Heidelberg Institute for Theoretical Studies

HITS

CSIRO

Crowdsourcing to GPUs

semi-automated classification of radio morphologies for 946,419 sources in FIRST with 299,266 IR counterparts in UKIDSS

Erica Hopkins[1]
Erica.Hopkins@h-its.org

Kai Lars Polsterer[1]
Kai.Polsterer@h-its.org

Tim Galvin[2]
Tim.Galvin@csiro.au

[1]Heidelberg Institute for Theoretical Studies, Astroinformatics, Heidelberg, Germany [2]CSIRO Astronomy and Space Science, Bentley WA, Australia

Abstract

We created a catalog with a probabilistic classification system for radio data, which was developed with a self organizing map (SOM), and applied it to all FIRST radio sources. This allows for a more flexible classification system without hard barriers to separate classes.

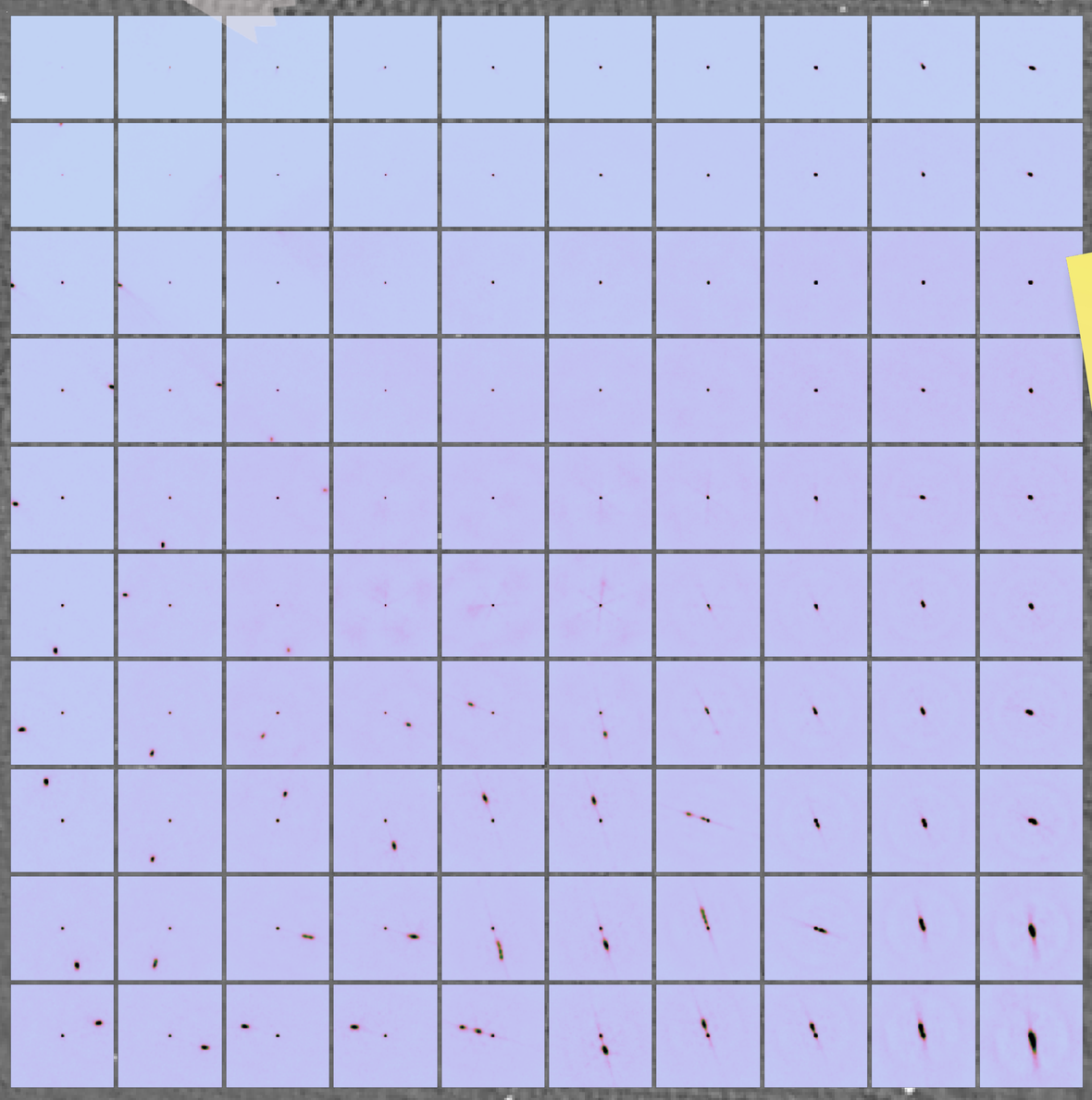
We created a second SOM using infrared data, which is used to determine if the radio source has an infrared counterpart. This provides information about the location of the center of the galaxy and can help to determine the nature of the radio galaxy and potentially help link multiple radio sources as part of the same galaxy.

Methods / Data

We created two SOMs, one based on radio data and the other based on infrared data. These were then annotated with our classes.

The radio SOM was created using FIRST data of ~900k images. We created a 10x10 SOM as that allowed for all common morphologies to be seen while also not allowing for an excess of repetitive neurons. The less common morphologies can be examined by analyzing the outliers.

The infrared SOM was created using the UKIDSS data which overlaps with FIRST, ~300k images. We created a 20x1 SOM which only focuses on the most central part of the radio source. We only required a one dimensional SOM due to the simplicity of the infrared data. As we only focused on the center, there could only be a centered source, an off center source, or no source. We decided to only focus on the central part due to the complexity of the infrared sky when compared to the radio sky. The infrared sky is much more crowded and the SOM will focus more on bright, unrelated sources than the source of interest when the area is too large.



Radio SOM

10x10 SOM created with FIRST data

Covers all common morphologies

Outliers allow us to find more interesting and unique objects

UKIDSS

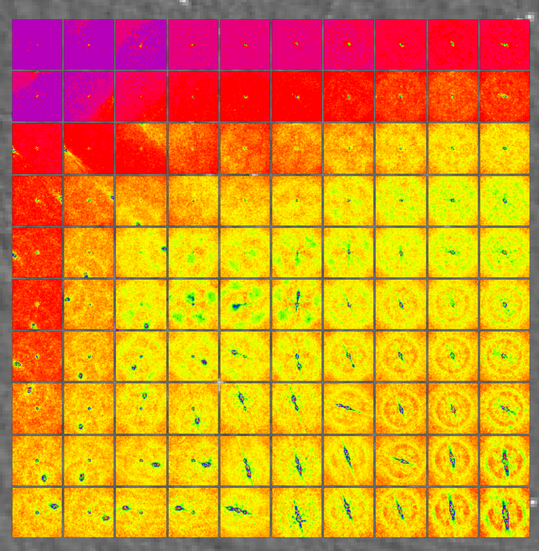
IR image of center of radio source

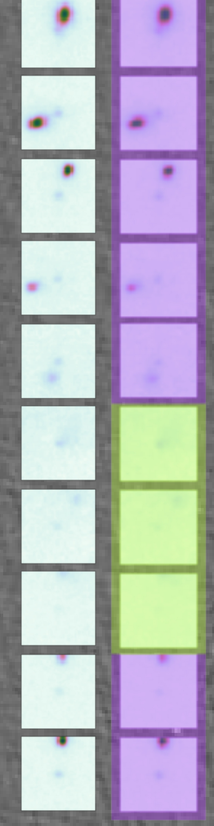
Can help determine the nature and location of radio source and central part of the galaxy

Divided into centered, off center, and no source

UKIDSS Classes

- Centered
- Off centered
- No source





Look what we found!

Remnant/PSF is visible

Usually too faint

Off center circles prove it's not due to rotation invariance

Can be used to improve calibration / image synthesis

$$dNorm_{i,j} = \frac{dMin_{i,j} - \min_j dMin_{i,j}}{}$$

$$\widetilde{L}_{i,j} = e^{-dNorm_{i,j}}$$

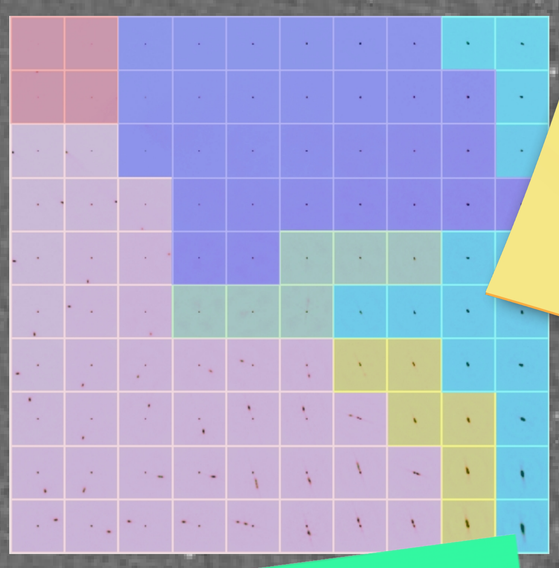
$$L_{i,j} = \frac{\widetilde{L}_{i,j}}{\sum \widetilde{L}_{i,j}}$$

$$P_m = \frac{\sum_m L_{i,j}}{\#neurons_j}$$

Heatmaps

Shows similarity between an image and each neuron

Are used as basis to transfer annotations

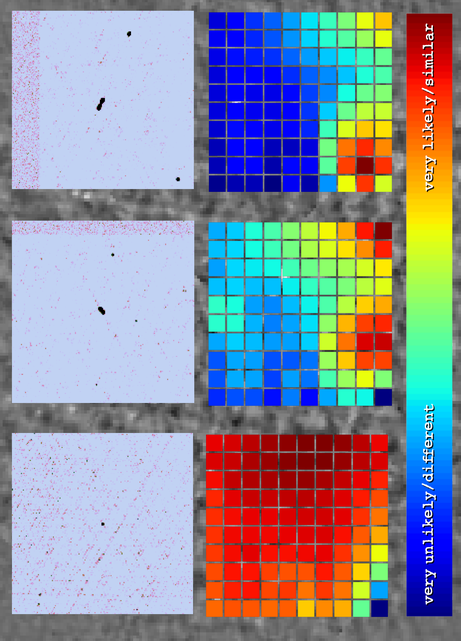


Classification System

SOM manually divided into 6 classes

Probability determined by the distance between the image and each neuron

- dim
- extended
- noisy
- concatenated
- point
- detached



Calibration issue

Connected?

Triple?

Artifact

Bent galaxy?

Outliers

Artifact coordinated can be cataloged as known issues

Improperly calibrated sources can aid in improving image synthesis

Interesting and rare objects can be analyzed with crowd sourcing projects, optimizing human efforts

I would like to acknowledge the Klaus Tschira Foundation for their financial support, TopCat, Aladain, and astropy for providing tools to assist in my analysis, and UKIDSS and FIRST for the data.

Erica Hopkins

github.com/HITS-AIN/PINK

h-its.org/people/erica-hopkins

Astroinformatics 2019, Pasadena, USA, June 24-28, 2019