NCRA-TIFR Project Proposal

Archit Sakhadeo
Rathin Desai
Shadab Shaikh
Shubhankar Deshpande

Mentor
Dr. Yogesh Wadadekar
Dr. C. H. Ishwar Chandra

1 Introduction

1.1 Morphological Classes of Radio Galaxies

Radio galaxies with active nuclei can be distinguished based on their radio luminosity or brightness of their radio emissions in relation to their hosting environment. Some of the basic morphological classifications include point sources, extended sources i.e. sources with extended contours, double radio sources, jets, and lobes.

1.2 Problems faced with current classification

Currently Radio astronomers manually classify galaxies based on visual inspection of the images which is a slow procedure, and increases the time to production of scientific results. Further, it introduces uncertainities in the classification procedure, both of which are problems which can potentially be mitigated by using an automated approach.

Contemporary algorithms classify radio sources into at most three different classes. Our aim is to build a robust model capable of handling more than 2 classes.

2 Objective

- Potentially discovering rare forms of radio sources by classification in different classes.
- Reduction in time to generate scientific results by radio astronomers.
- Deeper insight into topological representation of radio data during classification.

3 Approach

3.1 Source Modelling

The first step would be source extraction using the standard technique of gaussian modelling. We propose to do this using the robust PyBDSM pipeline used for fitting gaussian distributions to radio sources. The software contains a plethora of features, from which we would be using a small subset. This would mainly include:

1. Source extraction using gaussian modelling of radio data.
2. Generation of a catalog file containing details of radio sources (RA, DEC, Size of Gaussian (min, max), etc.)

3.2 Cutout Generation

The second step would be to convert the RA(Right Ascension) and DEC (Declination) values generated from the catalog, to their corresponding pixel values in the original image. Based on these pixel values we generate 10*10 px cutouts using as reference the co-ordinates of the center of the radio source. This involves a multistep procedure briefly including:

1. Using the astropy module to read the FITS image in the form of a matrix
2. Parsing through the generated catalog file using Pandas, and extracting data for each radio source such as RA, DEC, etc.
3. Converting the RA, DEC values using in-built functions in astropy to convert from WCS to pixel values.
4. Processing pixel values to account for difference in addressing between FORTRAN and C family of languages.
5. Slicing the image matrix assuming the reference pixel co-ordinates as the center of the source.
6. Storing the generated image cutout in a standard image format (JPEG), with name as (RA,DEC) values.

Prototype code for section *3.1* and section *3.2* has been written mainly for testing purposes. We used a sample image from the TGSS survey which was then processed using the first two steps of our pipeline to generate 470 cutout images. More details may be found at: https://github.com/NCRA-TIFR/radiogen.

3.3 Data Preprocessing

- Image Processing techniques

3.4 Analytical Approach

Two broad steps that we plan to use:

- Statistical modelling of data to manually extract features. We plan to employ Scale-invariant feature transform (SIFT) algorithm to detect the features.
- Classification of the radio galaxies based on these extracted features. Possible approaches : Naive Bayes, SVM and Random Forests.

3.5 Empirical Approach

We plan to deploy a Convolutional Neural Network model for classification which reduces the manual feature engineering part, and has achieved significant successes in object recognition and image classification tasks. (Give References to papers)

4 How we predict it will solve the problem

5 Timeline

- 26th April to 11th May - literature survey
- Mid-August to October - working on basic prototype model individually by trying out multiple approaches
- October to November - choosing the approaches which work, and implementing them on all data, validation of the results.
- November to December - Refining the system, cleaning and commenting the code

6 Conclusion

We would like to thank Dr.Yogesh Wadadekar [1], Dr.C. H. Ishwara Chandra [1] for their supportive presence during the process of brainstorming potential research ideas. Their constant guidance has been an invaluable source of inspiration for us, and we are eager to continue working with them.

[1] National Center for Radio Astrophysics - Tata Institute of Fundamental Research