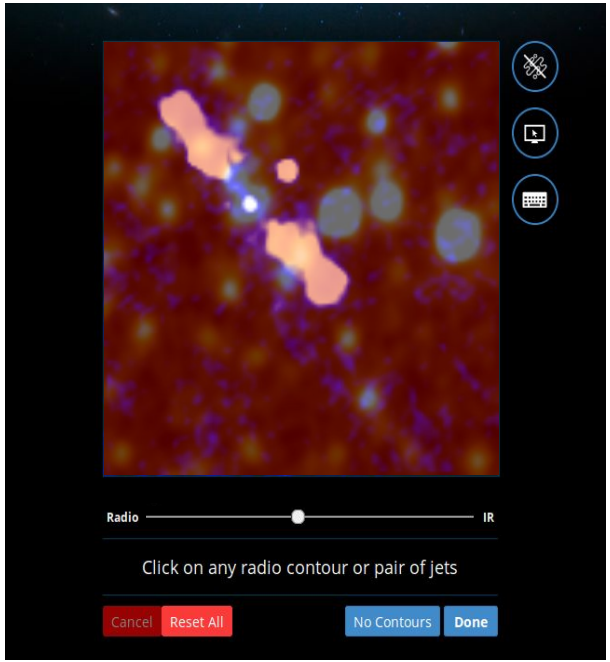


1.4 GHz radio
(FIRST)



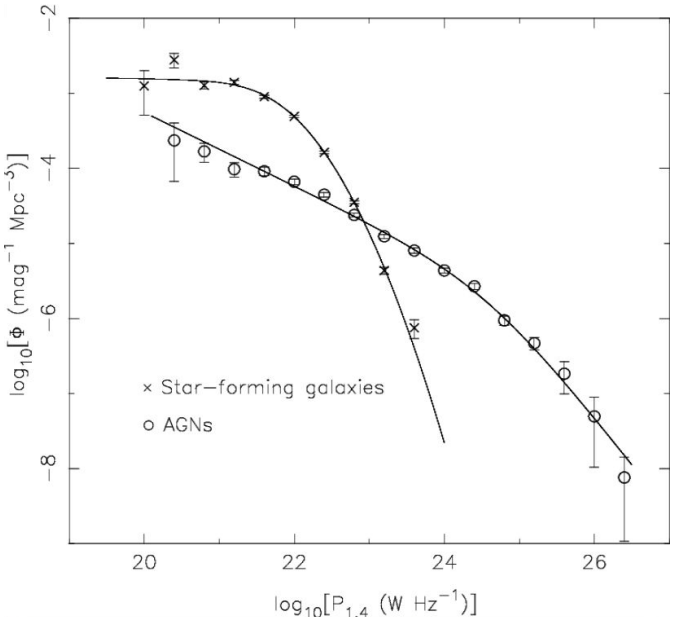
3.4 μm infrared
(WISE)

Citizen scientists cross-identify radio emission with infrared host galaxies



Radio luminosity functions

- Comoving density of radio sources as a function of radio luminosity
  - Units of dex^{-1} Mpc^{-3}
  - Distribution of radio source luminosities in a *physically meaningful* way
- Fractional radio luminosity functions
  - Luminosity distribution of physically-selected subsets may be different
  - Helps understand evolution and structure of radio galaxies



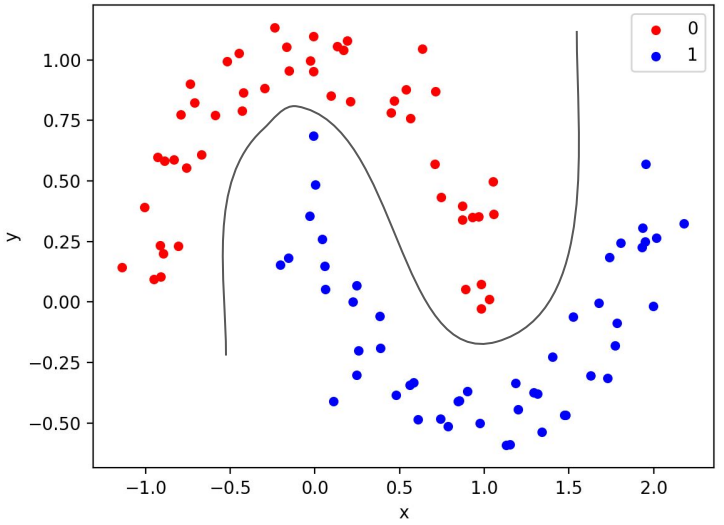
Radio luminosity function divided into radio due to star formation and radio due to active galactic nuclei.
Image: Mauch & Sadler (2007)

Binary classification

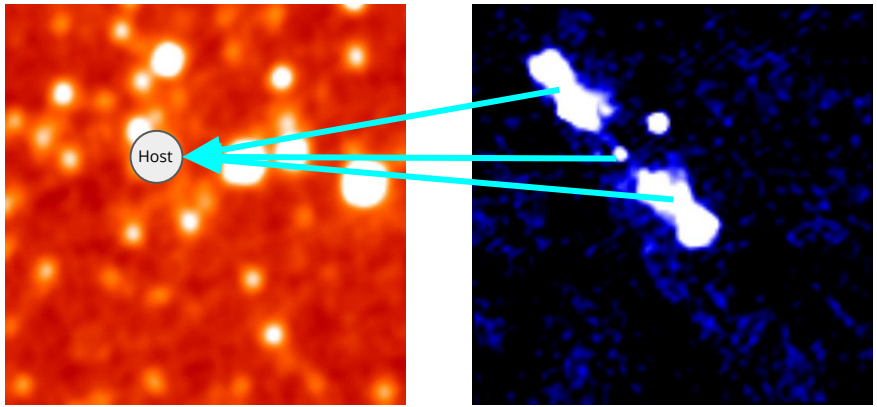
- Find a function that separates objects into two classes
- Well-understood

Equivalent:
$h(x) = g(x) > 0$
$g(x) = \sigma(f(x))$

$$\begin{cases} f : \mathbb{R}^d \to \mathbb{R} \\ g : \mathbb{R}^d \to [0, 1] \\ h : \mathbb{R}^d \to \{\top, \bot\} \end{cases}$$

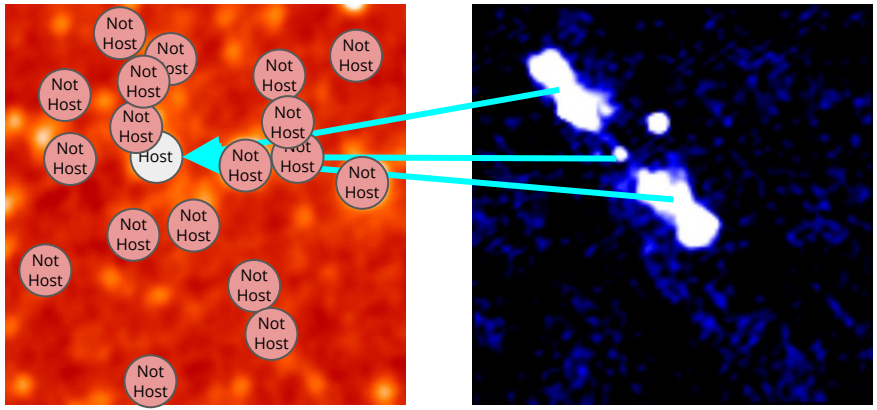


Learning from Radio Galaxy Zoo



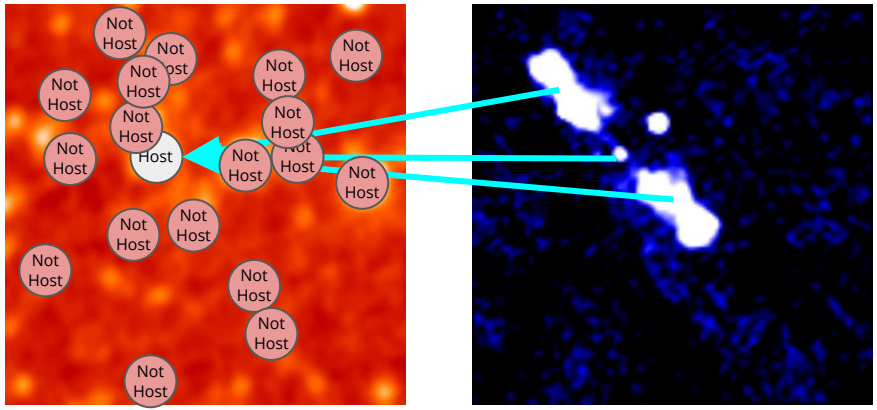
- Assign hosts positive labels

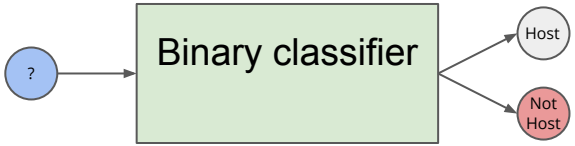
Learning from Radio Galaxy Zoo



- Assign hosts positive labels
- Assign everything else negative labels

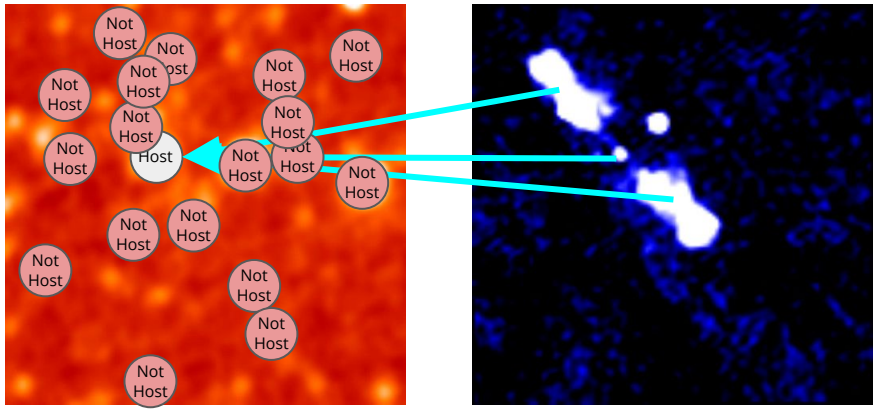
Learning from Radio Galaxy Zoo



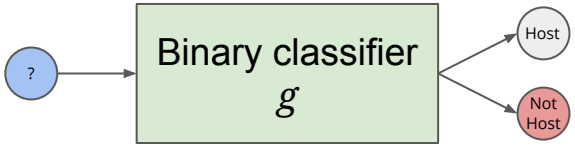


- Assign hosts positive labels
- Assign everything else negative labels
- Train classifier to identify *host* and *not host* classes

Learning from Radio Galaxy Zoo



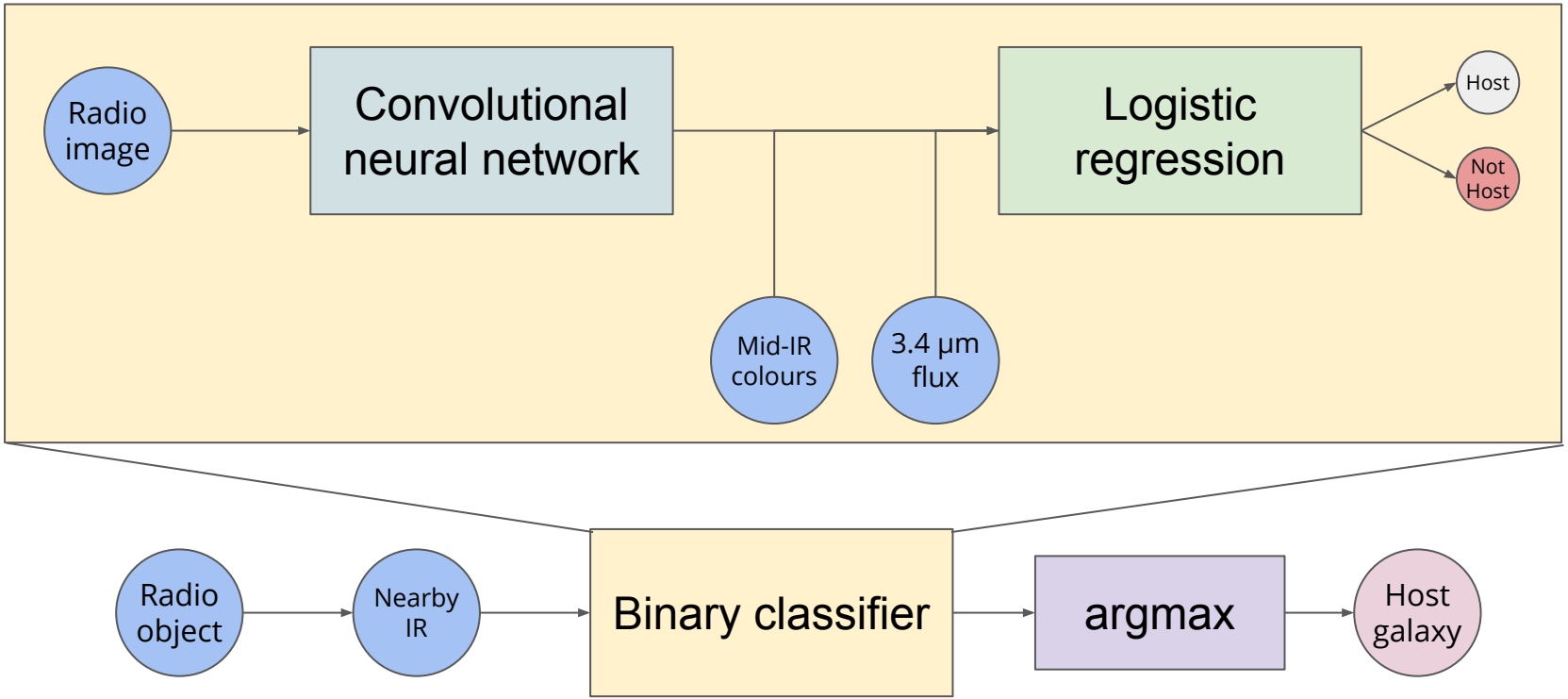
- Assign hosts positive labels
- Assign everything else negative labels
- Train classifier to identify *host* and *not host* classes



$$\text{xid} : \text{Radio} \rightarrow \text{IR}$$

$$\text{xid}(r) = \underset{i \in \text{IR objects}}{\text{argmax}}\ g(i)\ \mathcal{N}(r, i)$$

Binary classification model

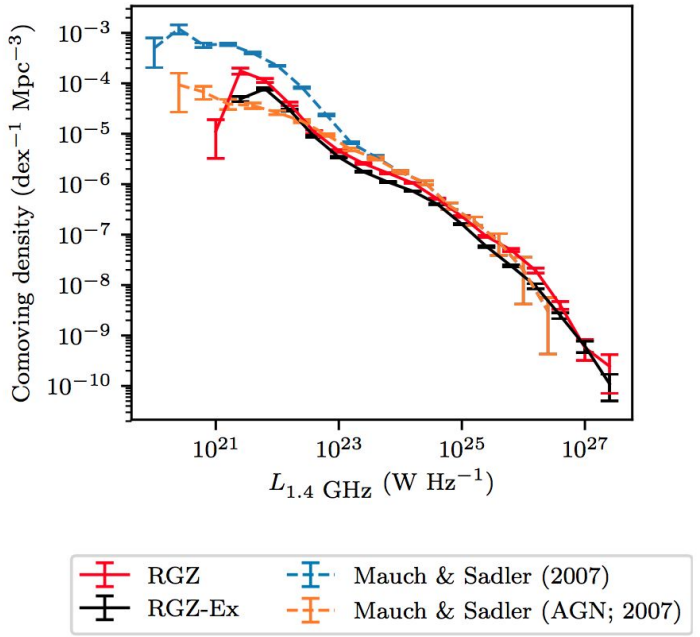


Luminosity function

- RGZ-Ex contains 157 007 cross-identified radio sources with 30 743 redshifts
- Large sample allows us to build a radio luminosity function of extended sources
  - Luminosities up to 10^{27} W/Hz
  - Close match to Mauch and Sadler (2007) radio AGN luminosity function

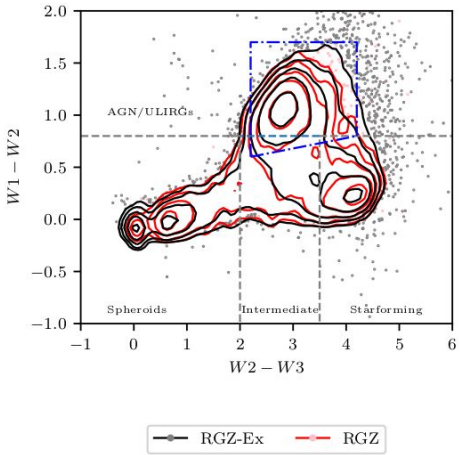
Radio Galaxy Zoo: radio luminosity functions of extended sources

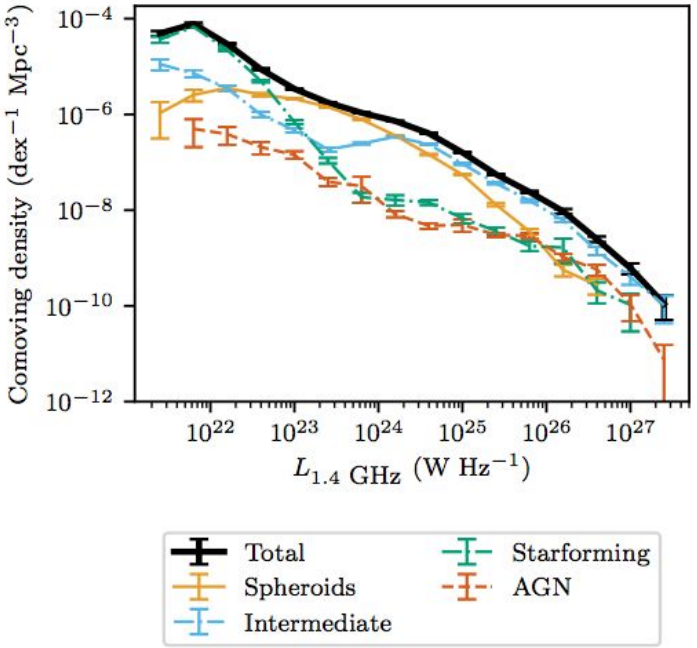
M. J. Alger[1,2]* et al.
[1]Research School of Astronomy and Astrophysics, The Australian National University, Canberra, ACT 2611, Australia
[2]Data61, CSIRO, Canberra, ACT 2601, Australia



Fractional luminosity function (Mid-IR)

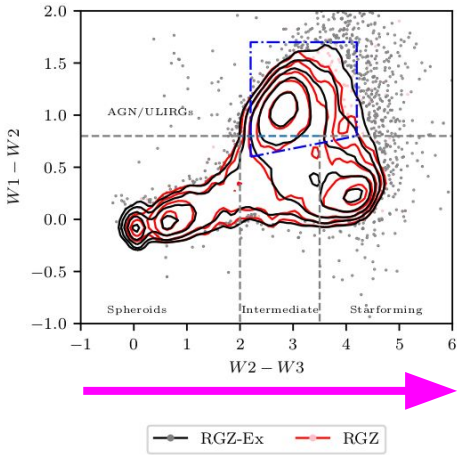
- Divide radio luminosity function based on mid-infrared host colours
  - “Extended” star-forming sources below 10^{23} W/Hz (visually verified)
  - Radio-loud sources dominated by “intermediate” galaxies

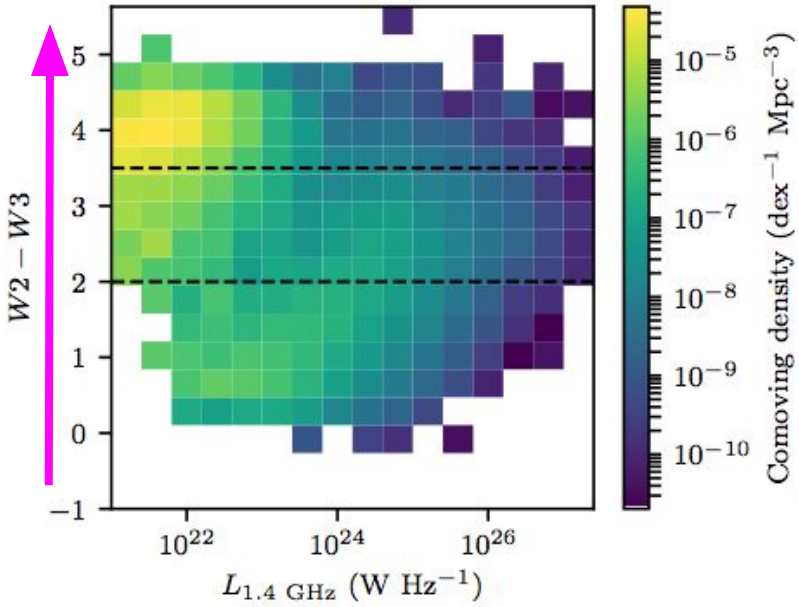




Bivariate luminosity function (Mid-IR)

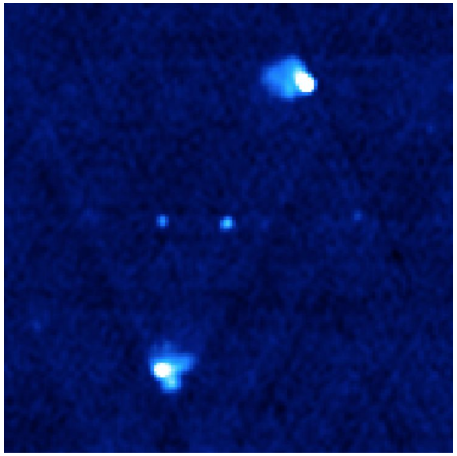
- Divide radio luminosity function based on 12 μm/4.6 μm colour
  - “Extended” star-forming sources below 10^{23} W/Hz (visually verified)
  - Radio-loud sources dominated by “intermediate” galaxies

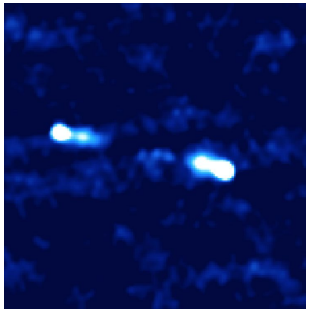


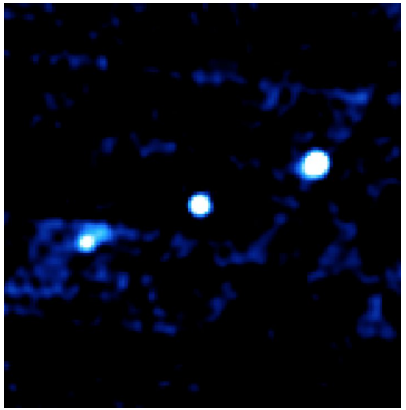


The RGZ-Ex catalogue

- Catalogue of 157 007 candidate radio sources and their hosts
- Large but noisy
- Contains around *sixty* previously unidentified giant radio galaxies (≥1 Mpc)

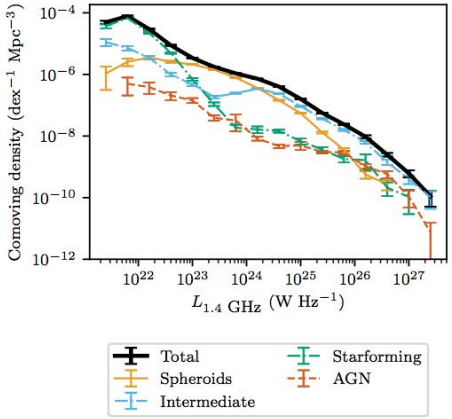


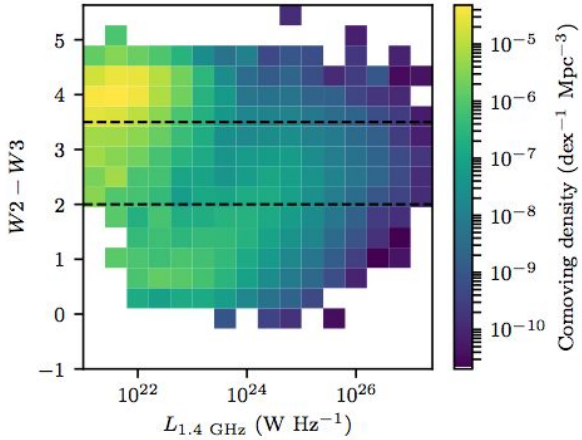


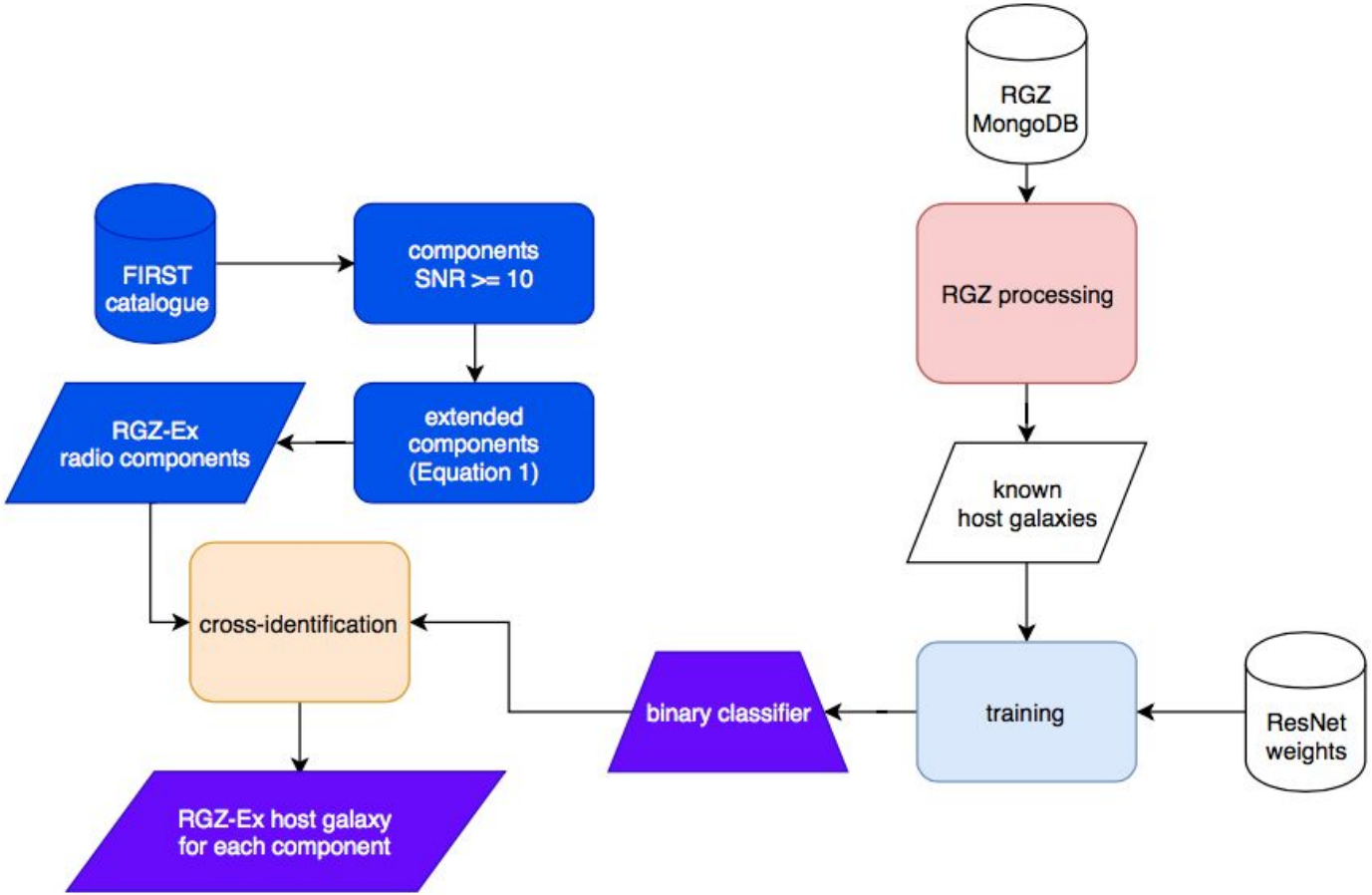


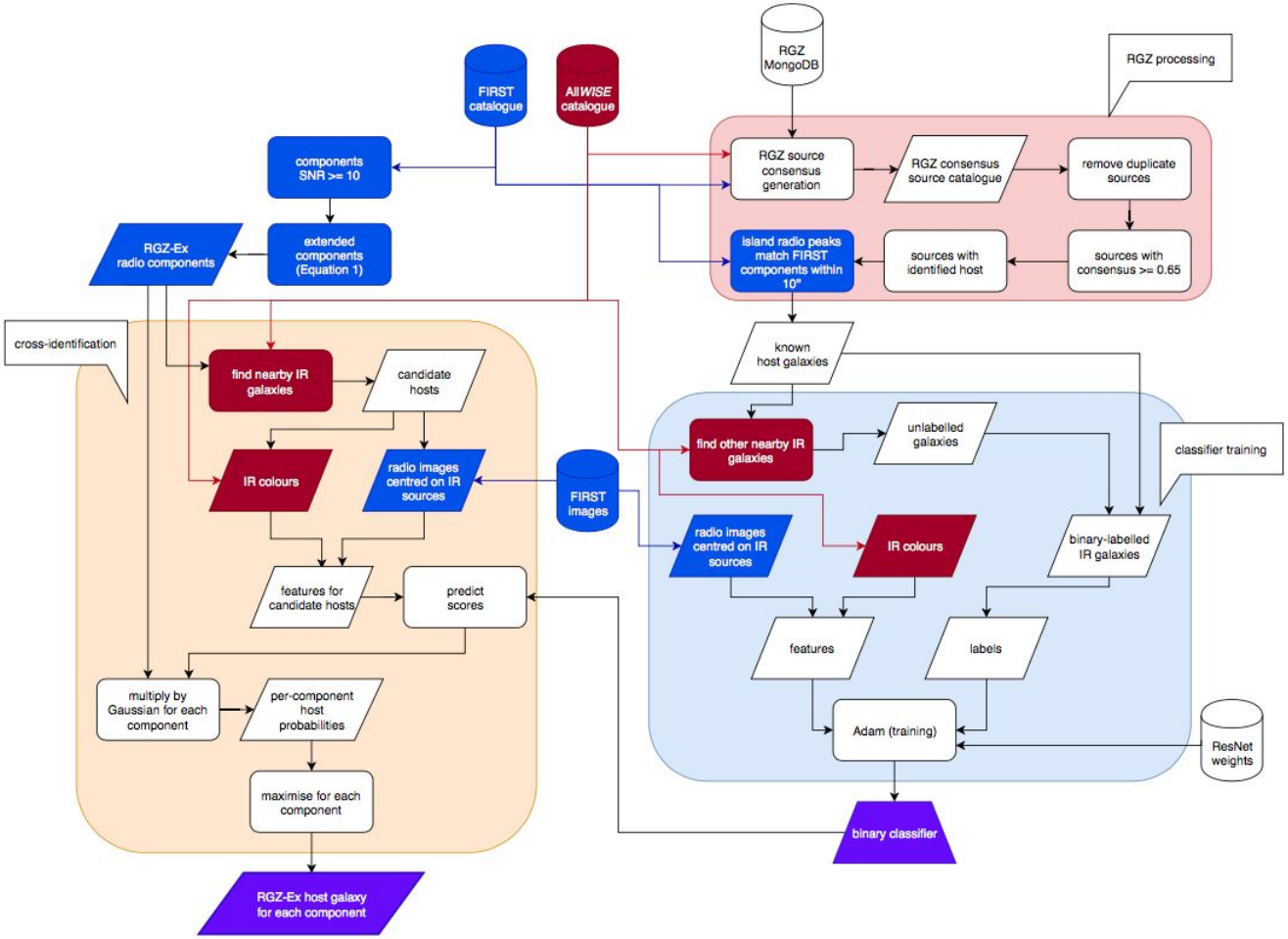
RGZ-Ex and luminosity functions

- We developed an automated, machine learning approach to radio-infrared cross-identification
- We created a huge catalogue of candidate radio sources and their hosts
- We estimated fractional radio luminosity functions of extended radio sources
- We found ~60 new giants



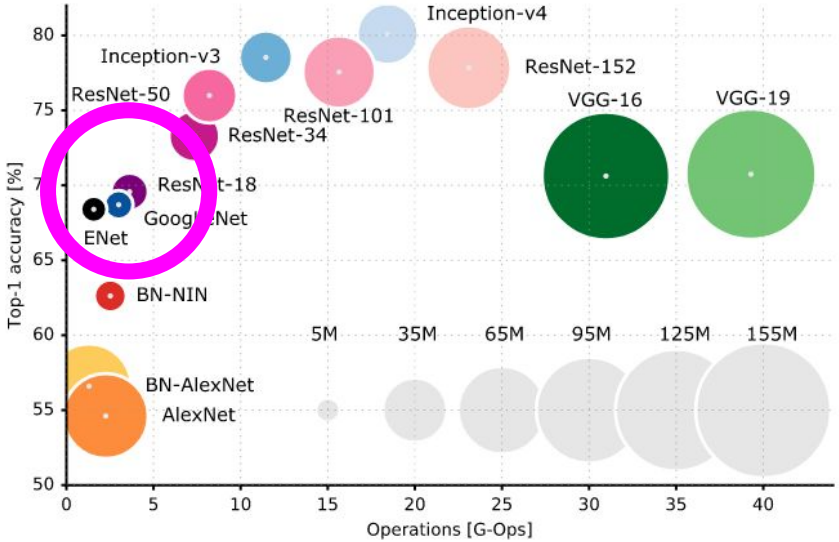



RGZ
MongoDB
RGZ processing
known
host galaxies
training
ResNet
weights
binary classifier
cross-identification
FIRST
catalogue
components
SNR >= 10
extended
components
(Equation 1)
RGZ-Ex
radio components
RGZ-Ex host galaxy
for each component


RGZ MongoDB
RGZ processing
FIRST catalogue
AllWISE catalogue
components SNR >= 10
RGZ source consensus generation
RGZ consensus source catalogue
remove duplicate sources
RGZ-Ex radio components
extended components (Equation 1)
island radio peaks match FIRST components within 10"
sources with identified host
sources with consensus >= 0.65
cross-identification
known host galaxies
find nearby IR galaxies
candidate hosts
find other nearby IR galaxies
unlabelled galaxies
classifier training
IR colours
radio images centred on IR sources
FIRST images
radio images centred on IR sources
IR colours
binary-labelled IR galaxies
features for candidate hosts
predict scores
features
labels
multiply by Gaussian for each component
per-component host probabilities
Adam (training)
ResNet weights
maximise for each component
binary classifier
RGZ-Ex host galaxy for each component

Binary classification model

- ResNet-18 (multiclass)
  - Good accuracy
  - Low complexity
  - Very fast to train and use
- Remove last layer and replace with a binary classifier
- Add non-image features
  - Mid-infrared colours
  - 3.4 μm flux
  - Room for improvement — e.g. add redshifts



Trade-offs between network complexity and accuracy on ImageNet.
Image: Canziani et al. (2016)

Cross-identification as binary classification

$$\text{xid} : \text{Radio} \rightarrow \text{IR}$$

$$\text{xid}(r) = \operatorname{argmax} f(i; r)$$

$i \in$ candidate IR hosts

where

$$f : \mathbb{R}^d \rightarrow \mathbb{R}$$

$$f(i) = p(\text{host} \mid i)$$

is a binary classifier