Supervised Machine Learning Techniques for Radio Source Classification

A project submitted in partial fulfilment of the requirements for the degree M.Sc. in the Department of Physics and Astronomy, as part of the National Astrophysics and Space Science Programme

University of the Western Cape

Author: Chaka Mofokeng

Supervisor: Prof. Mattia Vaccari

Co-Supervisor: Prof. Russ Taylor

Co-Supervisor: Prof. Mario Santos

March 2023

Declaration

I, Chaka Mofokeng know the meaning of plagiarism and declare that all of the work in the thesis titled “Supervised Machine Learning Techniques for Radio Source Classification” is my own, except for that which is properly acknowledged.

Acknowledgements

I would like first to express my deepest gratitude to my supervisors, Prof. Mattia Vaccari, Prof. Russ Taylor and Prof. Mario Santos, for their support and constructive advice over the years. In all sincerity, this journey was very challenging not only academically but in almost every aspect of life. However, they were there whenever needed to help and to provide resources and valuable insights that made it possible to endure this journey and also complete this research project successfully. Without their guidance, this study would not have been possible to carry out.

I would also like to thank the staff and my fellow post-graduate students of the Department of Physics and Astronomy at UWC, for their insightful comments and encouragements. Furthermore, I would like to thank NASSP (National Astronomy and Space Science Programme), IDIA (Inter-University Institute for Data Intensive Astronomy) and CRC (Centre for Radio Cosmology) for offering financial support for the duration of my MSc.

Also, thanks to Dr. O. Ivy Wong for granting me access to the first data release of the Radio Galaxy Zoo, her support and insightful conversations.

Thanks to Tekano Mbonani, Oarabile Moloko, Themba Gqaza, Orapeleng Mogawana, Michael Hlabathe and our (house) journal club members at large for their support and guidance in most aspects of life, and also making the duration of my MSc not only bearable but fun. Also thanks to Masego Peteke for her help with proofreading this thesis.

Lastly, I would like to give gratitude to my family, for their unfailing support and for being my constant source of inspiration, even through hard times, they were always there to support me.

Thanks to all!

Abstract

Classification is one of the most fundamental aspects of scientific investigation. Astronomers have thus developed several classification schemes to try and make sense of the evolving properties of planets, stars and galaxies. One of the most popular ways to classify galaxies is according to their shape, or morphology, which has long been performed visually to produce annotated galaxy catalogues. However, visual inspection and manual annotation by astronomers will not be able to keep up with the expected data flow from next-generation sky surveys.

In this context, the main objective of our study was to use deep learning to automate radio source characterization (that is detection, classification and identification) from image data efficiently. We adopted a pre-trained deep learning model called CLARAN (Classifying Radio Sources Automatically with Neural Networks) based on the Radio Galaxy Zoo Citizen Science Classification Project and applied it to a GMRT 610 MHz survey in the ELAIS-N1 region covering an area of 12.8 square degrees at a resolution of approximately 6 arcsec at a root-mean-square noise of about 40 μJy/beam.

We successfully applied transfer learning and confirmed via visual inspection that the completeness of our source characterization algorithm is better than the completeness of PyBDSF in most cases, and especially for faint and extended radio sources. Moreover, we computed an estimate of CLARAN's performance in detecting and correctly classifying extended radio sources and found that we achieved 78% completeness (recall) and 92% reliability (precision). Furthermore, we implemented a cross-identification algorithm to pinpoint the infrared counterparts of our radio sources. We thus turned a pre-trained deep learning model into a robust automated radio source characterization pipeline. Such a tool will be very useful when dealing with wide-area radio surveys such as VLASS and EMU and eventually SKA1.

Qapolo

Ketso ya ho arola dintho tse itseng ho ya ka dihlopha ka ho fapana ke emeng ya ditsela tsa motheo ha ho tluwa dipatlisisong tsa mahlale mme balepi ba dihlodilweng sepakapakeng ba ahile mekgwa e mmalwa e le ho leka ho arohanya leho utlwisisa maemo a mafatshe, dinaledi le dinkgume tsa dinaledi, marole le kgase. O teng mokgwa hara mekgwa ena oo ho ona dihlopha di arohangwa ka chebeho, eleng mokgwa o sa le o sebediswa ho tloha kgale mme o sebebedisetswa ho lekola le ho hlahisa lenane la dipalopalo tsa dinkgume. Le ha ho le jwalo, ketso ya balepi ba dihlodilweng tsena ya ho leka di arohanya ka ho di lekola ka bonngwe ka bonngwe, ebe ba hlahisa manane a dipalopalo ekeke ya kgona ho etswa ka potlako ho tshwana le sekgahla sa tsebo le dipalopalo tse lebelletsweng ho tswa dibonelaholeng tsa nako e tlang.

Ka moelelo ona, sepheo sa sehloho sa dithuto le dipatlisiso tsa rona e ne e le ho sebedisa ithuta-botebo (mokgwa wa ho ruta khomphuta ho phetha mosebetsi seka motho) ho leka hore tekolo le ho arohanya dihlopha tse fapaneng tsa dihlodilweng mahodimong, tse fanang ka mahlasedi a bonwang ka dibonelahole tsa maqhubu a radio e iketsahalle ka tsela ya mmankgonthe. Re sebedisitse se sebediwa sa ithuta-botebo se ileng sa rutwa pejana se bitswang CLARAN (Classifying Radio Sources Automatically with Neural Networks) e itshitleileng ho Radio Galaxy Zoo Citizen Science Classification Project, ho feta moo, re sebedisitse sebonelahole sa GMRT ho lekola mahlasedi a maqhubu a radio a 610 MHz ho tswa lebatoweng le ka kwahelang dikgato tse 12.8 ka bophara le bophahamo la ELAIS-N1, mme ka boleng ba setshwantsho bo lakanyeditsweng ho 6 arcsec le katiso-palohare e ka lekanyetswang ho 40 μJy/beam, e tswang mahlaseding a lerotho ho tswa lebatoweng lohle.

Re atlehile ka tshebediso ya thuto-neheletsano mme ra pakahatsa ka ho lekola ka bo rona, mme ra lemoha hore hangata moralo wa rona wa karohanyo ho fana ka sephetho se ntlafetseng hofeta sa PyBDSF, haholo ho dihlopha tse lerotho le tse phakalletseng. Ntle le moo, re lekantse tshebetso ya CLARAN bakeng sa fokisa le ho arahanya dihlopha tsa mahlasedi a radio tse phakellesteng, mme re fumane hore ka botlalo ba 78%, re fihlela botshepehi ba 92%. Ho feta moo, re kentse tseleng moralo o nepahatsang setho ka seng sa dihlopha tsena ho bomphato ba tsona ba fumanweng ka maqhubu a infrared. Ka hoo, re fetotse se sebediswa sa ithuta-botebo se retilweng pejana ho ba sesebediswa se iketsetsang karohanyo le ho nepahatsa ditho tse famanang ka mahlasedi a radio. Se sebediswa sa ho tshwana le sena se tla ba bohlokwa haholo bakeng sa dibonelahole tsa nako e tlang tsa ho lekola sebaka ka ho nama ha sona, hara tsona re ka qolla tsa mahlasedi a radio jwalo ka VLASS le EMU mme sethathong SKA1.

Contents

List of Figures

List of Tables

Chapter 1

1 Introduction

In the early days of scientific investigation, observing the sky with the "naked" eye and later with a small telescope gave rise to astronomy, where an observer would monitor and draw, or map, by hand, the positions and motions of celestial objects on the sky. Due to advances in technology over the years, telescopes have since improved, and more information about the observed sources can be gathered, such as their brightness and distance, and thus their luminosity. Today, ever larger telescopes and two-dimensional digital detector arrays are utilized to map or survey the sky and are thus transforming the way we do astronomy. Imaging surveys have become more effective, and are thus widely used to probe the universe near and far. Also, image data can be used to derive several additional properties for detected sources (Djorgovski et al., 2013). Most importantly, images show the spatial structure of the sources, which is often used to visually classify them according to their morphology.

The objects that astronomers study (e.g. stars, planets, and galaxies) often emit radiation at different wavelengths of the electromagnetic (EM) spectrum, depending on their physical properties such as e.g. temperature and density. A lot of unique and valuable information is thus carried by radiation from each part of the EM spectrum. For instance, when analyzing extragalactic sources, gamma-rays and X-rays are used to gather information about the high-energy processes (e.g. material accreting into a black hole). Optical wavelengths are best used to show the morphological structure of the different types of galaxies, while radio images e.g. show huge jets and lobes emanating from the centre of the galaxy. Infrared light is mostly used to see dusty star-forming regions. Therefore, multi-wavelength studies are crucial in order to understand the physical properties of the astronomical objects.

1.1 Radio Surveys

Sky surveys, i.e. coordinated observations of the sky over wide areas, have greatly benefited from recent technological advances, and have transformed the way astronomy is done by producing very large datasets which can be put to use for a variety of scientific purposes. A good review on sky surveys and their impact is provided by Djorgovski et al. (2013).

There are now sky surveys that map the universe in almost all wavelength regimes of the EM spectrum. As a result, they are categorized in terms of their scientific motivation, whether space- or ground-based, their depth and areal coverage, whether panoramic or targeted, among other properties.

Some of the most interesting sky surveys are those mapping a large area of the sky. This requires the collecting area and the resolution of the detector array to be correspondingly large. However, there is a limit to the level of detail that the images produced by sky surveys realized with an individual telescope can achieve. This limit was first derived by George Airy in 1831 by considering the wave nature of light and following the diffraction process of light from Young's double-slit experiment. Airy proposed that light emitted from a point-like source and observed with a telescope forms concentric rings that are bright at the center - the Airy disk, and dimmer along the radial distance from the center. The angular radius, θ_A – measured in radians, depends on the wavelength of light λ and the diameter of the given telescope D as:

$$\theta_A = \frac{1.22\lambda}{D} \tag{1}$$

Eq. 1 poses a problem to observe the sky at long wavelengths, since at these wavelengths the telescope diameter must be very large to detect small angular features. For instance in the optical (e.g. at a wavelength of 500 nm) a telescope must be of 12.6 cm diameter in order to resolve details with an 1 arcsec resolution. In the radio, e.g. at a wavelength of 21 cm corresponding to a frequency of 1.4 GHz, a single-dish telescope must be of 52.8 km diameter to obtain a resolution of 1 arcsec. Building such a huge single telescope is impossible, thus radio astronomers build telescope "arrays" to solve this. All telescopes, or "dishes", within a radio telescope array are linked together to make them work like a single large radio telescope. This technique is known as interferometry. Signals from all dishes are combined to form an output signal not unlike the one which would be obtained from a single much larger telescope. To a first approximation, the effective diameter of a telescope array equals the maximum separation between any two dishes in the array, the further the distance between them, the better the angular resolution. However, an array must also be made by a large enough number of dishes in order to produce high-quality image data. The process of combining these signals to form high-resolution images is known as aperture synthesis. Most modern radio telescopes are therefore built as interferometric telescope arrays, and the SKA will also be such an array.

The next-generation of radio surveys to be performed with the Square Kilometre Array (SKA) will build upon these developments. SKA "precursors" such as MeerKAT (Jonas and MeerKAT Team, 2016), the Australian SKA Pathfinder (ASKAP; Johnston et al., 2008) and the Murchison Wide-field Array (MWA; Tingay et al., 2013) are already in operation (Norris, 2017a), along with several SKA "pathfinders". Not only are they changing the way astronomy is done, but they are also transforming the way we process the data (Norris, 2011). For example, the ASKAP produces about 200 terabytes (TB) of data per day which is sent to a supercomputer for further processing. This process further increases the data rate to produce calibrated data, mostly in form of images and catalogues, of about 70 petabytes (PB) of data per year. As a result, these sky surveys bring challenges in terms of data processing. Traditional methods whereby astronomers interactively visualize and analyze the data will not be able to keep up with such a data flow. Therefore, automated and robust data processing methods are required.

In the following we introduce some of the most important SKA-mid precursors and pathfinders along with the large survey projects to be undertaken with them. These facilities are located around the world, and they help scientists prepare for the SKA in terms of science and technology. This is not a comprehensive view of all such facilities, but we focus on those whose data was used as part of this thesis and/or for which the techniques developed in this thesis may be useful.

1.1.1 JVLA Surveys

The Very Large Array (VLA) is an array of 27 radio telescopes (antennae) shown in Figure 1, located in New Mexico, USA (Thompson et al., 1980). The telescope was inaugurated in 1980 but has recently been upgraded, and is now also known as the Jansky Very Large Array (JVLA), named after the radio astronomy pioneer Karl Jansky. It is a radio interferometer whose telescopes are arranged in a Y-shaped array. The telescopes can be controlled by being moved across the rail tracks to a predefined set of configurations, known as A, B, C and D, and aperture synthesis can be performed with up to 351 baselines. The maximum and minimum baseline configuration is 36 and 1 km, respectively. It operates in a frequency range of 1 - 50 GHz, reaching a maximum angular resolution of 1.4 and 0.04 arcsec at 1.4 and 50 GHz, respectively (Perley et al., 2011).

The JVLA was the first modern powerful radio interferometer to enter operations, and it has therefore been used for a number of pioneering wide-area radio surveys.

Figure 1: Image of a few radio antennae (dishes) of the Jansky Very Large Array in Plains of San Agustin, New Mexico. The rail tracks on the ground are used to control the baseline (resolution) of the survey. Image courtesy of `http://www.vla.nrao.edu/`

The National Radio Astronomical Observatory (NRAO) Very Large Array (VLA) Sky Survey (NVSS; Condon et al., 1998) was the first one, covering the full sky visible from the VLA, that is 30,000 deg^2 square degrees. The NVSS observes the sky at 1.4 GHz with a sensitivity of about 0.45 μJy at an angular resolution of 45 arcsec.

The FIRST (Faint Images of the Radio Sky at Twenty-cm; Becker et al., 1995) followed, also observing at 1.4 GHz, while improving resolution and sensitivity with respect to NVSS and covering the same sky area covered by the Sloan Digital Sky Survey (SDSS), or 10,000 deg^2. FIRST maps the sky with an angular resolution of 5 arcsec at a typical root-mean-square (RMS) value of 0.15 mJy.

More recently, the Very Large Array Sky Survey (VLASS; Lacy et al., 2020) is currently being conducted by the JVLA, covering the same 30,000 deg^2 sky area covered by NVSS. VLASS observations began in 2017 and are expected to finish in 2024. VLASS cover a wide range of frequency bandwidth of 2 - 4 GHz and observations will be carried out at three epochs, separated by about 32 months with total sensitivity of 70 μJy at an angular resolution of 2.5 arcsec. Due to the fact that the survey will cover large portions of the sky that overlap with other SKA pathfinders and precursors thus allowing multi-frequency and multi-resolution studies of the same region of the sky to be carried out. Therefore this survey is complementary to the other SKA precursors and pathfinders to be discussed in sections to follow.

1.1.2 GMRT Surveys

The Giant Metrewave Radio Telescope (GMRT) is located in Pune, India. GMRT has 30 controllable parabolic dish antennae – each of 45 metres in diameter, spreading over up-to 25 km of distances shown in Figure 2. The 30 dishes results in 435 baselines. The dishes are set up to achieve high angular resolution and also have the ability to map diffuse extended regions. The array observes in six different frequency bands centred at 50, 153, 233, 325, 610 and 1420 MHz. GMRT achieves a wide range of resolutions depending on the frequency band, achieving about 60 arcsec at 50 MHz and 2 arcsec at 1.4 GHz.

The GMRT telescope has been undergoing upgrades since 2010, whereby the main goal is to improve the sensitivity and the frequency coverage. The recently inaugurated upgraded GMRT (uGMRT) has much-wider bandwidth and correspondingly higher sensitivity than the original GMRT (Gupta et al., 2017). The uGMRT maximum angular resolution in Band-3 (400 MHz) and Band-4 (700 MHz) is about 6" and 3", very close to the maximum angular resolution of MeerKAT L-Band (1.4 GHz) and S-Band (3.0 GHz) respectively.

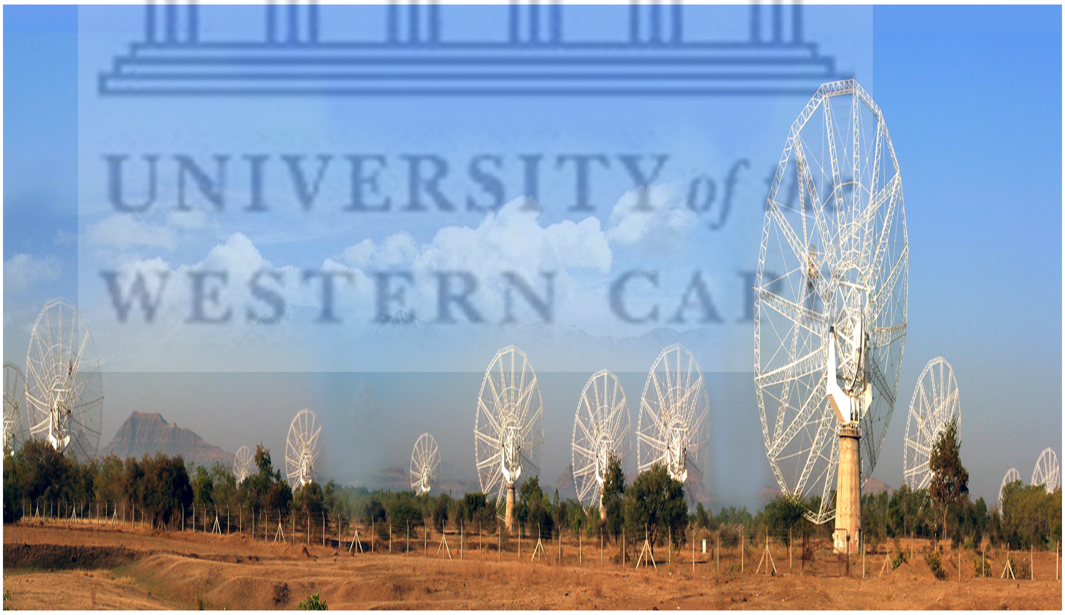


Figure 2: GMRT dish antennas in Pune, India. They cover a baseline distance of 25 km. Image courtesy of `https://www.skatelescope.org/news/indias-gmrt-telescope-becomes-ska-pathfinder/`.

1.1.3 ASKAP Surveys

The Australian SKA Pathfinder (ASKAP) is a radio telescope located in the Western Australian desert at Murchison Radio-astronomy Observatory (MRO) (Johnston et al., 2008). It has 36 parabolic dish telescopes shown in Figure 3, each of 12 metres in diameter, resulting in a total of 630 baselines. It operates in a frequency range of 700 - 1,800 MHz. The telescope reaches a maximum configuration baseline of ∼6 km, and it is able to reach an angular resolution of 10 arcsec at 1.4 GHz. The ASKAP feature outstanding multi-beam receiver arrays, which allows for the huge sky coverage and hence survey capability, for instance, rapid survey speed.

The Evolutionary Map of the Universe (EMU; Norris et al., 2011) survey will be carried out by the ASKAP to map the entire Southern Hemisphere and some parts of Northern Hemisphere (out to a declination of +30°). The EMU will be carried out at 1.3 GHz wavelength, at a sensitivity of $\sim 10\mu$Jy and an angular resolution of 10 arcsec.

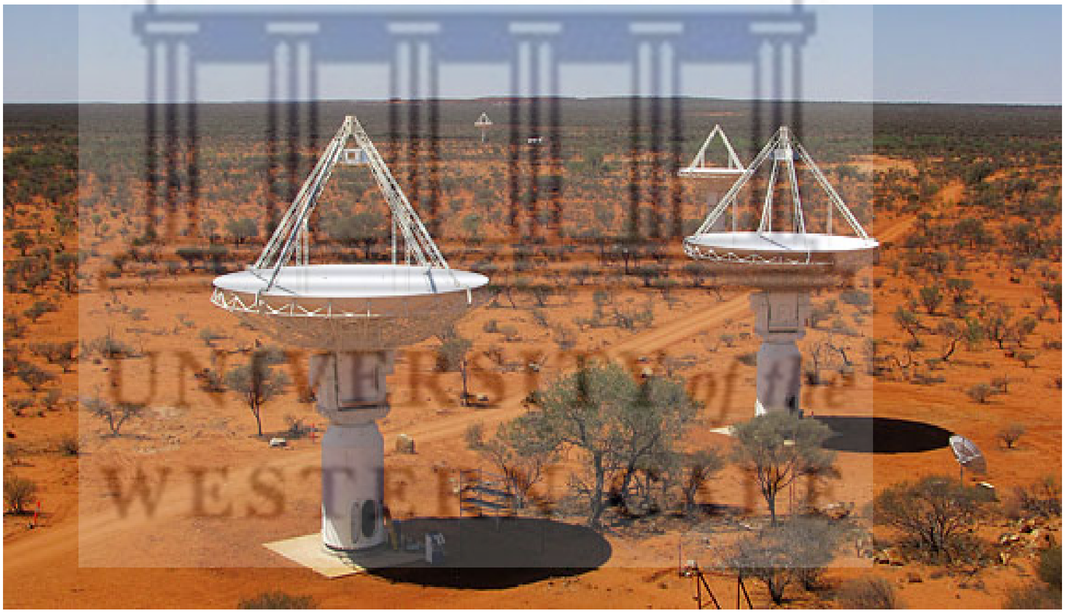


Figure 3: The ASKAP's antennae view. Image courtesy of https://www.skatelescope.org/australia/.

1.1.4 MeerKAT Surveys

The MeerKAT telescope, originally referred to as Karoo Array Telescope (KAT), is located within the Karoo desert in the Northern Cape Province of South Africa (Jonas and MeerKAT Team, 2016). MeerKAT currently has 64 dish telescopes shown in Figure 4, each of 13.5 metres. The 64 dishes amount to 2,016 baselines. It covers a frequency range of 0.5 - 4 GHz. It has minimum and maximum configuration baselines of 29 and 8,000 metres, respectively, and as such can deliver a resolution of 6 arcsec at 1.4 GHz. The array is fine-tuned for deep and high-quality imaging of low-brightness and diffuse emission (Jonas and MeerKAT Team, 2016). MeerKAT is an SKA1-mid precursor, and the current plan is for MeerKAT's 64 dishes to be incorporated within SKA1-MID which will consist of approximately 200 dishes in total.

The MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE; Jarvis et al., 2016) survey will map four well-studied fields in the southern hemisphere, namely XMM-LSS, ELAIS-S1, COSMOS and ECDFS, covering a total area of 20 degrees^2. The sensitivity will reach about 2 μJy in the L-Band (1.4 GHz), at an angular resolution of $\sim$ 6 arcsec.

Between other projects, the MeerKAT's MIGHTEE survey will be combined with the uGMRT survey, to map the same area on the sky. The resulting survey is called superMIGHTEE. SuperMIGHTEE will be an ultra-broad band survey with a frequency range from 0.25 - 2.7 GHz. superMIGHTEE's rms sensitivity in band-3 and band-4 is about 15 and 5 μJy/beam respectively. The survey will be one of the most sensitive surveys pre-SKA.

Figure 4: The MeerKAT telescope dishes arrangement in the Karoo, in the Northern Cape, South Africa. Image courtesy of `https://www.sarao.ac.za/gallery/meerkat/15/`.

Deep and wide-area sky surveys such as the ones mentioned above have had and will have a huge impact in our understanding of the Universe. They have also led to the rapid growth of data available to astronomers. Deep surveys cover smaller sky areas down to fainter levels, and as such they generally detect smaller number of sources compared to wide surveys. However only deep surveys can detect the faintest sources in the sky. Nowadays, surveys are often both deep and wide, therefore detect large number of sources with high sensitivities. Examples of deep and wide surveys are the SDSS and FIRST, which map about a quarter of the full sky. In order to be exploited to the full, the large quantities of data produced by these surveys must be analysed in a homogeneous and reliable manner. This has recently become increasingly challenging due to survey data rates, and it is now ever more important to develop automated data reduction pipelines. The reduction step is important to provide homogeneous data over large chunks of the sky to enable statistical studies of galaxy properties that are only limited by Poissonian errors.

1.2 Cosmic Radio Sources

Radio astronomy is mainly focused on studying emissions from astronomical sources spanning the frequency band between 10 MHz and 1 THz of the EM spectrum (Wang, 2017). The “birth” of radio astronomy was in 1932 when Karl Jansky reported that he observed a static radio signal using a 20.5 MHz antenna. This signal appeared to be coming from the plane of the Milky Way and it had a period of 23 hours 36 minutes (see Jansky, 1933). Reber, 1944 continued where Jansky left off, observing the radio sky at 160 and 680 MHz. He constructed the first contour maps of the radio sky and located the centre of our galaxy – the Milky Way – in Sagittarius.

1.2.1 Radio Emission

Emission of radio waves from astronomical objects in the sky is accounted for by the following processes:

- Thermal radiation - EM radiation from any object with temperature above absolute zero (0 Kelvin).

- Thermal bremsstrahlung radiation - EM radiation produced by electrons that are slowed down or deflected by atomic ions, causing electrons to lose kinetic energy and convert it into radiation.

- Synchrotron radiation - EM radiation from electrons traveling at speeds close to the speed of light (relativistic), in the presence of a magnetic field.

- Inverse Compton scattering - EM radiation produced by low energy photon scattered to high energies by relativistic electrons.

- Synchrotron self-absorption - EM radiation from electrons re-absorbing synchrotron radiation from within the sources.

- Atomic spin-flip - The hydrogen atom has protons and electrons that have a spin associated with them. As a result, they can be spinning in the same or opposite direction. When the direction of the spin is the same, it implies that they occupy a slightly high energy state than when the direction is opposite. About once in a few million years, an electron spin flips and as a result, emitting a radio photon with a wavelength of 21 cm.

- MASER (acronym for Microwave Amplification by Stimulation Emission of Radiation) - radiation from stimulated emission that excites atoms or molecules inside the gas clouds, producing a chain of reactions that amplify photons, thus radiation.

Radio astronomical observations were not very practical in their early days because a much bigger telescope was needed to match an optical telescope's angular resolution. Radio astronomy became more popular when it was indicated that antenna configurations could be set up similarly to Michelson's optical interferometer. As previously discussed in Section 1.1, in imaging the technique is known as aperture synthesis. Aperture synthesis was first introduced by Ryle and Vonberg, 1946 – who used this approach to measure the angular distance of the sunspots. Also, to produce high quality image data, different separations between different telescopes is required (one separation vector between two telescopes projected from a reference frame of a radio source is known as a baseline). The longer the baseline, the better the angular resolution (i.e., one is able to distinguish between two close radio sources). When the baseline is short, the resolution is less, however this gives information about the spatial distribution of the source in the sky. Therefore, most telescope arrays have tracks that allows astronomers to have many combinations of baselines and configurations. Astronomers gauge the quality of the data, by plotting virtual tracks that the telescopes trace out as the earth rotates, the plot is referred to as the uv-coverage plot. This flexibility of being able to use various telescopes and to change their positions and configurations became the standard approach in radio astronomy.

Observations done in the radio waveband have a unique advantage over observations done at other wavebands. Radio waves are not easily scattered or absorbed, but they can penetrate through thick layers of neutral gas and dust of the interstellar medium. Radio emission can penetrate from large extra-galactic distances, thus it has become an integral part of observational astronomy. Figure 5 shows the sky area versus sensitivity of modern (deep and wide) radio surveys, from the figure it is clear that modern radio surveys are exploring new parameter space, which will in fact increase the number of discoveries and thus provide us with large samples of galaxies. The large samples of galaxies produced by deep and wide radio surveys help us clarify the nature and evolution of different classes of cosmic radio sources.

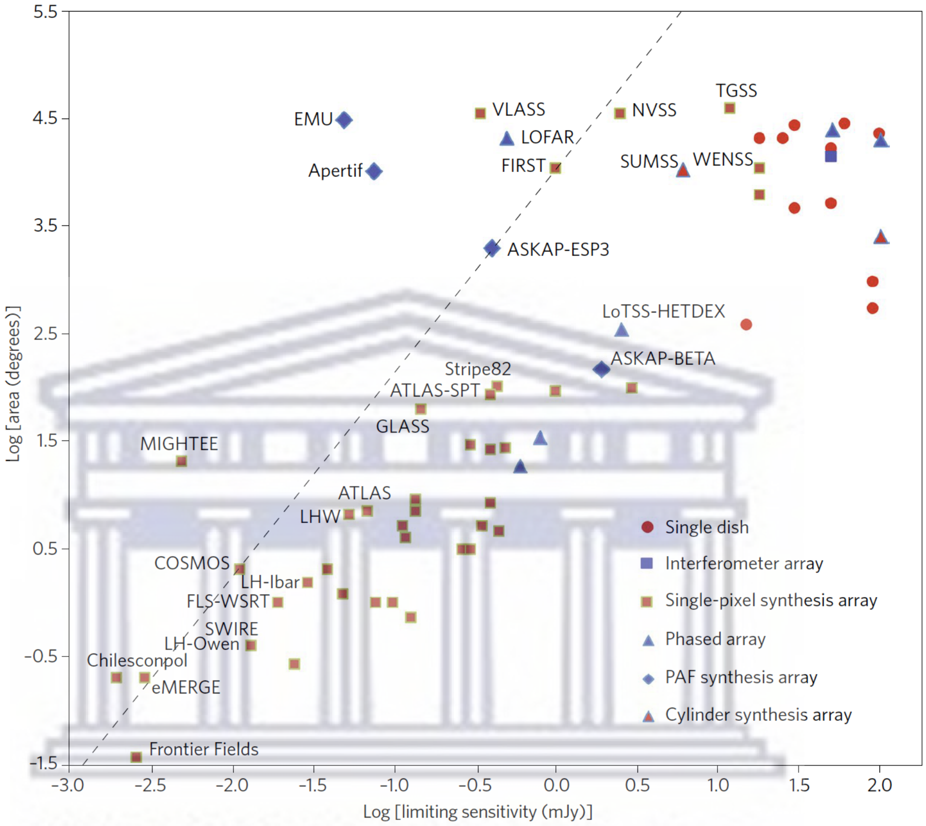


Figure 5: This figure shows the sky area as a function of sensitivity for modern telescopes. Sensitivity is given as either the quoted detection limit or five times sensitivity level. The dashed line indicate the limit of existing surveys (at the time of publication). The symbols represent the type of telescope used for the survey: red circle for a single dish; blue square for a non-synthesis interferometer array; red square for a conventional synthesis array; blue triangle for a phased array; blue diamond for a synthesis array using phased-array feeds (PAFs); red triangle for a cylindrical telescope; open circle for anything else. Image from Norris, 2017b.

1.2.2 Classes of Radio Sources

This subsection covers a brief review of the commonly-identified morphological classes of radio sources, their multi-wavelength properties, and the range of mechanisms involved in energy production (for more information, see Schneider, 2014, Padovani, 2016, Padovani, 2017b and Padovani, 2017a).

Radio sources display a radio spectrum that is complicated, involving a mix of thermal and non-thermal emission (see Section 1.2.1). At low frequencies, radio sources' spectrum follow a power law, defined as:

$$S \propto \nu^{-\alpha} \tag{2}$$

where S, ν and α represent the source flux density, the frequency and the spectral index, respectively. A spectral index (α) of 0.5 divides compact radio source from extended radio sources (Wang, 2017). Compact core sources display a flat spectrum, with $\alpha < 0.5$, whereas extended radio sources have a steep spectrum, $\alpha > 0.5$, usually associated with the synchrotron radiation from fast-moving electrons in a magnetic field. Compact sources' flat spectrum is a result of synchrotron self-absorption – where layers of the source become optically thick at some frequencies. There is an exception to this classification by spectral index when the sources are young radio sources, referred to as compact steep- and GHz peaked-spectrum radio sources. These sources are in their early phases of evolution. Therefore they will eventually evolve into extended radio sources. Some of the most common classes are detailed below.

In general radio sources to be discussed below are similar in terms of their central region, except for starforming galaxies, this region is where a large fraction of their energy comes from. Their source of energy found in central region is the supermassive black hole (SMBH) – with a mass of millions to billions times that of the Sun. Surrounding this central SMBH is dust and gas that forms a doughnut-like ring called a torus. The energy is emitted during accretion of matter from the torus onto the SMBH. As matter accretes onto a SMBH, its potential energy is converted into kinetic energy. Some parts of the kinetic energy are converted into heat due to friction. Subsequently, this heat is emitted as radiation that spans an extensive frequency range. Due to the origin of this powerful emission, such sources are referred to as active galactic nuclei (AGNs). Another important component of the AGN is the gas region responsible for the observed lines in the optical spectrum. There are two regions of gas that account for this near the SMBH, namely broad line region (BLR) and narrow line region (BLR). BLR is a region in which broad emission lines are produced whereas NLR is a region where narrow emission lines are produced. The BLR gas is located near the plane of the disk of the galaxy, whereas the NLR gas is located at large distances from the disk of the galaxy. These emission lines are often interpreted in terms of Doppler velocity. BLR gas have line widths of the order of $\sim$ 10,000 km/s, while the gas in NLR have line widths of the order of $\sim$ 400 km/s. The line width of gas in the BLR is called a Doppler broadening and this broadening is said to be due to strong gravitational fields. The emission lines in the NLR gas is due to UV-radiation, which ionizes this gas.

As matter accretes into a SMBH, this leads to high-velocity electrons being accelerated away from the centre of the black hole. These electrons are traveling at speeds close to the speed of light (relativistic), in the presence of the magnetic field. Thus, they emit synchrotron radiation that forms jet streams. These jet streams usually travel large distances and often interact with the surrounding gas forming radio lobes. Jets are slowed down by this process and are terminated to form hotspots on either side of the host galaxy. This often leads to AGNs having a morphologically complex structure. As a result, the classification of AGNs is complex and sometimes confusing. However, in general, different classes of AGNs are associated with their morphological appearance. The different classes are discussed below.

1.2.2.1 Seyfert Galaxies

The first AGN was detected in 1943 by Carl Seyfert (Seyfert, 1943). He studied several spiral galaxies and found that these galaxies have interestingly bright cores with faint arms as shown in Figure 6. These sources were later named after him, they are called Seyfert galaxies. The spectrum of the core of Seyfert galaxies shows broad and strong emission lines, broader than those of typical galaxies. The width of the emission lines is used to divide this class into two sub-classes, namely Seyfert 1 and 2. Seyfert 2 show narrow lines, whereas Seyfert 1 display both broad- and narrow-lines. Also, there are intermediate classes (e.g., Seyfert 1.5 and Seyfert 1.8) to this class, they are defined by the ratio of broad-to-narrow line flux. For this definition, these are sources that poses broad lines but having a smaller ratio broad-to-narrow line flux than Seyfert 1 galaxies.

1.2.2.2 Low-ionization Emission Line Regions (LINERs)

The low-ionization emission line regions (LINERs) represent the least luminous radio sources and the most abundant population of AGNs in the local Universe. These sources are classed based on their optical spectra that has low ionization energies on emission lines from neutral atoms and ions. LINER emission may also be associated with shock from central star-formation. Also, LINERs emission lines are narrower than the narrow emission lines from Seyfert galaxies. As one might expect, when using optical spectroscopic properties, LINERs fall under the class of LERGs, which will be discussed in the following paragraph.

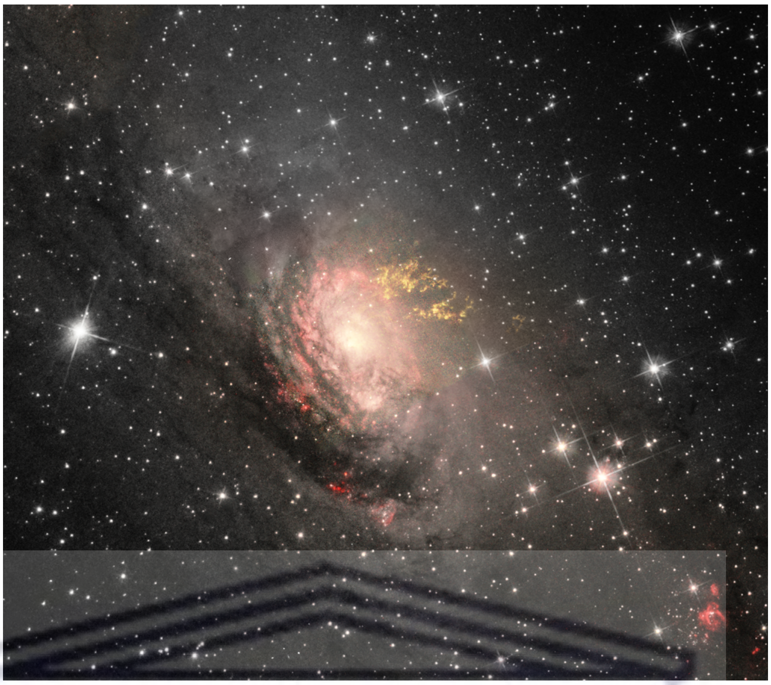


Figure 6: An optical image of Circinus galaxy from Hubble Space Telescope. It is located in the constellation of Circinus. This galaxy represent a spiral galaxy with faint arms, and a very bright reddish core at the centre. This galaxy belongs to type 2 Seyfert galaxy class. Image courtesy of https://www.wikiwand.com/en/Circinus_Galaxy.

1.2.2.3 Radio Galaxies (RGs)

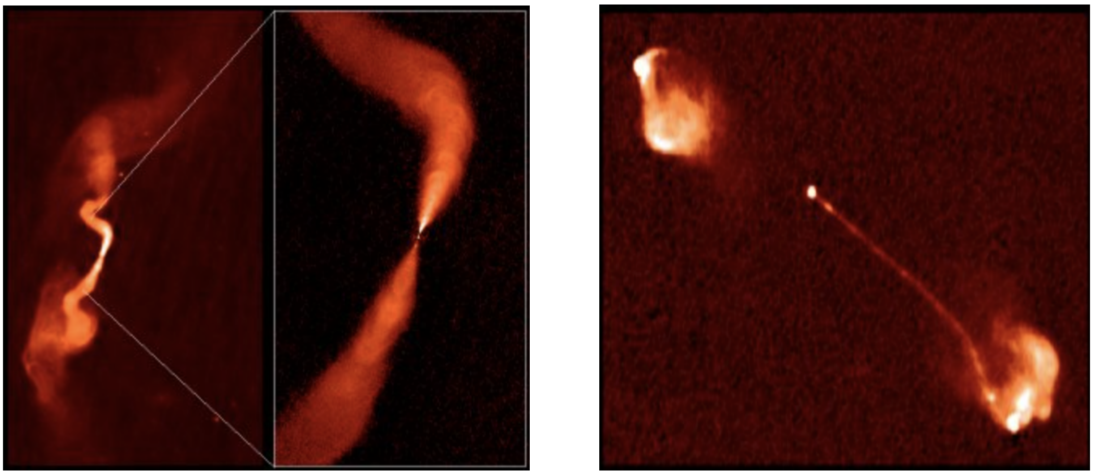
RGs are usually elliptical galaxies hosting an AGN. The AGN and the surrounding matter is often obscured by a dusty torus which absorbs radiation and re-emits it at other wavelengths. These sources are associated with relativistic jets extending well beyond the host galaxy and as a result they have radio powers $\gtrsim 10^{22}$ W Hz^{-1} when observed in the GHz waveband. This class is further divided according to spectral information at optical wavelength, resulting in two sub-classes, broad- and narrow-line RGs – BLRGs and NLRGs. As the names suggest, in BLRGs, broad emission lines are observed in their spectra. On the other hand, narrow emission lines are present in NLRGs spectrum. These are further attributed to the accretion rates of the black hole. The low accretion rate of the black hole produces low-excitation states in the NLR gas of the host galaxy, such an object is referred to as low-excitation RG (LERG). At the other extreme, the object is referred to as high-excitation RG (HERG). In LERGs, the accretion rate is low and thus the flows are radiatively inefficient. In contrast, the accretion rate is high and therefore fuels radiatively efficient flows in HERGs, and thus drives the ouflows efficiently. In recent literature, the two classes are referred to as jet-mode and radiative-mode AGNs. However, there is a significant overlap between these radio sources, as will become clear in paragraphs to follow. In paragraphs to follow, the morphology of different RGs when observed in radio waveband are discussed.

1.2.2.4 FR-I and FR-II Galaxies

Traditionally, RGs in this class have prominent emitting radio jets extending outwards, reaching radio powers $\gtrsim 10^{28}$ W Hz^{-1} at a wavelength of 1.4 GHz. These jets extend radially outwards from the central region (Fanaroff and Riley, 1974). Fanaroff and Riley, 1974 used the properties of the jets to morphologically classify such sources into two classes, namely FR-I and FR-II. FR-Is have low radio powers and are brighter near the center as shown in Figure 7a, while FR-IIs have higher radio powers and brighter edges as shown in Figure 7b. In other words, the morphology of these sources is correlated with their radio luminosity. Another distinction between the two classes is done on the basis of their radio power, where a dichotomy arises between the two classes at radio power of $\sim 10^{25}$ W Hz^{-1} at 1.4 GHz. In Figure 7b, which represents FR-IIs, the jets are faint but supersonic. Thus, the energy is efficiently transported to the lobes and terminates to form hotspots at the edges. Using optical spectroscopy, FR-IIs are observed to have high excitation lines, and thus they fall under HERGs. However, in Figure 7a, which represents FR-Is, the jets are bright and subsonic but inefficient energy transporters, and thus, the bright jets are observed at the center. As a result, using optical spectroscopy, they fall under LERGs. Some RGs have similar properties as FR-Is including the optical spectroscopic classification and nuclear luminosity, but instead, they are core dominated and lack extended radio emission. It was suggested that they represent a third class called FR-0 by Baldi et al., 2015, Padovani, 2016, Grandi et al., 2016 and Baldi et al., 2019. As previously discussed, FR-I sources usually have low luminosity jets terminating near their centers. This suggests that they interact with surrounding matter since they form plumes, and it is indicative that they are found in dense regions (near the centres of clusters). This effect is usually observed when a low luminosity radio source passes through a cluster of galaxies, resulting in a bent or warped shape of the radio galaxy and often referred to as wide-angle tail (WAT), narrow-angle tail (NAT), and X-shaped RG.

1.2.2.5 Quasars and Blazars

Quasars are compact radio sources that appear like stars in the optical sky, shown in Figure 8. This is due to their orientation in the sky. In Figure 8, the source appears very bright, even outshining the host galaxy. Also, there is a stream of



(a) 3C31

(b) 3C175

Figure 7: The VLA images of Fanaroff-Riley (FR) sources. Images (a) and (b) represent FR-I and FR-II sources, respectively. FRI source was observed at 1.4 and 8.4 GHz at a resolution of 5.5 and 0.3 arcsec, respectively. FR-II source was observed at 4.9 GHz at a resolution of 0.3 arcsec. Images taken from `https://www.cv.nrao.edu/~abridle/bgctalk/node4.html`.

particles (jet) top-right of the central quasar, the jet is usually oriented at a small angle with respect to the observer's line of sight. Observations support the fact that some RGs are seen to have larger projected jets sizes compared to quasars. Due to the orientation of the observer relative the the jet axis, and the fact that the jet may be relativistic, the apparent luminosity of the source will be modified – making it appear brighter than it really is. This process is known as beaming (or Doppler beaming/boosting). Also, there is evidence of Doppler broadened lines observed in their optical spectra. This is a confirmation that the observers view of the central region is not blocked and thus can see the BLR. The beamed radiation may also affect optical/UV spectrum. In such cases, the line emission may be completely outshone and the source will appear as a BL Lacertae. BL Lacertae objects (BL Lacs) are quasar-like objects with strong varying radiation and often featureless spectrum. They are also characterized by weak absorption derived from the host galaxy with $\alpha < 0.5$, showing a sign of radio compactness. In optical wavelength, the luminosity of BL Lacs varies over a long period of time, and sometimes during the period of low luminosity, their spectra show emission lines. During this period, the BL Lacs appear like an Optically Violent Variables (OVVs), also referred to as flat-spectrum radio quasars (FSRQs). Quasars, BL Lacs and OVVs are collectively called blazars. Blazars in general are another class of radio sources that have AGN hosting jets oriented a small angle ($<15°$ - $20°$) with the observer's line of sight. As a result, they have their central core brighter than the host galaxy. In terms of excitation states of the optical spectroscopy, blazars and Seyferts fall under HERGs.

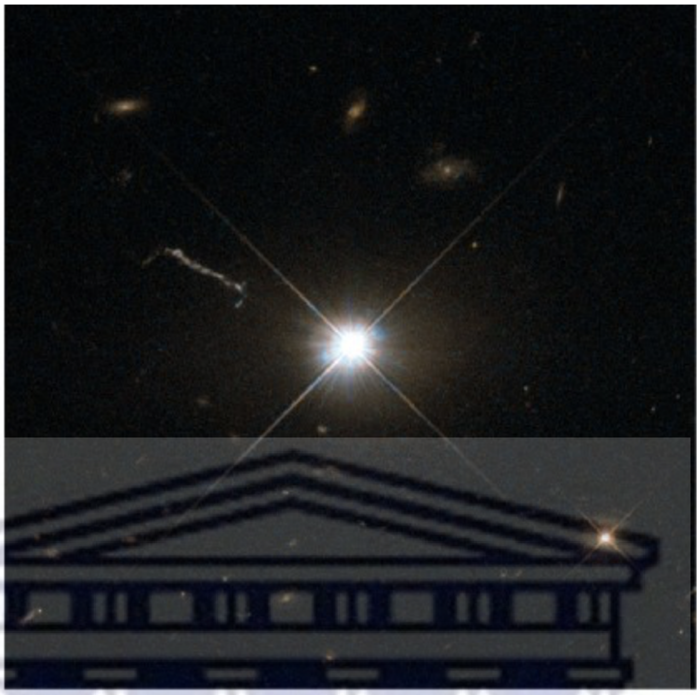


Figure 8: An optical image of a quasar 3C273, taken by Hubble Space Telescope. The central bright source is a quasar, and towards the top-left there is a jet being fired by this quasar. Image courtesy of https://www.nasa.gov/content/goddard/nasas-hubble-gets-the-best-image-of-bright-quasar-3c-273/#.XgWMWC17HOQ.

1.2.2.6 Radio-Quiet and Radio-Loud AGNs

When AGNs are studied in the radio, two classes can be distinguished – radio-loud (RL) and radio-quiet (RQ) AGNs. RL AGNs have been already discussed – FR-I, FR-II, flat and steep spectrum radio quasars, BL Lac, BLRG, and NRLG. All these radio sources emit a large amount of non-thermal radiation that is associated with relativistic jets. RL AGNs are also known as type 1 AGNs, while RQ AGNs are type 2 AGNs. RQ AGNs show faint core AGN emission but no strong radio jet(s). RQ AGNs have optical properties similar to those of quasars – they are observed at high redshifts and they show strong and broad emission lines. Note that these sources do show radio emission when observed with sufficiently sensitive instruments, thus the "quiet" in RQ may be misleading. RQ AGNs have luminosities higher than that of Seyfert galaxies. There is not much difference between Seyfert 1 galaxies and RQ AGNs, except the difference in the luminosity of their the cores. As a result, RQ AGNs and Seyfert 1 galaxies are collectively referred to as type 2 AGNs.

Extended radio galaxies that exhibit multi-radio components are sometimes not connected; this is due to the jets fading and resulting in distinct lobes. Padovani, 2016, suggested that RL and RQ labels should be dropped since they are misleading and RQ sources do indeed give off radio energy. They should instead be labeled as jetted and non-jetted AGNs. Because one of the main differences is that the jetted AGNs show strong, relativistic jets, whereas non-jetted AGN displays a radio structure that has small, weak, and slow jets.

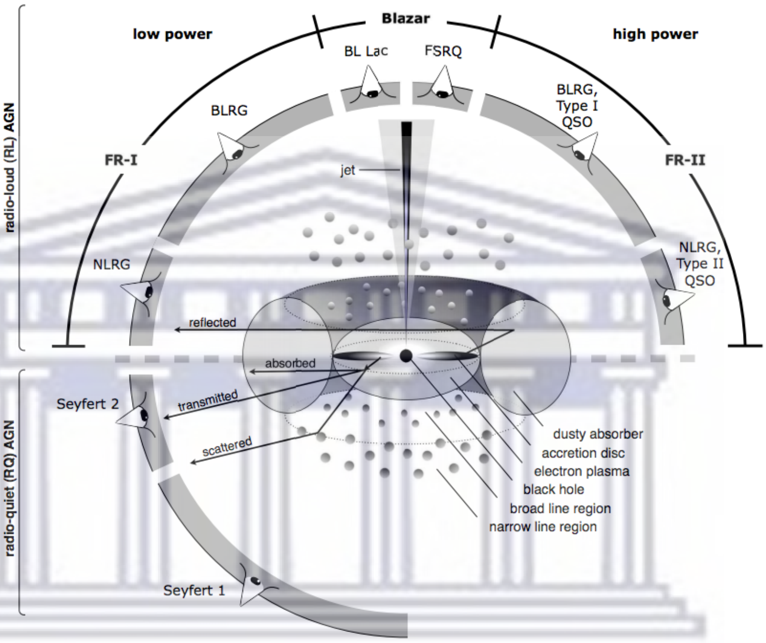


Figure 9: Diagram showing a representation of the unified model of AGNs. The distinction between most of the different classes is based on the viewing angle of the observer. Image from Beckmann and Shrader, 2012.

The same source generally powers AGNs, matter accreting onto the central super-massive black holes (SMBH, with mass $> 10^6 M_\odot$). Different viewing angles are used to unify different classes of AGN as shown by Figure 5. This figure shows different types of AGNs, previously discussed. It further shows that the classes depend on the viewing angle of the observer, whether the observer is placed closer to the disk, in its axis direction or closer to the plane of the disk. When the observer is near the plane of the disk, BLR gas is obscured, only the NLR gas is observed. As a result an observer will see a type 2 AGN. Type 1 AGNs are observed when an observer is placed in a direction closer to the axis of the disk, where an observer is exposed to the BLR gas. Also, the jet appearance depends on the closeness of the jet axis to the line-of-sight to an observer. The radiation from other components of the AGN is completely outshone by jet emission, in case the jet points to an observer. If the jet axis is at high inclination angles to an observer's line-of-sight, a pair of jets extending from the central source will be observed. The appearance of the AGN is also affected by the accretion rate as discussed previously. At high accretion rates, the radiation from the accretion disk dominates the luminosity of the AGN and thus the jet is launched efficiently. At low accretion rates, the radiation from the disk is not efficient and as such some of this energy in transferred inwards, subsequently launching an outflow. It is important to note – even if not indicated on the figure, that jetted AGN generally display jets on either sides of the accretion disk. However, few sources have faint lobes on one side. The energy radiated by the central engine is accounted for by different processes - which have been discussed previously, that covers several orders of magnitude in physical size as shown in Figure 10. In Figure 10, the cross-section of a quasar in log scale - distance given in gravitational radii of the central SMBH with mass $> 10^8 M_\odot$, together with log angular size of a luminous quasar at the redshift (z) of one are shown (see Moustakas et al., 2019).

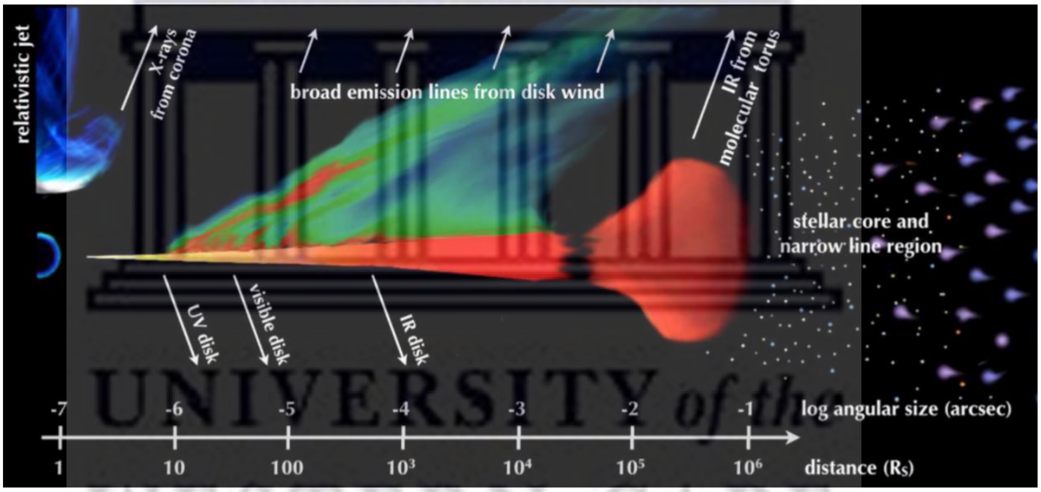


Figure 10: Diagram showing a multi-wavelength emission from a typical quasar at z=1, showing a cross-section (distance in gravitational radii of the central SMBH) and log angular size. Image from Moustakas et al., 2019.

1.2.2.7 Star-Forming Galaxies

Another class of sources that populate the radio sky are starforming galaxies (SFGs). SFGs are less powerful as compared to the radio sources discussed previously. Their radio powers reach an order of $\sim 10^{24}$ W Hz^{-1} at 1.4 GHz. Similar to powerful RGs, SFGs are observed to have a steep radio spectral index ($\alpha > 0.5$) which results from the synchrotron radiation from fast-moving electron in the presence of a magnetic field. However, the source of energy for SFGs is not the SMBH, but the supernova remnants as result of massive stars ($M > 8M_\odot$) that explode, ejecting material to the interstellar medium and resulting in shock waves. In optical wavelength, starforming galaxies are observed to be housed by spiral and irregular galaxies. Also, note that stars form in molecular clouds, as a result the optical spectra of SFGs have the presence of molecular emission lines (e.g. carbon monoxide, CO). In spiral galaxies, star formation occurs often, and it is observed even today (Schneider, 2014), e.g. the star formation rate of the Milky way is ∼1 $M_{\odot}$/year, similar to the Andromeda galaxy (M31). Figure 11 shows a spiral galaxy NGC 1559 as observed by the Hubble Space Telescope. The blue regions that follow the track of spiral arms in this figure represents regions of active star formation. Also, it is clear that some of these regions are obscured by dust which limits the application of optical wavelength as a probe to star formation. However, the dust in these regions absorbs UV radiation from young and hot stars and re-emits it in the far-IR (FIR) wavelength, also radio wavelength is able to penetrate dust. As a result, optical information must complemented with radio and IR information in order to fully understand SFG. In addition to the "normal" star formation rate in spiral galaxies, there exist a class where the star formation rate is of the order of ∼100 $M_{\odot}$/year, such galaxies are called starburst galaxies. The argument for such high star formation rate is proposed to be due to interacting or merging galaxies.

Radio sources are very large – extend up to mega-parsecs (Mpc) in size, however, they are observed at cosmological distances. As a result, they are observed in arcsec scales. Therefore, to classify these sources according to their morphology is a difficult task, as these small regions have to be visually inspected. Radio image data alone offers only part of the information about the physical properties of the source, as shown in Figure 10. However, because radiation from AGNs span most of the electromagnetic spectrum, radio information can be cross-identified with information from other wavelengths. As shown in in Figure 10, infrared (IR) radiation is mostly from dust and obscuring material – that absorbs radiation at optical and ultraviolet (UV) wavelengths, then re-emit it in IR wavelength. The optical and UV radiation is from the accretion disk, the X-ray emission is from the hot corona. Strong non-thermal radiation is usually associated with the jets and lobes is observed in the X-ray band. Thus, to fully understand any radio source, their respective radio data is usually cross-identified with observational data taken at other wavelengths of the EM spectrum – this is also known as data fusion. For example, the optical emission as probed by the SDSS survey is affected by the dust around radio sources, that absorb most of the emission, and this often leads to the misidentification of these sources. However, the FIRST survey was developed to map the same sky area as the SDSS so that complex radio sources with multiple components could also be studied in the optical, thus

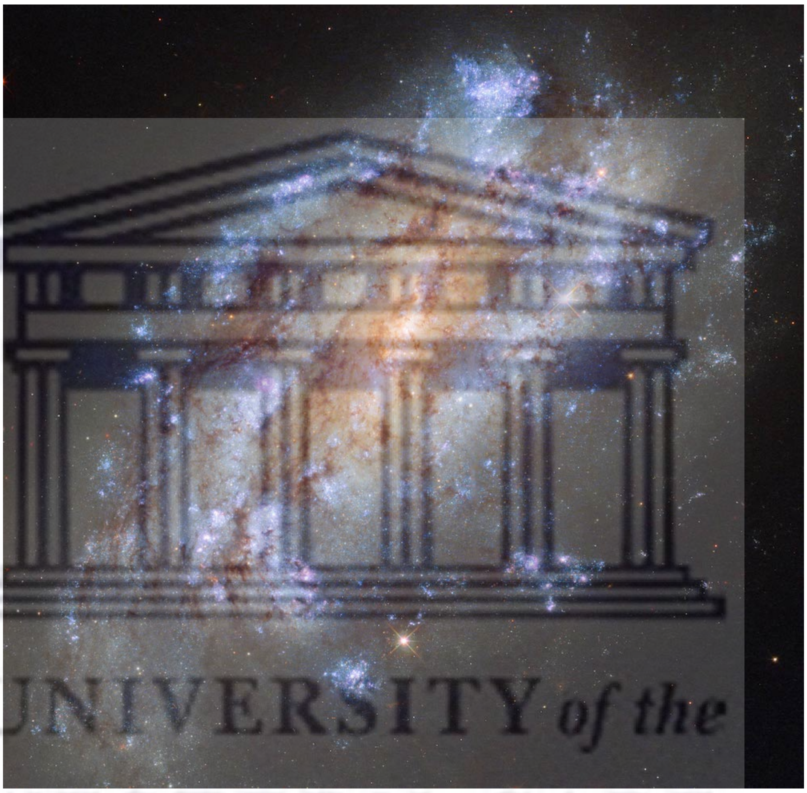


Figure 11: Spiral galaxy NGC 1559 as observed by the Hubble Space Telescope. This galaxy represents an example of a local starforming galaxy. The star formation regions are represented by the blue regions along the spiral arms of the galaxy. Image from `http://www.sci-news.com/astronomy/ngc-1559-spiral-galaxy-massive-star-forming-arms-05692.html`.

improving our understanding of radio sources.

Until recently, traditional methods of classifying radio sources were still being employed, whereby astronomers would visually inspect individual sources to classify them. Then the process that follows is cross-identifying with observational data taken at other wavelengths. For instance, infrared (IR) data would be used to find information about the host galaxy since the infrared emission from the host is galaxy is more concentrated and thus more easily pinpointed than the extended radio emission. Sometimes this task becomes very challenging when radio sources are large and complex like the FR-Is and FR-IIs. As a result, this process it is often time-consuming and thus unpractical for large samples. Citizen Science (Marshall et al., 2015) has recently become a compelling alternative for locating and classifying large samples of radio sources and cross-identifying them with their infrared counterparts, as e.g. done by the RGZ project (Banfield et al., 2015)

The next generation of radio interferometers will carry out deep and wide-area surveys expected to generate large volumes of image data. As a result, they will reveal millions of faint radio sources, and traditional methods of visual inspection will become extremely inefficient (Djorgovski et al., 2013; Goderya and Lolling, 2002). Therefore, there is an urgent need for efficient and automated algorithms to process the data in near real-time. Citizen science projects on their own will not be able to cope with the increasing data rates, as they take substantial time to be completed, and data often cannot be stored and served effectively at scale. Deep learning offers a mean to address this challenge.

1.3 Deep Learning

Machine Learning (ML) algorithms have recently become popular for automated data analysis tasks, where they can be used to find patterns in digital data and translate these patterns into predictive models (Ball and Brunner, 2010). ML algorithms can be divided according to whether they implement supervised learning or unsupervised learning. Supervised learning algorithms are provided with input-output pairs so that they learn the mapping (set of features) between the two. Unsupervised learning algorithms are not provided with the output data and they thus learn complex relationships by themselves.

Traditional machine learning algorithms have limitations when dealing with complex datasets. Most importantly, the data has to be simplified into a specific representation – selecting a few features to reduce its high-dimensionality. Representation learning is a subclass of machine learning techniques aiming to resolve this challenge by training algorithms to learn the representation automatically (i.e., extract relevant features from raw data themselves). In this context, deep learning is an approach where multiple "layers" are used to progressively extract higher level features from the raw input. Deep learning algorithms are thus very flexible, allowing a variety of tasks to be performed. Examples include reducing the dimensions of the data, classification and regression (Baron, 2019; Fluke and Jacobs, 2020). Thus, deep learning algorithms are gaining attention in Astronomy as a solution to many challenges in the era of big astronomical data. Most modern deep learning techniques build upon Artificial Neural Networks, and specifically Convolutional Neural Networks.

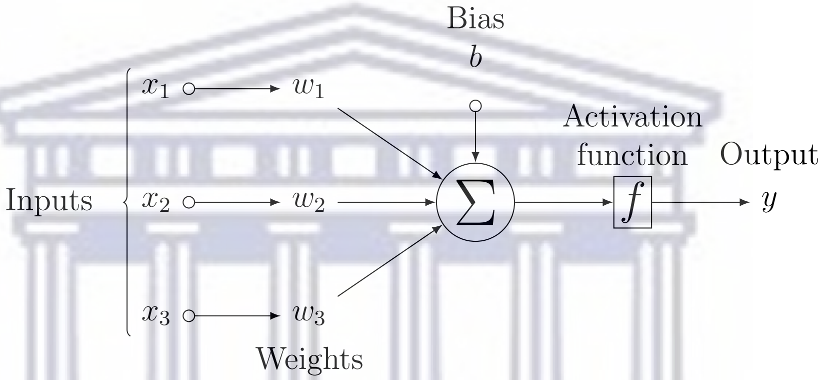


Figure 12: A simple model of an artificial neural network.

1.3.1 Artificial Neural Networks

To better understand deep learning algorithms, it is helpful to start by introducing the concept of Artificial Neural Network (ANN). The ANN is the basic building block of a deep learning algorithm. It is a system based on computations that try to mimic neural connections in human nervous systems (Rosebrock, 2017), which dates back to the 1940s (McCulloch and Pitts, 1943). A typical model of ANN is shown in Figure 12, where each input of vector $\vec{x}$ is connected to a neuron (Σ) via a weight vector $\vec{w}$, therefore each input has its associated weight. The neuron sums the weighted sum of the input, and a bias value b is introduced to each neuron. Then an activation function f is applied to determine if the neuron has essential information. Only non-linear activation functions allow artificial neural networks to compute nontrivial problems using a small number of nodes. The most frequently used activation functions are the following:

- the sigmoid function - $f(x) = \frac{1}{1+exp(-x)}$
- the hyperbolic tangent (tanh) function - $f(x) = \frac{exp2x-1}{exp2x+1}$

- the hard threshold function - $f_\beta(x) = 1_{x \geq \beta}$
- the Rectified Linear Unit (ReLU) function - $f(x) = max(0, x)$

In a nutshell, the activation function f checks if the output $y = f(\vec{w} \cdot \vec{x} + b)$ from the neuron is higher than some threshold, assigning a value of 1 if true and 0 if false. In recent work, ReLU has been the mostly used activation function. This function and its derivative are equal to zero for negative values; otherwise, it equals some positive value that results in some information at a given neuron.

ANNs became of practical interest when it was found that some limitations of a single neuron network can be overcome by multiple layers of interconnected neurons to create ANNs. This was first theorized as the universality approximation theorem, which states that ANNs with just three layers are capable of achieving desired levels of accuracy when modeling any function.

1.3.2 Multi-Layer Perceptron

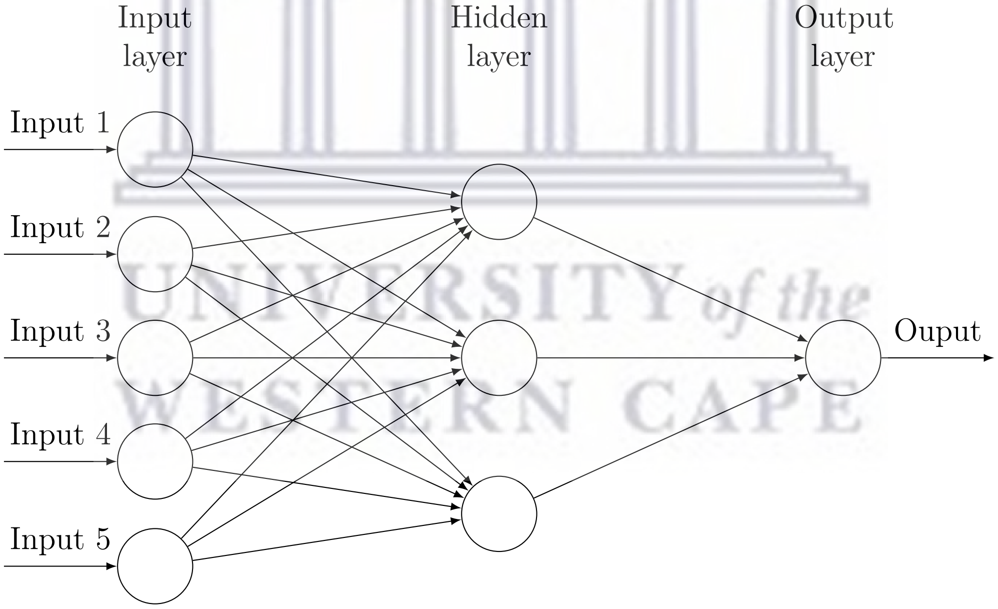


Figure 13: Representation of a multi-layer perceptron with one hidden layer. Each circle represents a neuron and weights are represented by arrow connection to other neurons on the next layer.

A multi-layer neural network, often referred to as a multi-layer perceptron, is a network that has several hidden layers of neurons. As shown in Figure 13, the output of a neuron from the previous layer becomes the input of a neuron on the next layer. The latter is referred to as a feed-forward network. On the output (last) layer, an activation function is usually applied, as well as on the hidden layers – which is dependent on the task at hand. In the case of binary classification, the output is a prediction with probability value in [0,1], whereas, for multi-class problems, the output layer contains the same number of neurons as the classes. For multi-class prediction, each output neuron has probability values for each class. The predicted class is the one having the highest probability values. Note that the classes are mutually exclusive, one example cannot be classified to belong to two classes at the same time. Thus, the sum of all those probability values equal one. In such a case, a softmax function is mostly used:

$$softmax(a)_i = \frac{exp(a_i)}{\Sigma_i exp(a_i)} \tag{3}$$

Equation 3 gives the probability values of a target class overall possible classes. This function ranges between [0,1], as a result, the output class is the one with the highest probability value.

The developments in technology have made available large labeled data ("Big Data") sets to the public. However, it is infeasible to reliably train an ANN that has an architecture, as shown in Figure 12. The progress in technology has also allowed for more specialized hardware, i.e., high-performance computing (HPC) systems – supercomputers. Moreover, neural networks have been improved to handle large amounts of training data by adding more layers of neurons, and these layers are connected in the form of a chain. The overall length of the chain determines the depth of the network - this gives rise to the term “deep learning” (Goodfellow et al., 2016). Essentially this means to increase the length of the chain of a network – Figure 13, more hidden layers are added between the input and output layer. The network uses a learning algorithm that decides how to use hidden layers to produces the best approximation of the desired output – given the input. The learning (training) algorithm is known as the back-propagation algorithm. It was introduced by Rumelhart et al., 1986. For each training example, the algorithm uses the feed-forward network to model the desired input. Starting with initial weights, and for every neuron in the consecutive layer, the output is computed. At the output layer, the network has the final model (predicted output) and then compares it with the input. At this layer, the network uses this information by measuring the total error of the network, which is contributed by the output error – the difference between the desired output (actual input) and the predicted output. The back-propagation algorithm then goes in reverse – starting with the last hidden layer, measures the contribution of the error in each layer and its connections to the output error. Finally, the algorithm then slightly changes the connection weight to reduce the error of the network.

Over the last few years, deep learning techniques have gained a lot of attention because they extract features automatically. Hidden layers are used to hierarchically learn abstract features, thus making deep learning techniques better generalizing algorithms. Deep learning has emerged as a better approach for achieving results that are promising especially in applications of image recognition. Many deep learning architectures/frameworks are being successfully applied to classify astronomy image data.

1.3.3 Convolutional Neural Networks

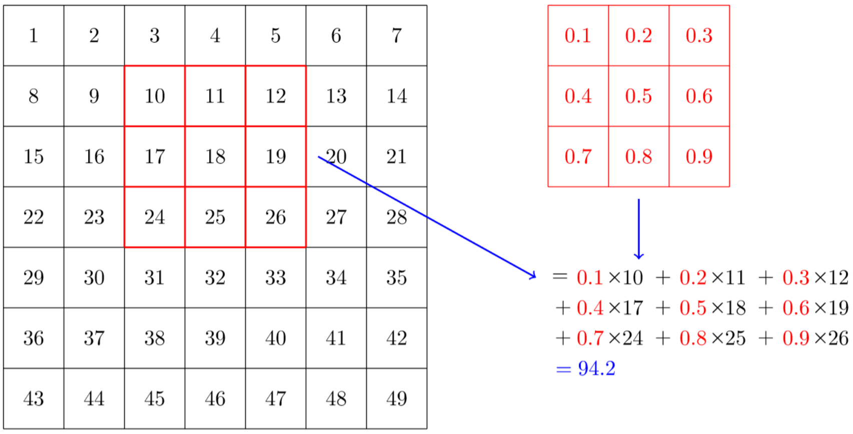
The most successfully applied deep learning architectures are the Convolutional Neural Networks (CNN; LeCun et al., 1989). CNNs fall under the class of supervised learning techniques. Therefore, the network is first trained with labeled data so that it learns a set of parameters (model) that best describes the input data. Then, the model is further tested with unseen data so that it can predict the target variables/labels.

CNNs are widely applied in the field of computer vision – dealing with how computers understand digital images and videos. For experiments presented in this thesis, image data was used. Therefore, it is important to understand image data representation. The representation of the image data in computers is a matrix of pixels or a grid of squares, each containing a single-pixel – a pixel represents color/intensity of light in a given square. Thus, an image is simply a matrix of width and height of pixels. In a grey-scale image, each pixel has an intensity value between 0 and 255, representing black and white colors, respectively. In the case of a color image, each pixel is represented by three intensity values (channels) ranging between [0,255] – indicating how red, green, and blue (i.e., RGB color) the given pixel is, combining these three colors captures the color of the pixel.

Typically, a CNN consists of three parts: (i) convolutional layer, (ii) pooling layer and (iii) fully connected layer.

1.3.3.1 Convolutional Layer

This layer makes use of a convolution operation by sliding the filter/kernel (i.e., weight vector) over the input image. The kernel slides horizontally and vertically over the input (image) vector. At each pixel position of the input image, elementwise multiplication with the kernel is computed. All those outputs are summed to get the elements of the output (feature) map, as shown in Figure 14. In simple



1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31	32	33	34	35
36	37	38	39	40	41	42
43	44	45	46	47	48	49

0.1	0.2	0.3
0.4	0.5	0.6
0.7	0.8	0.9

$$
\begin{aligned}
&= 0.1\times 10 + 0.2\times 11 + 0.3\times 12\\
&+ 0.4\times 17 + 0.5\times 18 + 0.6\times 19\\
&+ 0.7\times 24 + 0.8\times 25 + 0.9\times 26\\
&= 94.2
\end{aligned}
$$

Figure 14: Convolution operation. Left: 7×7 input image; Top-right: 3×3 filter/kernel. The output of convolution at location (i,j) is the sum of element-wise matrix multiplication, which is the value shown in blue at the bottom. Image courtesy of `https://sgugger.github.io/convolution-in-depth.html`.

terms, matrix multiplication applied in ANNs layers is replaced with convolution operation.

In a single layer of neurons representing different features, the convolution operation extracts N features based on a kernel of N filter, which results in N feature maps. For instance, if a layer has 64×64 grid of neurons operating on some input image, the convolution operation is going to extract 64×64 feature maps. Every input (x) from the input layer (often referred to as a plane) has its associated weight (w) connecting it with a neuron in the next layer. In addition, the very same neuron in the respective layer has connections with other neurons from the previous plane, forming a receptive field for that neuron (as shown in Figure 13). As a result, in an image, different neurons see different receptive fields. Thus, some neurons detect edges/corners of the image, while others detect objects around the centre. The extraction of features from images is done in those receptive fields, where this field forms a weight vector associated with some particular region from a previous layer. The output $f_{i,j}$ in the next plane at location (i,j) is computed using convolution operation by adapting equation $y = f(\vec{w} \cdot \vec{x} + b)$ to make it applicable to image data (2-dimensional), the convolution operation for discrete pixel becomes:

$$Y_{i,j} = f(\Sigma_{m,n} W(m,n) * X(i+n, j+m)) \tag{4}$$

where X represents the input given to that plane, W is the kernel – of size $m \times n$ pixels, that slides over the inputs, and * shows the convolution operation and f represents the activation function. The output of the convolution operation added with bias matrix are supplied to the activation function f that introduces non-linearity in the layer. In case of a color image – RGB image, the image data has an additional dimension, discussed previously, called a channel. A channel C results in an additional dimension of the matrix or a tensor, therefore equation 4 is updated to become:

$$Y_{i,j} = f(\Sigma_k \Sigma_m \Sigma_m W(k,m,n) * X(k, i+n, j+m) + B) \tag{5}$$

RGB color image has three (i.e., red, green and blue) channels, therefore k is defined in the range, [0,3], m and n represent pixel size or width and height of the feature map. B is the bias tensor which has the same dimension as the output feature map. The output feature map of one of the previous convolutional layers becomes the input of the next convolutional layer. In addition, neurons in the plane share the same weights from neurons in the previous plane, i.e., the same features occurring at different locations in the input data are easily detected. This also decreases the number of trainable parameters.

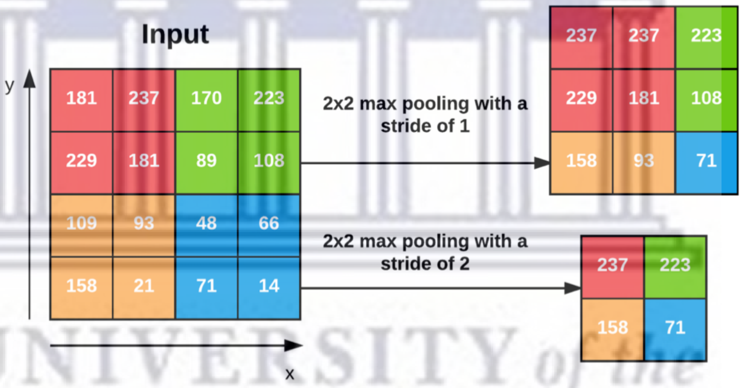


Figure 15: Left: a 4 × 4 pixels input image. Top-right: an output from a 2 × 2 pixels max-pooling layer applied to an input image with stride of 1. Bottom-left: a result from a 2 × 2 pixels max-pooling layer with stride of 2 also applied on the input image. Image from Rosebrock (2017).

1.3.3.2 Pooling Layer

This layer is used to control the over-fitting of the network, reduce the spatial size of the network, reduce the number of trainable parameters, and introduce translation in-variance by sub-sampling the given image. A pooling layer typically used is known as max-pool. This layer of defined pool size acts like a sliding window. It slides over the input image, where it takes only the most significant value. Moreover, the step size (number of pixels to skip) of the window is defined by a stride value. This window slides from left-to-right and from top-to-bottom across the image. Figure 15 shows a max-pooling operation applied over a 4×4 input image that is towards the left side of the figure. A 3×3 image that is towards the top right of the figure represents the output of a 2×2 max-pool applied to the input image, where some regions overlap because a stride of 1 was used. Towards the bottom right of the figure, an output from a 2×2 max pooling layer applied to the input images with stride 2 used. The stride of 1 pooling is often referred to as overlapping pooling, whereas stride of 2 is referred to as non-overlapping pooling (see Rosebrock, 2017).

Pooling layers reduces the dimensions of the input image, often zero-padding is used, which adds a margin of zero-valued pixels around the image, to control the size of the output of the network.

1.3.3.3 Fully Connected Layer

A fully connected (FC) layer is a layer that computes the output by using the dot product of weight and input vector and the output from the feature extraction phase (convolution and pooling over and over). This is the output layer of the network. Every neuron in this layer is connected to all the neurons in the previous layer. However, this can be computationally expensive. As a result, a dropout layer is added after an FC layer. The dropout layer is a regularization layer that helps reduce over-fitting and helps the network to generalize. It randomly drops connections between neurons from the preceding layer to the output layer that is below a given threshold probability. The last layer in the network is a soft-max activation layer (equation 3), which takes the output from the FC layer and result in a probability value for each class.

As with other supervised learning algorithms, the model needs to be trained in such a way to best represent the training set. Thus, the model needs to be validated whether it best fits the training data (i.e., the model needs to be optimized). For CNNs, a frequently applied optimization algorithm is known as Gradient Descent (GD). GD generally changes learnable parameters slightly, in order to reduce the cost function (output error) of the network. GD uses the cost function to compute the gradient in terms of the model parameters, then the step size, often referred to as the learning rate, is multiplied with the gradient vector. This vector product is then subtracted from the associated model parameters – backpropagation. The most common backpropagation algorithms are batch, stochastic, and mini-batch GD. Batch GD uses the entire training set to compute the gradient at every step of the training sample. As one might expect, it is prolonged for a large training dataset. Stochastic GD (SGD), as the name suggests, it uses random examples – at every step, from the training dataset and computes the gradient with only it. In mini-batch GD, gradients are computed only on small random sets of examples from the training set, called mini-batch. Both SGD and mini-batch GD have a better chance of finding the global minimum than batch GD. However, due to the randomness of SGD, it never settles well at a minimum. Nevertheless, the gradient is computed during the feed-forward phase. The output gradient is then backpropagated through the network in the direction of a decreasing gradient. For each epoch, the learnable parameters are updated and used to reduce the cost function. One epoch is completed when the feed-forward and backpropagation steps have been carried out for the entire dataset. This process is repeated until the network converges to a particular value of the cost function.

Over the last few years rapid advances in digital technology have led to the era of "big data" and to the corresponding increase in available computational power. This in turn has led to a widespread use of deep learning algorithms as a solution to addressing computer vision problems. Deep learning algorithms take leverage of this technological advancements, to result in more general and fast models. The most popular deep learning algorithms, as of recently, are CNNs. CNNs, as previously discussed in Section 1.3.3, are good at generalizing, the network only detect a small set of essential features to learn. It does that with kernels that only take a few number of pixels, irrespective of the number of pixels of the given image. Furthermore, the neurons share parameters during the learning phase, whereby each convolution kernel is used at every position of the input image. Thus, the network only learns a small set of learnable parameters, not a separate set of parameters at every position. Moreover, CNNs maintain the spatial relations between pixels because they use small filters to learn features from the input image. Thus, they are invariant to rotations and translations, sometimes scaling. Recently, it has been proven that they have transfer learning capabilities (Pan and Yang, 2010; Yosinski et al., 2014; Domínguez Sánchez et al., 2018). Transfer learning is a process where the model is allowed to take advantage of information from experience. The best model from previous training examples of a specific domain is stored. The best parameters of that model are re-used on a model built on a different domain to improve its accuracy.

1.4 Applications of Deep Learning in Astronomy

Recently, there have been several successful applications of deep learning algorithms, and specifically CNNs, in astronomy. Applications have often focused on classifying astronomical sources from images based on their shape and have been carried out across all the wavelengths of the EM spectrum, but most notably at optical and radio wavelengths as a result of the amount of available labeled data in these regimes. These efforts have been helped by the much-improved quality and quantity of labeled data generated thanks to the advent of "Citizen Science", or internet-enabled image classification by non-specialists, pioneered by projects such as Galaxy Zoo (Lintott et al., 2008) and Radio Galaxy Zoo (Banfield et al., 2015). The first Galaxy Zoo project (GZ1[1]) focused on the classification of galaxies as spirals or ellipticsls and their subclasses. The web site presented the citizen scientists with an interface showing a galaxy and they would be asked to determine if the galaxy shown is an elliptical or a spiral galaxy, and they would be asked to determine the rotation direction of a given spiral galaxy. The successor to the GZ1 project was GZ2[2] (Willett et al., 2013), which presented a subset of the brightest and largest galaxies to the volunteers and asked them to classify these galaxies using more detailed questions, including whether the respective galaxy has bars, spiral arms and/or bulges, whether it is an edge-on or a face-on galaxy, etc. One of the first applications of deep learning in astronomy actually arose from the so-called Galaxy Challenge, a data science competition run produced by the GZ team and partners on the Kaggle platform[3]. The challenge was for participants to write an algorithm that could learn from the classifications performed by citizens scientists and classify unseen galaxies from image data into different groups. The winning team (Dieleman et al., 2015a) applied a deep learning model based on translationally- and rotationally-invariant CNNs and demonstrated that such architectures allow us to automatically annotate large numbers of images, enabling quantitative studies of galaxy morphology on an unprecedented scale. This has led to the development of several astronomy-themed citizen science projects such as the Radio Galaxy Zoo[4] (Banfield et al., 2015), whose data products we will use in our work. These early projects have contributed to the establishment of the largest and most popular citizen science platform, known as the Zooinverse[5]. The Zooinverse operates similarly to the original Galazy Zoo project, however it has a wide range of projects, from galaxy classification to answering questions about historical records and even plants and animals. The latest incarnations of the Galaxy Zoo[6] projects currently available on the Zooinverse platform use a wide selection of images obtained at optical

[1] `http://zoo1.galaxyzoo.org/`

[2] `ttp://zoo2.galaxyzoo.org/`

[3] `https://www.kaggle.com/c/galaxy-zoo-the-galaxy-challenge/`

[4] `https://radio.galaxyzoo.org/`

[5] `https://www.zooniverse.org/`

[6] `https://www.zooniverse.org/projects/zookeeper/galaxy-zoo/` and Radio Galaxy Zoo at `https://www.zooniverse.org/projects/chrismrp/radio-galaxy-zoo-lofar`

and radio wavelengths as well as simulated images to help astronomers address some of the most challenging problems in contemporary astrophysics. This is complemented by an increased interest in applying Deep Learning algorithms, and specifically CNNs, to these problems. Dieleman et al. (2015b), Domínguez Sánchez et al. (2019), Huertas-Company et al. (2019), Ghosh et al. (2020) and Hausen and Robertson (2020) applied CNNs to classify galaxies from optical images according to the Hubble sequence and Abraham et al. (2018) used CNNs to detect and classify bar structure in galaxies from optical images. Burke et al. (2019) applied CNNs to accurately classify and deblend galaxies and stars from optical image data.

Another application of CNNs is the measuring of photometric redshift from optical image data (Hoyle, 2016; D'Isanto and Polsterer, 2018). For both the science applications (classification and photometric redshift estimation), CNNs achieve a good accuracy ($\geq$ 90%). One other application of deep learning in optical astronomy concerns the detection of gravitational lenses (Hezaveh et al., 2017; Petrillo et al., 2017; Metcalf et al., 2019; Jacobs et al., 2019; Petrillo et al., 2019a,b) from image data. Similarly, for radio astronomy, CNNs have been applied to classify extended radio sources from images into different classes, e.g. FR-I, FR-II and "bent" – also referred to as wide- and narrow-angle tailed galaxies, or WAT and NAT (Aniyan and Thorat, 2017; Alhassan et al., 2018; Lukic et al., 2018; Ma et al., 2019). However, applications in radio astronomy were until recently severely limited by the small samples of labeled data at our disposal. Thanks to RGZ, however, there is now a substantial amount of labeled radio data available to train CNNs. It is also important to point out the SKA Science Data Challenge (SDC) series whereby simulated images with specifications similar to those expected from the SKA are publicly released, encouraging the participation to the community at large, not just astronomers or developers of source finding algorithms. The fist data challenge (SDC1; Bonaldi and Braun, 2018), the public was invited to perform source finding and source characterization on the simulated images, and identify source populations. Similar to the SDC1, the second challenge invited the public to carry out the same procedure, perform source finding and source characterization, however, on simulated HI imaging data cube (SDC2; Hartley et al., 2023).

Using the RGZ dataset, Alger et al. (2018) and Wu et al. (2019) have used CNNs to cross-identify radio sources with their host galaxies using multi-wavelength data – radio and IR; this is of fundamental importance because multi-wavelength information is required to study the physical properties of radio galaxies. Recently, transfer learning has been successfully applied as well, showing an increase in the accuracy of the model when a pre-trained model is re-trained and/or tested on different input data (Vilalta, 2018; Domínguez Sánchez et al., 2018; Lukic et al., 2019; Vilalta et al., 2019; Tang et al., 2019; Wu et al., 2019). Many of these recent applications produce accurate results, as they make optimal use of a large amount of labeled data made available by the past and recent observations and the RGZ project. The large volume of data is required to be trained on, for the algorithm to generalize well on new data and avoid over-fitting. However, the downside of training on large volumes of data and using complex models, such as CNNs, is over-fitting. Over-fitting happens when the model learns the details and the noise of the training data. This results in a model that does not generalize well on the new data unseen data by the model during training, leading to inaccurate results.

Another major challenge for the acceptance of deep learning by the astronomical community is due to the fact that these models are considered, in general, to be “black boxes”. The black box nature of deep learning models is due to the fact that these models perform complex non-linear mapping, they are difficult to uncover and interpret or explain physically. This has led to the development of interesting research fields including interpretability and uncertainty estimation of deep learning neural networks. Interpretability focuses on the investigation the relations learned by the models, while uncertainty estimation focus on the inclusion of prior physical constraints to maintain known symmetries of the physical problems and control what the networks are extracting (see Huertas-Company and Lanusse, 2023). Although most of the work is still in the exploratory stage, it will be interesting to see the general consensus of the outcome of these researches — whether it will eradicate the perception that deep learning models are opaque black boxes. From a practical point of view, when applying an existing deep learning trained model to a new dataset (problem), it is important for the researcher to make sure that the new dataset (problem) is sufficiently “similar” to the one used to train the model in the first place. Where that is not the case, all care must be taken to evaluate model performance and avoid common pitfalls such as overfitting and underfitting.

The next-generation of radio sky surveys are expected to survey the sky with unprecedented sensitivities (Norris, 2011). As a result, they are expected to detect huge numbers of the common star-forming galaxies and non-jetted AGNs (see Padovani, 2016), as well as reveal large numbers of unusual extended radio sources. This is going to be a challenge for traditional source characterization schemes. Most of the source detection algorithms currently in use work by locating “islands” of pixels having higher emission than some user-defined threshold. One of the famous algorithms that follow such a procedure is known as

PyBDSF (the Python Blob Detection and Source Finder; Mohan and Rafferty, 2015). PyBDSF estimates the background noise of the image and the mean of the image, which creates the output image with the pixels with values higher than the threshold (known as sigma-clipping). It then fits 2-dimensional Gaussian profiles to those detected sources. It further produces a number of parameters such as the position and peak value of the maximum, position angle, the width of the Gaussian, amongst others. A list of detected sources with their corresponding parameter estimates is then produced after the Gaussians are grouped together. One of the main limitations of PyBDSF include the fact that sources are not Gaussians – it may work well to model point sources. However, when the source is resolved – even slightly (compact) resolved, or the source is extended with more complicated structures, PyBDSF may struggle to model the source. PyBDSF and other such methods will systematically underestimate the flux. Nevertheless, this is the source detection algorithm that we will use to compare the results of our algorithm against. However, Vafaei Sadr et al. (2019) applied CNNs to detect point sources on simulated radio images. Their algorithm enhances the SNR of the image by learning the correlated noise. As a result, when comparing their algorithm with PyBDSF, they found their algorithm outperforms PyBDSF in all metrics. However, in our work we employ PyBDSF because its setup is rather straightforward and its adoption is very common within the astronomical community. The detection step is essential for cross-identification since positional information is required to match the fields and sources from different surveys accurately.

All of this progress notwithstanding, there is still no "silver bullet" algorithm to characterize (i.e. detect, classify and identify) multi-source, multi-peaked radio sources. Also, most of these algorithms focus on image data from a single sky survey and they are thus unable to make the most of multi-wavelength data. Moreover, using cross-matched data (one catalogue to another) incorporates complications, such as resolution differences. However, from Section 1.2.2, it is clear that multi-wavelength data can be extremely useful to classify and cross-identify radio sources with their host galaxies. Cross-identifying helps resolve the difficulty of having to make sense of multiple radio components, e.g. whether they belong to an extended radio source having two lobes or rather represent radio emission from two separate galaxies undergoing star formation. Such cases can be solved by cross-identifying radio data with optical/IR data (Norris, 2017a). Modern radio surveys are likely to reveal several different and new kinds of radio sources. However, radio data alone will not be enough to clarify their nature. Moreover, using cross-matched data (one catalogue to another) incorporates complications, such as resolution differences. In preparation for these challenges, it is very useful to develop data analysis methods which make the most of multi-wavelength data in a coherent manner. Ideally, the method should be able to perform well on simulated data, as well as real data.

1.5 Objective

The main objective of this study was to build a "Radio Source Characterization" pipeline for upcoming radio surveys building on CLARAN as well as on some further in-house development, and to test it on a new deep and wide radio survey with the GMRT. Generally, characterization is the description of the distinct features of the object or source. In astronomy, characterization usually implies measuring source attributes (e.g., flux, size, surface brightness). For this study, source characterization is defined as the combined process of detection, classification and (multi-wavelength) identification.

1.6 Structure of the thesis

This thesis is structured as follows:

- **Chapter 1 - Introduction** An introduction to sky surveys, and in particular next-generation radio sky surveys. An introduction to radio astronomy, focusing on different classes of radio sources, emission processes involved, and the resulting structure (morphology). An introduction to deep learning concepts and its application to astronomy is presented. Starting with a brief history and proceeding to different types of neural networks and their architecture. Also, a brief overview of some recent literature on the application of deep learning to astronomy is presented.

- **Chapter 2 - Methodology** A brief overview of CLARAN and its architecture is presented along with its application to RGZ data. The adaptation of the CLARAN code to run on the ilifu facility and the application of transfer learning is presented. Also, an introduction to the source characterization pipeline is presented, furthermore providing details of the source characterization process.

- **Chapter 3 - Results** Output examples are presented. Furthermore, a subset of visually inspected cutouts per class is shown and discussed in order to evaluate the algorithm. Also, a few examples from the output dataset of the source characterization pipeline are presented.

- **Chapter 4 - Discussion** A discussion of the limitations of the source characterization pipeline. Also, insights on possible solutions to those limitations are discussed. Lastly, an example of CLARAN applied to a larger cutout is presented.
- **Chapter 5 - Conclusion and Future Work** A summary of the work and some final remarks are presented, including possible future work arising from this study.

Chapter 2

2 Methodology

Even though deep learning techniques such as CNNs have been widely applied, achieving excellent results for given tasks, model training to solve a new task is often challenging. Also, it can be CPU-intensive and thus time-consuming, especially if the model is not easy to optimize (Goodfellow et al., 2016). In such cases, a pre-trained deep learning model can often be tweaked to perform the desired task – either by directly testing or re-training the pre-existing model on new data. This process is known as transfer learning. In our case, the main interest is developing an automated and efficient algorithm to classify radio sources, and for this purpose we have adopted and adapted a recently-developed algorithm known as CLARAN. CLARAN was developed to classify multi-component and/or multi-peaked radio sources within a single cutout. It can also be adapted to identify radio sources with their infrared hosts, making it a powerful multi-wavelength algorithm – with the potential of being extended in other ways. CLARAN also locates and classifies sources in a larger cutout as compared to the cutout that was used to train it. Moreover, CLARAN performed well both on real data (RGZ Data Release 1, to be discussed in Section 2.2) and simulated data (SDC1, discussed in Section 1.4). The ICRAR (International Centre for Radio Astronomy Research) team was one of the nine teams, their method was primarily based on CLARAN version 0.28 prototype. The ICRAR team was the second best team in the competition based on CLARAN version 0.28 prototype's performance in the challenge. Considering all of its advantages, we developed and tailored CLARAN to our needs – aiming to use the pre-trained model on an original dataset obtained from another radio survey.

2.1 Overview of CLARAN

CLARAN (**Cla**ssifying **R**adio Sources **A**utomatically with **N**eural Networks) is a state-of-the-art algorithm that tries to provide a solution to the problem of classifying multiple sources in a single image cutout. CLARAN is a proof-of-concept algorithm developed by Wu et al. (2019) which classifies radio sources according to their morphology given some input image data. It uses a well known state-of-the-art deep learning object detection model called Faster Region-based CNN (FR-CNN; Ren et al., 2015). This model was fine-tuned to classify radio sources in an end-to-end manner – requiring no interaction from the user. CLARAN can also combine radio images with e.g. infrared images of the same region to more efficiently locate and identify compact as well as extended radio sources. CLARAN provides identification and classification of radio sources with a mean average precision (formally defined in Section 2.3.1) of 83.6% and an empirical accuracy above 90%.

Table 1: CLARAN architecture made up of 29 layers as represented by rows. Columns represents the layer number, the function applied to a layer (also given as an identifier to that layer), input image/filter size, type of activation function, output image/filter size and number of parameters in that layer. The architecture can be divided into three networks, ConvNet (layer 1 - 17), LocNet (layer 18 - 22) and RecNet (layer 23 - 29). see Table 4 from Wu et al. (2019).

Layer	Function	Input/Filter tensor size	Activation	Stride	Output tensor size	Number of parameters
1	Input	$600 \times 600 \times 3$	-	-	-	0
2	Conv1_1	$3 \times 3 \times 64$	ReLU	1	$600 \times 600 \times 64$	1,728
3	Conv1_2	$3 \times 3 \times 64 \times 64$	ReLU	1	$600 \times 600 \times 64$	36,864
4	MaxPool1_1	$2 \times 2 \times 64$	-	2	$300 \times 300 \times 64$	0
5	Conv2_1	$3 \times 3 \times 64 \times 128$	ReLU	1	$300 \times 300 \times 128$	73,728
6	MaxPool1_2	$2 \times 2 \times 128$	-	2	$150 \times 150 \times 128$	0
7	Conv3_1	$3 \times 3 \times 128 \times 256$	ReLU	1	$150 \times 150 \times 256$	294,912
8	Conv3_2	$3 \times 3 \times 256 \times 256$	ReLU	1	$150 \times 150 \times 128$	589,824
9	Conv3_9	$3 \times 3 \times 256 \times 256$	ReLU	1	$150 \times 150 \times 128$	589,824
10	MaxPool1_3	$2 \times 2 \times 256$	-	2	$175 \times 175 \times 256$	0
11	Conv4_1	$3 \times 3 \times 256 \times 512$	ReLU	1	$75 \times 75 \times 512$	1,179,648
12	Conv4_2	$3 \times 3 \times 512 \times 512$	ReLU	1	$75 \times 75 \times 512$	2,359,296
13	Conv4_3	$3 \times 3 \times 512 \times 512$	ReLU	1	$75 \times 75 \times 512$	2,359,296
14	MaxPool1_4	$2 \times 2 \times 512$	-	2	$37 \times 37 \times 512$	0
15	Conv5_1	$3 \times 3 \times 512 \times 512$	ReLU	1	$37 \times 37 \times 512$	2,359,296
16	Conv5_2	$3 \times 3 \times 512 \times 512$	ReLU	1	$75 \times 75 \times 512$	2,359,296
17	Conv5_3	$3 \times 3 \times 512 \times 512$	ReLU	1	$75 \times 75 \times 512$	2,359,296
18	RPN_Conv	$3 \times 3 \times 512 \times 512$	ReLU	1	$512 \times 37 \times 37$	2,359,296
19	Anchor_Cls_Conv	$1 \times 1 \times 512 \times 12$	-	1	$12 \times 37 \times 37$	6,144
	Anchor_Cls_Conv_Rs	$12 \times 37 \times 37$	-	-	$(6 \times 37) \times 37 \times 2$	0
20	Anchor_Cls_Softmax	$(6 \times 37) \times 37 \times 2$	-	-	$(6 \times 37) \times 37 \times 2$	0
	Anchor_Cls_Softmax_Rs	$(6 \times 37) \times 37 \times 2$	-	-	$37 \times 37 \times 12$	0
20	Anchor_Target	12×37^2, gt_box $\times 5$	-	-	$37^2 \times 12$, $37^2 \times 24$	0
19	Anchor_Reg_Conv	$1 \times 1 \times 512 \times 24$	-	1	$24 \times 37 \times 37$	12,288
21	RoI_Proposal	$37^2 \times 12$, 24×37^2	-	-	NMS_TopN $\times (4+1)$	0
22	RoI_Proposal_Target	NMS_TopN $\times 5$, gt_box $\times 5$	-	-	RoI_batch $\times 1$, RoI_batch $\times 28$	0
23	ST_RoI_Pool	$37 \times 37 \times 512$, RoI_batch $\times 5$	-	-	RoI_batch $\times 7 \times 7 \times 512$	0
24	FC_6	RoI_batch $\times 7 \times 7 \times 512$	ReLU	-	RoI_batch $\times 4096$	102,764,544
25	Droupout_6	RoI_batch $\times 4096$	-	-		
	RoI_batch $\times 4096$	0				
26	FC_7	RoI_batch $\times 4096$	ReLU	-	RoI_batch $\times 4096$	16,781,312
27	Droupout_7	RoI_batch $\times 4096$	-	-	RoI_batch $\times 4096$	0
28	FC_Cls_Score	RoI_batch $\times 4096$	-	-	RoI_batch $\times 7$	28,679
28	FC_Cls_Pred	RoI_batch $\times 4096$	-	-	RoI_batch $\times (7 \times 4)$	114,716
29	Cls_Softmax	RoI_batch $\times 7$	-	-	RoI_batch $\times 7$	0

The model has 29 layers, as shown in Table 1. These layers can be categorized into three networks – the ConvNet (layers 1 - 17), the LocNet (layers 18 - 22), and the RecNet (layers 23 - 29). ConvNet consists of typical convolution layers described in Section 1.3. In this network convolution operations, non-linear activation (ReLU) and max-pooling functions are used. The architecture of the first 17 layers is from VGG-16 (Configuration D) network (Simonyan and Zisserman, 2014). The weights in these layers were initialized by loading pre-trained VGG-16 weights from ImageNet (Russakovsky et al., 2015). However, in higher layers, the weights are set free in order for the model to learn high-level features. The parameters from the ConvNet weight layers are then shared by both the following networks, starting from layer 18 to 29. LocNet uses the output from ConvNet to propose regions of interest (boxes) of the given subject that contain a potential radio source. RecNet takes the output from ConvNet (feature maps) and LocNet (proposed boxes) and then classifies the detected sources. In this network, ReLU is the only activation function used. By summing the values in the last column – the resulting total number of parameters is 136,777,443. Therefore the model has over 1 million trainable parameters – evidence of how computationally intensive it is to train a model such as CLARAN. Moreover, it is clear from the number of trainable parameters that CLARAN is complex model, which may raise a few concerns about inherent weaknesses such as over-fitting. However we note that these issues were investigated, and CLARAN's over-fitting was evaluated (see Section 5.7 in Wu et al., 2019), and the outcome suggests that CLARAN is not over-fitting. The RGZ data was used to train CLARAN (to be discussed in the next section). This implies that CLARAN's "ground truth" was collectively produced by citizen scientists through visual inspection, which may not always reflect the "true" ground truth. Also, note that CLARAN research problem and method differ from other CNN methods (mentioned in Section 1.4), it is developed to perform source identification and morphology classification. In contrast, other CNN methods were developed to perform classification only, as a result, it is challenging to perform a direct comparison with CLARAN. Moreover, these methods use different training data (some using real data, while others are using simulated data), making it impractical to perform a direct comparison – simulations will be useful for the direct comparison. However, as noted in Section 2 above, CLARAN version 0.28 prototype performed very well in comparison with other source finding tools used by other teams — in terms of completeness and reliability, which resulted in the ICRAR team being ranked as the second best team in the competition.

2.2 The Radio Galaxy Zoo Project

Radio Galaxy Zoo (RGZ) was an online crowd-sourced platform where citizen scientists volunteered to classify radio galaxies and their host galaxies[7]. Participants were presented with a web-based interface showing a 3 arcmin × 3 arcmin figure representing a radio image overlaid on an infrared image, where the lowest contour level and shading of the radio image was preset. RGZ used radio data from the FIRST survey and the Australia Telescope Large Area Survey (ATLAS;

[7] The Radio Galaxy Zoo project is now archived at https://radio.galaxyzoo.org. An active more recent incarnation is the LOFAR Galaxy Zoo project at http://lofargalaxyzoo.nl/

Franzen et al., 2015 Data Release 3. To aid in the identification of the host galaxies of the radio sources, infrared data from the Wide-field Infrared Survey Explorer (WISE; Wright et al., 2010) and from the Spitzer Wide-Area Infrared Extra-galactic Survey (SWIRE; Lonsdale et al., 2003) are used respectively. For CLARAN, only the FIRST radio data and WISE infrared data have been used. The FIRST survey was briefly introduced in Section 1.1.1, while the WISE survey is briefly described below.

2.2.1 The WISE Survey

The Wide-field Infrared Survey Explorer (WISE; Wright et al., 2010) telescope is an all-sky survey – sky coverage shown in Figure 16, carried out by a space telescope with a 40 cm diameter mirror and a field of view (FoV) of 40 × 40 arcmin. It maps the sky in four different bands (W1, W2, W3 and W4) – 3.4, 4.6, 12 and 22 μm with an angular resolution of 6.1, 6.4, 6.5 and 12 arcsec, respectively.

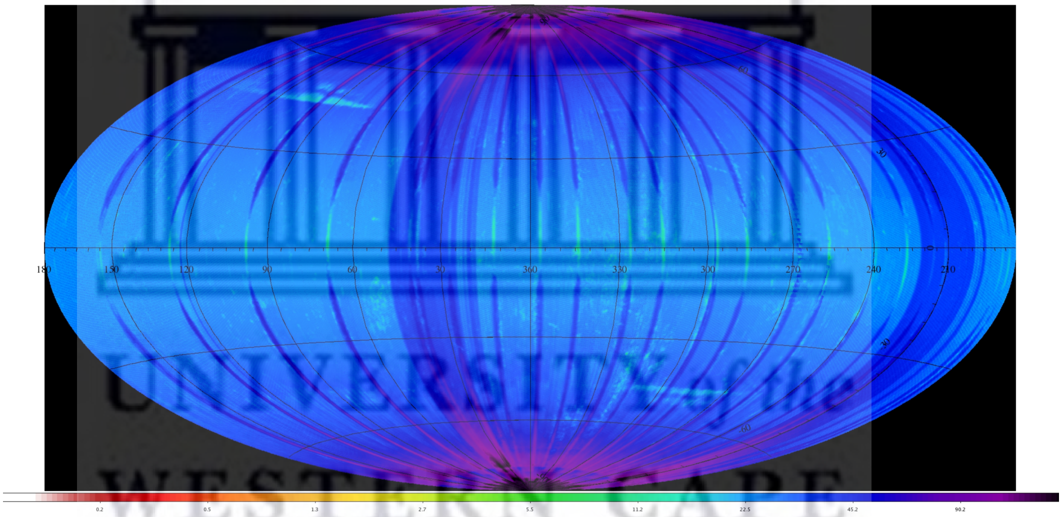


Figure 16: Sky coverage of WISE survey shown in ecliptic Aitoff projection. The colors represent the average number of individual 7.7/8.8 sec exposure frames within 15' × 15' spatial bins. The colorbar at the bottom shows the frame coverage depth. Image courtesy of `https://wise2.ipac.caltech.edu/docs/release/allwise/expsup/sec4_2.html`.

The RGZ browser-based graphical user interface brought together FIRST radio data and the overlapping WISE infrared (imaging) data. The participants started by going through a tutorial to help them correctly classify the sources. The tutorial showed the participants how: (i) to select contours that correspond to a radio source, (ii) select the corresponding infrared host galaxy that belongs to the selected radio contours, then (iii) either choose to continue to classify the remaining radio sources or continue to the following image (see Figure 3 from

Banfield et al., 2015). After completing the tutorial, participants were given a randomly selected image from the RGZ data to classify. Consensus levels among the participants were used as a measure for the reliability of the classifications provided for each source. For all the classifications, the position of the nearest IR host galaxy to the clicks by participants are recorded, together with the positions of the four corners surrounding the radio contours. From December 2013 to March 2016 RGZ project has had over 11,000 registered citizen scientists with over 75,000 source classifications (O. I. Wong, in preparation). The project thus aimed to provide a solution to the problem of distinguishing multiple components of single sources from multiple unrelated sources.

2.3 Application of CLARAN to Radio Galaxy Zoo Data

Deep learning algorithms require a large volume of data to be trained on, for it to generalize well and avoid under-fitting and over-fitting. The RGZ project has created one of the largest catalogues of extended radio galaxies and sources with disconnected lobes, cross-identified with their host galaxies, and the CLARAN software had been trained making use of this state-of-the-art dataset. It is worth noting that visual inspection is notoriously difficult and amateurs (public) will have wildly scattered results, even experts cannot agree on morphology. However, we note that (Wu et al., 2019) applied two selection criteria to select the highest quality data, by ensuring that most radio sources exposed to CLARAN are morphologically human-resolvable and that every radio source within the cutout has fewer than four components and four peaks.

In the RGZ classification scheme, sources were classified in terms of the number of components (C), and peaks (P) identified in the radio image only. Components indicate discrete (i.e. spatially separate) radio source components that are detected at 4σ flux-density threshold while peaks refer to the number of resolved peaks that can be identified, occasionally within the same components. CLARAN was trained using data from the RGZ Data Release 1 (DR1; O.I. Wong, in preparation).

2.3.1 Training and Evaluation

To effectively train CLARAN, only a subset of sources, or “subjects”, from RGZ DR1 was used, which was a result of the selection criteria applied. Firstly, only subjects that have consensus levels no less than 60% were chosen, implying that citizen scientists on average agreed in their classifications of the given source.

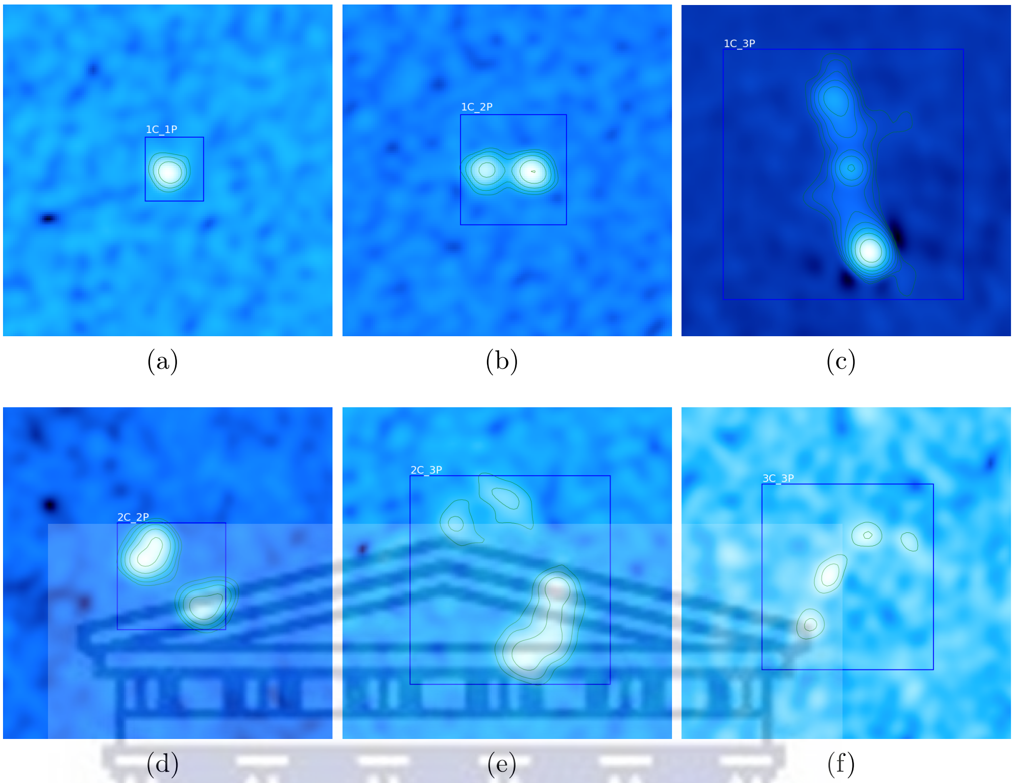


Figure 17: Example radio continuum images at 610 MHz, for each of the six morphological classes, overlaid with 5σ radio contours.

Secondly, a given subject had to contain no more than three components and no more three peaks – to avoid the issue of unbalanced classes (unequal distribution of classes) in the training and testing dataset. The two selection criteria result in a dataset containing 10,744 RGZ subjects distributed across six classes characterized by different number of components and/or peaks. Furthermore, this dataset was randomly split into two – training set containing 6,141 and testing set containing 4,603 subjects. The six morphological classes are 1C_1P, 1C_2P, 1C_3P, 2C_2P, 2C_3P, and 3C_3P, where C and P represent radio source components and peaks, respectively. Figure 17 shows example radio images for each of the six morphological classes, (a) shows a 1C_1P class that represents a single component (or compact) radio source, while (b) and (c) show a 1C_2P and a 1C_3P class that represent a single component radio source with two and three resolved peaks, respectively; (d) shows a 2C_2P class that represents a source with two discrete radio components, also referred to as a double; (e) shows a 2C_3P class that represents a source with two discrete components with three resolved peaks and (f) shows a 3C_3P class that represents a source with three discrete radio components.

As discussed previously in Section 2.2, the RGZ dataset contain radio and infrared image data from the FIRST and the WISE surveys shown as image F

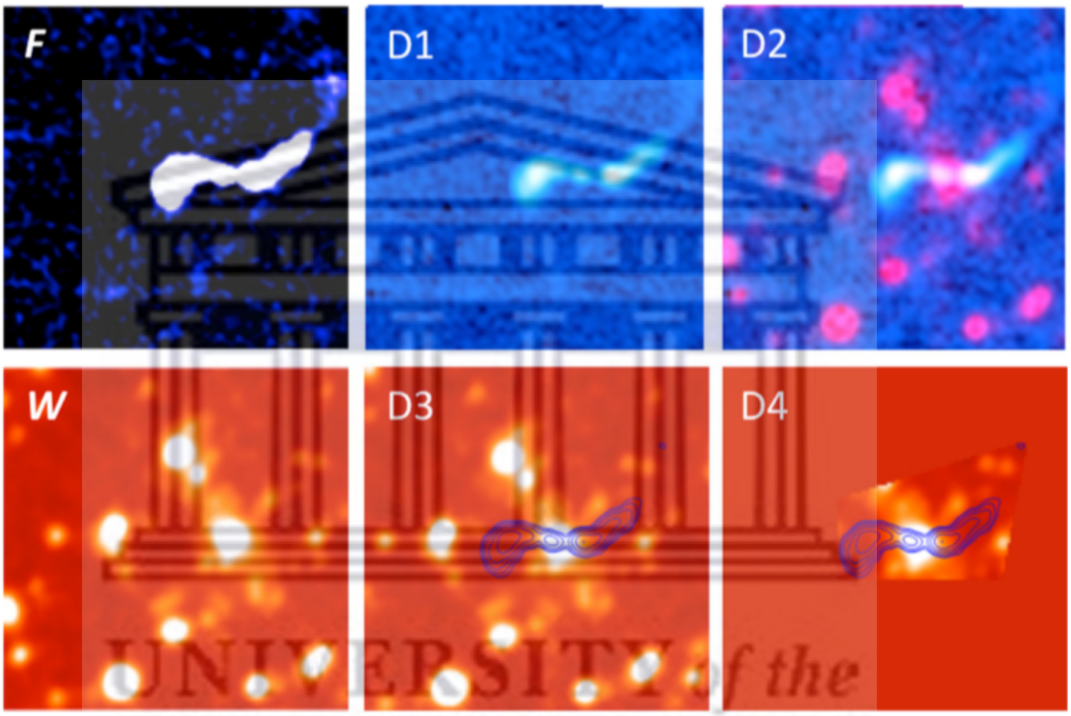


Figure 18: Combining radio and infrared images. F shows the radio (FIRST) image with an angular resolution of 5 arcsec and W shows the infrared (WISE W1 at 3.4 microns) image with an angular resolution of 6.4 arcsec. Different combinations of radio and infrared images are shown as D1, D2, D3, and D4. From Wu et al. (2019).

and W in Figure 18, respectively. Image F shows the radio image of an RGZ subject. The corresponding infrared image is shown underneath as image W. Also, Figure 18 shows different input images which can be used for CLARAN. All these maps were exported from FITS to PNG format as a three-channel (RGB) image by using DS9 (Joye and Mandel, 2003) and applying different functions. The raw radio (FIRST) image F uses the `linear zscale` and the `cool` colourmap. W uses the `linear zscale` and the `gist_heat` colourmap. As a result, the two images represent true flux values from the original FITS images. The additional four datasets (D1, D2, D3, and D4) were derived to effectively train CLARAN. The D1 image is the same as F, but using the `log zscale` scale. Although, this scale reveal the internal structures of the sources, it exposes more background noise to CLARAN. The D2 image is based on the D1 image but the red channel is enhanced with the corresponding values of the red channel of the W image. It is worth noting that the red channel mentioned previously, refers to one of three channels (including green and blue channels) that make up a digital image (or an RGB image). The latter is used to train CLARAN to learn how IR emission appear relative radio emission. The D3 image achieves the same goal, but by overlaying 5σ radio contours on the infrared image. However, the D3 image shows multiple infrared sources that may be unrelated with the overlaid radio source. As a result, the D4 image was produced by using a `convex_hull` to mask the pixels outside the union area of all radio contours (for more details, see Wu et al. (2019)), a method also known as clipping. The pixels outside the union area are assigned the mean value of each channel of the RGB image computed over all images in the training set. It is noted that the clipping method is necessary to expose the sufficient IR signals to capture the interplay between all radio sources/components. Evidence of the latter is provided by the D4 image in Figure 18, even with clipping, there remains multiple IR sources in the active region. It is also noted that the clipping cannot deal with special cases where a host galaxy is situated outside the union area and the proposed solution to deal with such cases is to use the D3 image.

Figure 19 shows the data flow during the training phase. The first part of the training phase was to preprocess the data, in which three different operations were done – zero centering, size scaling, and horizontal flipping. These preprocessing steps were motivated by (i) the acceleration of the convergence of the learning algorithm; in this case, Stochastic Gradient Descent (SGD). (ii) The receptive field of the last convolution feature map is larger than the input image, so the input image is scaled up as a result. (iii) Allow CLARAN to expect different orientations of the sources. The following step is feature extraction. It was performed at the end of the training phase – where 80,000 iterations were used to find the optimal values for all kernel weights (model convergence) and thus result in feature maps. The feature maps are then shared between the two networks, LocNet and RecNet for location and classification. The only learnable parameters were weights and bias tensors (see Section 1.3.3) that were updated using backpropagation SGD.

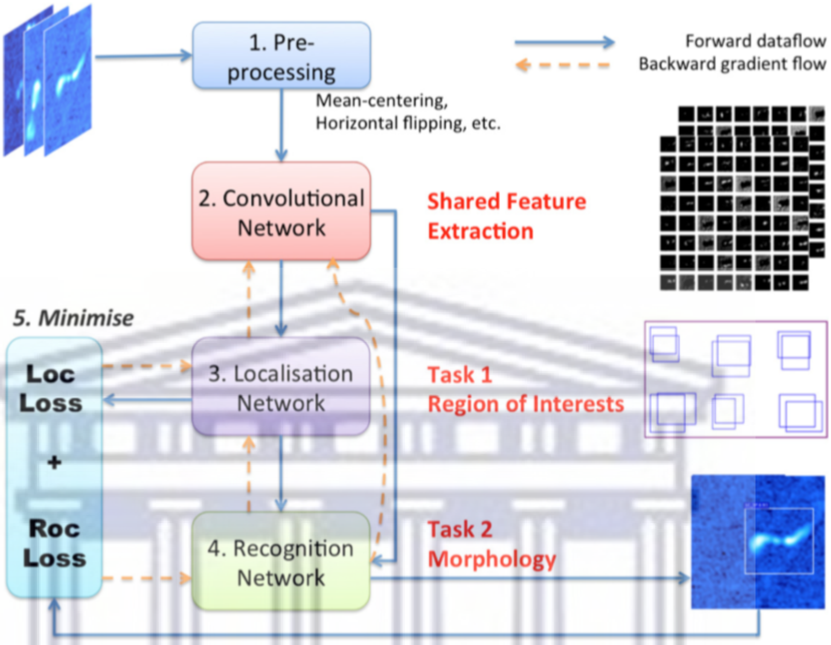


Figure 19: Data flow during training. Blue solid lines showing feed forward data flow and yellow dashed lines show data flow during backward error propagation - learning phase. Image adapted from Wu et al. (2019).

The evaluation of CLARAN on the RGZ test set was done using the average precision (AP) metric. This metric is a function of the most commonly used performance measures in machine learning – precision and recall. Precision measures the fraction of correctly identified sources to all real sources, while recall measures the fraction of correctly identified sources to all identified sources. Average precision is computed by averaging precision across all values of recall, which is the same as computing area under the precision-recall curve. In general, the average precision is given as:

$$AP = \sum_{k=1}^{S} P(k)\Delta r(k) \tag{6}$$

where S is the total number of images in the sample, $P(k)$ is the precision at k number of images and $\Delta r(k)$ is the change in recall measured between $k-1$ and k number of images. The mean AP (mAP shown in Table 2 below) score is then just the average of AP over all classes. These measures were used to evaluate the performance of CLARAN on the testing dataset, using different input images shown in Figure 18. The results are shown in Table 2, where one can see that the algorithm generally performs better when the D4 (and to a lesser extent D3) images are used as input, as shown in their measure of mean AP score. CLARAN outputs the following: (i) the predicted bounding box at the location of each detected radio source; (ii) the class/morphology of each detected radio source represented by number of resolved components and number of flux-density peaks; (iii) the probability value (score) of the predicted class of each detected radio source. It worth noting that CLARAN only performs two tasks, that is, detection and classification of radio sources. The discussion of how CLARAN's predictions were used to estimate the positions of the detected radio sources and cross-identify their IR counterpart is yet to follow in Section 2.7.3.

Table 2: Classification results from CLARAN applied to RGZ dataset with five different input image types indicated as columns and rows represent the average precision score of the respective image type for each class. Boldface numbers indicate the best performance, which is most often achieved when D4 (and to a lesser extent D3) images are used as input. Table adapted from Wu et al. (2019) (see Table 5).

Class	**F**	**D1**	**D2**	**D3**	**D4**
1C_1P	0.809	0.858	0.824	0.849	**0.879**
1C_2P	0.638	0.688	0.684	0.675	**0.707**
1C_3P	0.825	0.882	0.856	0.888	**0.894**
2C_2P	0.747	0.701	0.723	0.798	**0.820**
2C_3P	**0.809**	0.710	0.699	0.805	0.792
3C_3P	0.771	0.864	0.856	**0.942**	0.927
mean AP	78.5%	78.4%	77.4%	82.6%	**83.6%**

2.3.2 Limitations

Similar to other algorithms, CLARAN has some known limitations which restrict its application. The first limitation arises from the fact that CLARAN is sensitive to image noise. There are quite a number of detections where CLARAN estimated a large angular size of the source – even larger than the angular size of the image cutout itself. This is likely due to imaging artefacts around bright sources, these artefacts are amplified because of the logarithmic function used to scale up the pixel intensities. Also, the image noise resulted in CLARAN missing or misclassifying sources in such regions This may imply that CLARAN is confused by those image artefacts or that CLARAN is able to detect diffuse emission. A solution to mitigate these limitations, which we will explore in future, would be to compute the local noise and threshold from there. The second limitation is the angular size of the sources, in that sources extending beyond the adopted cutout size will not be correctly classified but rather broken into smaller components. The latter is called shredding – all characterization algorithms have this inherent weakness. The third limitation we discovered going through the output images was that in a few cases where CLARAN detected faint compact sources (source having a single contour line), it produced two overlapping detections that lie side-by-side which are not desirable. This suggests that CLARAN is over-fitting, which may be due to the large number of trainable parameters used by CLARAN (discussed in Section 2.1). However, we note that the impact of the large number of trainable parameter on CLARAN over-fitting was investigated by conducting two experiments. In the first experiment, the number of model parameters was reduced from 136 to 23 million, and in the second one, reduced to 18 million. The networks were trained using the same testing and training set. The outcome suggests that CLARAN is not in the over-fitting zone (see Section 5.7 in Wu et al., 2019).

Therefore, the overlapping detections that lie side-by-side may not be a result of CLARAN over-fitting. Another possible explanation for these overlapping detection may be the fact that CLARAN detected the noise or diffuse emission. Note that such detections are hard to suppress, in fact, a popular suppression algorithm (to be discussed in Section 2.7.1) struggled to deal with these detections because their intersection-over-union value was very low and as such the suppression algorithm deemed them as detections of separate sources. One possible way to get rid of such detections would be to use their central positions (computed by the source characterization algorithm to discussed Section 2.3) and compute their distance separation, then apply a threshold to say if the distance separation is below a given threshold, take a detection with the highest probability score and get rid of the other detection. Furthermore, It is important to note that CLARAN was trained using image cutouts with a single radio source at the centre. As such, CLARAN operates under the assumption of one source per image cutout. This results in a fourth limitation for CLARAN, whereby it might not be able to correctly identify physically distinct sources appearing next to each other on the cutout – true blended sources, in other words. These limitations will be discussed again in Section 4.1 where we will assess how they affect the source characterization pipeline presented in this study.

2.4 Adapting CLARAN

CLARAN is an open-source proof-of-concept code[8] that was developed in Python 2. Since the community support of Python 2 came to an end with its so-called "End of Life" on January 1st 2020, it was of paramount importance to port it to Python 3. As a result, as part of our work we took the existing python 2 code and upgraded it to Python 3. This task was rather challenging because of syntax differences and different library structures between the two. As a result, we had to go through a substantial portion of the CLARAN code base, update it where necessary and thoroughly test the updated version. After a good amount of code development and testing, the pipeline was running successfully and it was able to reproduce the results from Wu et al. (2019). The updated version can be found in our GitHub repository[9].

CLARAN was also adapted to take a monolithic input image from our radio survey of choice and use the given list of source positions – right ascension (RA) and declination (Dec), to generate radio cutouts of a given angular size and download the corresponding WISE cutouts. We note that archival WISE imaging is not at optimal native resolution, but slightly smoothed to accommodate better source extraction. Hence, more complex blending will occur – using improved imaging will be useful for our purpose (to be discussed in Section 4.1.2). Nevertheless, CLARAN also co-adds WISE images in cases where the requested cutouts extend over multiple WISE tiles. To do the latter and also to export input images to other formats, the code relied on external applications on the local machine. When doing this task from a remote machine such as the ilifu facility (See Section 2.6), it was necessary to access external apps from the local machine by using the X-forwarding command. However, there was a latency as a result of the internet speed connection. This latency was overcome by setting up a virtual display. As part of this process, additional python libraries were added to our python environment (software container) to allow us to run the full workflow within the same environment. This was achieved with the help of the team that developed CLARAN, and in particular Chen Wu from the International Centre for Radio Astronomy Research (ICRAR) in Perth, Australia. This also resulted in a more versatile software container for general use on the ilifu facility.

2.5 GMRT and WISE Data

We used data from a deep wide-area survey carried out with the GMRT at 610 MHz and centered on the sky area known as European Large Area ISO Survey

[8] `https://github.com/chenwuperth/rgz_rcnn`

[9] `https://github.com/Mofokeng-C/rgz_rcnn_py3`

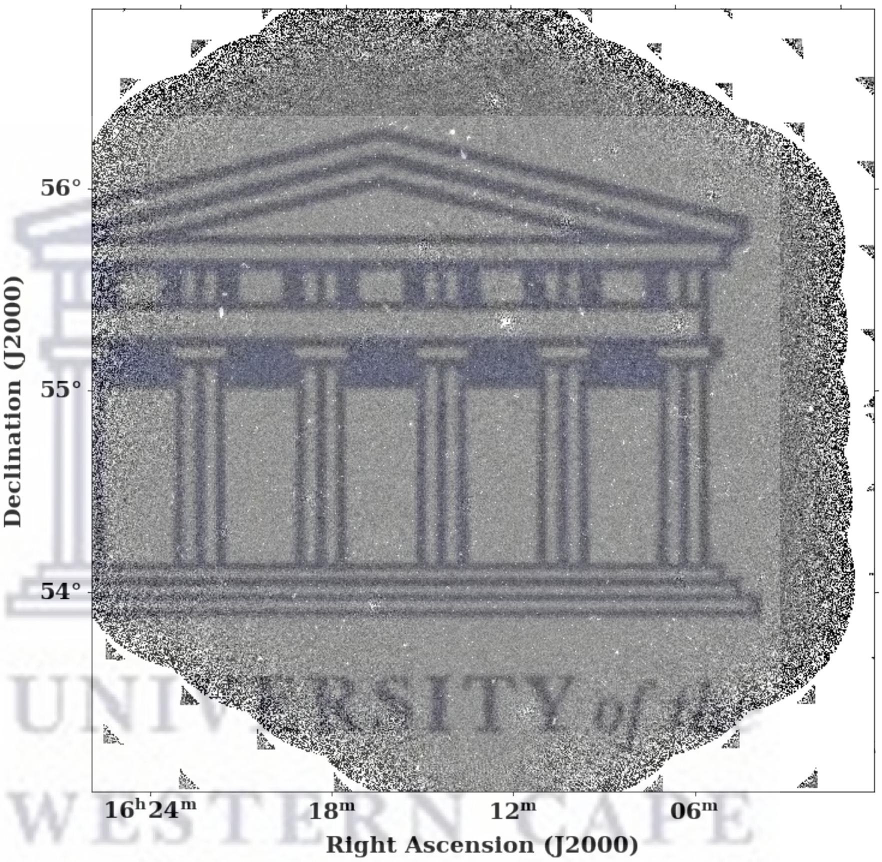
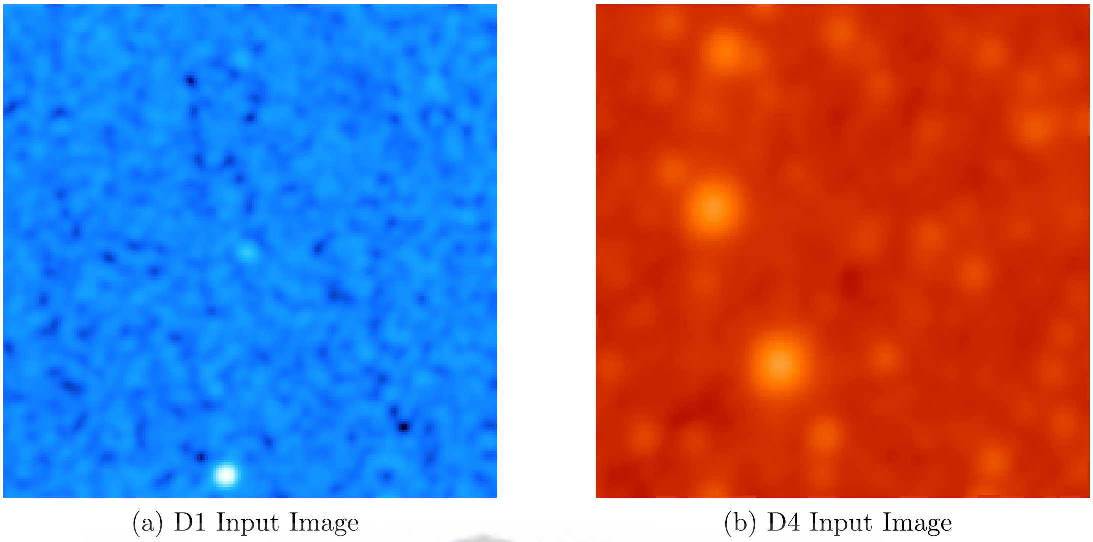


Figure 20: The final "cleaned" output image of 12.8 square degrees area in the ELAIS-N1 field, as observed by the GMRT telescope at 610 MHz wavelength (Ishwara-Chandra et al., 2020).

North 1 (ELAIS-N1). These observations cover an area of 12.8 degrees^2 at a synthesized beam resolution of ~ 6 arcsec and a root-mean-square (RMS) noise of ~ 40 μJy/beam (Ishwara-Chandra et al., 2020). Figure 20 shows the area covered for these observations. The figure shows a cleaned image, and hence it has the artifact features inherent to recovering the beam. Also note the additional noise at the edges, the crust, which will create problems with source detection – resulting in too many false detections. Following the production of the final radio image, a source catalogue was created using a popular source detection algorithm known as PyBDSF (briefly discussed in Section 1.4). A total of 6,400 sources were catalogued above the flux-density threshold of 5σ. To detect sources with PyBDSF, a user must define the size of the rms_box – a typical scale where the background varies substantially. Often a user must define multiple boxes in order to handle artifacts around bright sources and the edges of the image. Also, to detect extended emission and fit it accurately, a specific configuration of the software must be adopted, which may then require a long time to develop. An ideal algorithm should be able to perform source characterization over the large area covered by our GMRT observations in an automated manner while correctly identifying point-like as well as extended emission. While such an algorithm does not exist as yet, CLARAN comes close to solving some of these challenges over relatively large cutouts (see e.g. Figure 15 in Wu et al., 2019 and Section 4.2 of this thesis). Thus, we adopted and adapted CLARAN so as to be able to apply it to our GMRT dataset.

Based on some exchanges with the CLARAN developers and on a preliminary visual inspection of our radio sample, we evaluated that 3 arcmin was a suitable angular size for the cutouts to be extracted and fed to CLARAN. Since the pixel scale of the GMRT image is 1 arcsec/pixel, 3 arcmin corresponds to 180 pixels. We thus generated 181 × 181 pixel cutouts – an example is shown in Figure 21a – centered on the pixel containing the radio source position determined with PyBDSF. This resulted in 6,400 radio images that were preprocessed to meet CLARAN's requirements.

In order to obtain matched infrared cutouts, we used the GMRT positions from the PyBDSF catalogue to get infrared cutouts of the same angular size. Infrared cutouts were obtained from WISE (discussed in Section 2.2.1). As noted previously (and as shown by Figure 21b), the WISE image looks fuzzy compared to the radio image (Figure 21a) because the archival WISE images are smoothed – optimized for source detection, which may be detrimental to CLARAN and source identification. However, we note that we are using native W1 resolution (6 arcsec), implying the two beams match – which suggests that the WISE data


(a) D1 Input Image

(b) D4 Input Image

Figure 21: A 3 arcmin × 3 arcmin image for (a) radio and (b) IR data image from GRMT and WISE surveys, respectively. The two images are centred at the same position. While the color scales of the two images are different, white represents bright emission in both cases.

may still be useful for our study. Also, we aim to test the impact of applying transfer learning on CLARAN – by using the same IR data as (Wu et al., 2019) (i.e., WISE data) and only changing radio data. For our study, we only use the W1 band (3.4 microns) images. Similar to radio cutouts, infrared cutouts were preprocessed and bad pixels were filled with the mean value of the cutout. An example cutout is shown in Figure 21b. At first we followed the CLARAN preprocessing pipeline in applying a contrast value of 0.684039 to the IR PNG images. However, in so doing faint IR sources appeared washed-out, making it difficult to distinguish them when visually inspecting the output. To try and improve upon this, mean background subtracted image data was used, and the contrast values was decreased to 0.30 in DS9. The Legacy Survey Sky Viewer[10] was used as a reference to compare our IR PNG images and make sure that they displayed fainter IR sources correctly. After determining the optimal choice of contrast, the output IR PNGs were then used with CLARAN. This resulted in improved probability scores in most examples. Also, the process of mean-subtraction (3σ-clipping) and decreasing the contrast of the images resulted in fainter IR sources being enhanced relative to the noise. Consequently, the number of predictions decreased further. For this reason, the study presented in this thesis contain results of the IR image data that was subjected to mean background subtraction and also the contrast of the IR PNG image decreased using DS9.

[10] http://legacysurvey.org/viewer

The data analysis process required large computational and storage resources for preprocessing, classification with CLARAN and post-processing. Data analysis was carried out on the ilifu cloud computing research facility developed by the Inter-University Institute for Data Intensive Astronomy (IDIA).

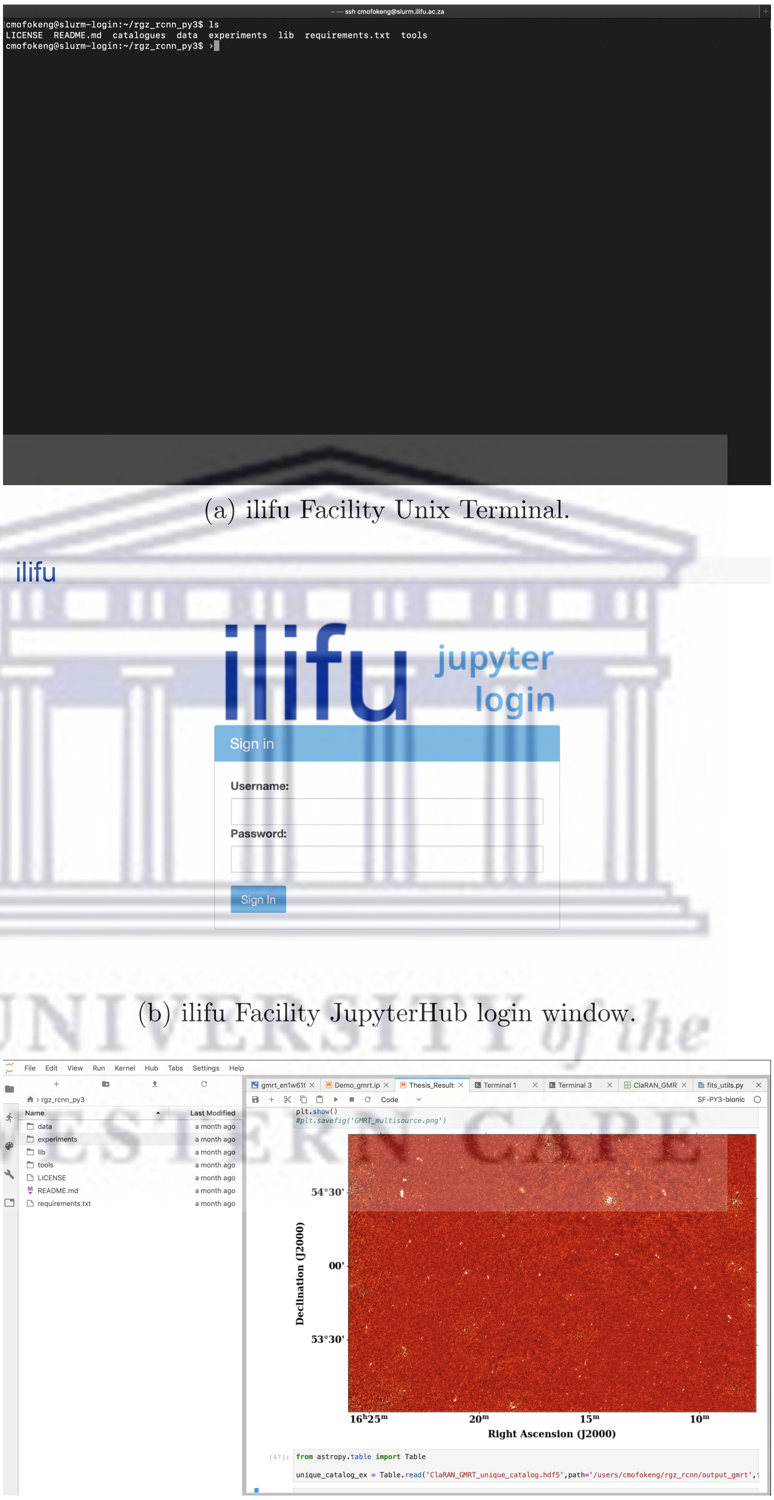
2.6 The ilifu Cloud Computing Research Facility

The Inter-University Institute for Data Intensive Astronomy (IDIA[11]) is a consortium of three South African universities: the University of Cape Town (UCT), the University of Pretoria (UP) and the University of the Western Cape (UWC). The three institutes partnered to tackle challenges expected from large volumes of data from the next generation of radio sky surveys. One of the main goals of IDIA is to develop tools to store, re-process, and analyze data from the MeerKAT telescope and other SKA precursors and pathfinders. In order to achieve this goal, IDIA developed a cloud computing research facility for data intensive astronomy. The original IDIA facility had 40 compute nodes, each having a 2.6 GHz Xeon Processors, with 32 cores and 256 GB RAM. The facility also had a few graphical processing units (GPU), which are increasingly being used for data processing in this era of big data. GPUs substantially accelerate certain computational workloads and are thus preferred to CPUs for some tasks. The IDIA facility therefore had 2 × NVidia P100 GPUs in 4 of its nodes. It had a storage capacity of more than 1 PB, and was connected to SANReN (South African Research Network) at 10 Gb/s. The technologies underpinning the IDIA facility have been receiving a lot of attention from researchers in other fields. Thanks to investment from UCT CBIO (Computation Biology) group and from DIRISA (Data Intensive Research Initiative of South Africa), the IDIA facility's computing and storage capabilities have recently been expanded upon by a factor of about 3, and the resulting system is now known as the ilifu cloud computing research facility[12].

The ilifu facility allows for real-time analysis for users across the globe, thus allowing researchers to collaborate by running tasks on the same science platform. The science platform allows its users to deploy different processing environments at the same time, based on the 'containerization' technology powered by Singularity. Although the facility can support custom environments, the most commonly used environments are the Unix terminal and the JupyterHub GUI, shown in Figure 22a and 22b. The Unix bash shell is used for non-interactive tasks such as running processes taking several hours to execute. For interactive work and for data visualization, JupyterHub is mostly employed.

[11] https://www.idia.ac.za

[12] http://www.ilifu.ac.za



```
~ — ssh cmofokeng@slurm.ilifu.ac.za
cmofokeng@slurm-login:~/rgz_rcnn_py3$ ls
LICENSE  README.md  catalogues  data  experiments  lib  requirements.txt  tools
cmofokeng@slurm-login:~/rgz_rcnn_py3$ ›
```

(a) ilifu Facility Unix Terminal.

(b) ilifu Facility JupyterHub login window.

```
/ rgz_rcnn_py3
Name                 Last Modified
data                 a month ago
experiments          a month ago
lib                  a month ago
tools                a month ago
LICENSE              a month ago
README.md            a month ago
requirements.txt     a month ago
```

```
plt.show()
#plt.savefig('GMRT_multisource.png')
```

```
[47]: from astropy.table import Table

      unique_catalog_ex = Table.read('ClaRAN_GMRT_unique_catalog.hdf5',path='/users/cmofokeng/rgz_rcnn/output_gmrt',
```

(c) ilifu Facility JupyterHub virtual environment.

Figure 22: Accessing the ilifu cloud computing research facility through the Unix terminal and the JupyterHub virtual environment.

Although the Unix terminal was used to execute a variety of tasks, since our work focused on image data, visualization played a vital role. As previously discussed and as shown in Figure 22, the ilifu facility allows users to deploy different environments and thus run different tasks in a visually-rich environment. In Figure 22c, a JupyterHub virtual environment is shown. The left window pane shows the available directories inside the user's home directory. To the left of the pane is the sidebar, which on top, a directory icon is shown. This icon is related to the directories and files that are shown. Below it is an icon of a running human being. It is used to show tasks currently being run by the user (e.g., Jupyter-notebooks, bash terminals, etc.). The rest of the icons are used to format the environment to user-specific settings. To the right of the figure, a Jupyter-notebook is open. On top of this figure, there is a taskbar. As shown in by the taskbar in this figure, a user can have several tasks open at the same time. In this case, two Jupyter-notebooks are running; two bash terminals are open and as well as a couple of files (an image, python script file, and a text file). The possibility to run several tasks and workflows at the same time allowed us to rapidly prototype and execute our data analysis pipeline.

2.7 Applying CLARAN to GMRT Data

The term transfer learning is used to describe the process of using what was learned from data in one domain and apply it to make predictions on data from another domain (Goodfellow et al., 2016). This process takes advantage of the information extracted from the data in the first domain and uses it in the second domain when learning or when directly making predictions. In this context, a different domain can refer to different problems in different science fields or to different problems in the same science field. In our case, we apply transfer learning to use what CLARAN learned from the RGZ data and apply it to make predictions for the GMRT data described in Section 2.5. The resolution of GMRT observations is similar to the resolution of the FIRST survey used by CLARAN's pre-trained model (see Section 2.2), albeit with different depths and different frequency bandwidths. Another key difference is the lower frequency of GMRT, which will have many more sources that have steep indices, it will also have brighter features, notably the extended hotspots and jets. Thus, another aim is to provide proof that using transfer learning predictions can be made accurately on data from different surveys.

While pre-processing the radio and infrared input image datasets, an error was encountered when generating radio contours for a few sources. Such cutouts were visually inspected, and it was noted that one or more of the following applied: a) some of the cutout pixels were blank — may be due to cutouts that were close to the edge of the field; b) cutouts were characterized by high RMS noise values. While improved pre-processing could have been carried out to fix some of these issues, we simply decided to remove these cutouts from our sample. The resulting dataset consisted of a total of 6,330 images for both the D1 and the D4 datasets, with a total of 70 cutouts excluded from the input dataset. The cutouts that were not part of the input data are shown in Figure 23, as blue square boxes. While there are some clusters of sources at the edge of the field where the image is generally noisier, by and large 'missing' sources are distributed somewhat homogeneously across the field. There is one square box with a source in it, and it was noted that the source is towards the edge of the 3 arcmin cutout and the source extends well beyond the 3 arcmin size, which may explain why CLARAN may have not detected this source. However, for the most part the blue boxes are empty.

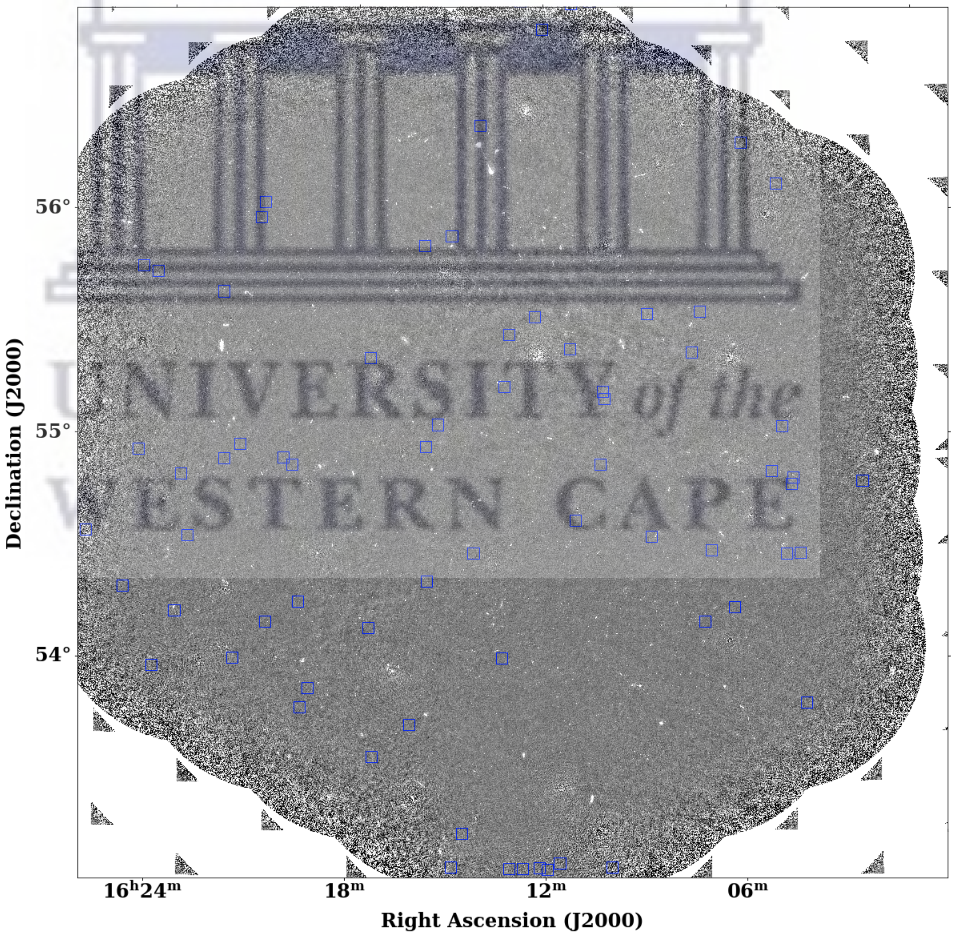
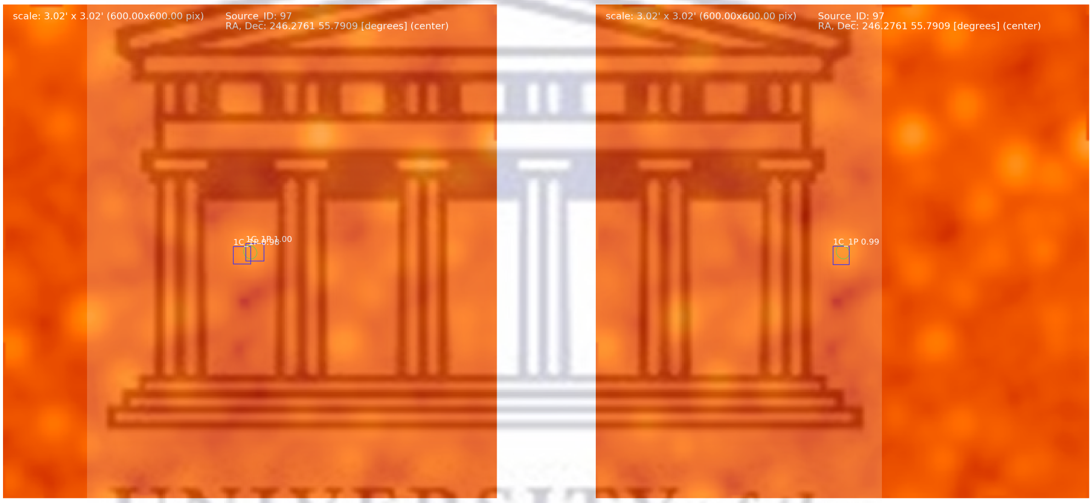


Figure 23: The final output image of 12.8 square degrees area in the ELAIS-N1 field, as seen by the GMRT telescope at 610 MHz wavelength. The blue square boxes indicate the cutouts excluded from the input dataset. Note that the size of each box in this figure represents the original 3 arcmin (181 pixels) cutout.

2.7.1 Cutouts Overlap

Since we generated 3 arcmin-side image cutouts for all PyBDSF detections, several radio sources will appear in multiple cutouts and will thus be classified more than once by CLARAN. In order to produce a catalogue of "unique" sources from CLARAN's output, a filtering process to remove duplicate detections is thus required. The problem of having overlapping predictions from a CNN is a well-known issue in computer vision, however there are dedicated algorithms used to reduce such detections. The most popular algorithm used for this task is known as the Non-maxima suppression (NMS). The NMS algorithm compare the predicted bounding boxes belonging to a single source by computing the degree of the overlap (intersection-over-union, IOU) between these boxes (Rothe et al., 2014). Using a pre-defined overlap threshold, the algorithm keeps only the boxes with IOU less than the pre-defined threshold. As a result, the NMS algorithm has got some limitations. For instance, even if a detection has a high probability score, it will be suppressed if it has an IOU higher than the threshold, and as such if detections are lying side-by-side one of them will generally be removed (as shown in Figure 24a). However, the algorithm occasionally fails when it comes to some examples showing faint sources displaying a single peak. Figure 24a shows such an example, where it predicts bounding boxes lying side-by-side. On the other hand, if a detection has a low probability score but with an IOU less than the threshold, such a detection will be kept. As a result one can fail to detect nearby sources. However, due to the fact that our image cutouts were generated at each source position from the catalogue, all the sources are expected to be observed in multiple cutouts, and as such the limitation of the NMS algorithm described above is presumably unlikely to significantly affect our results.

To avoid complications when applying this algorithm, we used the pixel coordinates. We adapted predefined hyper-parameters from the pipeline, with the overlapping threshold set at 0.2 (previously set at 0.5), while for the score threshold the predicted score of each box was used. Figure 24b shows the output of using the threshold of 0.2, whereby the algorithm is able to remove a less reliable detection of the same source lying side-by-side. Even with this different choice of parameters, visualizing all the detections on the full mosaic on the observation (shown in Figure 20), we realized that we get rid of a substantial amount of true detections that were suppressed by the NMS algorithm either because they are lying side-by-side with other detections or because they had a low probability score compared to other close-by detections (above mentioned limitations of the NMS algorithm). Therefore to remove duplicate detections as a result of the cutouts overlap, we only used the new value of the NMS threshold when detecting the sources and opted to implement a new approach to handle duplicate detections predicted by CLARAN. We implemented a simple approach that relies on the geometric centres of the boxes. For each predicted bounding boxes, we searched for other predictions having the centres within a respective bounding box (i.e overlapping predictions). Such predictions likely predict the same source if the predicted classes are the same, if that is the case we retain a detection with the highest probability and remove the rest, else we keep all the detections. This approach was very effective except for when the overlapping bounding box predicted different classes. However, this approach combined with the approach to be discussed in the following paragraph, were able to remove most of the duplicate detections efficiently.


scale: 3.02' x 3.02' (600.00x600.00 pix)
Source_ID: 97
RA, Dec: 246.2761 55.7909 [degrees] (center)
1C_1P 0.98
1C_1P 1.00
scale: 3.02' x 3.02' (600.00x600.00 pix)
Source_ID: 97
RA, Dec: 246.2761 55.7909 [degrees] (center)
1C_1P 0.99

(a) NMS threshold of 0.5

(b) NMS threshold 0.2

Figure 24: Images of a field size of 3-acrmin centred at source of ID 97. The two images show a single source detected using two different values of the NMS threshold of the proposal network when running the pipeline; (a) 0.5 and (b) 0.2.

2.7.2 Edge Detections

The input dataset was derived by generating cutouts around each source position based on a predefined input list. As a result, our focus is mostly on detections at the centre of the cutout. Another filtering algorithm we implemented is thus based on the position of the predicted bounding box. In Section 2.3.2, we highlighted that CLARAN is affected by the size of the image cutout, whereby elongated sources extending well beyond the size of the image cutout will be missed or misclassified. Consequently, sources at the edges of the cutouts are most likely to be misclassified, as they might be extending beyond the given image cutout. As a result, detections of sources at the edge of the cutouts were rejected. To define the edge of the cutout, distributions of the size of the sides of the bounding boxes per class were visualized. As expected, the distribution of the 3C_3P sources had the widest ranges that covered all of the other distributions (this will be covered in Section 3.4). This distribution of the size of the bounding box sides for the 3C_3P class has a median value just below 120 pixels. We used this median value as the threshold. All the detections that are beyond a radius of 60 pixels from the centre of the cutout are regarded as edge detections. In other words, only detections within a radius of 60 pixels ($\sim$ 1 arcmin) from the centre of the cutout are retained. Removing edge detections is not likely going to affect our detections because all the sources have a particular cutout at which they are located at the centre of the cutout. Moreover, the sources close to the centre of the cutout may have better characterization compared to their "dupe" brother at the edge of another cut-out. Hence, it does matter which duplicate you end up using, those from the centre of the cut-out is best.

2.7.3 Source Characterization Pipeline

The development of CLARAN was driven by the need to classify radio sources based on their morphology. However, in this work we set out to develop CLARAN into a tool for radio source characterization, which we defined as the combination of source detection, classification and identification. The details of each step are provided below.

2.7.3.1 Detection

For the detection step, pixel/flux values within each bounding box identified by CLARAN were used to compute the flux-weighted center-of-mass of the bounding box, which is then used as the source position of the source detected by CLARAN. In computing the center-of-mass, the first attempt was using the PNG images and then masking a region given by the predicted bounding boxes. However, the central positions were always biased toward the center of the bounding box. This was mainly due to the logarithmic function adopted to scale flux values when these images were created. We then changed the approach and used the actual pixel values from the original FITS images and computed the centre-of-mass positions within each bounding box. This was done for both radio and infrared images. For infrared images, the center-of-mass at first showed up mostly at the center of the bounding box. Visually inspecting a few cutouts and performing statistics on them (mean, standard deviation, etc.), it was soon realized that while our radio images have zero-mean because of the way FIRST data reduction was carried out, our infrared images from WISE have a mean value which is measurably greater than zero, skewing the center-of-mass calculation. Mean background subtraction was thus applied to both the radio and infrared images before computing the center of mass. The mean subtraction did not significantly affect radio data since the mean is basically zero in this case. Mean subtraction however caused significant changes when computing the center-of-mass for infrared data. This helped in particular when the the infrared image was dominated by a single bright source. However, most infrared cutouts clearly showed several infrared sources within the bounding boxes. Thus, computing the correct infrared center-of-mass for the corresponding radio source is particularly challenging and the result often is not very helpful. For this reason, only the radio image data was eventually used to compute the center-of-mass position, which we will hereafter simply refer to as Radio Centre or RC. The RC is thus the center-of-mass of the radio signal within the bounding box determined by CLARAN. Upon visually inspecting the results of the RC calculation for a few cutouts, we realized that in a few cases the RC was quite some way off from the expected center-of-mass position. This was mostly observed in cutouts where CLARAN detected three or more radio source components within a large bounding box. In such cases the center-of-mass calculation is likely rather unreliable. As shown by Figure 25, the histogram of flux values within a random 3 arcmin image is characterized by a high peak at low flux values due to the background/noise signal and by a low tail at high flux values due to the signal from actual radio sources. The two red solid lines represent the lower and upper boundary values of the 3σ-clip threshold. This is indicative that, to perform source detection and positioning accurately – similar to most source detection algorithms, the background pixels must first be removed and only the pixels associated with the source must be employed. To identify the background pixels, the median of the respective cutout was measured (shown as the black dashed line in Figure 25) and the associated standard deviation (σ) then used to determine the detection threshold. Note that the normal standard deviation is affected by extremely high and low flux-density values. In such cases, the median absolute deviation (MAD) is the best estimator of the spread of the data. As a result, σ refers to the MAD of the flux-density values in the given cutout. The signal threshold to flag background pixels was determined by adding 3σ to the median signal – the upper boundary value in Figure 25. All pixels having flux values below the threshold were then removed (i.e., assigned a value of zero) before computing the RC.

Figure 26 shows a 3-arcmin cutout of the flux values represented in Figure 25. CLARAN detected an extended source with the bounding box overlaid as a white rectangular box. The region bounded by the box was extracted and used to compute the RC position. The RC positions before and after removing the back-

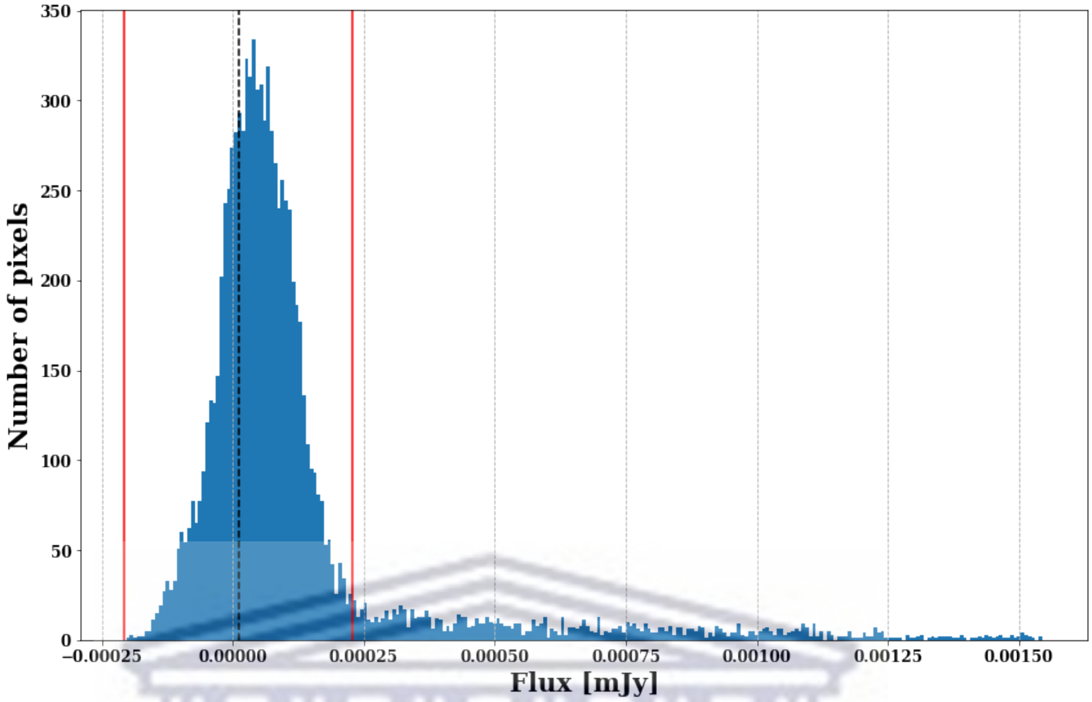


Figure 25: Distribution of flux values from a randomly selected 3 arcmin cutout. The red solid lines indicate the lower and upper boundaries of the 3σ clip interval centered on the median value (black dashed line).

ground noise is shown in Figure 26. In this figure, the red and the blue crosses represent the RC position for the "raw" and for the 3σ-clipped image. It is clear that the red cross is slightly off from the expected RC position. In some cutouts this effect was significant and the RC position was observed to be way off from the expected position. The sigma-clipped data result in better positions of the RC. Thus, the RC presented from here onward, is the position that was computed after the application of 3σ clipping on the respective 3-arcmin cutout.

2.7.3.2 Classification

Source classification is one of the most fundamental aspects in astronomy. As such there are several schemes defined by astronomers to try and make sense of the radio sources, such as the ones discussed in Section 1.2. In our case, we stuck with using the original classification scheme used to compile the RGZ dataset and thus adopted by CLARAN. Since one of the main aims of this study is to prove the effectiveness of transfer learning, it follows that we must adopt the same classification scheme in order to re-use the pre-trained CLARAN model and apply it to our GMRT data. In this scheme, classes are represented by the number of components and the number of peaks. For example, one component-one peak 1C_1P represents a compact radio source, whereas two components-two peaks 2C_2P represents a radio source with two disconnected components sufficiently

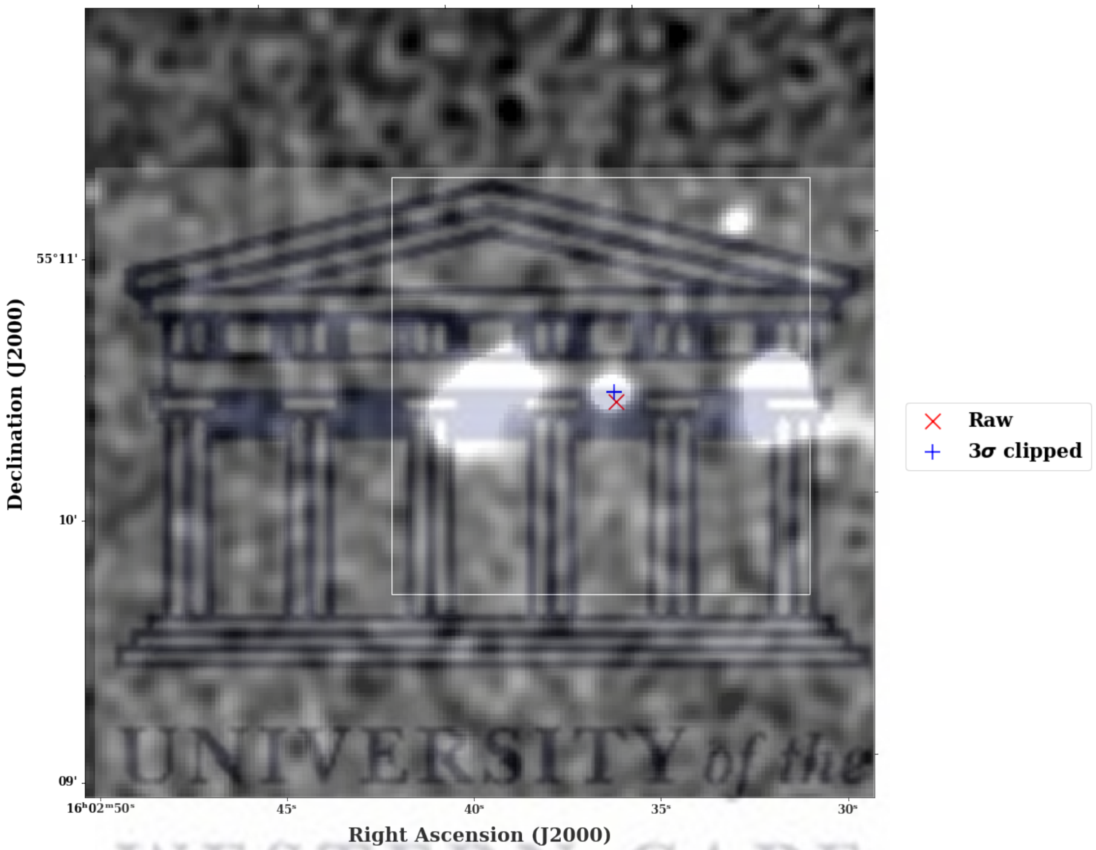


Figure 26: A 3 arcmin × 3 arcmin cutout viewed using DS9 zscale. A bounding box from CLARAN is shown by a white rectangular box. The histogram in Figure 25 represent flux values of this cutout. The white blobs indicate the source signal. The crosses are used to indicate the RC position the raw and the 3σ clipped data. The red cross represents the RC computed from the raw data, while the blue cross represents the RC computed from the 3σ clipped data.

close to one another to be regarded as a single radio source. A three components-three peaks 3C_3P source represents a radio source with three disconnected components close enough to be considered as components of same source.

2.7.3.3 Identification

To study the potential infrared host galaxy, we followed the same approach as the RGZ project (Banfield et al., 2015; Alger et al., 2018). The infrared host galaxy is approximated as the catalogued infrared source closest to centre of the identified radio source(s). Firstly, we searched the AllWISE (Cutri et al., 2013) catalogue to get the positions of infrared sources within the cutout, which will be used to identify the IR host. We then used D1 and D4 maps as input images and overlaid the positions of the infrared sources on the given input image, furthermore using World Coordinate System (WCS) information of the given radio image to transform between sky and pixel coordinates. This process was somewhat challenging because of the discrepancy in axis / pixel ordering between FITS files and Python packages for computer vision and visualization. Lists and arrays in python use zero-based indexing whereas FITS files use one-based indexing. In addition, in python the top-left pixel in an N-dimensional array is represented by [0,0] indices, which marks the "origin", and these indices are row-major ordered, whereas for FITS files the bottom-left pixel marks the origin [1,1]. Therefore, y- and x-axis are represented by a row and a column, respectively–in a row-major ordered array. For this study, python was used for resampling the arrays and for visualizing the output. Therefore, positions of the infrared sources had to be transformed to follow the row-major ordering. At the end of this process, the angular separation between the RC position and the infrared sources within the given cutout was computed. The infrared source within the CLARAN bounding box closest to the RC position was taken as the most likely infrared host of the radio source. While this approach is satisfactory in most cases, it is crucially dependent on the astrophysics since the radio emission and infrared hosts may not be co-aligned, making the cross-identification process difficult with completely different wavelength. Moreover, the approach also depends on the quality (i.e. the completeness and reliability) of the infrared source catalogue as well as on the accuracy of the RC positional estimate.

2.7.3.4 Validation

In machine learning, validation is generally performed with respect to an existing test set for which the target variables (labels) are already known. Validation is then the process of feeding the test set to the machine learning algorithm and then comparing the predicted target labels with the known labels of the test data. Note that in our work we use test data and validation data interchangeably. Evaluation metrics such as precision (often referred to as reliability) and recall (often referred to as completeness) can then be used to optimize the hyper-parameters of the model, to search for the best features and to evaluate the performance of model. That is, to balance the completeness versus reliability and that is most difficult, and sometimes one is relaxed to improve the other, depending on the survey goals. However, for our GMRT dataset, while we have a radio source catalogue created with PyBDSF, the classification of extended sources has not been carried out via visual inspection. Since one of the fundamental challenges CLARAN tries to address with the RGZ dataset is distinguishing the components and peaks of a given radio source from those belonging to other radio sources, one of the first things we decided to carry out was a visual inspection of CLARAN's output and use it to compute an estimate of CLARAN's performance.

Namely, we set out to estimate the completeness (recall) and reliability (precision) of CLARAN's predictions, defined as follows:

$$completeness(recall) = \frac{tp}{tp + fn} \tag{7}$$

and,

$$reliability(precision) = \frac{tp}{tp + fp} \tag{8}$$

where

- a true positive (tp) is defined as a correct classification of an extended source (i.e. a source that is not 1C_1P, or a point source)
- a false positive (fp) is a point source classified as an extended source
- and a false negative (fn) is a missed extended source (i.e an extended source which was either not detected by CLARAN or incorrectly classified as a point source).

In essence we compute an estimate of CLARAN's performance when it comes to detecting and correctly classifying extended radio sources.

A similar process can in principle be carried out to estimate the performance of the source characterization pipeline, with the focus on the estimated positions of the IR hosts. However, this is more problematic since several radio sources will not be visible in WISE images. This is to be expected since deep radio images see very high redshift (z) sources (because of the negative k-correction), where the infrared is sensitive to galaxies with z ≤ 0.5. This task is also challenging for cases when there are a handful of IR sources clustered together. Thus we settled on evaluating the performance of the source classification step.

2.7.3.5 Codebase

The source characterization pipeline can be found on our github repository `https://github.com/Mofokeng-C/rgz_rcnn_py3/tree/sc_pipeline`.

Chapter 3

3 Results

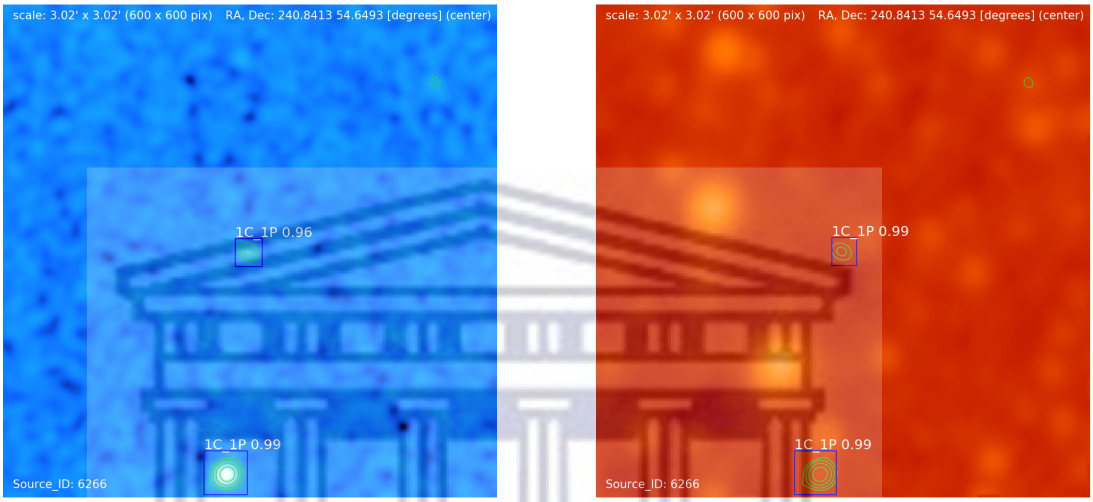
In this work, we applied a pre-trained model to directly classify unseen image data from GMRT 610 MHz observations of the EN1 field coupled with WISE infrared images of the same field. In this Chapter we detail the results we achieved on the GMRT dataset.

3.1 CLARAN Output

Our input dataset, after excluding rare failures during the pre-processing stage, is made up by 6,330 image cutouts of 3 arcmin × 3 arcmin. As input, both the D1 and the D4 datasets were used and the relative performance we achieved was compared. As previously discussed, the D1 dataset shown on the left panel 27a is only based on the radio image, while the D4 dataset shown in the right panel 27b is a fused image combining the radio and infrared images. Figure 27 shows a typical CLARAN output from the D1 and D4 datasets, specifically for source ID 6266, previously shown in Figure 21. For each input cutout of size 181 × 181 pixels, CLARAN outputs a resampled image of 600 × 600 pixels, overlaid with predicted bounding boxes and classes and probability scores annotated on top of the bounding boxes. Also, the pipeline was tweaked to overlay information about the size of the image (3 arcmin × 3 arcmin in our case), the coordinates and source ID of the source at the centre of the cutout. In the following sections, examples from the output datasets are shown for each input image type.

Figure 27a shows an example output of CLARAN when supplied with a D1 input image. In this example, CLARAN located and classified two 1C_1P (compact) radio sources. As seen on the figure, the central source is faint compared to the source below it, and as such CLARAN assigned a probability score of 96% to the faint source. However the source detected below the central source is the brightest in this field. Thus, CLARAN is confident about the classification assigning a higher probability score of 99% to the prediction. The reason for the discrepancy in the predicted probability score, as it will become clear in other examples to follow, is the background signal and the surface brightness of the source.

Figure 27b shows the corresponding output from the D4 input image. It is clear in both examples that CLARAN produced similar predictions for the locations of the radio sources and the predicted classes. The only difference is the predicted probability score of the faint central source. For the D4 example, CLARAN returns a high probability score (99%) that the source is compact (1C_1P) for both the central and bottom source. One of the reasons for this difference in the classifications between the two datasets, as highlighted in the previous paragraph, is that the D1 dataset exposes more background signal to CLARAN. Also, the surface brightness of the sources affects CLARAN's predictions. However, this effect is less significant when using multi-wavelength information by overlaying radio contours on the IR background image, resulting in better predictions from CLARAN.


scale: 3.02' x 3.02' (600 x 600 pix) RA, Dec: 240.8413 54.6493 [degrees] (center)
1C_1P 0.96
1C_1P 0.99
Source_ID: 6266
scale: 3.02' x 3.02' (600 x 600 pix) RA, Dec: 240.8413 54.6493 [degrees] (center)
1C_1P 0.99
1C_1P 0.99
Source_ID: 6266

(a) D1 Output Image

(b) D4 Output Image

Figure 27: An example CLARAN output image. Output image for radio source ID: 6266 from (a) D1 and (b) D4 input images are shown. The predicted sources are indicated with a blue bounding box, accompanied by the predicted class and probability score on top of the bounding box.

3.2 Detection

While the detection of excess signal is performed by CLARAN, our source characterization pipeline also allows us to estimate a source position from CLARAN's bounding box for every source that was detected by CLARAN. Figure 28 shows the output of the source characterization pipeline, namely the source positions (Radio Centre, or RC) estimated are shown as blue crosses. In this figure, lime crosses represent source positions from PyBDSF, whereas the black crosses represent catalogued IR source positions in this cutout. From this figure, it is clear that for both detected sources, the RC source positions are very close to the source positions from PyBDSF. This suggests that the characterization pipeline can accurately estimate the source positions of point sources in a similar way to PyBDSF. It is important to note that only as long as sources are detected by CLARAN the source positions can successfully be estimated by our pipeline.

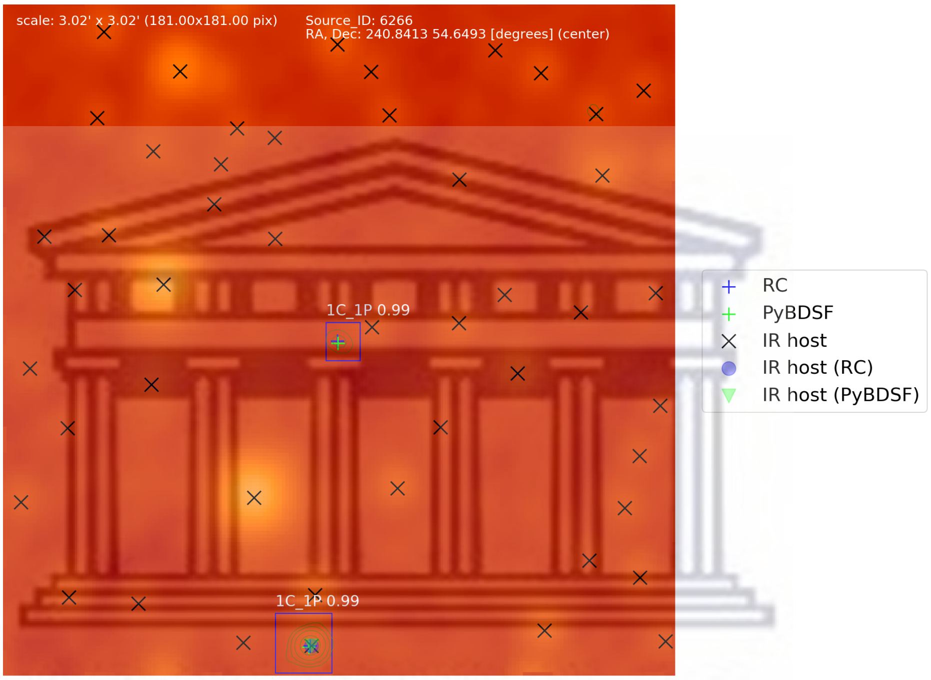


Figure 28: An example output image of size 3 arcmin × 3 arcmin, centred at source of ID: 6266, showing the detection algorithm's output. Black crosses represent catalogued positions of the IR sources in the cutout, lime crosses represents the source positions of the cutout from GMRT catalogue, while blue crosses represents the radio centre of the bounding box. The output from the identification algorithm is also shown by the blue shaded circle and the green shaded triangle which indicates the estimated positions of the IR host as determined from the RC and PyBDSF central positions, respectively.

In addition to cutouts which were removed from the input data (see Section 2.3), there are a few images where CLARAN "missed" the sources detected by the PyBDSF algorithm. Figure 29 shows a distribution of flux values for all detected sources by PyBDSF (shown as a blue distribution) and the sources "missed" by CLARAN (for both the D1 and the D4 datasets) shown as green and orange distributions, respectively. From this figure, it is clear that although rare bright sources are being missed, the vast majority of the missed sources is at the faint end. The rare bright sources might have been missed because CLARAN's training examples did not have enough sources of this nature, as highlighted previously that the GMRT data is expected to have more bright sources. As previously mentioned, the background noise and the surface brightness of the sources affect CLARAN's capabilities. From the figure it is also clear that the D4 dataset outperforms the D1 dataset in terms of detection, whereby 5,701 (90.1%) D4 images returned a detection compared to 4,212 (66.5%) D1 images.

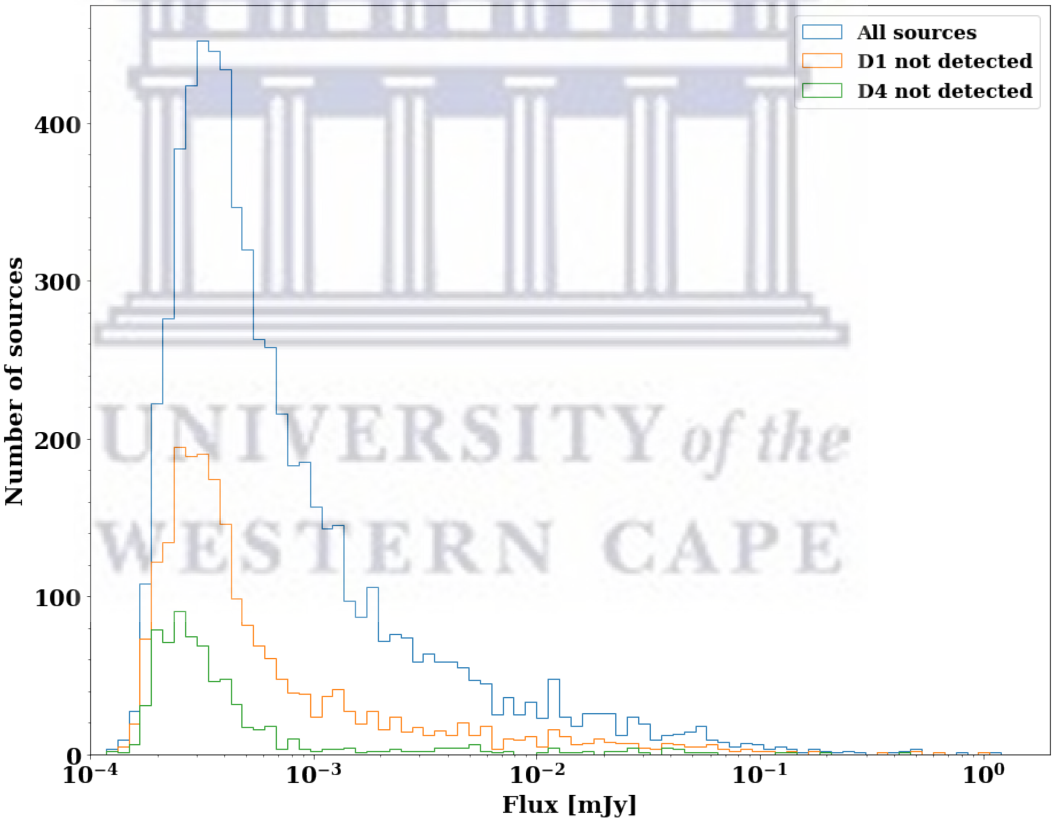
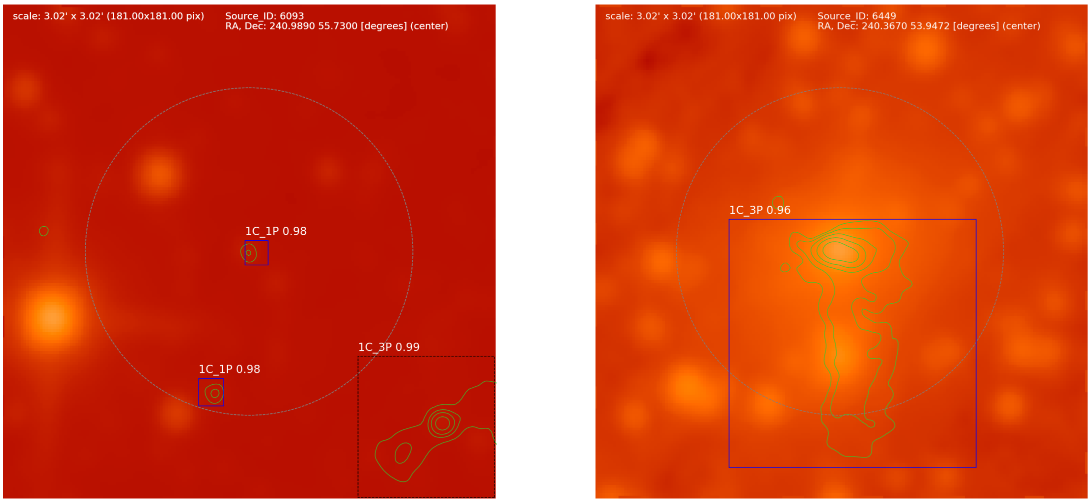


Figure 29: A bar graph showing the distribution of flux densities for detected sources. The blue bars represent flux values for all sources detected by PyBDSF, while the orange and green bars represent flux values of sources not detected by CLARAN, for the D1 and the datasets, respectively.


scale: 3.02' x 3.02' (181.00x181.00 pix)
Source_ID: 6093
RA, Dec: 240.9890 55.7300 [degrees] (center)
1C_1P 0.98
1C_1P 0.98
1C_3P 0.99
scale: 3.02' x 3.02' (181.00x181.00 pix)
Source_ID: 6449
RA, Dec: 240.3670 53.9472 [degrees] (center)
1C_3P 0.96

(a)

(b)

Figure 30: Example images of the filtering process applied to each image in our catalogues. In both images, the grey dashed-line circle represents the 60 pixels (∼ 1 arcmin) radius used retain detections within it. The detections beyond the preset radius, for instance in (a) the detection indicated by the black dashed-line rectangle) are removed from the final catalogues. (b) shows an elongated source whose central position is within the radius, thus the detection is retained.

3.3 Filtering

Note that at this stage our catalogues still contain several duplicate predictions arising from multiple (physically distinct) radio sources appearing in most cutouts, e.g. the D4 catalogue has over 11,000 entries, or about twice the number of individual cutouts. As a result, two filtering approaches were applied as previously discussed in Section 2.7. Figure 27 and 28 are resulting images from the catalogues before applying the filtering approaches, thus the compact source at the edge is still part of the catalogue. Figure 30 shows output images post the application of the filtering process. In these images, retained detections are shown with blue solid-line rectangles while removed detections are shown with black dashed-line rectangle. The grey dashed-line circle represents the definition of the edge; detections within this circle of radius 60 pixels are retained, for instance, in Figure 30a the two detections of compact sources are retained, while the detection towards the lower right corner is removed from the catalogue. Figure 30b shows a detection of a complicated source extending well beyond the circle, however because the central position of this source is within the circle, this detection is correctly retained. Retained detections after the filtering process are shown in Table 3, where the number of predictions (entries) in the D1 and the D4 output catalogue after the application of each filtering process is shown. The first row shows the total number of detections from the output catalogue. The second row represents the total number of detections after removing the overlapping detections of the same source. Thus, for the D1 catalogue 489 detections were rejected by this algorithm, while for D4 catalogue 922 detections were rejected. The NMS algorithm has in principle a known limitation of perceiving nearby detections arising from different objects as the same object, and as such it suppresses them. However, by overlaying detections on the full mosaic as DS9 region files, we concluded that in our case, presumably because of the moderate areal density of radio sources, this limitation of the algorithm did not adversely affect the results in a significant manner. As a result, we suppressed overlapping detections by their central positions and the predicted class and probabilities.

Our catalogues required another filtering process to remove the detections at the edge of the cutout as discussed in Section 2.7. The third row in Table 3 shows the output number of detection after removing detections at the edges. This process removed 2,131 and 3,929 detections were removed from the D1 and the D4 respectively, by the edge filtering algorithm, post the application of the other filtering process based on positions and the predicted classes and probabilities of the sources. The two filtering algorithms removed about 3,000 and 5,000 detections in the D1 and the D4 catalogues, respectively. As a result, the output D4 catalogue has about 200 more number of detections as PyBDSF detections (6,400), whereas the output D1 catalogue has over 1,200 less number of detections than PyBDSF. However, visualizing the source positions from PyBDSF, it was realized that PyBDSF resulted in many discrete source positions (i.e., detected more disconnected components) for some of the extended sources. The two filtering algorithms therefore achieved the goal of removing duplicate detections while "unique" individual classifications were recovered. As such, it would be expected for CLARAN to have a smaller number of detections as compared to PyBDSF due to the fact that CLARAN is better at detecting extended radio sources than PyBDSF. However, while visualizing the detections after the filtering processes, it appeared that a smaller number of the sources still had more than one bounding box overlaid on them (see Figure 41 in Section 4.2). Therefore, the filtering algorithm we implemented will have to be improved to efficiently recover unique individual detections.

3.4 Classification

For the purpose of classification, we rely on CLARAN's results using its pretrained model based on the Radio Galaxy Zoo dataset and applying it to our dataset. For each input image, a catalogue of all individual predictions was produced. The catalogue lists some of the properties from the original GMRT

Table 3: The output number of sources recovered from the D4 output catalogue after the application of two filtering processes, removing overlapping detections and the detections at the edge of each cutout.

	Number of sources	
Method	**D1**	**D4**
CLARAN Original Detections	7,849	11,444
After Removing Overlapping Detections	7,360	10,523
After Removing Edge Detections	5,229	6,594

catalogue together with CLARAN predictions, including the following columns: (1) `Source_ID`, from the GMRT PyBDSF catalogue, represented by a numerical integer value; (2) `Class`; (3) `Scores`; (4, 5, 6 and 7) as `x1`, `y1`, `x2` and `y2` – bounding box coordinates; (8 and 9) as `RC_RA` and `RC_Dec` – Radio Centre (RC) coordinates of the given bounding box; (10 and 11) `PyBDSF_RA` and `PyBDSF_Dec`, J2000 right ascension and declination from the GMRT PyBDSF catalogue, both in decimal degrees. The RC pairs of coordinates are a result of CLARAN's ability to locate sources in a given cutout, as such this ability was used to turn CLARAN into a source detection algorithm whereby source positions of the detected sources are computed on the fly, parallel to the detection and classification processes. An example of the output catalogue for the D1 image dataset is shown in Table 4, where entries were sorted according to the Source_ID column. If CLARAN detected multiple sources within a given cutout, the resulting multiple classifications will be assigned the same Source_ID. This is shown in Table 4.

3.4.1 Source Classes

We will now examine the overall CLARAN results in terms of classification. Figure 31 displays histograms for the overall distribution of predicted classes for the D1 and D4 dataset after removing duplicate detections. Blue and red bars represent the D1 and the D4 datasets, respectively. Similar to many other observations, CLARAN predicted that the area covered by the GMRT observations used in this study is dominated by compact radio sources (1C_1P). However, when using the D1 dataset as input a much larger (and unrealistic) number of 3C_3P sources are detected than when using the D4 dataset, For instance, when using the D1 dataset images, about 19% of sources are classified as 3C_3P sources, whereas for the D4 dataset, only about 3% are classified as such. A cursory visual inspection revealed that most 3C_3P candidates returned by the D4 dataset are actually spurious, as one might expect, the 19% from the D1 dataset are even more spurious. Visually inspection of a subset of these spurious detections revealed that they were just chance alignment due to faint sources being close to one another

Table 4: An example output catalogue, showing the first 20 entries with their corresponding properties consisting of 11 columns. The table was sorted by the Source_ID column. The first and the last two columns are from the original GMRT source catalogue and the rest of the columns are from CLARAN's predictions. More details in Section 3.6

Source_ID	Class	Scores	x1	y1	x2	y2	RC_RA	RC_Dec	PyBDSF_RA	PyBDSF_Dec
6	1C_1P	0.9960	246.4876	54.1299	246.4806	54.1347	246.4843	54.1320	246.484	54.1319
7	1C_1P	0.9946	246.5356	55.0467	246.5243	55.0541	246.5304	55.0503	246.5303	55.0500
11	3C_3P	0.9563	246.4544	54.2176	246.4310	54.2332	246.4440	54.2277	246.4473	54.2295
12	1C_1P	0.9832	246.4388	54.1458	246.4296	54.1519	246.4342	54.1486	246.4341	54.1485
12	1C_1P	0.8591	246.4287	54.1586	246.4236	54.1621	246.4262	54.1600	246.4341	54.1485
12	2C_2P	0.9435	246.4444	54.1458	246.418	54.1626	246.4331	54.1502	246.4341	54.1485
13	1C_1P	0.9906	246.4371	54.1237	246.4283	54.1298	246.4325	54.1265	246.4246	54.1377
13	1C_1P	0.9306	246.4271	54.1364	246.4223	54.1396	246.4246	54.1378	246.4246	54.1377
13	2C_2P	0.9490	246.4424	54.1237	246.4165	54.1404	246.4315	54.1281	246.4246	54.1377
14	1C_1P	0.9926	246.5122	55.2448	246.5070	55.2484	246.5096	55.2466	246.5096	55.2464
16	1C_1P	0.9936	246.4765	55.0841	246.4701	55.0884	246.4734	55.0863	246.4732	55.0861
18	1C_1P	0.9507	246.3973	54.1742	246.3885	54.1802	246.3931	54.1770	246.3931	54.1774
18	1C_1P	0.9355	246.3728	54.1688	246.3674	54.1725	246.3708	54.1702	246.3931	54.1774
19	3C_3P	0.8363	246.5151	55.5547	246.4962	55.5666	246.5060	55.5581	246.5052	55.5577
20	1C_1P	0.9843	246.4789	55.1138	246.472	55.1182	246.4758	55.1161	246.4649	55.1012
21	1C_2P	0.9957	246.4356	54.7305	246.4218	54.7399	246.4283	54.7347	246.4281	54.7351
22	1C_1P	0.9941	246.4244	54.5836	246.4176	54.5880	246.4209	54.5856	246.4211	54.5856
28	3C_3P	0.9710	246.5356	55.8199	246.5011	55.8422	246.5153	55.8309	246.5110	55.8286
30	1C_1P	0.9918	246.4264	55.1384	246.4180	55.1436	246.4223	55.1408	246.4466	55.1414

with no clear IR counterpart(s). Nonetheless, it is evident that the addition of multi-wavelength data dramatically improves the performance of CLARAN, as shown by the figure, and the D4 dataset yields substantially better results than the D1 dataset.

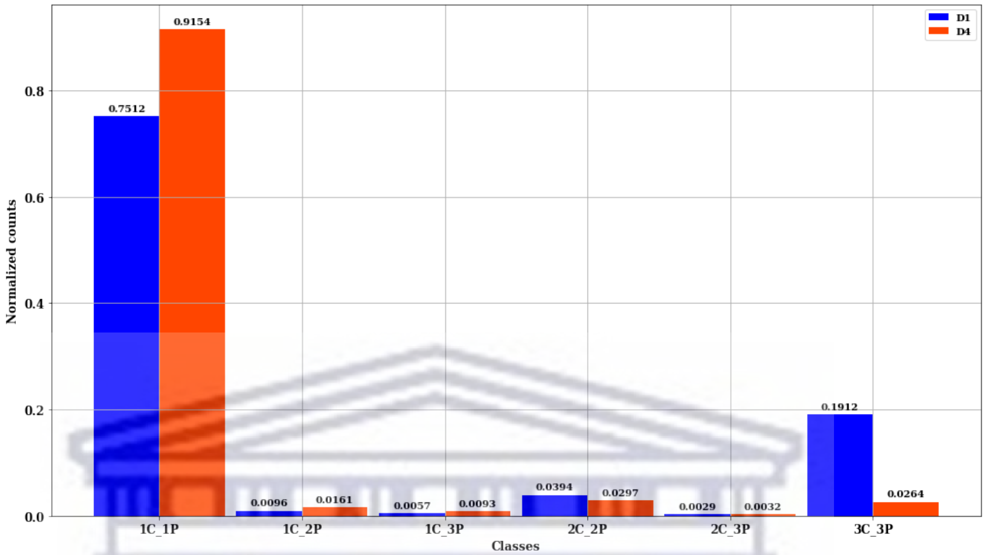
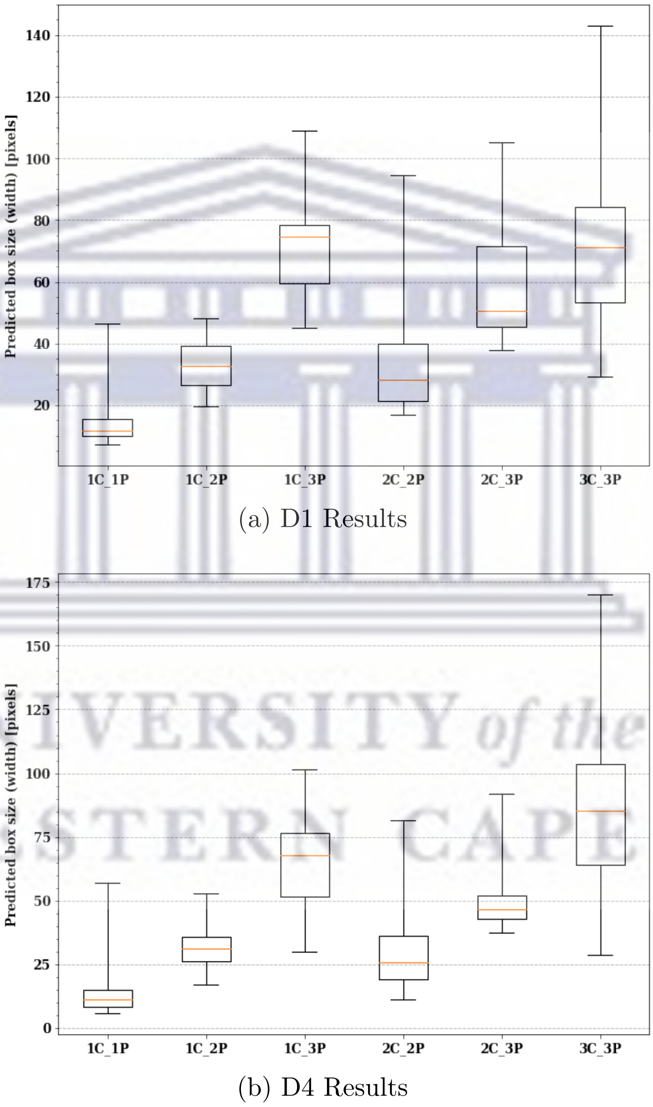


Figure 31: Bar graph showing the distribution of classes predicted by CLARAN from GMRT observations of the EN1 field. Blue and orange-red bars represent D1 and D4 images respectively. The numerical value on top represent the height of the bar.

Furthermore, a box plot was used to visualize the distribution of the bounding boxes and the probability scores. Box plots are used to show the distribution of the given data by using five statistical measures, minimum and maximum, median, first, and third interquartile. A rectangular box is drawn using the first and third interquartile, and a line inside the rectangle marks the median. The top and bottom horizontal lines on either side box represent the lower and the upper boundaries of the box plot. The lower boundary is defined as the difference between the first interquartile and 1.5 times the box size (the difference between the first and the third interquartiles), while the upper boundary is a sum of the third quartile and 1.5 times the box size. The vertical lines connecting these horizontal lines to the box are known as whiskers. Another point is the outliers; they are defined as data points that fall outside the lower and the upper boundaries of box plot.

3.4.2 Bounding Boxes

Figure 32 shows the distribution of bounding box sizes for different source classes. Note that all bounding boxes were predicted to have size < 181 pixels (3 arcmin) as shown in Figure 32. This figure shows the distribution of the bounding box sizes per class, in the D1 and the D4 output catalogues, shown by Figure 32a and 32b, respectively. In these two figures, it is clear that the predicted bounding box sizes in both datasets give a similar distribution per class. Moreover, for class 1C_1P and 1C_2P, the distributions are tight around their respective medians, suggesting fewer variations in the size of the predicted bounding boxes as expected because these are compact sources. For other classes, the sizes vary, as it will BE shown in Section 3.5, where we visually inspect output examples of various classes, shapes and sizes.



(a) D1 Results

(b) D4 Results

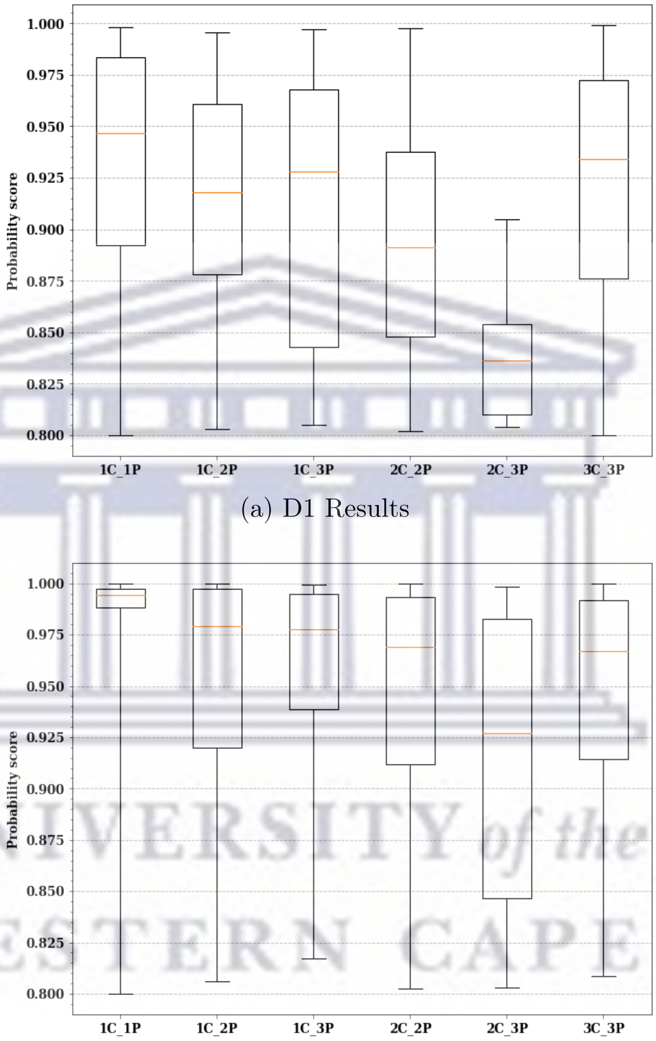
Figure 32: The distribution of the size of the bounding boxes predicted by CLARAN per class for (a) the D1 and (b) the D4 datasets. The median size is indicated by the orange horizontal line in the box and the edges of the box represents the first and the third quartile size.

3.4.3 Probability Scores

Each classification by CLARAN has an associated probability score, which estimates the probability (P-value) that the detected source belongs to the respective morphology class. Figure 33 shows the distributions of the probability score per class. Note that the configuration threshold of CLARAN was previously set to 80%. Thus detections that had a probability score less than this threshold were discarded by the classifier. The latter is also evident in Figure 33. This figure shows the distribution of the probability scores across different classes. Figure 33a and 33b represent the distribution for the D1 and the D4 datasets, respectively. It is clear that probability scores are spread over a broader range for the D1 than the D4 dataset for all the classes, except 2C_3P class. Moreover, despite the widespread distributions for both the datasets, note that the D4 dataset returns a higher probability score on average. This is determined by the span of the box, which is between 85.0%-99.5% for all the classes for the D4 dataset, unlike the D1 dataset where the span is between 81%-98%. Therefore, as also shown in Wu et al. (2019) – CLARAN performs better when used on the D4 dataset. Once again, using the D4 dataset (i.e multi-wavelength images) results in a substantial advantage in terms of performance.

3.5 Validation via Visual Inspection

Techniques such as the ones we developed in this thesis face a challenge when evaluating their performance. While the PyBDSF catalogue we refer to in our comparisons is a good example of a radio source catalogue which can be used for science exploitation (albeit with a few caveats), in most cases for observational datasets there is no “ground truth” we can use to compare our results against. While this is not completely satisfactory from a theoretical point of view, since we are mostly interested in reliably applying these techniques to real observational datasets, we decided to proceed in this manner rather than applying our algorithms to synthetic datasets which are likely to be somewhat unrealistic. As a first indication of the goodness of our results, and to assess the goodness of results obtained with D1 images against those obtained with D4 images, we decided to resort to the visual inspection of a subset of our sample. The subset was selected as all sources which were classified as “extended” (i.e. not 1C_1P) in both the D1 and the D4 output catalogue, for this we used the full catalogues (output catalogues before applying filtering algorithms). We have thus inspected and compared results obtained from the D1 and D4 images so as to compare the relative effectiveness of the two approaches. The resulting subset contains 478 sources, and Table 5 lists the number of predictions per class. However, all the



(a) D1 Results

(b) D4 Results

Figure 33: The distribution of the predicted probability score by CLARAN per class for (a) the D1 and the (b) the D4 datasets. The first and third quartile sizes are represented by the edges of the box and the median is indicated by the orange horizontal line in the box.

output files (catalogues and images) in this section and sections to follow, are made publicly available on our google drive directory `https://drive.google.com/drive/folders/116KTrOhfNCtOl-WdliNAnZ_zBxCEFLxd?usp=sharing`.

Table 5: Distribution of subset of data used for evaluation. Each row shows the predicted classes as well as the number of predictions for that class per dataset.

Class	**D1**	**D4**
1C_2P	55	77
1C_3P	42	49
2C_2P	87	160
2C_3P	12	7
3C_3P	282	185
Total	478	478

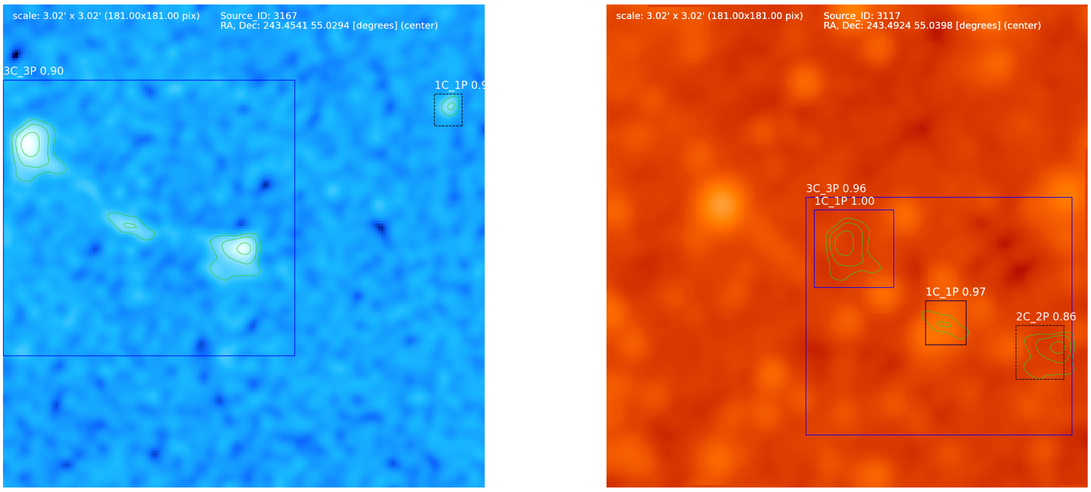
In the following we show some examples from the different classes of "extended" (i.e. not classified as 1C_1P) sources. As noted in the previous section, one radio source may appear in multiple cutouts and there may be multiple detections within one cutout, but here we focus on predictions for sources at the centre of the given cutout. It is also important to note that all images shown in the following sections were obtained following the filtering process. Thus removed detections are going to be shown with black dashed-line rectangles.

3.5.1 Source Characterization

3.5.1.1 Detection and Classification

Figure 34 displays output images from CLARAN that show the same region of the sky generated from different positions, Figure 34b centred at the northern component and Figure 34a centred at the southern component. Both the images show three disconnected radio components, two outer components with a core. The two outer components do not seem to coincide with any clear bright IR sources, only the central core appears to have a clear IR source associated with it. This is also confirmed by overlaying AllWISE catalogue source positions. The two outer components appear to be edge-brightened, the northern component being the brightest of the two, which is typical for a FR-II sources - two edge brightened components (lobes) with a central core. CLARAN's results are in agreement with this, whereby it regarded the three discrete components as a single source belonging to a 3C_3P class in both the images with high probability scores, 90% in Figure 34b and 96% in Figure 34a. CLARAN further differentiates between two smaller components as shown in Figure 34b.

Visually inspecting CLARAN's results per class, it is clear that CLARAN is efficient when dealing with cutouts that have few extended sources and is


scale: 3.02' x 3.02' (181.00x181.00 pix)
Source_ID: 3167
RA, Dec: 243.4541 55.0294 [degrees] (center)
3C_3P 0.90
1C_1P 0.9
scale: 3.02' x 3.02' (181.00x181.00 pix)
Source_ID: 3117
RA, Dec: 243.4924 55.0398 [degrees] (center)
3C_3P 0.96
1C_1P 1.00
1C_1P 0.97
2C_2P 0.86

(a) D1 Output Image

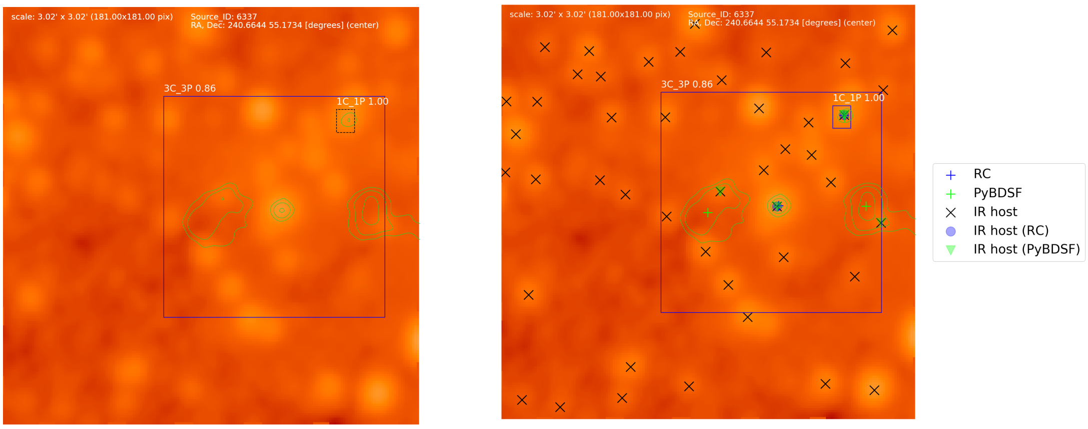
(b) D4 Output Image

Figure 34: Cutouts showing the same region on the sky, centred at different positions. Images (a) and (b) are centred at a northern and southern components of ID 3167 and 3117, respectively. CLARAN predicted this source to belong to class 3C_3P in both images.

challenged by cutouts with multiple sources, some of which are multi-components. Also, CLARAN was challenged when dealing with IR image crowded with sources (e.g., near the Galactic Plane, or near some local stellar cluster). This challenge of crowded fields is also one of the reasons why CLARAN predicted that most extended radio sources belong to 3C_3P class as shown in Figure 31, which was also observed when visually inspecting the images. Although going through the images, it was realized that CLARAN classified most of these sources more than once because they showed up as separate entries in the GMRT catalogue. For instance, image cutouts are generated at every position of the radio source from the GMRT catalogue as a result of a source with three disconnected radio components – e.g., Figure 34b, the same source is going to be classified three times since it will appear in three different cutouts, generated from the position of each source. Visual inspection also revealed that other reasons for questionable 3C_3P class predictions is the fact that CLARAN was confused by unresolved radio components that are very close to one another and (ii) CLARAN was challenged by complex bright and/or extended diffuse sources, and sources at the edges of the field of the GMRT observations.

3.5.1.2 Identification

As shown previously, as long as a radio source is detected by CLARAN, the pipeline will provide a corresponding radio source position (RC). The pipeline will then identify the most likely IR host galaxy of a given radio source as the


scale: 3.02' x 3.02' (181.00x181.00 pix)
Source_ID: 6337
RA, Dec: 240.6644 55.1734 [degrees] (center)
3C_3P 0.86
1C_1P 1.00
scale: 3.02' x 3.02' (181.00x181.00 pix)
Source_ID: 6337
RA, Dec: 240.6644 55.1734 [degrees] (center)
3C_3P 0.86
1C_1P 1.00
RC
PyBDSF
IR host
IR host (RC)
IR host (PyBDSF)

(a) D1 Output Image

(b) D4 Output Image

Figure 35: Output examples from (a) CLARAN and (b) the source characterization pipeline for radio source ID: 6337. Black crosses represent catalogued positions of the IR sources in the cutout, lime crosses represents the source positions of the cutout from GMRT catalogue, while blue crosses represents the radio centre of the bounding box. The blue shaded and the green shaded triangle indicates the estimated positions of the IR host as determined from the RC and PyBDSF central positions, respectively

catalogued IR source closest to centre of the identified radio source, as long as it falls within the bounding box associated with the radio source. The importance of using the RC position as the central position of the source becomes apparent when the source is extended and thus has multiple components and/or peaks. In such a case, two different hosts might be identified via the RC and PyBDSF source positions. Alger et al. (2018) tackled the challenges of finding the host galaxy by cross-identification. They also trained their CNN using RGZ data, given a 2 arcmin x 2 arcmin input centred on radio component that is overlaid on IR image. Furthermore, they apply a sliding window that uses a Gaussian kernel to weigh point sources on the IR image that are within 1 arcmin of the given radio component. Their algorithm works under the assumption that each cutout represents a single extended radio source. Thus, the algorithm breaks when having multiple radio sources within the cutout. However, CLARAN offers solutions for that, for instance, detecting and classifying radio components that are less than 2 arcmins from one another. However, our pipeline performs better for such, where the RC position would determine the host galaxy. In the following examples, the source position from the GMRT catalogue, denoted as PyBDSF, is represented by lime crosses while blue crosses indicate the RC position. For each of the two central positions, the likely IR host galaxy is overlaid as a shaded shape, filled with the same color as the corresponding central position used to determine the host position. Thus, lime shaded-triangles represent the IR hosts as determined by the source positions from PyBDSF and blue shaded-circles represent the positions of the hosts as determined by the RC positions. Note that examples to follow are presented in no particular order in terms of number of components and peaks. Furthermore, the size of the figure was increased (not altering the original output image) to indicate the positions as clearly as possible. From the previous sections, it is clear that CLARAN performs better when using the D4 dataset. Therefore, we only consider the D4 dataset for this discussion.

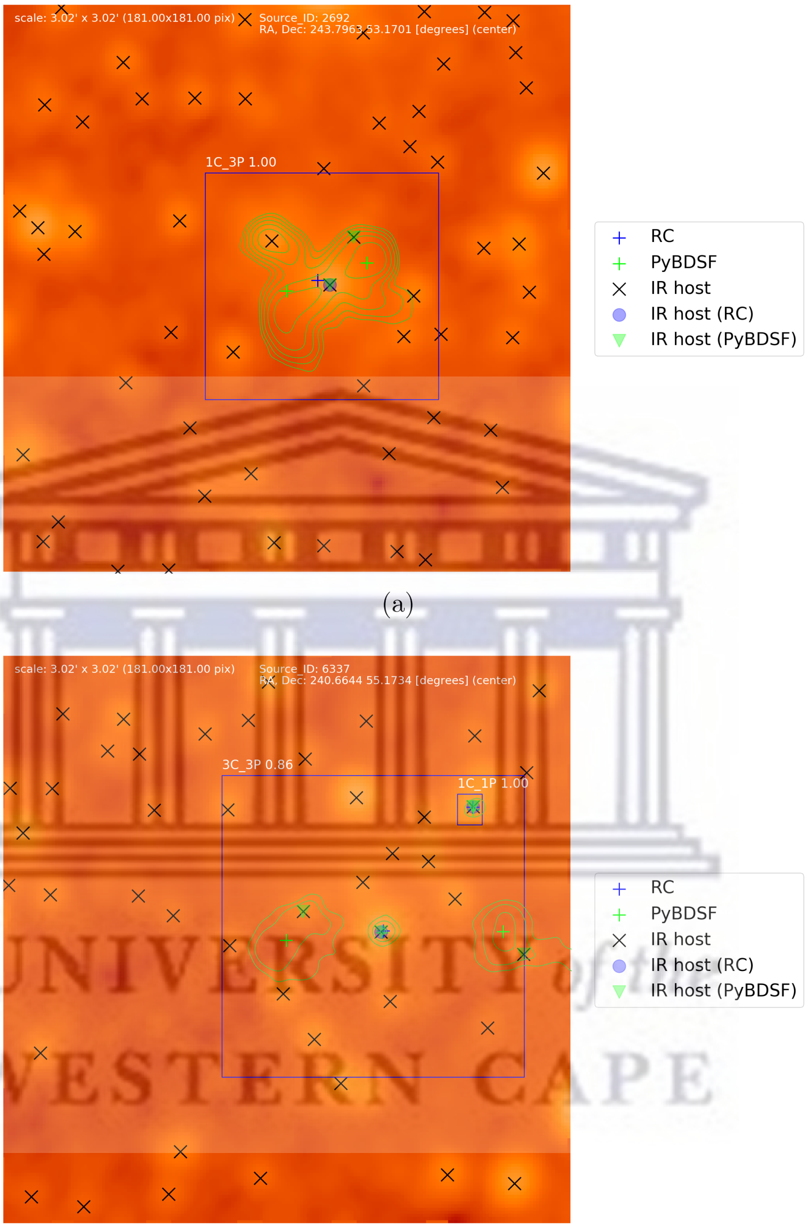
Another elongated source that appears to be an FR-II is shown in Figure 35a. In this case, it is clear that the radio component on either side of the central source, are components of the same source – lobes showing hotspots at the edges. In this case the pipeline produced an accurate source position than PyBDSF, that is right at the centre of this source and thus an estimate IR host position is also correct. As indicated, the source position from the PyBDSF is positioned at the lobe, also note that it detected three disconnected components for this source (to be discussed in the following chapter). It is clear from the examples provided that the source characterization algorithm is advantageous based on the fact that CLARAN is reliable to distinguish between different multiple radio sources. Therefore the detection algorithm adapted to CLARAN is very effective as shown. It is able to locate different host positions in case of extended radio sources. This is of importance in order to distinguish between star-forming galaxies and lobes of FR-II sources that are usually unresolved and may appear as two different sources in the sky. Moreover, the adapted detection algorithm can efficiently work and estimate host positions even if there is no known radio source position in that cutout. Despite the fact that there might be chance alignments between IR and radio sources, whereby the density of IR sources is such that some will land on the radio source by chance. This is a well-known challenge for source identification algorithms, and may require astronomer's intervention. Nevertheless, we can safely say, using CLARAN with the identification algorithm, multiple radio components in a single cutout can be reliably detected, classified, and cross-identified with their host galaxies. Thus, making this a very robust pipeline to produce a catalogue of properties of the detected radio sources from a given cutout, in this case, 3 arcmin × 3 arcmin.

The effectiveness of the source characterization pipeline as compared PyBDSF is illustrated by Figure 36. The central position of the source from our pipeline is shown as blue crosses (RC) and the detected source positions for this cutout from the GMRT catalogue are shown as lime crosses (PyBDSF). Note that we showing all the predictions from CLARAN for this cutouts and also the source positions from PyBDSF. These two positions from each algorithms were further used to estimate the position(s) of the IR host the galaxy. Thus, the blue shaded-circle represents the corresponding host of the RC source position, while the lime shaded-triangle represents the IR host position by the PyBDSF source position. In Figure 36a shows an example of a complex radio source, it is clear that PyBDSF detected two sources, thus giving two different positions for this source. However, our characterization pipeline adapted to CLARAN gives an accurate central position of this source. Another examples is shown in Figure 36b, where an elongated radio source with multiple components is presented, PyBDSF detected three components for this sources, one at the center and the other two on either side of the central component. For this case, the source positions from PyBDSF will result in three different host galaxy positions, however CLARAN detected a single source with multiple components associated with it. Thus, CLARAN results in a reliable central position of the source. These two cases are proof that despite the limitations faced by the pipeline, it produces better radio source and IR host galaxy positions as compared to the PyBDSF algorithm.

3.5.2 Performance Evaluation

The evaluation of a ML algorithm's performance is generally done against a test dataset for which the labels are already known. In our study this is not the case, since we have no predefined labels, and as such it is challenging to directly use popular evaluation metrics to assess the performance of CLARAN. However, quantifying the performance of an algorithm is very important. as discussed in Section 2.7.3.4, we used completeness (recall) and reliability (precision). For this task we defined a true positive as a correct classification of a complex source, a false positive as a point source classified as a complex source and a false negative as a missed complex source. To do this, we used a subset of our dataset containing 1,000 D4 images, visually inspected each one of them and noted down the outcome.

We then evaluated CLARAN's performance by counting true positives, false positives and false negatives as previously mentioned. In so doing we counted 78 true positives, 7 false positives and 22 false negatives. This results in a completeness (recall) of 78% and a reliability (precision) of 92%. While the completeness/recall is not particularly high, the rather high reliability/precision, including the fact that CLARAN performed well both on real data (RGZ DR1, similar to our data) and simulated data (SKA DC1), makes CLARAN well-suited at reliably identifying samples of extended/complex radio galaxy candidates e.g. in future large radio surveys.


scale: 3.02' x 3.02' (181.00x181.00 pix)
Source_ID: 2692
RA, Dec: 243.7963 53.1701 [degrees] (center)
1C_3P 1.00
RC
PyBDSF
IR host
IR host (RC)
IR host (PyBDSF)
(a)
scale: 3.02' x 3.02' (181.00x181.00 pix)
Source_ID: 6337
RA, Dec: 240.6644 55.1734 [degrees] (center)
3C_3P 0.86
1C_1P 1.00
RC
PyBDSF
IR host
IR host (RC)
IR host (PyBDSF)

(b)

Figure 36: Output examples showing the capabilities of the source characterization pipeline. Similar to previous examples, the black crosses, blue and green crosses indicate the positions of the IR sources, the RC and the PyBDSF source positions, respectively. The respective IR hosts for the RC and the PyBDSF source positions are represented by the blue shaded-circle and the green-shaded triangle.

3.6 Final Catalogue

For all input images, individual source characterizations were grouped into a final catalogue. The final catalogue of all predictions contain 20 columns, where the following 9 columns were added: (12) `PyBDSF_host`, IR host ID predicted using the PyBDSF source position associated with the cutout; (13 and 14), `PyBDSF_host_RA` and `PyBDSF_host_Dec`, coordinates of the `PyBDSF_host`; (15) `PyBDSF_separation`, the distance/separation (in arcsec) between the PyBDSF source position and the IR Host position; (16) `RC_host`, IR host ID predicted using the Radio Centre (RC) from our detection algorithm; (17 and 18) `RC_host_RA` and `RC_host_Dec`, coordinates of the `RC_host`; (19) `RC_separation`, the distance/separation (in arcsec) between the RC position and the IR host position; (20) `file_path`, the file path linked to the output file from CLARAN. Note that the file paths point to files on the ilifu cloud facility. However, all the catalogues are available on our google drive directory, in a sub-directories named "D1" and "D4" for each of the final catalogue from the D1 and the D4 datasets.

For the purpose of reproducibility and evaluation, this google drive directory contains sub-directories for: output images from the final catalogues for each of the D1 and the D4 datasets ("D1" and "D4"), all the catalogues pre- and post-filtering processes ("catalogues"), output images from the source characterization pipeline ("SC_output") and a subset output images showing sources removed from the final catalogues ("Filtering examples").

Chapter 4

4 Discussion

In this study, the CLARAN code was adapted and developed to run on our GMRT dataset. By computing performance metrics and visually inspecting the two output datasets – the radio only (D1) one and the fused (radio + infrared; D4) one, we found that transfer learning was successfully applied. While most recent studies are trained to work exclusively with radio data (Aniyan and Thorat, 2017; Alhassan et al., 2018; Lukic et al., 2018; Tang et al., 2019; Lukic et al., 2019) and focus on classifying sources into traditional classes of radio sources (i.e., compact, FR-I, FR-II and bent), we demonstrated that CLARAN can make the most of multi-wavelength data as well as classify more complicated shapes and sizes. We provided evidence that CLARAN is an accurate classifier, especially when using the composite D4 dataset. Furthermore, using the capability of CLARAN to locate and classify multiple sources in the given cutout, we implemented a source characterization pipeline. The pipeline performs three tasks, detection, classification and identification. For detection, the pipeline produces the central position of the source (source position) – by computing the centroid of the data within the bounding box, and for identification, the source position is used to estimate the position of the IR host (by taking the closest IR host to the source position). In summary, we found that our source characterization pipeline provides distinct advantages with respect to conventional tools such as PyBDSF, especially for extended and complex radio sources. However, the source characterization pipeline based on CLARAN also shows some limitations in its current form. In this Chapter we review some of these limitations and show an output image from CLARAN when applied on a 15 arcmin cutout.

4.1 Limitations

As previously discussed in Section 2.3.2, CLARAN is limited by the angular size of the cutout and performs somewhat poorly when classifying multiple sources per cutout. Consequently, this limits the application of the source characterization pipeline in its current form. The limitations are discussed further in Section 4.1.1 and 4.1.2 below.

4.1.1 Angular size of the cutout

The angular size of the cutout determines the extent to which sources are seen by CLARAN. That is to say, sources extending beyond the 3 arcmin angular size will

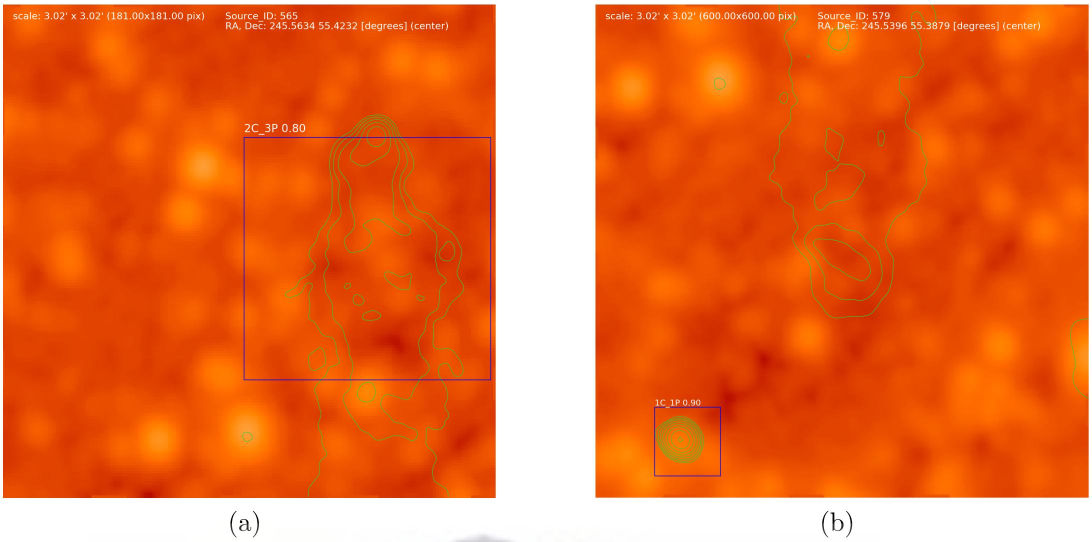
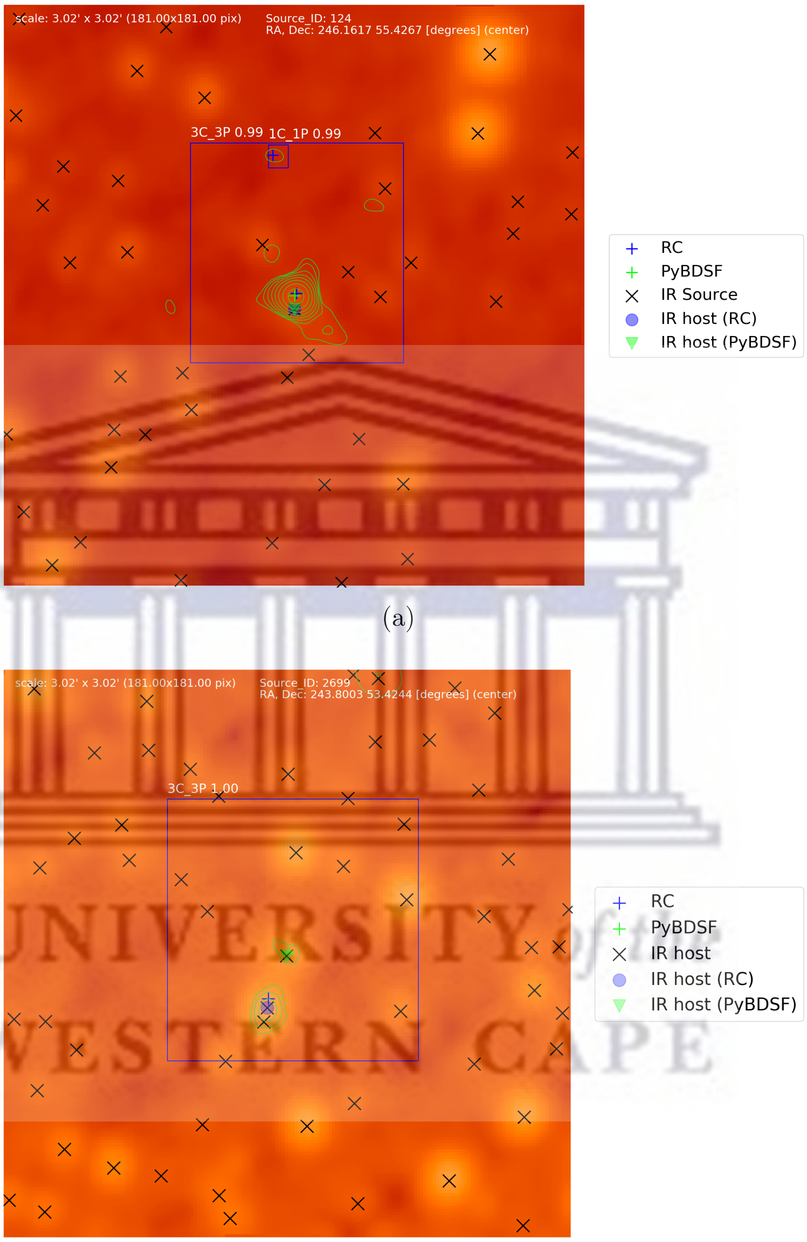


Figure 37: Images of a field size of 3-acrmin centred at source of ID (a) 565 and (b) 579. The two actually show a single source that extends well beyond the 3-arcmin field. CLARAN was able to detect the top half part shown in (a), however it 'missed' the bottom half part in (b) – the extended source in the upper center does no show any indication of a detection.

most likely be misclassified, as shown in Figure 37. The two images in this figure display the same source, that extends beyond the 3-arcmin field size. PyBDSF detected two separate sources, as a result two cutouts were generated and as such this source is cut into two halves, the top half part shown in Figure 37a, was detected by CLARAN, while it 'misses' the bottom half part shown in Figure 37b. As a result, CLARAN will predict less accurate boxes. This significantly affects the estimated RC position and thus the estimated position of the IR host galaxy. Also, the angular size determines the amount of background noise exposed to CLARAN. As previously discussed, some of the detections are significantly affected by this, and often it results in less accurate detection, classification and identification. Figure 38 show such examples, where CLARAN detected an extended radio sources with multiple peaks. As shown, some of the peaks are faint but close enough for CLARAN to regard them as a single source. As a result, it will detect spurious sources – creating false extended sources, resulting in a less accurate classification of the radio source. However, it appears that CLARAN may have been affected by the faint, diffuse emission around the peaked components, of which, when observed at this scale, may appear as noise. This will be shown in sections to follow, where a larger cutout was generated from the same source position as the cutout in Figure 38a. However, the RC position is likely to be accurate because it is found by computing the center-of-mass of the flux values within the predicted bounding box, thus most likely to be towards the position of the bright spot. Therefore, the identification is also likely to be correct, as it will estimate the IR source close to the RC position. At this scale, several elongated and very bright sources are missed as well by CLARAN as discussed previously (see Section 3.2). Also, it missed sources around clumpy regions, which are likely to be star-forming regions. Although, the latter may be due to the fact that CLARAN was not trained to classify such sources. Thus, it might be solved by training CLARAN to expect such sources.

4.1.2 Multiple sources per cutout

One of the assumptions made during training was one source per field. Although, from the previous sections, it is clear that CLARAN is capable of detecting and classifying most of the radio sources per cutout. However, in some cases CLARAN is confused by large sources that have got many components/peaks, whereby it will break these sources into smaller components, as an example is shown in Figure 34b. Figure 34b shows an output example where CLARAN predicted 4 boxes, one large box enclosing all the components of the elongated radio source, three other focus on smaller components (one box is filtered out by the filtering algorithm), as such, the source characterization pipeline estimated separate RC positions for each box. However, looking at Figure 34b this clarifies that all radio contours belong to the same radio source located at the centre of the cutout. Also because of CLARAN's capabilities to detect and classify most of the radio sources per cutout, it is challenging to distinguish these multiple classifications as a single multi-component source or multiple classifications of single sources. As a result, it was very challenging to produce a 'unique' catalogue even with the most applied suppression algorithm. Consequently, this result in multiple plausible RC positions for multiple sources and in turn multiple plausible positions of the IR host galaxies. This problem can be mitigated by using IR images from unWISE (Schlafly et al., 2019), CATWISE2020 (Marocco et al., 2021) and Spitzer/IRAC (Fazio et al., 2004) which have got an improved mid-infrared angular resolution and depth with respect to WISE. An example of this improvement is shown in Figure 39 where Figure 39a and Figure 39b shows an output image from CLARAN using IR data from WISE and Spitzer/IRAC, respectively. When an improved resolution and depth IR image is used, CLARAN performs better at differentiating between the components, predicting that the two sources are separate in Figure 39b as compared to Figure 39a when a low resolution and depth image in which CLARAN got confused by the IR sources that appear washed-out and thus predicted the source as a single source with two components and peaks. Another possible solution for obtaining unique catalogue is to use to the the IR host IDs estimated by the RC positions, whereby detections are grouped by IDs.

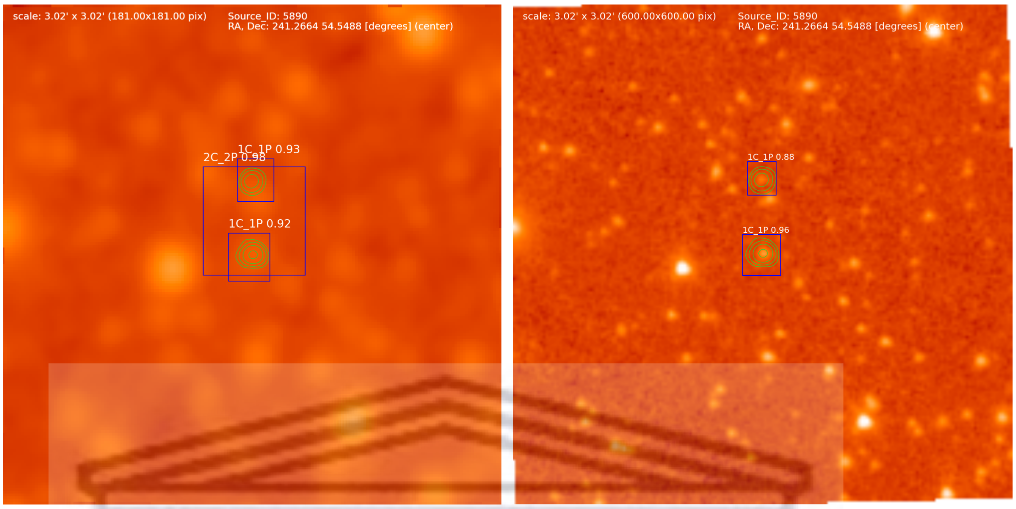


(a)

(b)

Figure 38: Two output examples of cutouts that resulted in ’false’ detections from CLARAN. Cutout (a) is centred at a radio component of ID: 124, whereas (b) is centred at a component of ID: 2699. The source positions by RC and PyBDSF are indicated by the blue and green crosses, respectively.

Therefore, detections of the same source will most likely have the same IR host ID. However, this will only work if there is an IR source within the bounding box.


scale: 3.02' x 3.02' (181.00x181.00 pix)
Source_ID: 5890
RA, Dec: 241.2664 54.5488 [degrees] (center)
1C_1P 0.93
2C_2P 0.98
1C_1P 0.92
scale: 3.02' x 3.02' (600.00x600.00 pix)
Source_ID: 5890
RA, Dec: 241.2664 54.5488 [degrees] (center)
1C_1P 0.88
1C_1P 0.96

(a) AllWISE (b) Spitzer/IRAC

Figure 39: Output images from CLARAN showing the effects of using IR data with improved angular resolution and depth for the same region on the sky. Image (a) shows CLARAN's results when using IR data from WISE, whereas image (b) shows output result when Spitzer/IRAC data is used.

Another solution is to use to the the IR host IDs estimated by the RC positions, whereby detections are grouped by IDs. Therefore, detections of the same source will most likely have the same IR host ID. However, this will only work if there is an IR source within the bounding box.

The limiting cases presented in the two previous subsection are mainly due to the angular size of the cutout. Also, this affects the reliability of the results from CLARAN because for the same source, the classifications will be different for different angular sizes (see Section 4.2 below).

4.1.3 Bounding box size

In addition to limiting cases discussed above, that CLARAN faces, the source characterization algorithm also has limiting cases. The size of the bounding box limits the pipeline due to the fact that for larger bounding boxes, the pipeline is exposed to a large background signal. The more the background signal the higher the clipping threshold. When the threshold is high, often some of the source signal will be lost as well. As a result, the RC position is likely to be less accurate. Another limiting case is multiple IR sources within the predicted bounding boxes, for instance when a source is elongated with three components, two lobes and a core, in some cases the brightness of the lobes is not the same. As such, the RC position will be more towards the brightest component, consequently the estimated IR counterpart will be incorrect. Also, the pipeline only works when the IR host is within the predicted bounding box. Therefore, if the IR host galaxy is distant from its radio emission, the host galaxy estimated by the pipeline might be biased. However, this is also challenging to non-automated methods, as a result, requires efforts of scientists to accurately cross-identify.

4.2 Larger Image Cutout

Previously, it was noted that CLARAN can be used on even a larger cutout, 15 arcmin × 15 arcmin, in this subsection one test case is shown. Similar to Wu et al. (2019), a 15 arcmin × 15 arcmin cutout was generated from a given central position as in shown in Figure 40. As shown in this figure, CLARAN is able to be applied on a larger cutout. In this cutout, there are two radio sources, and both were detected and classified by CLARAN. The left most source is classed as a 1C_1P source with a probability score of 96%. This cutout is centred the same coordinates as the cutout in Figure 38a, where CLARAN classed the central source as 3C_3P. However, on a larger sale, the central source is classed as 1C_1P with a probability score of 83%. The latter is a result of the test scale, which determines the resampling scale of the image, for this example we used the same test scale as Wu et al. (2019). As a result, a radio component may be missed on one test scale but then detected on another. This may be solved by classifying each cutout in multiple scales, and then using a filtering algorithm to get the best classification. However, this is left for possible future work. Nevertheless, CLARAN's capabilities combined with the source characterization algorithm, makes this pipeline a promising characterization pipeline, but still mostly unproven to date.

In closing this discussion, Figure 41 shows all detections post the application of the filtering algorithms overlaid on the full mosaic from the GMRT observations. The rectangular blue patches represent the bounding boxes of the detected sources, while lime dots represent PyBDSF source positions. Clearly, on the current scale of the figure, it is very challenging to see these boxes/patches. Zooming in on this figure, it will be clear that CLARAN – although it missed some of the sources at the edges of the FOV, as well as extended, clumpy, and very bright structures — it detected and classified most of the radio sources across the FOV. PyBDSF detected point sources mostly, as expected. In general, both PyBDSF and CLARAN performed comparably well for the case of point sources. Moreover, PyBDSF struggled to model complex, resolved and extended sources – which results in no detection made or inaccurate source position. While CLARAN may

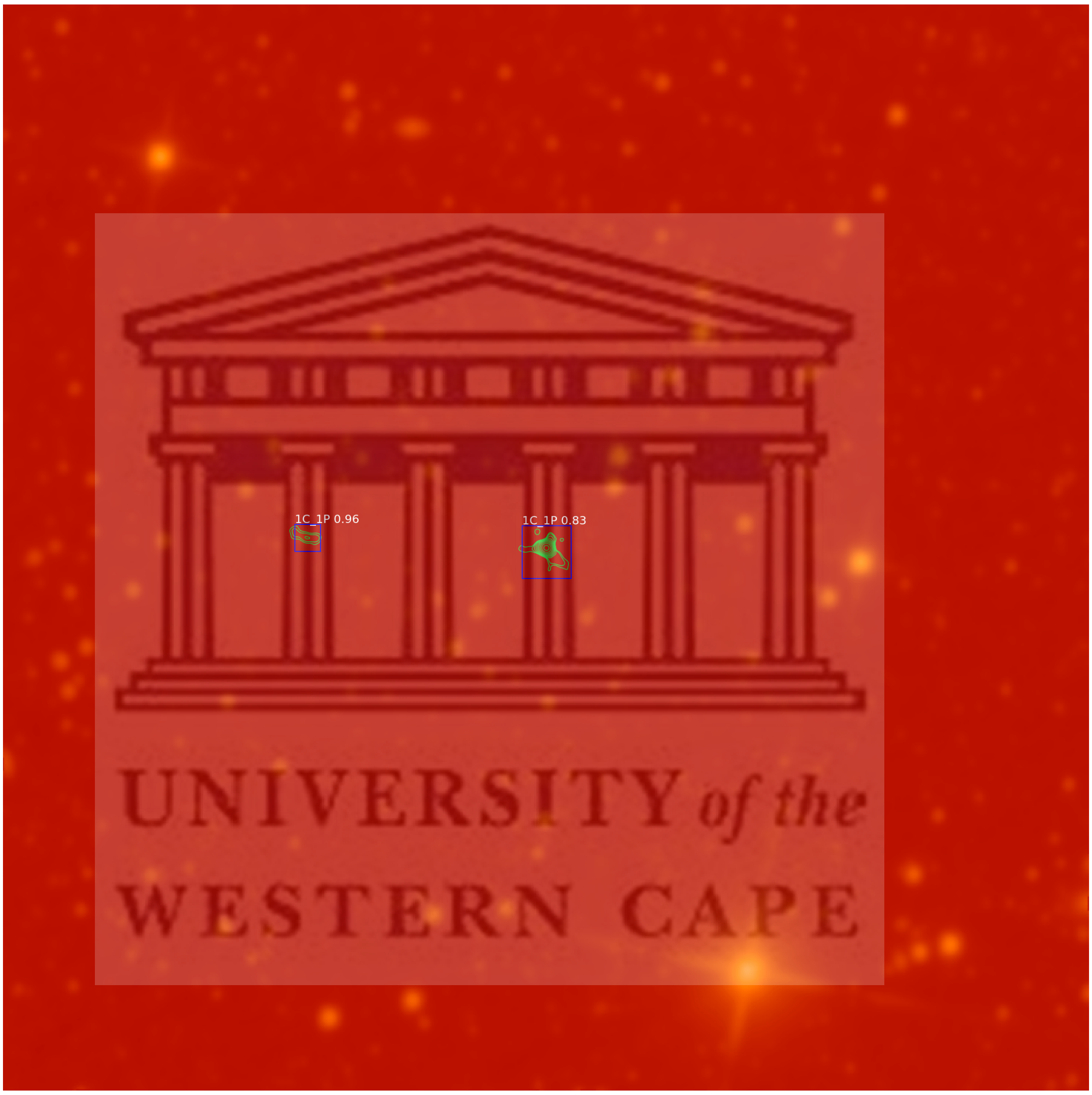


Figure 40: A 15 arcmin × 15 arcmin cutout centred at a source of ID: 124; same as in Figure 38. CLARAN located and classified two compact radio sources.

have missed a few complex, resolved and extended source, and detected a spurious extended sources, it is evident that it detected most of these resolved, complex and extended better than PyBDSF. Some of the bounding boxes overlap as a result of having duplicate predictions that could not be successfully removed by the two filtering approaches we applied. Nevertheless, the bounding boxes of those detections were further used in the source characterization pipeline to cross-identify their IR counterparts (provided the counterpart is within the predicted bounds). Another aim of this study was to apply an efficient source characterization algorithm that is fast enough to keep up with the data flow from the current- and the next-generation of radio sky surveys. The source characterization pipeline built on CLARAN is an end-to-end pipeline performs source characterization (detection, classification and identification) automatically and fast. To put this into context, we measured execution time taken by CLARAN to classify each cutout measured on the ilifu cloud facility. For both the D1 and the D4 datasets, CLARAN took about 3 seconds on average for each cutout. Moreover, for a handful of tests performed on the 15 arcmin cutouts, CLARAN took 10 seconds on average. However, it is important to note that Wu et al. (2019) run CLARAN with a GPU, as such it took approximately 200 milliseconds to locate and classify sources on each cutout. As a result, in future we will utilize the GPUs on the ilifu facility to try and accelerate the execution time. Furthermore, it is important to note that CLARAN locates and classifies radio sources in terms of components and peaks – not traditional classes. As shown in the previous chapter, CLARAN works well and it is a promising framework to develop robust data processing pipelines – such as the source characterization pipeline presented in this thesis, to expect the "unknown-unknowns" from the next-generation of radio sky surveys.

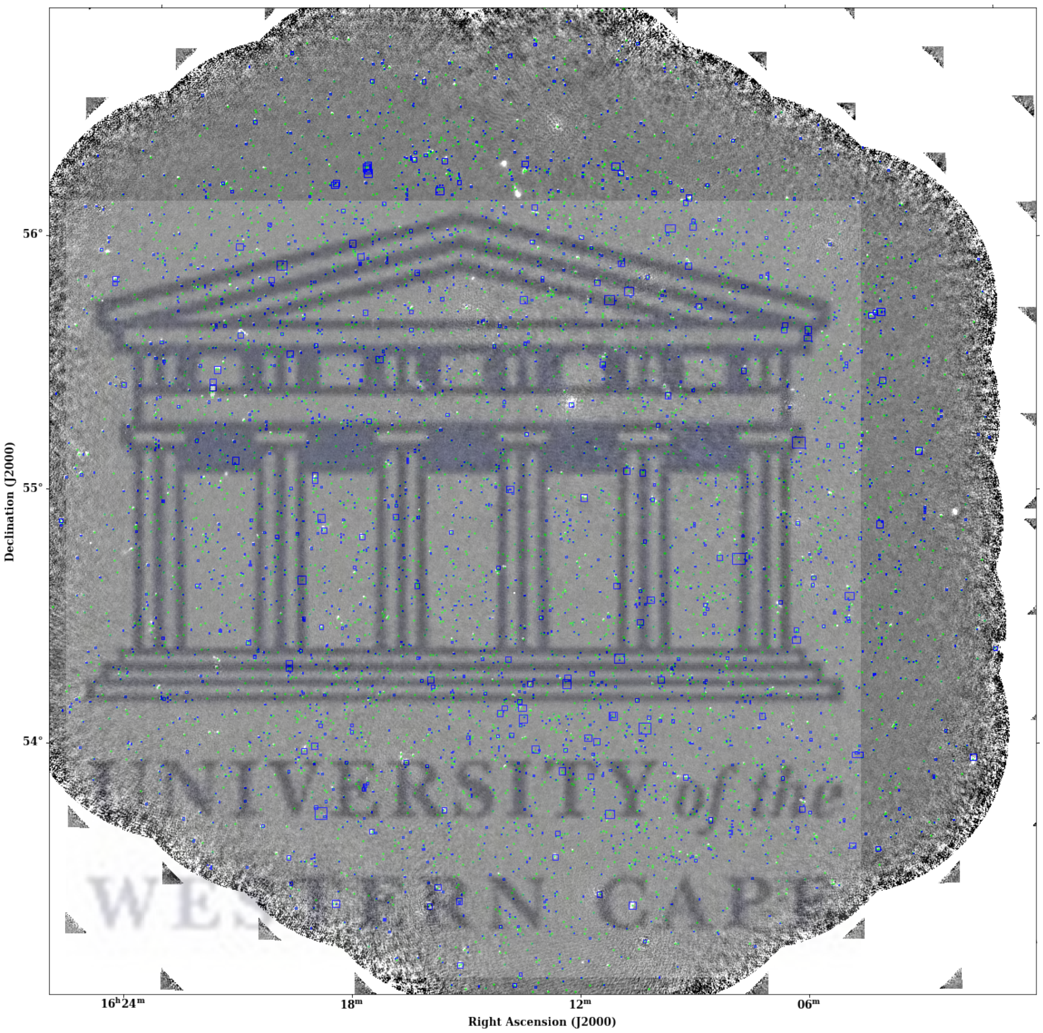


Figure 41: The final output image of 12.8 square degrees area in the EN1 field, overlaid with blue rectangular boxes that represent sources detected and classified by CLARAN and lime dots that represent PyBDSF source positions.

Chapter 5

5 Conclusion and Future Work

The next generation of radio sky surveys with Phase 1 of the Square Kilometre Array (Braun, 2015; Braun et al., 2019) is expected to generate a large volume of data, with data rates from the mid-frequency dishes of over 1 petabits per second and 10 petabits per second from the low-frequency phased-arrays. This poses substantial challenges to traditional methods of astronomical data processing. Therefore, tasks such as radio source characterization need to be automated to deal with the expected data flow in an effective manner.

The main objective of this study was to apply a pre-trained deep learning model as a solution to the challenge of efficient detection, classification and identification of radio sources. A pre-trained model known as CLARAN was the best candidate for this task. CLARAN was trained on image data from the FIRST and WISE surveys classified by citizen scientists via the Radio Galaxy Zoo project. It detects and classifies radio sources in a single image based on the number of connected components and emission peaks detected in a "fused" radio-infrared image.

In this study the pre-trained CLARAN model was applied to the classification of image data from GMRT observations of the ELAIS-N1 field and corresponding WISE images. Results produced by CLARAN using the radio-only (D1) images and the "fused" radio-infrared (D4) images were compared. The comparison showed that CLARAN is substantially more accurate when the D4 dataset is used, showing the power of multi-wavelength data for source classification. Furthermore, visual inspections indicate that CLARAN is very sensitive to the background noise (RMS). As a result, it is making inaccurate predictions when the noisier radio-only D1 dataset is used. Also, the process of generating cutouts starting from a pre-existing source list resulted in some of the radio sources appearing in more than one cutout, producing multiple classifications of those sources. Some of those multiple predictions were removed using suppression algorithm adapted from the NMS algorithm, which produces unique predictions by getting rid of some of the overlapping bounding boxes. Also, the predictions were further improved by removing detections at the edges of the cutouts. Moreover, we computed an estimate of CLARAN's performance using 1,000 D4 images by estimating reliability and completeness when it comes to detecting and classifying complex sources, and found that we achieved a completeness (recall) of 78% and a reliability (precision) of 92%. We also investigated how CLARAN is limited by the test scale and, thus, the size of the input image. Nevertheless, in most respects transfer learning was successfully applied, and as expected CLARAN performs better on the “fused” radio-infrared D4 dataset than on the radio-only D1 dataset.

Building on CLARAN, we devised a full radio source characterization pipeline to detect, classify and identify sources in an efficient manner. Comparing radio source detections obtained with our algorithm and with PyBDSF shows very similar results in the case of compact radio sources. However, PyBDSF is likely to predict more than one radio component for most of the complex (e.g. elongated or multi-component) sources, even if all the components are connected. Thus, our algorithm provides an advantage in such cases, providing estimates of the central position as well as the position of the IR host galaxy. In cases where an extended source is detected by CLARAN due to diffuse and faint radio emission around a bright source, this will affect the size of the predicted bounding box of the source and the IR host position. However, this effect may be negligible since the detection/identification process will produce a source position that it is more towards the brightest part of this source. CLARAN is fairly fast when tested on CPUs, taking on average 3 seconds to classify radio sources in a single cutout of 3 arcmin × 3 arcmin. Furthermore, CLARAN performed well both on real data (RGZ DR1, similar to our data) and simulated data (SKA DC1). With our additions and modifications, CLARAN results in a very powerful source detection, classification and identification algorithm compared to PyBDSF. We have thus successfully turned a pre-trained deep learning algorithm into an efficient source characterization pipeline.

In future work, we will run CLARAN using graphical processing units (GPUs) to try and accelerate the classification tasks. We will also address the problem of detecting sources without a pre-existing source list, by regularly “gridding” large fields into different cutouts and optimizing the cutout size for classification performance and speed. Large survey projects such as ASKAP’s EMU (Norris et al., 2011)and JVLA’s VLASS (Lacy et al., 2020) will greatly benefit from such work. EMU will be a 1.4 GHz wide-area survey with 10 arcsec resolution, while VLASS will be a 3 GHz wide-area survey with 2.5 arcsec resolution. EMU and VLASS will overlap over a large portion of the sky around the celestial equator and will thus allow multi-frequency and multi-resolution studies of radio sources. For this work, the unWISE maps and catalogues produced over the full sky by Schlafly et al. (2019) will provide improved angular resolution and depth. Over deeper and smaller areas, we will also make use of Spitzer/IRAC (Fazio et al., 2004) images which have got an improved mid-infrared angular resolution and depth with respect to unWISE. This will be particularly useful for MeerKAT's MIGHTEE survey (Jonas and MeerKAT Team, 2016; Jarvis et al., 2016), whose footprint is fully covered by Spitzer/IRAC observations. MIGHTEE will be strongly affected by confusion, which will pose a new challenge, and the Spitzer/IRAC higher-resolution data may therefore provide a substantial advantage.

The source characterization pipeline developed as part of this work yields promising results to tackle the challenges of source detection, classification and identification. In the era of SKA1, its unprecedented sensitivity will make such challenges even more important. Thus, applying transfer learning, developing and testing tools such as CLARAN on data from SKA precursors and pathfinders will be of great importance to prepare for the SKA.

References

Abraham, S., Aniyan, A. K., Kembhavi, A. K., Philip, N. S., and Vaghmare, K. (2018). Detection of bars in galaxies using a deep convolutional neural network. *MNRAS*, 477(1):894–903.

Alger, M. J., Banfield, J. K., Ong, C. S., Rudnick, L., Wong, O. I., Wolf, C., Andernach, H., Norris, R. P., and Shabala, S. S. (2018). Radio Galaxy Zoo: machine learning for radio source host galaxy cross-identification. *MNRAS*, 478(4):5547–5563.

Alhassan, W., Taylor, A. R., and Vaccari, M. (2018). The FIRST Classifier: compact and extended radio galaxy classification using deep Convolutional Neural Networks. *MNRAS*, 480(2):2085–2093.

Aniyan, A. K. and Thorat, K. (2017). Classifying Radio Galaxies with the Convolutional Neural Network. *ApJS*, 230(2):20.

Baldi, R. D., Capetti, A., and Giovannini, G. (2015). Pilot study of the radio-emitting AGN population: the emerging new class of FR 0 radio-galaxies. *A&A*, 576:A38.

Baldi, R. D., Capetti, A., and Giovannini, G. (2019). High-resolution VLA observations of FR0 radio galaxies: the properties and nature of compact radio sources. *MNRAS*, 482(2):2294–2304.

Ball, N. M. and Brunner, R. J. (2010). Data Mining and Machine Learning in Astronomy. *International Journal of Modern Physics D*, 19(7):1049–1106.

Banfield, J. K., Wong, O. I., Willett, K. W., Norris, R. P., Rudnick, L., Shabala, S. S., Simmons, B. D., Snyder, C., Garon, A., Seymour, N., Middelberg, E., Andernach, H., Lintott, C. J., Jacob, K., Kapińska, A. D., Mao, M. Y., Masters, K. L., Jarvis, M. J., Schawinski, K., Paget, E., Simpson, R., Klöckner, H. R., Bamford, S., Burchell, T., Chow, K. E., Cotter, G., Fortson, L., Heywood, I., Jones, T. W., Kaviraj, S., López-Sánchez, Á. R., Maksym, W. P., Polsterer, K., Borden, K., Hollow, R. P., and Whyte, L. (2015). Radio Galaxy Zoo: host galaxies and radio morphologies derived from visual inspection. *MNRAS*, 453(3):2326–2340.

Baron, D. (2019). Machine Learning in Astronomy: a practical overview. *arXiv e-prints*, page arXiv:1904.07248.

Becker, R. H., White, R. L., and Helfand, D. J. (1995). The FIRST Survey: Faint Images of the Radio Sky at Twenty Centimeters. *ApJ*, 450:559.

Beckmann, V. and Shrader, C. (2012). The AGN phenomenon: open issues. In *Proceedings of "An INTEGRAL view of the high-energy sky (the first 10 years)" - 9th INTEGRAL Workshop and celebration of the 10th anniversary of the launch (INTEGRAL 2012). 15-19 October 2012. Bibliotheque Nationale de France*, page 69.

Bonaldi, A. and Braun, R. (2018). Square Kilometre Array Science Data Challenge 1. *arXiv e-prints*, page arXiv:1811.10454.

Braun, R. (2015). The Square Kilometre Array: Current Status and Science Prospects. In *IAU General Assembly*, volume 29, page 2252814.

Braun, R., Bonaldi, A., Bourke, T., Keane, E., and Wagg, J. (2019). Anticipated Performance of the Square Kilometre Array – Phase 1 (SKA1). *arXiv e-prints*, page arXiv:1912.12699.

Burke, C. J., Aleo, P. D., Chen, Y.-C., Liu, X., Peterson, J. R., Sembroski, G. H., and Lin, J. Y.-Y. (2019). Deblending and classifying astronomical sources with Mask R-CNN deep learning. *MNRAS*, 490(3):3952–3965.

Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B., and Broderick, J. J. (1998). The NRAO VLA Sky Survey. *AJ*, 115(5):1693–1716.

Cutri, R. M., Wright, E. L., Conrow, T., Fowler, J. W., Eisenhardt, P. R. M., Grillmair, C., Kirkpatrick, J. D., Masci, F., McCallon, H. L., Wheelock, S. L., Fajardo-Acosta, S., Yan, L., Benford, D., Harbut, M., Jarrett, T., Lake, S., Leisawitz, D., Ressler, M. E., Stanford, S. A., Tsai, C. W., Liu, F., Helou, G., Mainzer, A., Gettings, D., Gonzalez, A., Hoffman, D., Marsh, K. A., Padgett, D., Skrutskie, M. F., Beck, R. P., Papin, M., and Wittman, M. (2013). Explanatory Supplement to the AllWISE Data Release Products. Explanatory Supplement to the AllWISE Data Release Products.

Dieleman, S., Willett, K. W., and Dambre, J. (2015a). Rotation-invariant convolutional neural networks for galaxy morphology prediction. *MNRAS*, 450(2):1441–1459.

Dieleman, S., Willett, K. W., and Dambre, J. (2015b). Rotation-invariant convolutional neural networks for galaxy morphology prediction. *MNRAS*, 450(2):1441–1459.

D'Isanto, A. and Polsterer, K. L. (2018). Photometric redshift estimation via deep learning. Generalized and pre-classification-less, image based, fully probabilistic redshifts. *A&A*, 609:A111.

Djorgovski, S. G., Mahabal, A., Drake, A., Graham, M., and Donalek, C. (2013). *Sky Surveys*, page 233. Springer.

Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., Kaviraj, S., Fischer, J. L., Abbott, T. M. C., Abdalla, F. B., Annis, J., Avila, S., Brooks, D., Buckley-Geer, E., Carnero Rosell, A., Carrasco Kind, M., Carretero, J., Cunha, C. E., D'Andrea, C. B., da Costa, L. N., Davis, C., De Vicente, J., Doel, P., Evrard, A. E., Fosalba, P., Frieman, J., García-Bellido, J., Gaztanaga, E., Gerdes, D. W., Gruen, D., Gruendl, R. A., Gschwend, J., Gutierrez, G., Hartley, W. G., Hollowood, D. L., Honscheid, K., Hoyle, B., James, D. J., Kuehn, K., Kuropatkin, N., Lahav, O., Maia, M. A. G., March, M., Melchior, P., Menanteau, F., Miquel, R., Nord, B., Plazas, A. A., Sanchez, E., Scarpine, V., Schindler, R., Schubnell, M., Smith, M., Smith, R. C., Soares-Santos, M., Sobreira, F., Suchyta, E., Swanson, M. E. C., Tarle, G., Thomas, D., Walker, A. R., and Zuntz, J. (2019). Transfer learning for galaxy morphology from one survey to another. *MNRAS*, 484(1):93–100.

Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., Tuccillo, D., and Fischer, J. L. (2018). Improving galaxy morphologies for SDSS with Deep Learning. *MNRAS*, 476(3):3661–3676.

Fanaroff, B. L. and Riley, J. M. (1974). The morphology of extragalactic radio sources of high and low luminosity. *MNRAS*, 167:31P–36P.

Fazio, G. G., Hora, J. L., Allen, L. E., Ashby, M. L. N., Barmby, P., Deutsch, L. K., Huang, J. S., Kleiner, S., Marengo, M., Megeath, S. T., Melnick, G. J., Pahre, M. A., Patten, B. M., Polizotti, J., Smith, H. A., Taylor, R. S., Wang, Z., Willner, S. P., Hoffmann, W. F., Pipher, J. L., Forrest, W. J., McMurty, C. W., McCreight, C. R., McKelvey, M. E., McMurray, R. E., Koch, D. G., Moseley, S. H., Arendt, R. G., Mentzell, J. E., Marx, C. T., Losch, P., Mayman, P., Eichhorn, W., Krebs, D., Jhabvala, M., Gezari, D. Y., Fixsen, D. J., Flores, J., Shakoorzadeh, K., Jungo, R., Hakun, C., Workman, L., Karpati, G., Kichak, R., Whitley, R., Mann, S., Tollestrup, E. V., Eisenhardt, P., Stern, D., Gorjian, V., Bhattacharya, B., Carey, S., Nelson, B. O., Glaccum, W. J., Lacy, M., Lowrance, P. J., Laine, S., Reach, W. T., Stauffer, J. A., Surace, J. A., Wilson, G., Wright, E. L., Hoffman, A., Domingo, G., and Cohen, M. (2004). The Infrared Array Camera (IRAC) for the Spitzer Space Telescope. *ApJS*, 154(1):10–17.

Fluke, C. J. and Jacobs, C. (2020). Surveying the reach and maturity of machine learning and artificial intelligence in astronomy. *WIREs Data Mining and Knowledge Discovery*, 10(2):e1349.

Franzen, T. M. O., Banfield, J. K., Hales, C. A., Hopkins, A., Norris, R. P., Seymour, N., Chow, K. E., Herzog, A., Huynh, M. T., Lenc, E., Mao, M. Y., and Middelberg, E. (2015). ATLAS - I. Third release of 1.4 GHz mosaics and component catalogues. *MNRAS*, 453(4):4020–4036.

Ghosh, A., Urry, C. M., Wang, Z., Schawinski, K., Turp, D., and Powell, M. C. (2020). Galaxy Morphology Network: A Convolutional Neural Network Used to Study Morphology and Quenching in ~100,000 SDSS and ~20,000 CANDELS Galaxies. *ApJ*, 895(2):112.

Goderya, S. N. and Lolling, S. M. (2002). Morphological Classification of Galaxies using Computer Vision and Artificial Neural Networks: A Computational Scheme. *Ap&SS*, 279(4):377–387.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. The MIT Press.

Grandi, P., Capetti, A., and Baldi, R. D. (2016). Discovery of a Fanaroff-Riley type 0 radio galaxy emitting at γ-ray energies. *MNRAS*, 457(1):2–8.

Gupta, Y., Ajithkumar, B., Kale, H. S., Nayak, S., Sabhapathy, S., Sureshkumar, S., Swami, R. V., Chengalur, J. N., Ghosh, S. K., Ishwara-Chandra, C. H., Joshi, B. C., Kanekar, N., Lal, D. V., and Roy, S. (2017). The upgraded GMRT: opening new windows on the radio Universe. *Current Science*, 113(4):707–714.

Hartley, P., Bonaldi, A., Braun, R., Aditya, J. N. H. S., Aicardi, S., Alegre, L., Chakraborty, A., Chen, X., Choudhuri, S., Clarke, A. O., Coles, J., Collinson, J. S., Cornu, D., Darriba, L., Delli Veneri, M., Forbrich, J., Fraga, B., Galan, A., Garrido, J., Gubanov, F., Håkansson, H., Hardcastle, M. J., Heneka, C., Herranz, D., Hess, K. M., Jagannath, M., Jaiswal, S., Jurek, R. J., Korber, D., Kitaeff, S., Kleiner, D., Lao, B., Lu, X., Mazumder, A., Moldón, J., Mondal, R., Ni, S., Önnheim, M., Parra, M., Patra, N., Peel, A., Salomé, P., Sánchez-Expósito, S., Sargent, M., Semelin, B., Serra, P., Shaw, A. K., Shen, A. X., Sjöberg, A., Smith, L., Soroka, A., Stolyarov, V., Tolley, E., Toribio, M. C., van der Hulst, J. M., Vafaei Sadr, A., Verdes-Montenegro, L., Westmeier, T., Yu, K., Yu, L., Zhang, L., Zhang, X., Zhang, Y., Alberdi, A., Ashdown, M., Bom, C. R., Brüggen, M., Cannon, J., Chen, R., Combes, F., Conway, J., Courbin, F., Ding, J., Fourestey, G., Freundlich, J., Gao, L., Gheller, C., Guo, Q., Gustavsson, E., Jirstrand, M., Jones, M. G., Józsa, G., Kamphuis, P., Kneib, J. P., Lindqvist, M., Liu, B., Liu, Y., Mao, Y., Marchal, A., Márquez, I., Meshcheryakov, A., Olberg, M., Oozeer, N., Pandey-Pommier, M., Pei, W., Peng, B., Sabater, J., Sorgho, A., Starck, J. L., Tasse, C., Wang, A., Wang, Y., Xi, H., Yang, X., Zhang, H., Zhang, J., Zhao, M., and Zuo, S. (2023). SKA Science Data Challenge 2: analysis and results. *arXiv e-prints*, page arXiv:2303.07943.

Hausen, R. and Robertson, B. E. (2020). Morpheus: A Deep Learning Framework for the Pixel-level Analysis of Astronomical Image Data. *ApJS*, 248(1):20.

Hezaveh, Y. D., Perreault Levasseur, L., and Marshall, P. J. (2017). Fast automated analysis of strong gravitational lenses with convolutional neural networks. *Nature*, 548(7669):555–557.

Hoyle, B. (2016). Measuring photometric redshifts using galaxy images and Deep Neural Networks. *Astronomy and Computing*, 16:34–40.

Huertas-Company, M. and Lanusse, F. (2023). The Dawes Review 10: The impact of deep learning for the analysis of galaxy surveys. *PASA*, 40:e001.

Huertas-Company, M., Rodriguez-Gomez, V., Nelson, D., Pillepich, A., Bottrell, C., Bernardi, M., Domínguez-Sánchez, H., Genel, S., Pakmor, R., Snyder, G. F., and Vogelsberger, M. (2019). The Hubble Sequence at z $\sim$ 0 in the IllustrisTNG simulation with deep learning. *MNRAS*, 489(2):1859–1879.

Ishwara-Chandra, C. H., Taylor, A. R., Green, D. A., Stil, J. M., Vaccari, M., and Ocran, E. F. (2020). A wide-area GMRT 610-MHz survey of ELAIS N1 field. *MNRAS*, 497(4):5383–5394.

Jacobs, C., Collett, T., Glazebrook, K., Buckley-Geer, E., Diehl, H., Lin, H., McCarthy, C., Qin, A., Odden, C., Escudero, M. C., et al. (2019). An extended catalog of galaxy–galaxy strong gravitational lenses discovered in des using convolutional neural networks. *The astrophysical journal supplement series*, 243(1):17.

Jansky, K. G. (1933). Radio Waves from Outside the Solar System. *Nature*, 132(3323):66.

Jarvis, M., Taylor, R., Agudo, I., Allison, J. R., Deane, R. P., Frank, B., Gupta, N., Heywood, I., Maddox, N., McAlpine, K., Santos, M., Scaife, A. M. M., Vaccari, M., Zwart, J. T. L., Adams, E., Bacon, D. J., Baker, A. J., Bassett,

B. A., Best, P. N., Beswick, R., Blyth, S., Brown, M. L., Bruggen, M., Cluver, M., Colafrancesco, S., Cotter, G., Cress, C., Davé, R., Ferrari, C., Hardcastle, M. J., Hale, C. L., Harrison, I., Hatfield, P. W., Klockner, H. R., Kolwa, S., Malefahlo, E., Marubini, T., Mauch, T., Moodley, K., Morganti, R., Norris, R. P., Peters, J. A., Prandoni, I., Prescott, M., Oliver, S., Oozeer, N., Rottgering, H. J. A., Seymour, N., Simpson, C., Smirnov, O., and Smith, D. J. B. (2016). The MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) Survey. In *MeerKAT Science: On the Pathway to the SKA*, page 6.

Johnston, S., Taylor, R., Bailes, M., Bartel, N., Baugh, C., Bietenholz, M., Blake, C., Braun, R., Brown, J., Chatterjee, S., Darling, J., Deller, A., Dodson, R., Edwards, P., Ekers, R., Ellingsen, S., Feain, I., Gaensler, B., Haverkorn, M., Hobbs, G., Hopkins, A., Jackson, C., James, C., Joncas, G., Kaspi, V., Kilborn, V., Koribalski, B., Kothes, R., Landecker, T., Lenc, E., Lovell, J., Macquart, J. P., Manchester, R., Matthews, D., McClure-Griffiths, N., Norris, R., Pen, U. L., Phillips, C., Power, C., Protheroe, R., Sadler, E., Schmidt, B., Stairs, I., Staveley-Smith, L., Stil, J., Tingay, S., Tzioumis, A., Walker, M., Wall, J., and Wolleben, M. (2008). Science with ASKAP. The Australian square-kilometre-array pathfinder. *Experimental Astronomy*, 22(3):151–273.

Jonas, J. and MeerKAT Team (2016). The MeerKAT Radio Telescope. In *MeerKAT Science: On the Pathway to the SKA*, page 1.

Joye, W. and Mandel, E. (2003). New features of saoimage ds9. In *Astronomical data analysis software and systems XII*, volume 295, page 489.

Lacy, M., Baum, S. A., Chandler, C. J., Chatterjee, S., Clarke, T. E., Deustua, S., English, J., Farnes, J., Gaensler, B. M., Gugliucci, N., Hallinan, G., Kent, B. R., Kimball, A., Law, C. J., Lazio, T. J. W., Marvil, J., Mao, S. A., Medlin, D., Mooley, K., Murphy, E. J., Myers, S., Osten, R., Richards, G. T., Rosolowsky, E., Rudnick, L., Schinzel, F., Sivakoff, G. R., Sjouwerman, L. O., Taylor, R., White, R. L., Wrobel, J., Andernach, H., Beasley, A. J., Berger, E., Bhatnager, S., Birkinshaw, M., Bower, G. C., Brandt, W. N., Brown, S., Burke-Spolaor, S., Butler, B. J., Comerford, J., Demorest, P. B., Fu, H., Giacintucci, S., Golap, K., Güth, T., Hales, C. A., Hiriart, R., Hodge, J., Horesh, A., Ivezić, Ž., Jarvis, M. J., Kamble, A., Kassim, N., Liu, X., Loinard, L., Lyons, D. K., Masters, J., Mezcua, M., Moellenbrock, G. A., Mroczkowski, T., Nyland, K., O'Dea, C. P., O'Sullivan, S. P., Peters, W. M., Radford, K., Rao, U., Robnett, J., Salcido, J., Shen, Y., Sobotka, A., Witz, S., Vaccari, M., van Weeren, R. J., Vargas, A., Williams, P. K. G., and Yoon, I. (2020). The Karl G. Jansky Very Large Array Sky Survey (VLASS). Science Case and Survey Design. *PASP*, 132(1009):035001.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551.

Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M. J., Nichol, R. C., Szalay, A., Andreescu, D., Murray, P., and Vandenberg, J. (2008). Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *MNRAS*, 389(3):1179–1189.

Lonsdale, C. J., Smith, H. E., Rowan-Robinson, M., Surace, J., Shupe, D., Xu, C., Oliver, S., Padgett, D., Fang, F., Conrow, T., Franceschini, A., Gautier, N., Griffin, M., Hacking, P., Masci, F., Morrison, G., O'Linger, J., Owen, F., Pérez-Fournon, I., Pierre, M., Puetter, R., Stacey, G., Castro, S., Polletta, M. d. C., Farrah, D., Jarrett, T., Frayer, D., Siana, B., Babbedge, T., Dye, S., Fox, M., Gonzalez-Solares, E., Salaman, M., Berta, S., Condon, J. J., Dole, H., and Serjeant, S. (2003). SWIRE: The SIRTF Wide-Area Infrared Extragalactic Survey. *PASP*, 115(810):897–927.

Lukic, V., Brüggen, M., Banfield, J. K., Wong, O. I., Rudnick, L., Norris, R. P., and Simmons, B. (2018). Radio Galaxy Zoo: compact and extended radio source classification with deep learning. *MNRAS*, 476(1):246–260.

Lukic, V., Brüggen, M., Mingo, B., Croston, J. H., Kasieczka, G., and Best, P. N. (2019). Morphological classification of radio galaxies: capsule networks versus convolutional neural networks. *MNRAS*, 487(2):1729–1744.

Ma, Z., Xu, H., Zhu, J., Hu, D., Li, W., Shan, C., Zhu, Z., Gu, L., Li, J., Liu, C., and Wu, X. (2019). A Machine Learning Based Morphological Classification of 14,245 Radio AGNs Selected from the Best-Heckman Sample. *ApJS*, 240(2):34.

Marocco, F., Eisenhardt, P. R. M., Fowler, J. W., Kirkpatrick, J. D., Meisner, A. M., Schlafly, E. F., Stanford, S. A., Garcia, N., Caselden, D., Cushing, M. C., Cutri, R. M., Faherty, J. K., Gelino, C. R., Gonzalez, A. H., Jarrett, T. H., Koontz, R., Mainzer, A., Marchese, E. J., Mobasher, B., Schlegel, D. J., Stern, D., Teplitz, H. I., and Wright, E. L. (2021). The CatWISE2020 Catalog. *ApJS*, 253(1):8.

Marshall, P. J., Lintott, C. J., and Fletcher, L. N. (2015). Ideas for Citizen Science in Astronomy. *ARA&A*, 53:247–278.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133.

Metcalf, R. B., Meneghetti, M., Avestruz, C., Bellagamba, F., Bom, C. R., Bertin, E., Cabanac, R., Courbin, F., Davies, A., Decencière, E., Flamary, R., Gavazzi, R., Geiger, M., Hartley, P., Huertas-Company, M., Jackson, N., Jacobs, C., Jullo, E., Kneib, J. P., Koopmans, L. V. E., Lanusse, F., Li, C. L., Ma, Q., Makler, M., Li, N., Lightman, M., Petrillo, C. E., Serjeant, S., Schäfer, C., Sonnenfeld, A., Tagore, A., Tortora, C., Tuccillo, D., Valentín, M. B., Velasco-Forero, S., Verdoes Kleijn, G. A., and Vernardos, G. (2019). The strong gravitational lens finding challenge. *A&A*, 625:A119.

Mohan, N. and Rafferty, D. (2015). PyBDSF: Python Blob Detection and Source Finder.

Moustakas, L., O'Dowd, M., Anguita, T., Webster, R., Chartas, G., Cornachione, M., Dai, X., Fian, C., Hutsemekers, D., Jimenez-Vicente, J., Labrie, K., Lewis, G., Macleod, C., Mediavilla, E., Morgan, C. W., Motta, V., Nierenberg, A., Pooley, D., Rojas, K., Sluse, D., Vernardos, G., Wambsganss, J., and Yong, S. Y. (2019). Astro2020 Science White Paper - Quasar Microlensing: Revolutionizing our Understanding of Quasar Structure and Dynamics. *arXiv e-prints*, page arXiv:1904.12967.

Norris, R. P. (2011). Data Challenges for Next-generation Radio Telescopes. In *Sixth IEEE International Conference on eScience*, pages 21–24.

Norris, R. P. (2017a). Discovering the Unexpected in Astronomical Survey Data. *PASA*, 34:e007.

Norris, R. P. (2017b). Extragalactic radio continuum surveys and the transformation of radio astronomy. *Nature Astronomy*, 1:671–678.

Norris, R. P., Hopkins, A. M., Afonso, J., Brown, S., Condon, J. J., Dunne, L., Feain, I., Hollow, R., Jarvis, M., Johnston-Hollitt, M., Lenc, E., Middelberg, E., Padovani, P., Prandoni, I., Rudnick, L., Seymour, N., Umana, G., Andernach, H., Alexander, D. M., Appleton, P. N., Bacon, D., Banfield, J., Becker, W., Brown, M. J. I., Ciliegi, P., Jackson, C., Eales, S., Edge, A. C., Gaensler, B. M., Giovannini, G., Hales, C. A., Hancock, P., Huynh, M. T., Ibar, E., Ivison, R. J., Kennicutt, R., Kimball, A. E., Koekemoer, A. M., Koribalski, B. S., López-Sánchez, Á. R., Mao, M. Y., Murphy, T., Messias, H., Pimbblet, K. A., Raccanelli, A., Randall, K. E., Reiprich, T. H., Roseboom, I. G., Röttgering, H., Saikia, D. J., Sharp, R. G., Slee, O. B., Smail, I., Thompson, M. A., Urquhart, J. S., Wall, J. V., and Zhao, G. B. (2011). EMU: Evolutionary Map of the Universe. *PASA*, 28(3):215–248.

Padovani, P. (2016). The faint radio sky: radio astronomy becomes mainstream. *A&A Rev.*, 24(1):13.

Padovani, P. (2017a). Active Galactic Nuclei at all wavelengths and from all angles. *Frontiers in Astronomy and Space Sciences*, 4:35.

Padovani, P. (2017b). On the two main classes of active galactic nuclei. *Nature Astronomy*, 1:0194.

Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359.

Perley, R. A., Chandler, C. J., Butler, B. J., and Wrobel, J. M. (2011). The Expanded Very Large Array: A New Telescope for New Science. *ApJ*, 739(1):L1.

Petrillo, C. E., Tortora, C., Chatterjee, S., Vernardos, G., Koopmans, L. V. E., Verdoes Kleijn, G., Napolitano, N. R., Covone, G., Kelvin, L. S., and Hopkins, A. M. (2019a). Testing convolutional neural networks for finding strong gravitational lenses in KiDS. *MNRAS*, 482(1):807–820.

Petrillo, C. E., Tortora, C., Chatterjee, S., Vernardos, G., Koopmans, L. V. E., Verdoes Kleijn, G., Napolitano, N. R., Covone, G., Schneider, P., Grado, A., and McFarland, J. (2017). Finding strong gravitational lenses in the Kilo Degree Survey with Convolutional Neural Networks. *MNRAS*, 472(1):1129–1150.

Petrillo, C. E., Tortora, C., Vernardos, G., Koopmans, L. V. E., Verdoes Kleijn, G., Bilicki, M., Napolitano, N. R., Chatterjee, S., Covone, G., Dvornik, A., Erben, T., Getman, F., Giblin, B., Heymans, C., de Jong, J. T. A., Kuijken, K., Schneider, P., Shan, H., Spiniello, C., and Wright, A. H. (2019b). LinKS: discovering galaxy-scale strong lenses in the Kilo-Degree Survey using convolutional neural networks. *MNRAS*, 484(3):3879–3896.

Reber, G. (1944). Cosmic Static. *ApJ*, 100:279.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99.

Rosebrock, A. (2017). *Deep learning for computer vision with Python*. PYIMAGESEARCH, 1 edition.

Rothe, R., Guillaumin, M., and Van Gool, L. (2014). Non-maximum suppression for object detection by passing messages between windows. In *Asian conference on computer vision*, pages 290–306. Springer.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252.

Ryle, M. and Vonberg, D. D. (1946). Solar Radiation on 175 Mc./s. *Nature*, 158(4010):339–340.

Schlafly, E. F., Meisner, A. M., and Green, G. M. (2019). The unWISE Catalog: Two Billion Infrared Sources from Five Years of WISE Imaging. *ApJS*, 240(2):30.

Schneider, P. (2014). *Extragalactic astronomy and cosmology: an introduction*. Springer.

Seyfert, C. K. (1943). Nuclear Emission in Spiral Nebulae. *ApJ*, 97:28.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Tang, H., Scaife, A. M. M., and Leahy, J. P. (2019). Transfer learning for radio galaxy classification. *MNRAS*, 488(3):3358–3375.

Thompson, A. R., Clark, B. G., Wade, C. M., and Napier, P. J. (1980). The Very Large Array. *ApJS*, 44:151–167.

Tingay, S. J., Goeke, R., Bowman, J. D., Emrich, D., Ord, S. M., Mitchell, D. A., Morales, M. F., Booler, T., Crosse, B., Wayth, R. B., Lonsdale, C. J., Tremblay, S., Pallot, D., Colegate, T., Wicenec, A., Kudryavtseva, N., Arcus, W., Barnes, D., Bernardi, G., Briggs, F., Burns, S., Bunton, J. D., Cappallo, R. J., Corey, B. E., Deshpande, A., Desouza, L., Gaensler, B. M., Greenhill, L. J., Hall, P. J., Hazelton, B. J., Herne, D., Hewitt, J. N., Johnston-Hollitt, M., Kaplan, D. L., Kasper, J. C., Kincaid, B. B., Koenig, R., Kratzenberg, E., Lynch, M. J., Mckinley, B., Mcwhirter, S. R., Morgan, E., Oberoi, D., Pathikulangara, J., Prabu, T., Remillard, R. A., Rogers, A. E. E., Roshi, A., Salah, J. E., Sault, R. J., Udaya-Shankar, N., Schlagenhaufer, F., Srivani, K. S., Stevens, J., Subrahmanyan, R., Waterson, M., Webster, R. L., Whitney, A. R., Williams, A., Williams, C. L., and Wyithe, J. S. B. (2013). The Murchison Widefield Array: The Square Kilometre Array Precursor at Low Radio Frequencies. *PASA*, 30:e007.

Vafaei Sadr, A., Vos, E. E., Bassett, B. A., Hosenie, Z., Oozeer, N., and Lochner, M. (2019). DEEPSOURCE: point source detection using deep learning. *MNRAS*, 484(2):2793–2806.

Vilalta, R. (2018). Transfer Learning in Astronomy: A New Machine-Learning Paradigm. *arXiv e-prints*, page arXiv:1812.10403.

Vilalta, R., Dhar Gupta, K., Boumber, D., and Meskhi, M. M. (2019). A General Approach to Domain Adaptation with Applications in Astronomy. *PASP*, 131(1004):108008.

Wang, C.-Y. (2017). Essential radio astronomy, by james j. condon and scott m. ransom. *Contemporary Physics*, 58(3):278–279.

Willett, K. W., Lintott, C. J., Bamford, S. P., Masters, K. L., Simmons, B. D., Casteels, K. R. V., Edmondson, E. M., Fortson, L. F., Kaviraj, S., Keel, W. C., Melvin, T., Nichol, R. C., Raddick, M. J., Schawinski, K., Simpson, R. J., Skibba, R. A., Smith, A. M., and Thomas, D. (2013). Galaxy Zoo 2: detailed morphological classifications for 304 122 galaxies from the Sloan Digital Sky Survey. *MNRAS*, 435(4):2835–2860.

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., Ressler, M. E., Cutri, R. M., Jarrett, T., Kirkpatrick, J. D., Padgett, D., McMillan, R. S., Skrutskie, M., Stanford, S. A., Cohen, M., Walker, R. G., Mather, J. C., Leisawitz, D., Gautier, Thomas N., I., McLean, I., Benford, D., Lonsdale, C. J., Blain, A., Mendez, B., Irace, W. R., Duval, V., Liu, F., Royer, D., Heinrichsen, I., Howard, J., Shannon, M., Kendall, M., Walsh, A. L., Larsen, M., Cardon, J. G., Schick, S., Schwalm, M., Abid, M., Fabinsky, B., Naes, L., and Tsai, C.-W. (2010). The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. *AJ*, 140(6):1868–1881.

Wu, C., Wong, O. I., Rudnick, L., Shabala, S. S., Alger, M. J., Banfield, J. K., Ong, C. S., White, S. V., Garon, A. F., Norris, R. P., Andernach, H., Tate, J., Lukic, V., Tang, H., Schawinski, K., and Diakogiannis, F. I. (2019). Radio Galaxy Zoo: CLARAN - a deep learning classifier for radio morphologies. *MNRAS*, 482(1):1211–1230.

Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? In *Advances in neural information processing systems*, pages 3320–3328.