UNIVERSITY OF KWAZULU-NATAL

Using Machine Learning Techniques to Identify Strong Gravitational Lensing Systems in DES

by

Annarien Gertruida Headley

214526675

Submitted in fulfilment of the academic requirements for the degree of Master of Science in the

School of Mathematics, Statistics and Computer Science,

College of Agriculture, Engineering and Science,

University of KwaZulu-Natal,

Durban, South Africa

July 2021

As the candidate's supervisor I have approved this thesis for submission.

Name: _____ Signature: _____ Date: _____

I would like to dedicate this project to my Lord, Jesus Christ.

In loving memory of my Ouma Bettie

(1941 - 2020)

"Have I not commanded you? Be strong and courageous. Do not be terrified; do not be discouraged, for the Lord your God will be with you wherever you go."-Joshua 1:9

Abstract

Gravitational lensing systems enable astronomers to look into the distant universe by magnifying distant objects. Strong gravitational lensing systems are an incredibly rare phenomenon, with only a total about $\sim 1,000$ having been discovered. The use of machine learning (ML) has enabled the search for these systems, in the vast sky surveys that currently exist, to be narrowed down. This work investigates the use of ML techniques to identify strong gravitational lensing systems within the Dark Energy Survey (DES). We use the ML technique of convolutional neural networks (CNNs), a deep neural network architecture, as they are able to perform various image processing tasks efficiently. We generated a dataset of $96,768$ images to train and validate our CNN, half of which contains images from DES and the other half containing simulated lenses. The images from DES are scored with a 0, and the simulated lenses with a 1. Our CNN gained an accuracy of $99.73 \pm 0.07\%$ and a mean loss of 0.81 after being evaluated against an unseen dataset similar to that of the training data. We also evaluated our CNN against 389 real lenses, and gained an accuracy of $11.92 \pm 2.75\%$ and a mean loss of 10.49. Our CNN correctly predicted $\frac{57}{389}$ (14.65%) lenses. In this thesis, we present our CNN and the lenses that were correctly identified. In previous studies using CNNs to identify gravitational lensing systems, accuracies between 20% to 40% were achieved, thus the accuracy of 15% achieved by our CNN is competitive. Our results could be improved by training our CNN on the all morphologies of known lensing systems, not only those containing bright arcs. Results could also be improved by ensuring that all

the images are centred on the lens, and not the source images. These results and limitations are presented and discussed in the thesis. By looking at accurately simulated, as well as real lenses, one can train the CNN to be more precise.

Declaration of Authorship

I, Annarien Headley (214526675), declare that this thesis titled, 'Using Machine Learning Techniques to Identify Strong Gravitational Lensing Systems in DES' and the work presented in it are my own. I confirm that:

1. The research reported in this thesis, except where otherwise indicated, is my original research.

2. This thesis has not been submitted for any degree or examination at any other university.

3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from another person(s).

4. This thesis does not contain the writings of another person(s), unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

   (a) Their words have been rewritten, but the general information attributed to them has been referenced

   (b) Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

5. This thesis does not contain text, graphics or tables copied and pasted from the internet, unless specifically acknowledged, and the source being detailed in the thesis and in the Bibliography section.

Signature: _____ Date: _____

Acknowledgements

I would like to thank the following people for helping me in undertaking this Masters project:

- My supervisors Dr. Matt Hilton, Dr. Ilya Sinayskiy and Mr. Anban Pillay for all your wisdom, guidance, encouragement and teaching me how to become a better scientist. Thank you for the weekly meetings and the constant feedback, which made this project manageable.
- The University of KwaZulu Natal for allowing me to do this project.
- CSIR for giving me the opportunity to do my Masters project by funding me with a bursary.
- My husband Simon Headley, thank you for your endless support and encouragement.
- My family for all your support and prayers.

List of Figures

1.1 A *Hubble Space Telescope* image of gravitational lenses. 12

1.2 Diagram of gravitational lensing. 13

1.3 Diagram of the light path between the source, lens and observer. 14

1.4 Images of different types of gravitational lenses morphologies. 18

1.5 Diagram showing the subgroups of AI. 25

1.6 Diagram showing the subgroups of ML. 27

1.7 Diagram showing the connections of a typical ANN. 30

2.1 Diagram showing the flow of retrieving the necessary data. 39

2.2 Images showing the process of retrieving the negative data images. 41

2.3 RGB composite images of the negative images. 42

2.4 Background sky images of DES. 43

2.5 Images showing the process of retrieving the positive simulated data. 45

2.6 RGB composite images of the positive simulated images. 46

2.7 RGB composite images of the unseen real lenses. 48

2.8 The architecture of the CNN model that is used. 51

2.9 A diagram of k-fold cross validation. 54

3.1 Explanation of the loss and accuracy graphs. 57

3.2 Loss and accuracy graphs of the CNN model. 58

3.3 RGB images of all our datasets. 61

3.4 Loss and accuracy graphs of all runs. 67

3.5 Loss and accuracy graphs of all runs, continued 68

4.1 Experiment 1 loss and accuracy graphs. 73

4.2 Experiment 2 loss and accuracy graphs. 75

C.1 Histogram of the indentified lenses in the g-band. 103

C.2 Histogram of the indentified lenses in the r-band. 104

C.3 Histogram of the indentified lenses in the i-band. 105

List of Tables

1.1 Table showing the different sky surveys with their sky area, and the data volumes expected after completion. 24

3.1 Table showing the results obtained from our CNN, for the various unseen datasets. 63

3.2 Table showing the mean loss and accuracy scores obtained for the different datasets, during each of the five runs of our CNN. 64

3.3 Table showing the confusion matrices of 5 runs of our CNN. 65

4.1 Table showing a few different instrumental aspects of DES, Euclid and LSST. 79

B.1 Table showing the 42 unique strong gravitational lensing systems that our CNN was able to identify correctly across all five runs. 100

Contents

Abstract **1**

Declaration of Authorship **3**

Acknowledgements **4**

List of Figures **5**

List of Tables **7**

1 Introduction **10**

1.1 Gravitational Lensing 11

1.2 History of Gravitational Lensing 21

1.3 Artificial Intelligence and Big Data in Astronomy 23

1.3.1 Artificial Intelligence 26

1.3.2 Machine Learning 26

1.3.3 Deep Learning Neural Networks 29

1.4 Literature Review 32

2 Methodology **37**

2.1 Negative Data 40

2.2 Positive Data 43

2.3 Unseen Data 46

2.4 Training Convolutional Neural Networks 48

3 Results **55**

4 Discussion **69**

4.1 Results Using Alternative CNN Architectures 72

4.2 Possible Improvements 76

5 Conclusion **80**

Bibliography **84**

A Explanation of the Different Datasets **98**

B Table of Identified Strong Gravitational Lenses **100**

C Histograms of the Identified Lenses Magnitudes **103**

Chapter 1

Introduction

Gravitational lensing systems enable us to look into the distant universe. The phenomenon of gravitational lensing occurs as the photons emitted by a distant object are deflected as they pass the vicinity of another object with a large enough mass. If the observer is positioned in such a way that both objects are aligned along the line of sight, the most distant object will be magnified and distorted (Treu, 2010; Kneib and Natarajan, 2012; Diehl et al., 2017). Gravitational lensing systems are quite rare, with only a few thousand having been discovered and only a few hundred thousand expected to be found in the upcoming large sky surveys such as Euclid, *Roman Space Telescope*, Vera C. Rubin Observatory (Laureijs et al., 2011; Davies et al., 2019; Ivezić et al., 2019; Poberezhskiy et al., 2021). Studying gravitational lenses offers a unique opportunity to observe objects in the distant universe and to measure their physical properties (Bouwens et al., 2019), adding to our comprehension of the structure and formation of galaxies (Wambsganss, 1998; Davies et al., 2019; Li et al., 2020).

In this project, we use machine learning (ML) techniques to automate the search for these strong lenses within the large area (5000 sq. degrees) covered by the Dark Energy Survey

(DES) (Abbott et al., 2018). ML is a computational system that is able to learn from a given input to produce an expected output, through repeated application of the system, gaining experience (Bell, 2020).

We begin this chapter with an introduction to gravitational lensing in Section 1.1. We present a brief history on this subject in Section 1.2. We describe the various ML techniques and its applications in astronomy in Section 1.3. We end this chapter with a review of the literature surrounding the use of ML to search for gravitational lenses, and the motivation for this thesis in Section 1.4.

1.1 Gravitational Lensing

Photons that travel from a distant object (source) are diverted due to a large mass object (lens) between the source and the observer, when the source, the lens and the observer are aligned along the same line of sight (Schneider et al., 1992; Kneib and Natarajan, 2012; Treu, 2010; Diehl et al., 2017). This distortion of the background source is known as gravitational lensing. Objects with a large enough mass, such as galaxies and galaxy clusters allow for gravitational lensing to take place (Soucail et al., 1987; Schneider et al., 1992; Zalesky and Ebeling, 2020), producing multiple magnified images of distant background galaxies. Some of these images can be seen as bright arcs, as shown in Figure 1.1.

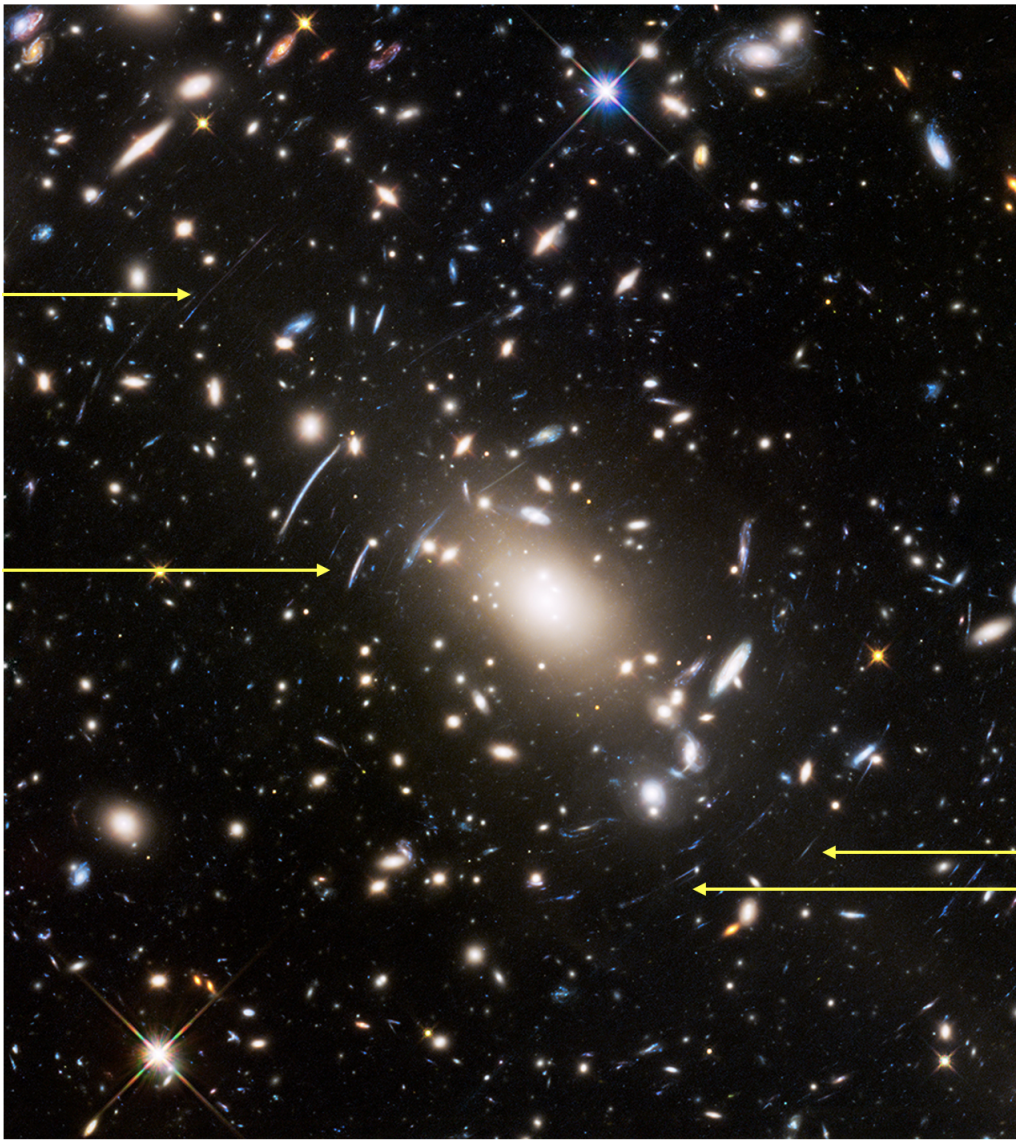

Figure 1.1: This is a *Hubble Space Telescope* image of the galaxy cluster Abell S1063. Four strong gravitational lensing arcs are shown with the yellow arrows around the galaxy cluster. This image comes from the Hubble Frontier Fields project (Koekemoer et al., 2017).

Figures 1.2 and shows the formation of gravitational lenses as the light path from the source curves around the lens creating multiple distorted images of the source. The lens is labelled as **A**, and the source is labelled as **B**. The lens is an object with a large enough mass that can curve light emitted by a distant source, and thus, creating distorted images

of the source. As seen in the right panel, the observer can not directly observe the source but rather its images (Treu, 2010; Metcalf et al., 2019). The geometry of these systems is seen in Figure 1.3.

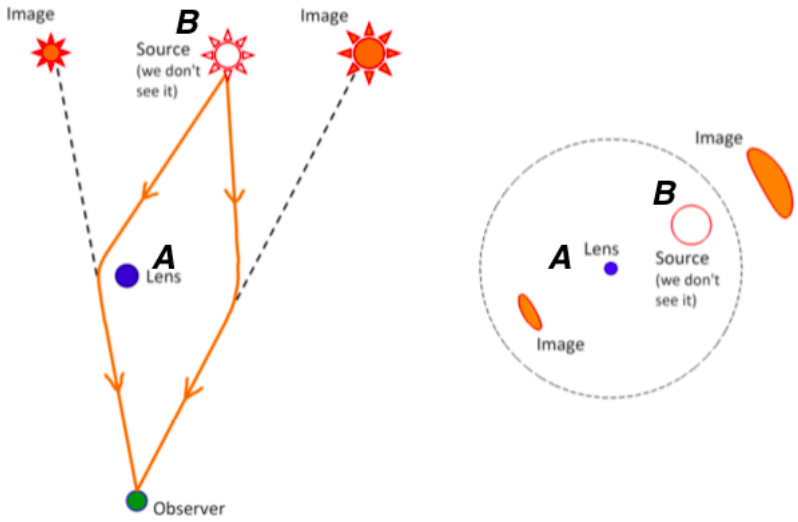

Figure 1.2: (Left panel): This diagram shows how the multiple images of the source are created through gravitational lensing. The large mass of the lens, A, causes the path of the photons from the source, B, to bend around it. (Right panel): This causes the observer to observe the source as multiple images (Bisnovatyi-Kogan and Tsupko, 2017).

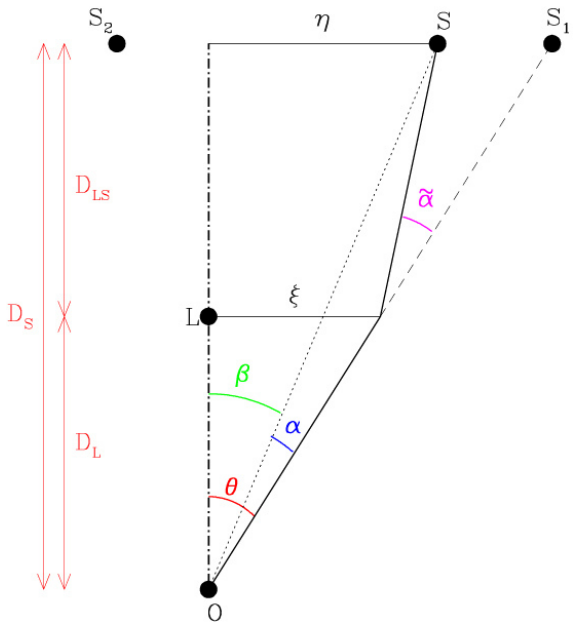

Figure 1.3: Diagram showing the geometry between the source (S), the lens (L), the observer (O). The images of the source ($S_{1/2}$) are created by light waves curving around the lens. β is the angle between the lens and the source, and θ is the angle between the lens and the source images. The angle α is the angle between the source and its observed images (Wambsganss, 1998).

Figure 1.3 shows the geometry of the curvature of the light from the source (S), as it passes through the lens (L), creating images ($S_{1/2}$). D_S, D_L and D_{LS} are the angular diameter distances between the observer (O), the lens (L) and the source (S). It is important to note that at large cosmological distances $D_S \neq D_L + D_{LS}$ (Wambsganss, 1998). The observer will observe the source at the position of the images ($S_{1/2}$). The angle of deflection of the light ray ($\widetilde{\alpha}$) passing a lens with mass M, at a distance of r, is calculated by

$$\widetilde{\alpha} = \frac{4GM}{c^2}\frac{1}{r}, \tag{1.1}$$

where G is the gravitational constant and c is speed of light (Wambsganss, 1998). The relation between the scaled deflection angle (α) and the angle of deflection of the light ray

$(\widetilde{\alpha})$ is given by,

$$\alpha(\theta) = \frac{D_{LS}}{D_S}\widetilde{\alpha}(\theta), \tag{1.2}$$

where α is the angle between the source (S) and its image (S_1) and θ is the angle between the lens and the source's image, as seen by the observer (Treu, 2010; Meneghetti, 2016). The angle between the lens and the source is $\beta = \theta - \alpha$. If θ, β and α are small, then the observed positions of the source (S) and its images $(S_{1/2})$ can be calculated using the **lens equation** (see review papers by Treu, 2010; Meneghetti, 2016), as

$$\theta D_S = \beta D_S + \widetilde{\alpha} D_{LS}, \tag{1.3}$$

and can be rewritten in terms of the lensing potential in two dimensions (ψ) as,

$$\beta = \theta - \bigtriangledown\psi = \theta - \alpha. \tag{1.4}$$

The relationship between the angular diameter distances of the source (S), the lens (L), the observer (O) and the angle between the lens, the source and its images $(S_{1/2})$ is calculated using the following equation,

$$\beta = \theta - \frac{D_{LS}}{D_S D_L}\frac{4GM}{c^2\theta}. \tag{1.5}$$

The shape of the source images $(S_{1/2})$ depends on the solutions for the lens equation (equation (1.3)), at all points of the source (Meneghetti, 2016). Strong gravitational lensing will cause a source to appear as multiple images when equation (1.4) has several solutions. Using the Poisson Equation $(\bigtriangledown^2\psi = 2\kappa)$, where κ is the projected surface mass density of the source, also called convergence, (Treu, 2010), the equation of the distorted images created through

gravitational lensing can be calculated as

$$\frac{\partial \beta}{\partial \theta} = \delta_{ij} - \frac{\partial^2 \psi}{\partial \theta_i \partial \theta_j} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}. \tag{1.6}$$

The matrix given in equation (1.6), shows that due to convergence (κ) and shear (γ), the light rays emitted by the source passing through the lens are magnified and distorted. Convergence is the projected surface mass density in terms of the critical density of the universe and causes the image of the source to be stretched isotropically by a constant factor. Shear is used to describe the distortion of the lensed images of the source. Both the convergence and shear are dimensionless (see review papers by Bartelmann, 2010; Treu, 2010). If both κ and γ are non-zero, a circular source appears as an ellipse, where a and b are the semi-major and semi-minor axes, and r is the radius of the source (Meneghetti, 2016), as seen in the following equation,

$$a = \frac{r}{1 - \kappa - \gamma}, b = \frac{r}{1 - \kappa + \gamma}. \tag{1.7}$$

Equation (1.6) represents the inverse magnification of a lensed source. The magnification of a source can be calculated in terms of the convergence (κ) and shear (γ_1, γ_2) as seen in equation (1.8), where μ represents the magnification (see review by Treu, 2010).

$$\mu = \frac{1}{(1 - \kappa)^2 - \gamma_1^2 - \gamma_2^2} \tag{1.8}$$

There are two types of gravitational lensing regimes, strong and weak. **Weak gravitational lensing** is the minute distortions of the image of the source which is observed through the use of statistics (Treu et al., 2012). This can be utilised to investigate the statistics of the matter distribution between an observer and a distant object giving insight to various cosmological

parameters and models (Bartelmann and Schneider, 2001). **Strong gravitational lensing** occurs when the light emitted by the source bends due to large mass densities, creating multiple images (or arcs) of the source in the background (Treu, 2010), providing insight to the structure of galaxies and clusters. These arcs reveal interesting findings of galaxy clusters: that the mass distributions of clusters are not symmetric; the abundance of dark matter in clusters may not be connected to the galaxies; and since the arcs are thin, sharp density profiles should be observed for clusters (see review by Bartelmann, 2010). Various types of lensing systems can be seen in Figure 1.4. For this project we will be looking at strong gravitational lensing systems associated with individual galaxies, rather than clusters.

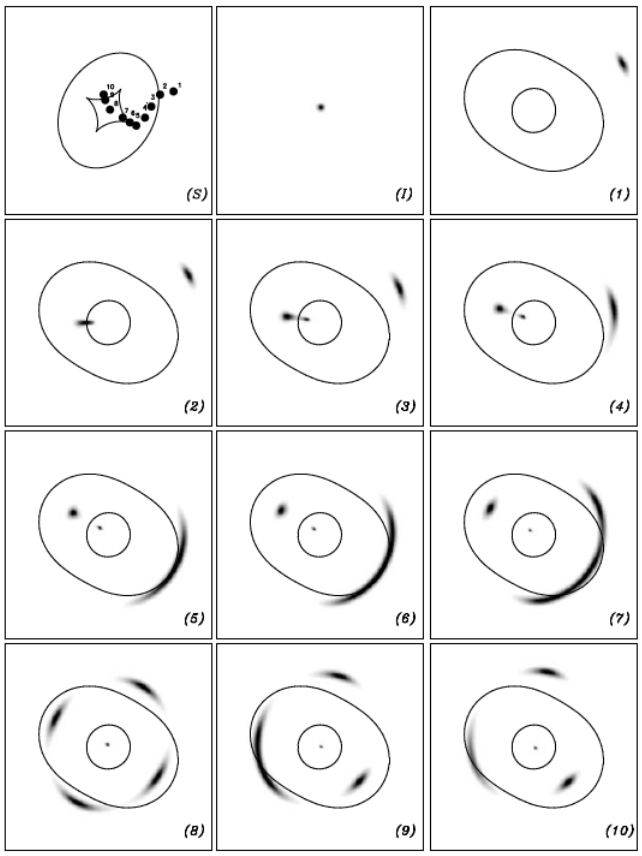

Figure 1.4: These images shows the different possible configurations or morphologies of gravitationally lenses systems. This shows the positions of the source (black point) in relation to the caustic lines (solid lines). Panel (I) shows the source without a lens present. The panels (1-10) shows the different lensing morphologies that show various lensed images of the source that could occur (Kneib and Natarajan, 2012).

In Figure 1.4, the bright arcs represents the distorted magnified images of the background source. The caustic lines of the lens systems in this figure, show the area relative where magnification of the source will increase massively, allowing for the observance of the lensed source images. The systems present in panels (3),(6),(8) and (10) for instance corresponds to a radial arc, a cusp arc, an Einstein cross and a fold arc respectively (Kneib and Natarajan, 2012). In our project we will be looking at strong lensing systems with very bright arcs, similar to the cusp arcs in panels (5),(6) and (7).

There are three main properties of strong gravitational lenses: (1) the observable parameters depend on the lens gravitational potential, resulting in parameters such as positions, flux and time delays between images; (2) the ability to observe these systems depend on the geometry between the observer, the lens and the source as seen in Figure 1.3; (3) the source has a brightness an order of magnitude higher than that of the lens (see review by Treu, 2010).

The impact of gravitational lensing systems in both astronomy and cosmology is vast, since the source is magnified and its surface brightness does not diminish as its light undergoes a distortion when passing the lens. For this reason, faint distant objects can be observed (eg. Diehl et al., 2017). Gravitational lensing systems are used to study cosmology, including the investigation of dark matter and dark energy. Dark matter does not emit or absorb light, but can only be visualised or mapped thanks to its gravitational influence on objects (Fraknoi et al., 2016). Dark energy is believed to be responsible for the observed acceleration of the expansion of the universe (Riess et al., 1998; Perlmutter et al., 1999; Fraknoi et al., 2016), and could represent up to 70% of the mass-energy content of the universe (Bennett et al., 2013; Planck Collaboration et al., 2020)

In 2012, the dark energy equation of state was assumed to be $w = -1$, derived from the combination of results using the Sloan Digital Sky Survey Quasar Lens Search (SDSSQLS) with baryon acoustic oscillations and the Wilkinson Microwave Anisotropy Probe (WMAP) (Oguri et al., 2012). Several cosmological parameters in a flat standard cosmological model were measured by using gravitationally lensed quasars. This was done by measuring time delays of three multiple imaged quasar systems (Courbin et al., 2016). The Hubble constant was measured to be $H_0 = 71.9^{+2.4}_{+3.0}$ km s^{-1} Mpc^{-1} and the energy density due to a cosmological constant as $\Omega_\Lambda = 0.62^{+0.24}_{-0.35}$. The Hubble constant represents the expansion rate of the universe, with the latest values being 67.4 ± 0.5 km s^{-1} Mpc^{-1} (Planck Collaboration et al.,

2020). It can be constrained by measuring the time delays of the deflected photons around the lens, as opposed to when they are not deflected (Schechter, 2004).

Most strong galaxy-galaxy lensing systems consist of early type galaxies (ETGs) as lenses (Shu et al., 2017; Jacobs et al., 2019b). Since ETGs contribute to a large amount of the stellar mass in the universe and are amongst the oldest objects in the universe, by studying them, a foundation for understanding the formation of the universe may be formed (Kochanek et al., 1999; Koopmans and Treu, 2004; Jacobs et al., 2019b). Strong gravitational lenses allow for the observation of objects at high redshift that would otherwise not be visible (Marshall et al., 2007), providing relevant information of the internal mass distribution within galaxies as well as their dark matter content (Refregier, 2003; Bolton et al., 2011). Gravitational lensing systems are used to calculate masses of distant objects, and observe quasars (Schneider et al., 1992; Kelly et al., 2018), as well as giving the mass distribution of the lens, which is usually a galaxy or a cluster of galaxies (Petrillo, 2019). Through the use of strong lensing, the mass substructure and the interstellar mediums of galaxies can be investigated (Schneider, 2003). Strong lensing systems have also been used to explore the dark matter halos, and substructures of galaxies outside our local group (Diaz Rivero and Dvorkin, 2020; Shajib et al., 2020). Through the use of these systems, it is possible to test general relativity, which in turn influences various fields of science (Collett et al., 2018). For the purposes of this thesis, only a few applications of strong gravitational lensing have been discussed.

Initial studies of gravitational lensing systems were performed using small research teams of around 10 people to visually inspect large amounts of data from sky surveys to identify lenses. Later on the use of crowd sourcing became common practice. Using more than $200,000$ people to visually inspect large surveys, the rate at which lenses were discovered increased (Marshall et al., 2016). However, in recent years, automation to find these systems through the use of ML techniques has increased, due to the rapid improvements in new tech-

nology and an overall growth of large sky surveys (Lanusse et al., 2017), as will be discussed in Section 1.3.

1.2 History of Gravitational Lensing

Many scientists have contributed to the study of the gravitational lensing phenomenon as progresses in this field took place over several decades. A few contributions and the progression made by various scientists in understanding gravitational lensing, are discussed in this section.

In 1780, Jean-Paul Marat stated that light, upon nearing the edge of objects, is diffracted from the light path, bending either away or towards an object (Valls-Gabaud, 2006, 2009). Henry Cavendish proposed the idea of gravitational attraction, in his 1783/1784 manuscript stating that when light passes near the surface of a body, it bends because of the attraction of that body (Valls-Gabaud, 2006; Cervantes-Cota et al., 2019). In 1801, Johann Georg von Soldner was able to calculate that stars that are behind the Sun, but appear close to the edge of the Sun, had a deflection of 0.84 arcseconds (Jaki, 1978; Ellis, 2010).

In 1907-1911, Albert Einstein rejuvenated the idea of light bending around objects due to the attraction of those objects, concluding that the gravitational field of the object must be responsible for the curvature of the light rays (Einstein, 1911). In 1919, Stanley Arthur Eddington undertook an experiment to investigate the Suns influence on light rays, to answer the question of how gravity and light interact with one another. He performed his experiments during an eclipse so that he could observe light rays near the Sun. Eddington observed deflection of light from stars in accordance with Einstein's theory of relativity, which was as a result of the gravitational field of the Sun (Eddington, 1919; Dyson et al., 1920). In 1936,

Einstein was able to describe this gravitational lensing effect as light curving around a star when in front of another star. However, he calculated that it would be very unlikely that we will be able to observe this phenomenon. He stated that the apparent brightness of the source would merge with that of the lens, and the observer would have great difficulty in distinguishing the two objects from each other (Einstein, 1936).

Inspired by Einstein's work, Frantisek Link was able to create images through the deflection of light around objects, as well as determining the image positions and brightness. He realised that a faint star may be seen with this curvature if it were faint and faraway (Valls-Gabaud, 2009). In 1936 and 1937, Link was able to explain the many aspects of gravitational lensing, being very optimistic that it would expand the future of astronomical research (Link, 1937; Valls-Gabaud, 2009). In 1937, Fritz Zwicky explained that an object with the mass of a star ($M_{\odot}$) would be too small to produce the lensing effect, but that a nebula sized object ($M_N = 10^9\ M_{\odot}$) has enough mass for the formation of gravitational lenses. This was verified through the observations of the Virgo cluster ($M_{Virgo} = 4 \times 10^{14}\ M_{\odot}$) (Smith, 1936; Zwicky, 1937).

The first observation and confirmation of a gravitational lens outside our solar system was done by D. Walsh, R.F. Corswell and R.J. Weymann in 1979 (Walsh et al., 1979; Treu, 2010). They observed a quasar (0957 + 561 A, B) and were baffled by it being a twin quasar. They had difficulties in describing the two quasars as two different distinct objects, and investigated the possibility of a single quasar appearing as multiple images due to a gravitational lens (Walsh et al., 1979).

In 1987, G. Soucail and her team observed the first arc formed due to a gravitational lensing effect, confirming the suspicion that lensing arcs can be observed (Soucail et al., 1988). Since then, the study of gravitational lenses has evolved dramatically and the amount of data from

sky surveys and telescopes has accumulated. At the time of writing this, there are about 675 gravitational lensing systems identified in the Master Lens Database [1]

The understanding of gravitational lenses and its applications has escalated profoundly over the last few decades. The application also highlights the significance of these objects to astronomy and cosmology as a whole. Several of these applications have been discussed in Chapter 1.1. If interested, the reader may look at applications discussed in detail by Bartelmann, Ellis, and Treu (Bartelmann, 2010; Ellis, 2010; Treu, 2010).

1.3 Artificial Intelligence and Big Data in Astronomy

The use of artificial intelligence (AI) in the field of astronomy is growing rapidly due to the exponential accumulation of data from modern technologies. This includes large ground and space-based sky surveys, resulting in the astronomy field to enter the big data era (Zhang and Zhao, 2015). The data volumes of the upcoming sky surveys is expected to range from a few terabytes (TB), to petabytes (PB) and exabytes (EB), as shown in Table 1.1. With this growth in data, there is an urgent need to analyse and process data automatically, taking advantage of the computational power presented by AI algorithms. This will allow astronomers to analyse certain objects quicker than manually searching through an enormous dataset (Ball and Brunner, 2010).

[1] admin.masterlens.org

Table 1.1: Table showing the different sky surveys with their sky area, and the data volumes expected after completion. The references are as follows: [1] de Jong et al. (2012), [2] Quinn et al. (2015), [3] Abbott et al. (2018).,[4] Ivezić et al. (2019)

Sky Survey Name	**Sky Area (sq.degrees)**	**Expected Data Volume**	**Years Of Operation**
The Kilo-Degree Survey (KiDS)[1]	1500	$\sim$ 15 TB	2011 - Present
The Dark Energy Survey (DES)[3]	5000	$\sim$ 500 TB	2013 - 2019
The Rubin Observatory Legacy Survey of Space and Time (LSST) [4]	$\sim$ 18000	$\sim$ 500 PB	Start 2023
The Square Kilometre Array (SKA)[2]	$\sim$ Southern Hemisphere	$>$ 1 EB	Start $\pm$ 2025

AI is the ability for a machine or system to learn to do a specific task from a given set of data without being specifically programmed to do it (Hurwitz and Kirsch, 2018). ML is a form of AI, in which machines are able to perform specified tasks through the use of certain software which can analyse statistics of the given data (Mohammed et al., 2016). Figure 1.5, shows that neural networks are part of deep learning (DL), a subfield of ML.

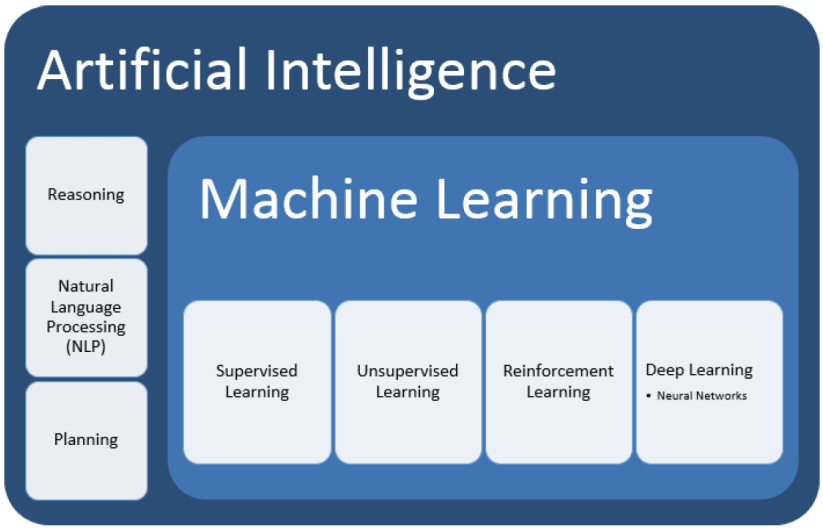

Figure 1.5: This diagram shows the different groups that fall into the category of AI. Neural networks are part of DL, which in turn is part of the larger ML class. ML is a branch of AI (Hurwitz and Kirsch, 2018).

AI gives a machine or system the ability to perform reasoning and planning through constant learning and training, similarly to neurons in our brains when performing complex tasks (Shalev-Shwartz and Ben-David, 2014; Iliashenko et al., 2019). AI is discussed in Section 1.3.1. ML (Section 1.3.2) is a sub-class of AI in which different types of learning, mainly supervised, unsupervised, reinforcement learning, and DL, can be implemented. AI and ML techniques include various artificial neural networks (ANNs), which contain multiple neuron connections (Section 1.3.3). DL has two or more layers than those seen in ANN (Kurata, 2018). The complexity of DL increases with the addition of layers in the neural network (Section 1.3.3). The ML technique that is used in this project is convolutional neural networks (CNNs) and this is discussed in Section 1.3.3.

1.3.1 Artificial Intelligence

AI is the ability of a machine to make intelligent decisions independently based on the given input and the required output. As seen in Figure 1.5, ML is a branch of AI, in which continuous learning and training are done on a selected algorithm in order to perform tasks associated with human abilities (Moolayil, 2018).

In 1957, Frank Rosenblatt identified that with the ability to view a large dataset from all possible angles, an AI system could recognise complex patterns and distinguish future images with ease (Rosenblatt, 1957). This is similar to teaching a baby that a shoe is a shoe, by repeating the object's name and showing the image multiple times. Subsequently the baby will learn to differentiate between a shoe and other objects. AI algorithms trained on a system obtain the required output, rather than doing this manually which is time consuming. This process is more data intensive and, subsequently, is more expensive (Jordan and Mitchell, 2015).

1.3.2 Machine Learning

ML is the automatic improvement through the experience gained as a result of training (Jordan and Mitchell, 2015). In 1959, Arthur Samuel defined ML as the ability for a computer to learn new things without being programmed specifically to know such things. Samuel is known for his self-learning computer programme that learns how to play checkers, and continuously learns with each new move (Samuel, 1959; Bell, 2020). Tom Mitchell defined ML as a computer programme that learns from the obtained experience when doing tasks. The performances of those tasks are measured, and increased as more tasks are performed (Mitchell, 1997; Bell, 2020). Each person has a different definition, but the basic idea remains the same; a computer that learns how to perform tasks on its own without being directly programmed

to do so through the use of repetition and large volumes of data. For an ML algorithm to perform similarly to humans, it needs a large amount of information to do the simplest tasks (Correspondent, A, 1969).

The data that a ML system needs varies based on the type of problem to be solved. There are a few ways in which a ML system can learn and be trained. This includes the way in which the ML interprets the given data, with the most common being supervised and unsupervised learning, and the less common ones being semi-supervised learning and reinforcement learning, as seen in Figure 1.6.

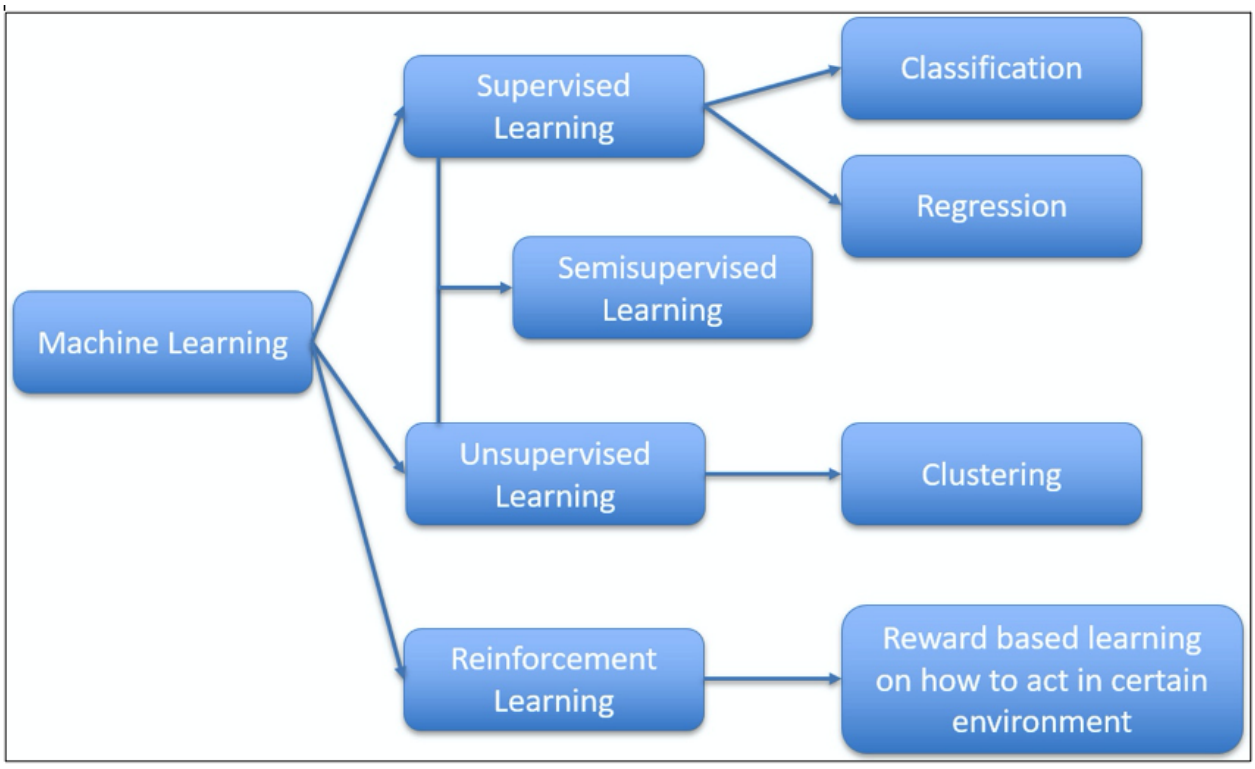

Figure 1.6: This diagram shows the subsets of ML. ML has three main types of learning, supervised, unsupervised and reinforcement learning. The classification and regression learning types are part of supervised learning (Liu et al., 2020).

Supervised learning is when the input and expected output are predetermined, through the use of labelled data (Bell, 2020). Supervised learning determines a general rule or pattern from the given input to obtain the correct output (Liu, 2017). The benefit of supervised learning is that a model can be trained to produce accurate outcomes when given unseen dataset. Supervised learning is used in various ways, for example: in face recognition, weather forecasting, patient diagnosis and bank card transactions (Shalev-Shwartz and Ben-David, 2014; Liu, 2017; Moolayil, 2018). Supervised Learning also has two subsets, regression and classification (Liu, 2017).

Regression is used when modelling is based on a continuous response by taking into consideration the previous results to make future predictions. Regression is most common when doing statistical analysis (Hurwitz and Kirsch, 2018). An example of when regression is used, is when determining petrol prices: the previous prices and the continuous inflation or deflation rates are taken into account. Classification is when the modelling is based on the appropriate labels for classes or categories etc., for example: an animal is either a dog or a cat (Liu, 2017). Within this project, supervised learning with classification is used to determine whether a strong gravitational lens is detected within an image or not.

Unsupervised learning occurs when the algorithm observes a pattern within a given dataset. This method is most often used in data mining cases (Bell, 2020). Unsupervised learning is done when the given data has no indicative descriptions (unlabelled data) and the algorithm finds a pattern in the data in question. Unsupervised learning is mostly used to find anomalies, outliers, different groupings or clusters with similar characteristics (Liu, 2017).

Semi-supervised learning is used when there is a mixture of labelled and unlabelled data (Mohammed et al., 2016). In most instances there is a larger set of unlabelled data than labelled data (Liu, 2017). The goal of this type of learning is to train the model to learn and then

predict future data better than using only labelled data (Mohammed et al., 2016). Reinforcement learning is when the data being learnt provides feedback to improve the system by adapting to certain conditions (Liu, 2017). This uses information collected through interaction or observation of the data to train the algorithm (Mohammed et al., 2016).

1.3.3 Deep Learning Neural Networks

DL contains a higher complexity than ANN, as it has more layers (Kurata, 2018). DL has been created to gain key characteristics representing certain data automatically from a given dataset (Najafabadi et al., 2015). DL algorithms consist of multiple layers, and is able to extract underlying features from the data. The DL algorithm uses these underlying features, and is able to create patterns describing the data. When more unseen data is given, the DL algorithm can create new descriptions and patterns, increasing the amount of pattern recognition as more features are learnt (Najafabadi et al., 2015). During the training of the DL, the simplest solutions are determined first, then the more complex solutions, with the most complex solutions determined at the very end (Schaub, 2019). When using DL, a high accuracy rate is often obtainable, especially when working with a large dataset and both unlabelled data or unsupervised learning is concerned (Arnold et al., 2011; Najafabadi et al., 2015).

Artificial Neural Networks

ANN is a general term to define various AI architectures containing multiple neuron connections (Bailer-Jones et al., 2002). The basis of an ANN is to simply learn from a given input, to optimise for an expected output (O'Shea and Nash, 2015), and to gain improved results through the use of training and experience (Davies et al., 2019).

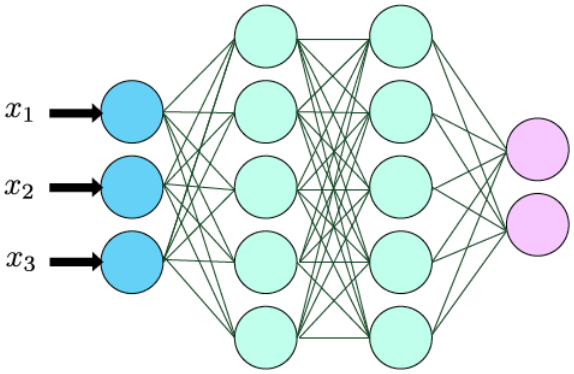

Figure 1.7: This diagram shows the connections between the input (blue), hidden (green) and output layers (pink) in a typical ANN (Jacobs et al., 2019b).

A typical ANN has three basic layers, the input (blue), the hidden (green) and output (pink), as seen in Figure 1.7. There exist multiple neurons in each layer (the circles) and the input is received by the neurons in the first layer, the input layer. The neurons in the input layer produce a result which is received by the neurons in the hidden layers. This is repeated until the last layer, the output layer. Within the layers there exists various neurons and activation functions. An activation function receives the output of a previous layer as input and applies a function to it, creating a new output, which is used as input for the next layer (Moolayil, 2018). This leads to a weight describing the relation between two succeeding neurons. This weight describes the impact of the receiving input to the produced output of a neuron and eventually the entire output of the neural network, allowing for correct predictions (Moolayil, 2018). The depth and complexity of an ANN algorithm is directly related to the amount of the hidden layers.

Convolutional Neural Networks

Within DL, there exist different types of neural networks. One specific type is CNNs, which contains certain features within its layers such as convolution, max-pooling, dropout etc. CNN is a form of an ANN where a group of neurons are connected to other neurons, sending messages or signals from one group to the next, using information from the given input set to produce an appropriate output (Moolayil, 2018).

CNN algorithms have been used for various research projects and has improved over time. CNN applications vary widely, some include bipolar patterns being recognised, handwriting recognition (Atlas et al., 1988), drug discovery (Wallach et al., 2015) and graphing of data (Niepert et al., 2016).

CNNs are highly effective in recognising patterns in images, as they join groups of neurons and calculate a group weight, therefore, shrinking the amount of trainable weights whilst keeping the visual data (Jacobs et al., 2019b). CNNs are able to extract features of given images during training (Baron, 2019), and because of this advantage it is used in this project, as we try to identify gravitational lensing arcs in images. CNNs are equal or faster than humans in accurately classifying images and are known to outperform other known ML techniques when processing two dimensional images (Lecun et al., 1998; Davies et al., 2019; Metcalf et al., 2019).

The current substantial progress in the usage of neural networks on large datasets became possible due to two main advancements, which is the back propagation algorithm and the ability to effectively implement it through the use of a graphics processing unit (GPU). Back propagation allows the training process of the full neural network to be reformulated as a sequence of matrix multiplications (Mitchell, 1997). This is efficiently executed on a GPU.

GPUs have the ability to perform linear algebra operations, such as matrix multiplication and matrix addition, very efficiently (Moolayil, 2018).

1.4 Literature Review

We now end this chapter with a review of some of the recent literature in which searches for gravitational lensing systems have been presented. The current number of gravitational lenses identified is $\sim 1,000$ (Sonnenfeld et al., 2020), in total discovered to date. There is an increasing need for an automated method of finding gravitational lenses in large surveys, and through the use of ML this can be achieved. ML, especially CNNs, have been used in multiple areas of astronomy and in particular to find gravitational lensing systems in large sets of data.

Gravitational lensing systems are rare, due to a close linear alignment between the observer, lens and source needed, as well as the lens needing to have a large enough, so that the lensed source images may be viewed. Therefore, CNNs have been applied to large optical surveys to search for these rare objects (Shu et al., 2017). Lenses also appear differently due to their various morphologies as seen in Figure 1.4. CNNs are used to identify features pertaining to those searched for in lenses and whether or not an image contains a lens.

Sonnefeld et al. (2017) used three separate methods to identify lensing systems in a dataset containing $43,000$ luminous red galaxies (LRGs). These images are gathered from the Baryon Oscillation Spectroscopic Survey (BOSS). LRGs are large luminous early-type galaxies with very little activity in them, such as star formation (Hoshino et al., 2015). Their first method was to use a ML algorithm called YattaLens, designed to identify arc-like traits located around large mass galaxies, and to determine the possibility of these arcs being lensed galaxies. The YattaLens algorithm was evaluated against previously identified lenses, and iden-

tified 15/17 correctly. The second method was to use a model-based algorithm, called CHITAH. The third method was to look for lenses by investigating the spectroscopic emissions of objects in galaxies. Using all three methods, 51 highly probable lensing candidates were identified. Some arcs appeared to be in a close proximity to that of the foreground lens, and thus, the brightness of the source merged with that of the lens, complicating the ability of the CNN to differentiate between the two objects (Sonnenfeld et al., 2017).

In Sonnefeld et al. (2020), two methods were used to identify possible gravitational lenses in a set of $300,000$ galaxies from the Hyper Suprime-Cam survey. The first method was to use crowd sourcing, and the second was to use the YATTALENS algorithm. The method of crowd sourcing makes use of the assistance of a large group of people to gather, provide and/or analyse a large dataset. Both methods combined found 143 highly probable gravitational lenses, of which 70 have been identified in previous studies (Sonnenfeld et al., 2020). They also noted that using crowd sourcing to find lensing systems achieves a high quality sample of lensing candidates. Crowd sourcing methods may ensure that more lenses can be correctly identified. Using crowd sourcing in co-operation with the YATTALENS can ensure higher performances are achieved (Sonnenfeld et al., 2020).

Petrillo et al. (2019) used CNNs to detect LRGs with a low redshift of ($z < 0.4$) in the Kilo-Degree Survey (KiDS). Their CNN was trained using images containing simulated lenses and galaxies from KiDS. In a set of $88,327$ LRGs, 89 galaxies were detected as possible candidates for strong gravitational lenses. In an additional study, they used a sample of $930,651$ galaxies, and identified 30 candidates of strong gravitational lensing systems (Petrillo et al., 2019). They suggest that the hypothetical number of strong lenses in KiDs which follows their magnitude and low redshift criteria is 450. However, their CNN only retrieved $\sim 20\%$ of the hypothetical strong lensing systems. They stated that with more realistic simulations in the training data, and fine tuning of their CNN, their results could be improved (Petrillo

et al., 2019).

Metcalf et al. (2019) investigated the best method to classify a dataset of $100,000$ possible lensing candidates in order to find gravitational lenses. These methods included using either visual inspection, ring and arc finding algorithms, support vector machines or CNNs. The best CNN method found 22% of testing dataset correctly. The CNNs and support vector machines outperformed the humans as those systems were able to detect underlying differences within the images (Metcalf et al., 2019). When observing real objects such as Einstein rings, the CNN struggled to predict them correctly. They listed improving their results by adding more realistic images and to include more data of each image as well, including colours in multiple bands and various resolutions (Metcalf et al., 2019).

Huang et al. (2020) used a residual neural network (RNN) to detect strong lensing systems in the Dark Energy Camera Legacy Survey (DECaLS). A RNN is a neural network which takes the output of previous layers and adds it to the output of a later layer, ensuring that little information is lost whilst the data is processed (Chollet, 2018). Previous observations in surveys were used as training data to train the RNN. About 40% of their testing dataset were correctly identified by the RNN, and 335 new highly probable strong lensing candidates were detected in the Legacy Surveys. They stated that using a wider variety of lensing and non lensing images would teach the RNN to reject the large amount of images not containing a lens. They also state that using more resources such as more GPUs or a larger team, they would be able to train and test their RNN more effectively (Huang et al., 2020).

In Jacobs et al. (2019), CNNs were used to identify strong gravitational lensing systems within DES. Using the LensPop code (Collett, 2015), they created $250,000$ simulated lenses with a redshift range of $0.8 < z < 2$. Their CNN ran twice, to learn to reject any images

that does not contain visible lenses, and to learn the difference between objects that look like lenses but are not (eg. spiral galaxies). The accuracy scores obtained for the validation dataset, which is similar to the training data, is $98.6 \pm 0.1\%$ and $99.4 \pm 0.1\%$ for each run of the CNN. In a set of 1.1 million DES images, 84 images were detected to contain highly possible lensing systems. They stated that possible improvements include to expand the type of galaxies seen within the training set, and to include more realistic seeing in the simulations. Understanding the criteria that humans use to determine which images contain a lens can help in developing various tools to inspect these images. This can lead to a more developed CNN, which may be able to detect lenses not detected by humans by looking at the underlying features of the image in the same way a human would (Jacobs et al., 2019b).

Despite the vast amount of objects in large surveys, only around $\sim 1,000$ lenses have been identified so far (Sonnenfeld et al., 2020). The need for an accurate automated way of finding these systems has grown and the use of ML and neural networks seem to be the way forward. However, the training data is almost always simulated as there are not enough observed lenses available. This shows that the CNNs will always have the problem of learning features that do not correctly represent what we are looking for. Through the use of CNNs, we should be able to identify strong gravitational lenses within DES, given enough training data and variety.

The aim of this Masters project is to identify strong gravitational lensing systems in DES through the use of ML techniques. This project has two objectives to ensure that our aim is met: (1) create a ML algorithm in the form of a CNN and train it with a dataset containing both images from DES and images containing simulated lenses - we test the performance of the CNN against another dataset, similar to the training dataset but untouched during training and validation by the CNN; (2) test the performance of our CNN algorithm against real data, containing strong gravitational lensing systems that are presented in previous

studies and not used to train or validate the CNN.

In this thesis, gravitational lensing as a whole, has been presented in Section 1.1, as well as a brief history of this topic in Section 1.2. Artificial intelligence has been discussed as a whole in Section 1.3. A literature review has also been provided in Section 1.4. The methodology followed to create the relevant data and the CNN algorithm is described in Chapter 2. In Chapter 3, the results are presented, which are then discussed in Chapter 4, including the various tests performed and possible future improvements. In Chapter 5, this project is concluded and the limitations are discussed.

Chapter 2

Methodology

In this chapter, the application of ML techniques to identify strong gravitational lensing systems in DES is described. We use a CNN algorithm with a supervised learning classification method. The strong gravitational lenses we wish to detect appear as arclets in the g, r, and i-band images. To train the CNN, we use a combined set of images, which includes images from DES with little or no lenses visible and images containing simulated lenses where arcs are visible. Our CNN is tested against unseen images containing real lenses to determine how well it performs. These datasets are discussed below in Sections 2.1, 2.2 and 2.3. The reader can also refer to the Appendix A, for brief explanations on these datasets.

In Figure 2.1, the work flow diagram shows the process followed to retrieve appropriate images to form the different datasets in order to train and test the CNN algorithm. All data is created following a similar method: creating the data in the form of 100×100 pixel g, r, and i-band FITS images, and ensuring that Gaussian normalization is applied to the images. Gaussian normalisation is applied with the formula of $\frac{X - \mu}{\sigma}$ where X is the image data, μ is the mean of the image and σ is the standard deviation of the image. Ensuring that all the images have followed a similar setup is necessary to maintain similarity between the

different datasets, and to minimise the number of unnecessary features learnt by the CNN algorithm. This will in turn improve its performance. For viewing and clarification purposes red, green and blue (RGB) composite images of the g, r and i-band images are created. Here the g-band magnitude corresponds to the colour blue, the r-band magnitude corresponds to green, and the i-band magnitude corresponds to red. An example of these RGB images for each dataset is seen in Figures 2.3, 2.6 and 2.7.

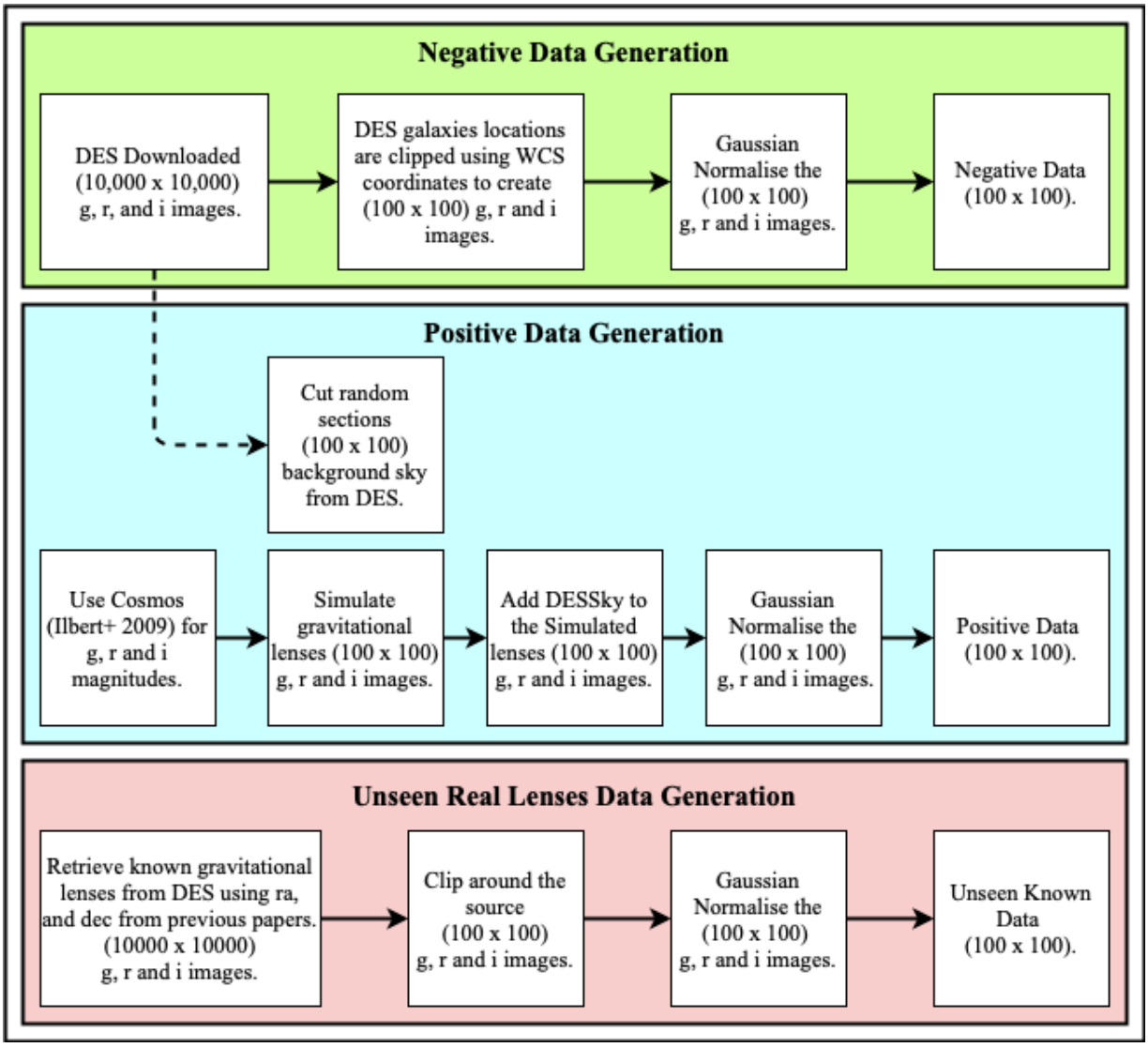

Figure 2.1: A work flow diagram showing the process of retrieving the images for the negative, positive simulated and unseen real lensing systems datasets. All three datasets undergo a similar process of retrieval to maintain as much similarity amongst all data as possible. The general process is to download or create images, clip them to a size of 100×100 pixels, and apply a Gaussian normalisation.

We collect our data from the DES catalogue for our different datasets. DES covers an area of $5,000$ square degrees in the southern galactic hemisphere, from which a large amount of galaxies ($\sim$ 300 million) is observed (Abbott et al., 2018). DES is an optical wavelength

survey designed to observe and collect data of galaxies and in turn enabling the measurements of dark energy. DES has an average limiting magnitude for the different wavelength bands of $g = 24.28$, $r = 23.95$ and $i = 23.34$.

2.1 Negative Data

We downloaded $10,000 \times 10,000$ pixel (0.7306×0.7306 degrees) tile images in the g, r and i-bands from DES Data Release 1 (DESDR1) [1], as seen in the first block of the *Negative Data Generation* in Figure 2.1. The relation between pixels and degrees is 1 pixel = 0.263" (Abbott et al., 2018). We subsequently clipped random sections from these images to provide realistic backgrounds for the simulated lensed systems (see Section 2.2), to provide a negative dataset to train the CNN.

These g, r and i-band FITS images are clipped using the World Coordinate System (WCS) around galaxies to have dimensions of 100×100 pixels (26.33×26.33 arcseconds) as seen in the second block of the *Negative Data Generation*. To generate the negative images, we applied the colour-magnitude cuts of $18 < i < 22$ and $g < 24$ to these images, ensuring that all images of galaxies from DESDR1 match the brightness of the simulated gravitational lenses. An example of these WCS clipped images can be seen in the first row of Figure 2.2. These g, r and i-band clipped images undergo a Gaussian normalisation (third block of the *Negative Data Generation*). An example of these normalised images can be seen in the second row in Figure 2.2. These normalised images construct the dataset collectively known as the negative data, containing a set of galaxies with no known gravitational lenses detected. Nine random RGB composite images of these negative images can be seen in Figure 2.3.

[1] https://des.ncsa.illinois.edu/releases/dr1

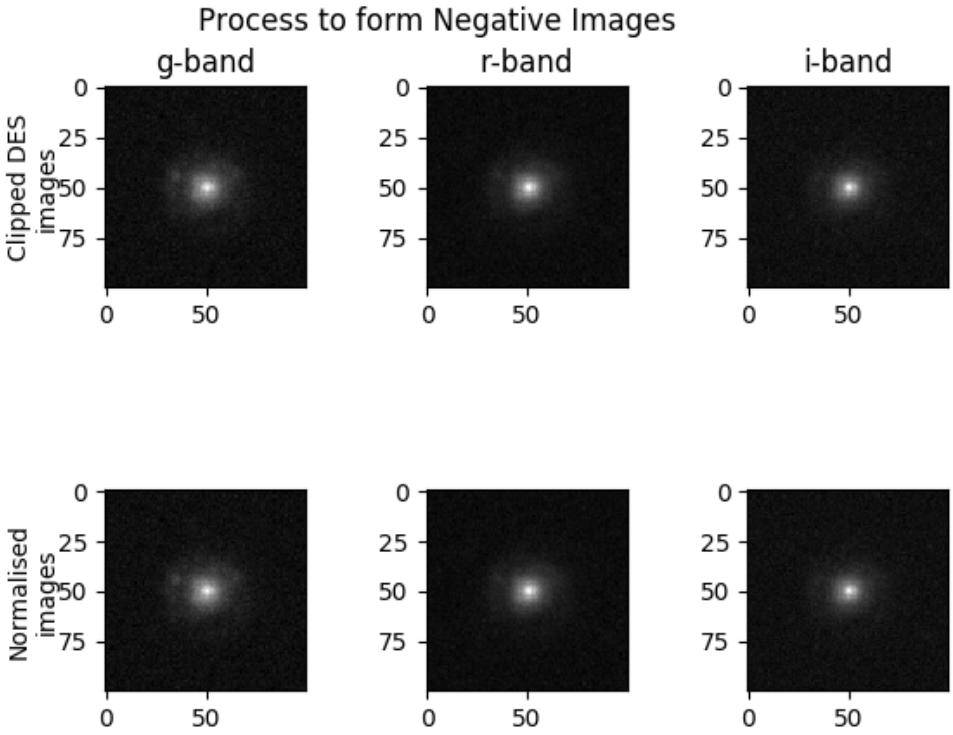

Figure 2.2: The process of images for the negative data, consisting of DES data without gravitational lenses. The first row shows the clipped WCS images, the second row shows these images Gaussian normalised. Each column indicates the g, r and i-bands respectively.

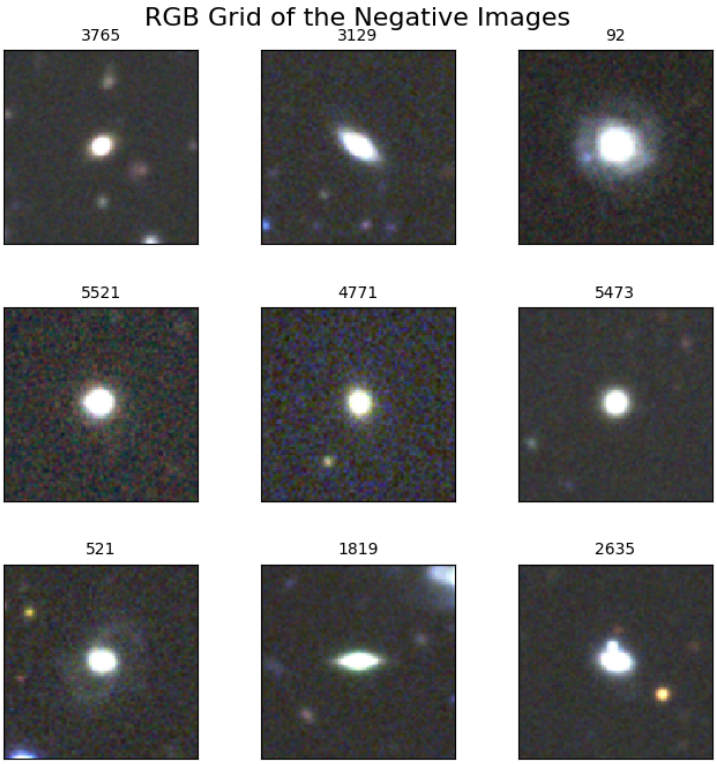

Figure 2.3: RGB composite images of nine random sources of the negative normalised images. These images contain various objects from DES. Their corresponding object ID numbers, that we provided for reference purposes, is seen above each image.

We clipped 100×100 pixel sections that were randomly chosen from the original $10,000 \times 10,000$ pixels images from DESDR1, and this is referred to as the DESSky images. These DESSky images are saved separately from the negative data. These images are added to the positive simulated images, in order to make the positive simulated gravitational lenses more realistic with the addition of real background noise, as discussed in Section 2.2. An example of these background images can be seen in Figure 2.4.

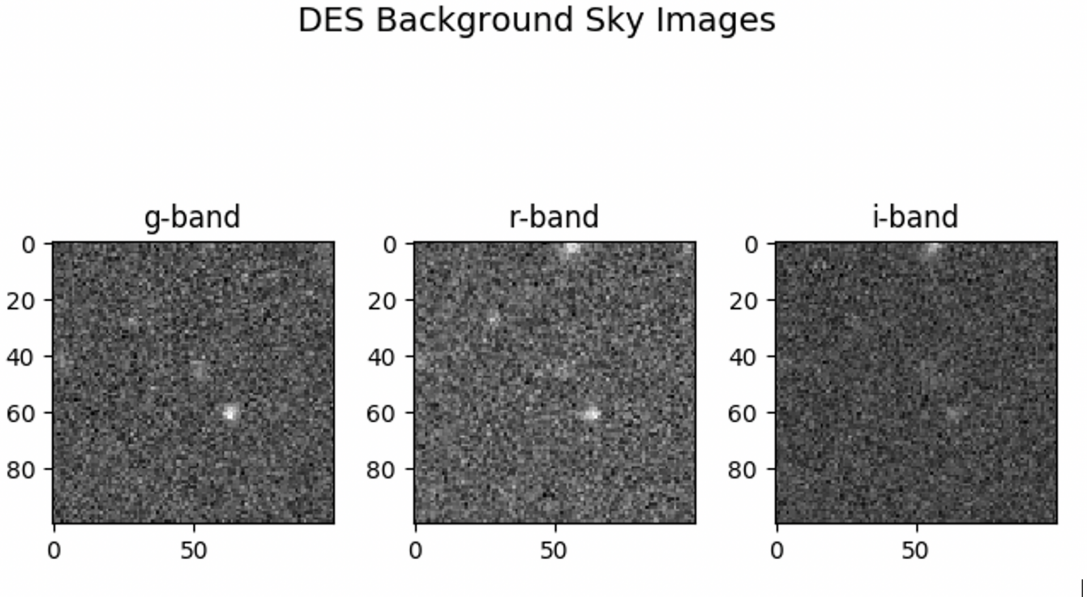

Figure 2.4: The clipped images of 100×100 random sections of the original DES images in g, r and i-bands creating our background noise for the positive images.

2.2 Positive Data

The positive dataset consists of simulated strong gravitational lenses and is used to train our CNN. The g, r and i-band magnitudes are set using information from the Cosmic Evolution Survey (COSMOS) photometric redshift catalogue (Ilbert+, 2009)[2] (Ilbert et al., 2008) as seen in the first block of the *Positive Data Generation* in Figure 2.1, ensuring that the colours of the simulated lenses are realistic. COSMOS is a sky survey with a two square degree field with an i-band magnitude limit of $i < 27.8$, and is located on the celestial equator to include observations from ground-based facilities covering various wavelengths. We set an upper limit of the r-band magnitude to $r < 22$, so that the sources are detectable in DES. Sources and lenses are separated in terms of redshift to create realistic parameters similar to that of real lenses and sources. The redshift for the sources is limited to $1 < z_{sources} < 2$, and

[2]http://cosmos.astro.caltech.edu/page/photoz

for the lenses $0.1 < z_{lens} < 0.3$. The simulated lenses have an i-band magnitude range of $18 < i < 22$. These redshift ranges correspond to that of the galaxy lensing sample obtained in the Strong Lensing Legacy Survey (Sonnenfeld et al., 2013).

To create simulated lensing systems in the g, r and i-bands, we used the LensPop code (Collett, 2015). We changed the magnitudes of the simulated lenses to include the low redshift. These images are defined within this project, as positive noiseless images since they are completely smooth simulated lenses without background noise, with a size of 100×100 pixels (second block in *Positive Data Generation*). An example of these g, r and i-band noiseless images is shown in the first row of Figure 2.5.

The third block in the *Positive Data Generation* shows that in order for these smooth images to be realistic, background noise is added (DESSky images), as discussed in Section 2.1. An example of the DESSky images added to the simulated lenses, can be seen in the second row of Figure 2.5. Now these simulated images undergo a Gaussian normalisation (fourth block in *Positive Data Generation*), maximising the algorithm's ability to learn to detect the lensing arcs. An example of these positive normalised images is seen in the last row of Figure 2.5. These normalised images make up the positive simulated images dataset. Nine RGB composite images of these positive simulated gravitational lenses can be seen in Figure 2.6.

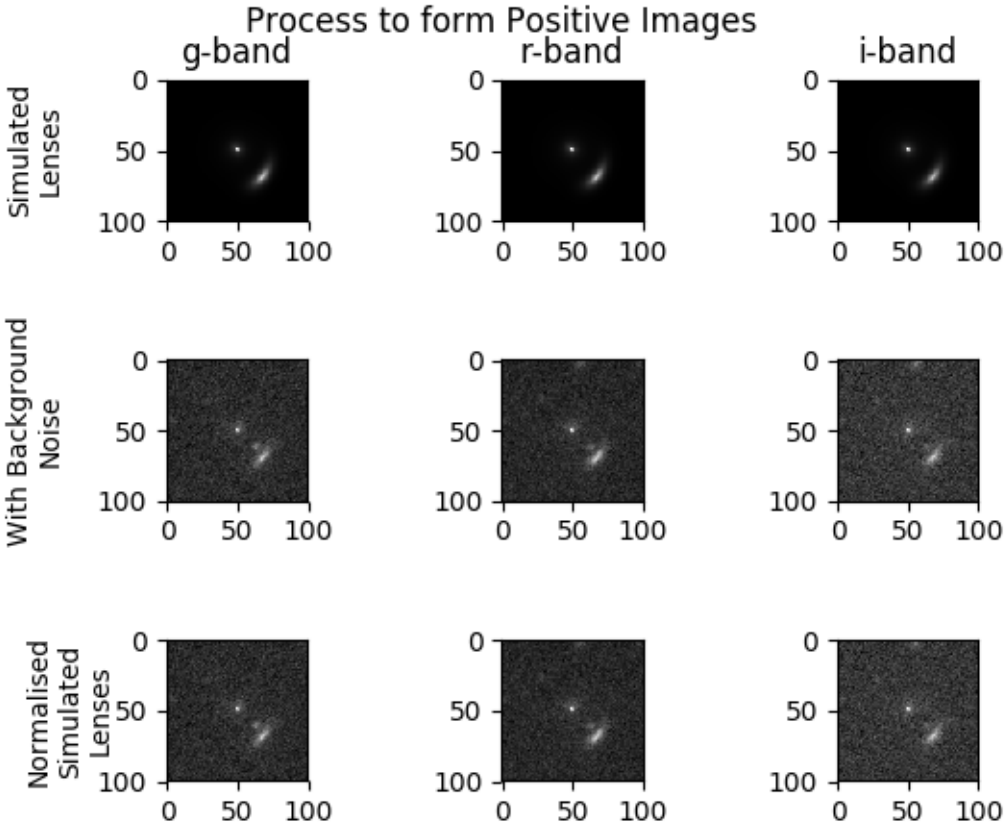

Figure 2.5: The process of images for the positive simulated data. The first row shows the raw simulated data, the second row shows the simulated data combined with the clipped DESSky (Figure 2.4), and the third row shows the positive simulated images with the background sky normalised. Each column indicates the respective g, r and i-bands.

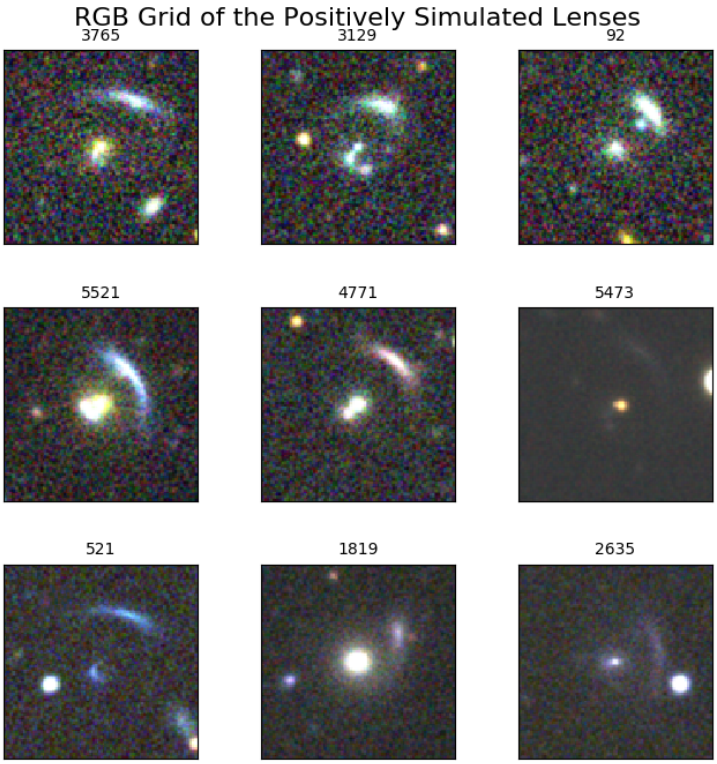

Figure 2.6: RGB composite images of nine random sources of the positive simulated normalised images. These images contain bright simulated arcs around the lens. Each lensing system has been given an object ID number for reference purposes, and this can be seen above the images.

2.3 Unseen Data

The unseen data is used to test the performance of the CNN algorithm and is not used during training or validation. There are two different types of unseen data: the unseen testing data and the unseen real data. The first type, the unseen testing dataset, consists of positive simulated images and negative images from the respective manufacturing processes of those seen in Sections 2.1 and 2.2. This is purely used to test how well the CNN model performs on data that it has not seen before.

The second type is the unseen real lenses dataset and the process of retrieving this is described in Figure 2.1, under the *Unseen Real Lenses Data Generation*. This data includes 389 images of gravitational lensing systems retrieved from various papers including the DES Bright Arcs Survey paper (Diehl et al., 2017), the CFHTLS legacy survey (More et al., 2012), the CASSOWARY survey (Stark et al., 2013), papers studying the lensed objects in the Hyper Suprime-Cam survey (Wong et al., 2018; Jaelani et al., 2020), a catalogue of strong lensing systems from the COSMOS survey (Faure et al., 2008), and a catalogue of high redshift ($z > 0.8$) strong lenses identified in DES (Jacobs et al., 2019b). We downloaded the $10,000 \times 10,000$ DES tile images containing these lensed systems, and created 100×100 pixel postage stamp images centred on these lensed systems. These images undergo a Gaussian normalisation maintaining a likeness between the positive, negative and unseen images. RGB composite images of nine random unseen real gravitational lenses are seen in Figure 2.7. The necessity of all images to be from DES and to undergo a Gaussian normalisation ensures that all the images have the same noise properties, angular resolution and seeing, limiting the learning of unnecessary details instead of the strong lensing arc properties.

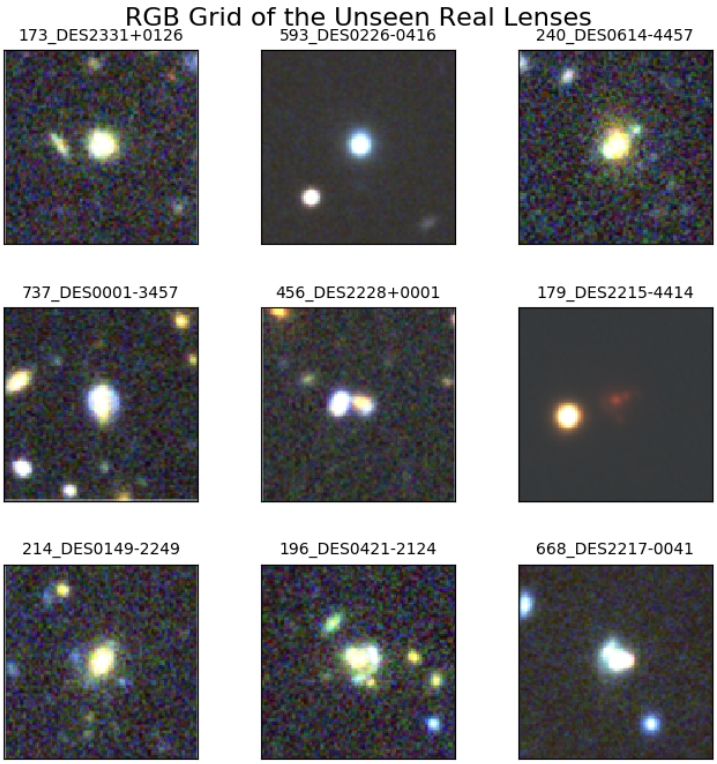

Figure 2.7: RGB composite images of nine random unseen real lenses. Their object ID numbers (for reference purposes) and system names from DES, are given above each image.

2.4 Training Convolutional Neural Networks

In order for strong gravitational lenses to be identified, we use a CNN algorithm with a *Keras*[3] *TensorFlow* [4] implementation. The data consists of FITS images in the g, r and i-bands for the negative, positive and unseen datasets as discussed in Sections 2.1, 2.2 and 2.3. The data is labelled binarily, where images containing gravitational lenses are labelled as 1, and those without lenses as 0. The dataset consists of $24,192$ low redshift simulated lenses (positive data) and DES images (negative data) each, and is multiplied by a factor of

[3]https://keras.io
[4]https://www.tensorflow.org

two through the use of data augmentation, to create a total of 48,384 each. Therefore, the total size of the dataset is 96,768, containing a shuffled combination of positive and negative datasets. Refer to the Appendix A, to be reminded of the various datasets used during our CNN.

The data is multiplied through the use of an augmentation algorithm in which a copy of the original images is created. These transformations are done using the *ImageDataGenerator* module from the *Keras* library, and includes a feature-wise centering, a feature-wise standard normalisation, a rotation from 0 to 90 degrees, width and height shifts, as well as horizontal and vertical flips. This augmented dataset is added to the original images, thereby multiplying the original amount and expanding the variety of images, when only using a small amount of data to begin with. This data augmentation allows for the lenses to be observed in various ways. The original data and augmented data together are referred to as the total dataset.

This total dataset is split into 80% (77,414) for the training data and 20% (19,355) for the validation data using a 0.2 train-validation split. The CNN algorithm is trained using the training data, leveraging the different layers seen in Figure 2.8, where the architecture of our CNN is seen. We use the validation data to determine the CNN's performance. The CNN is then evaluated against the unseen testing dataset to determine how well our CNN learns. Our CNN is also evaluated against the unseen real gravitational lenses in order to determine how many lenses are correctly identified. These unseen datasets have been discussed in Section 2.3.

The CNN architecture was built from the ground up, using other CNNs for guidelines. After seeing that our CNN was not performing optimally, we narrowed down the ineffectiveness by looking at CNNs of previous projects and applying similar layers. These layers

were either adjusted, removed or added, using substantial experimentation to gain a more optimally performing CNN when identifying strong gravitational lensing systems. One CNN used as a guideline is the CNN in Jacobs et al. (2019). Our resulting CNN contained 14 hidden layers whereas theirs contained 21 hidden layers. We also have a larger amount of trainable weights of $55,746,689$ compared to their $8,833,794$ (Jacobs et al., 2019b).

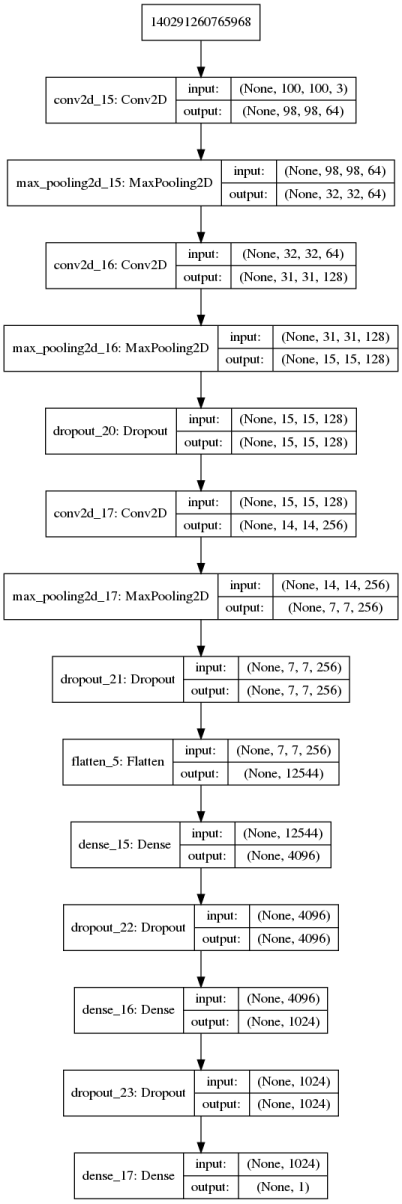

Figure 2.8: This is the architecture of the sequential CNN model used in this project, where the first layer contains the input with images of size $(100 \times 100 \times 3)$, and the last layer is a Dense Layer of size 1. This architecture has $55,746,689$ trainable weights.

The CNN is trained with the architecture seen in Figure 2.8, with $55,746,689$ trainable weights through the use of a sequential model with a *TensorFlow* back-end. We use the sequential model, as there is one input that is processed throughout the hidden layers in the

CNN resulting in one expected output (Chollet, 2018), which is either that a gravitational lens has been detected or not. The CNN model has multiple hidden layers, each with input and output tensors to increase the performance of the algorithm. The various hidden layers add complexity to the neural network and in turn improves its accuracy. The hidden layers of the CNN consists of convolutional, max-pooling, dropout, flatten and dense layers. The convolutional layer creates a convolution of the given input to produce a tensor of the output data (Chollet et al., 2015). The convolution operation applies a multiplication product between an array of weights (called a filter) and the input from the previous layer, resulting in a single value. If done enough times, a two-dimensional array creates a feature map from which certain identifiable features can be detected (Brownlee, 2019a).

The max-pooling layers ensures faster convergence and is able to identify certain features, improving the generalization of the CNN algorithm (Ciresan et al., 2011). Max-pooling creates a smaller dataset of the given input, by taking the maximum value within a dimensionally defined grid of pixels (Chollet et al., 2015). The dropout layer is used to limit over-fitting in the early stages of training through the use of regularization and generalization (Moolayil, 2018; Pourrahmani et al., 2018). The dropout layer ensures that a random number of features in a given input are removed which are of no importance in the CNN (Chollet, 2018; Bell, 2020). The flatten layers are used to transform the pixels from the multi-dimension input to a single vector, in order to insert this data into a fully connected layer preceded by the CNN model (Chollet et al., 2015; Jeong, 2019). The dense layer ensures that all the neurons in a previous layer are connected to all the neurons in the next layer (Moolayil, 2018; Bell, 2020). In the final dense layer, all the neurons are arranged to connect to only one neuron, allowing the use of a binary classification of 0 and 1 with a sigmoid activation, and the output is, therefore, a probability between 0 and 1 (Chollet, 2018). The sigmoid function provides a boolean classification of the data, either true or false, or in our case 0 or 1 (Moolayil, 2018). If the CNN model identifies a lens in the given image, the sigmoid activation function produces

a 1, otherwise it will produce a 0. The CNN model is optimised with an *Adam optimiser* with a learning rate of 0.0002. The *Adam optimiser* is a stochastic optimization used in ML applications and is effective when large datasets are used (Kingma and Ba, 2014). The learning rate of the *Adam optimiser* shows how fast the learning occurs during training of the CNN model, with a small value indicating the CNN model learns slowly (Brownlee, 2017).

In order to increase the model's performance, we use a few *Keras* callbacks, including the model checkpoint, and early stopping. Callbacks are functions with access to the modelled data and its performance, and are able to interrupt the training, save the model and its weights, and is able to load the different weights or change the model as needed (Chollet, 2018). The model checkpoint callback is used to save the weights throughout the training of our CNN model (Chollet, 2018). The early stopping callback discontinues the training when the validation loss no longer improves through the use of the patience variable, which indicates when the training is to be interrupted (Chollet, 2018).

The k-fold cross validation is used to gain an overall accuracy rate of the CNN model. The k-fold cross validation is performed by splitting the given dataset into k compartments, where one is the validation set, and the rest is the training set. These sets undergo training $k-1$ times, and an average is gained of the accuracy scores, where each validation set is a new k compartment, as shown in Figure 2.9 (Mitchell, 1997; Chollet, 2018). Using this method of k-fold cross validation, one can calculate the error of the algorithm by calculating the standard deviation across the folds (James et al., 2013). This will ensure that the performance of the CNN is accurately measured.

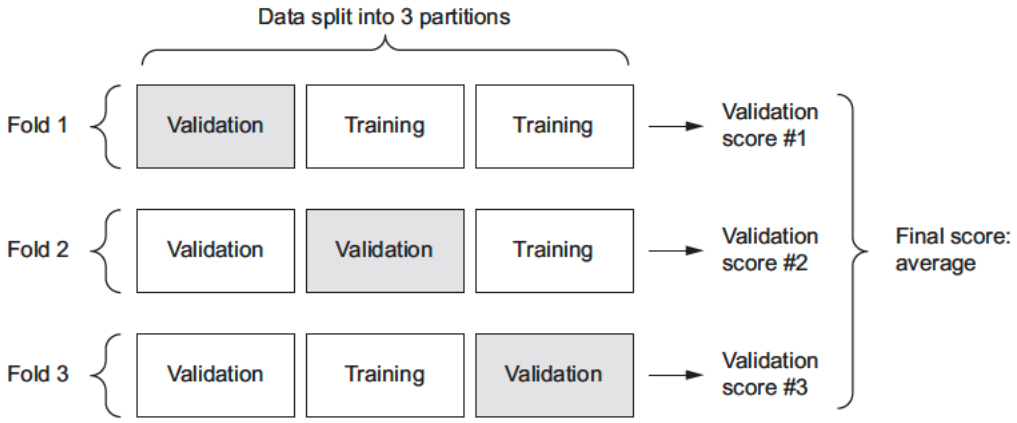

Figure 2.9: A diagram showing how k-fold cross validation works and how the validation and training are split for each fold. The CNN is trained and validated a k number of times, each time a different fold is used. In each fold a different validation dataset is chosen, to those which were previously used, and the training data includes previous validation datasets (Chollet, 2018).

Chapter 3

Results

The aim of this project is to identify strong gravitational lensing systems within DES through the use of ML. To train the CNN algorithm, negative and positive simulated data is used, as discussed in Chapter 2. The ML technique of CNN is used, with a *Keras Tensorflow* implementation. The CNN algorithm includes the various hidden layers as discussed in Section 2.4, with its architecture shown in Figure 2.8. For a clear description of the datasets used, please refer to the descriptions in the Appendix A.

The CNN is setup with a batch size of 128, 50 epochs, an early stopping on the validation loss scores, with a patience of 10, and all data is shuffled before training. The epochs indicate the amount of times the entire training dataset is passed through the CNN algorithm. After each iteration, the CNN is able to compute the respective weights and update them, ensuring that it is able to generalise well, receiving lower loss or error values (Chollet, 2018). The data consists of $24,192$ negative and $24,192$ positive simulated images, and a data augmentation multiplier is applied to increase this amount by a factor of 2, forming a total dataset of $96,772$. This dataset undergoes a train-test split of 0.2, resulting in $77,417$ images used as training data and $19,355$ images used as validation data. The unseen testing

data consists of $6,034$ negative images and $6,034$ positive simulated images, untouched by the CNN algorithm during training and validation. The unseen real lenses, which contain previously identified lenses, consists of 389 images. The unseen real lenses is used to evaluate the CNN's ability at identifying real lenses. The CNN was trained with the architecture in Figure 2.8, and a k-fold cross validation was applied, where $k = 5$, to get a more accurate representation of the results, as described in Section 2.4.

The training and learning of the CNN model is analysed through graphs showing the loss and accuracy of the training data and validation data. The training and validation loss and accuracy graphs indicate visually how well the model performs. Through these graphs, it is possible to deduce how well the model is able to learn from a given dataset and generalise from a given test set. The loss function determines how well the CNN is able to correctly predict an object, compared to what the object actually is. The loss function is comprised of scores that show how correct the CNN is. The lower the loss score, the better the CNN is at correctly predicting the given objects (Chollet, 2018). The accuracy metric shows how often the CNN is able to correctly predict an object, and is defined as the amount of objects correctly predicted divided by the total amount of objects as seen in the equation (3.1).

$$Accuracy = \frac{Amounts\ of\ Objects\ Correctly\ Predicted}{Total\ Amount\ of\ Objects\ to\ Predict} \tag{3.1}$$

The loss and accuracy metrics indicates whether or not there is possible over-fitting, under-fitting or a good-fitting occurring during training (Chollet, 2018). An example of these graphs is seen in Figure 3.1. Over-fitting can be identified when the training loss decreases, but the validation decreases until it reaches a certain point and then starts to increase again (Brownlee, 2019b). Under-fitting occurs when the model is unable to decrease the

training error (loss) (Zhang et al., 2020). This can be identified when the training loss graph remains unchanged or if it decreases until the training has stopped (Brownlee, 2019b). A good-fit occurs when the error between the training and validation data is minimal and can be identified when both the training and validation loss decreases to a certain point and starts to flatten out (Brownlee, 2019b).

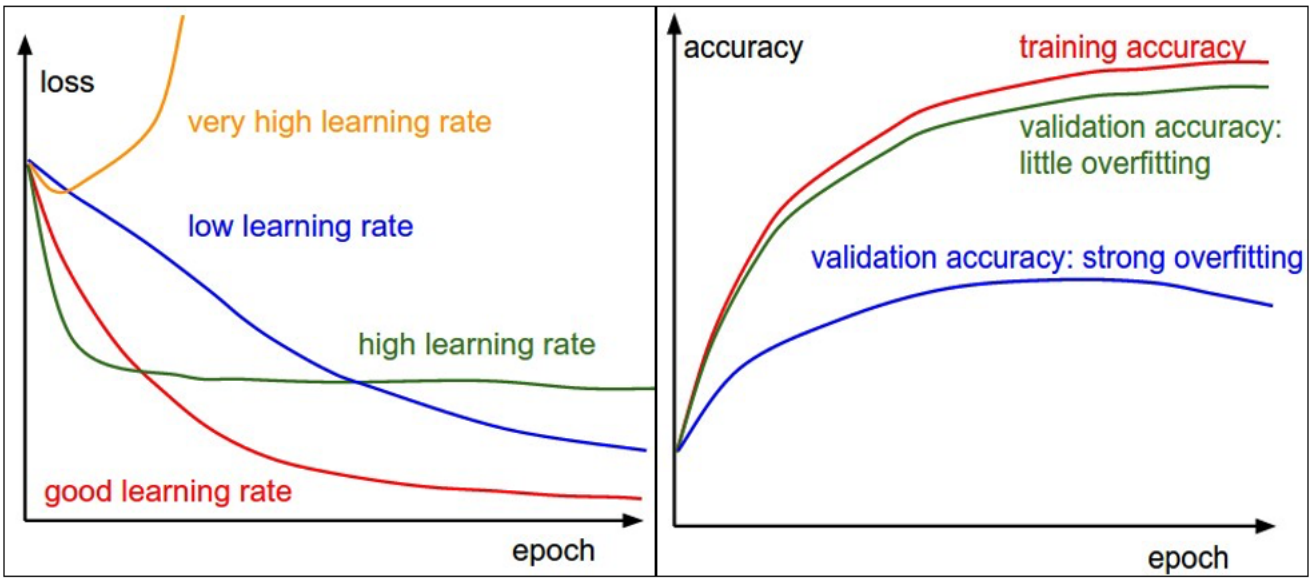

Figure 3.1: Interpretation of the loss (*left*) and accuracy (*right*) graphs for the training and validation data as a CNN algorithm learns and trains. A good loss rate (*red*), shows a rapid decrease in error and then flattens out as training continues. A good accuracy rate (*green*) shows a rapid increase in accurate prediction and a small difference is seen between both the training and validation accuracies (Karpathy, 2020).

When evaluating our CNN with the training data, an accuracy and loss score of $97.15 \pm 2.81\%$ and 0.07 was obtained respectively. When evaluating with the validation data, an accuracy and loss score of $97.03 \pm 1.55\%$ and 0.08 was obtained respectively. An accuracy and loss is obtained, of $99.73 \pm 0.07\%$ and 0.81 respectively, when evaluating the CNN against the

unseen testing data, which shows that the CNN is able to distinguish between the positive and negative images. When evaluating against the unseen real lenses, an accuracy and loss of $11.92 \pm 2.75\%$ and 10.49 was obtained. The loss and accuracy graphs of the training and validation data during the training of the CNN model is seen in Figure 3.2.

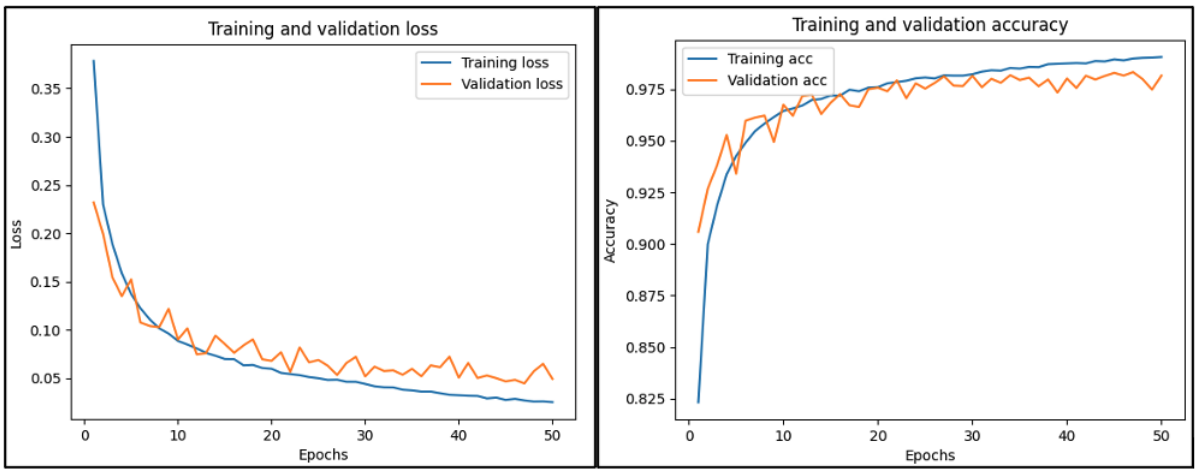

Figure 3.2: Graphs showing the loss (*left*) and accuracy (*right*) as epochs increase for the training (*blue*) and validation(*orange*) datasets.

The training and validation loss and accuracy graphs are shown in Figure 3.2. The training loss and accuracy curves show how well the CNN is learning from the given training dataset with every epoch. The validation loss and accuracy curves show how well our CNN generalises and performs with the given validation set with every epoch (Brownlee, 2019b). As seen in Figure 3.2, the training loss decreases with more epochs. This shows that with each epoch that the training data is passed through, the error decrease and the CNN is learning to generalise from the training data. The validation loss also decreases with each epoch, showing that the CNN is able to predict the images more and more accurately with each epoch. Both loss and accuracy graphs have a small gap between them at the end of

training, demonstrating that a good fit is being reached.

The amount of sources that the CNN has predicted correctly or incorrectly against their true values, can be determined using a confusion matrix as the one seen in the following matrix,

$$\begin{bmatrix} & & \textbf{Predicted} & \textbf{Outcome} \\ & & 0 & 1 \\ \textbf{Actual} & 0 & \textit{True Negative } (TN) & \textit{False Positive } (FP) \\ \textbf{Outcome} & 1 & \textit{False Negative } (FN) & \textit{True Positive } (TP) \end{bmatrix}.$$

True negative (TN) represents the amount of sources the CNN correctly predicted as negative. False positive (FP) represents the amount of sources incorrectly predicted as positive. False negative (FN) represents the amount of sources that were incorrectly predicted as negative. True positive (TP) represents the amount of sources correctly predicted as positive. In the matrices the negative connotation corresponds to the label 0, representing no strong gravitational lenses identified in the image, and the positive connotation corresponds to 1, representing the identification of a strong gravitational lens within the image. The following confusion matrices shows the amount of lenses correctly and incorrectly identified by the CNN, in the unseen testing dataset (*left*) and the unseen real dataset (*right*),

$$\begin{bmatrix} TN = 6,013 & FP = 21 \\ FN = 0 & TP = 6,034 \end{bmatrix} \qquad \begin{bmatrix} TN = 0 & FP = 0 \\ FN = 355 & TP = 34 \end{bmatrix}.$$

The left matrix shows that the CNN correctly predicted the $\frac{6,013}{6,034}$ and $\frac{6,034}{6,034}$ of the unseen testing data as negative data and positive simulated lenses respectively. The right matrix shows that the CNN correctly predicted $\frac{34}{389}$ and incorrectly predicted $\frac{355}{389}$ of the real unseen lenses.

These confusion matrices, combined with a consistent training and validation loss graph (Figure 3.2), we came to the conclusion that the simulated data contains identifiable features which differentiates themselves from the negative training data, but does not accurately represent all possible lensing morphologies. This leads to the CNN being substantially less effective when analysing the unseen real lenses than the simulated lenses. After investigating the data thoroughly, it is possible to state that the unseen real lensing data does not appear to be similar to the positive simulated lenses, as seen in Figure 3.3.

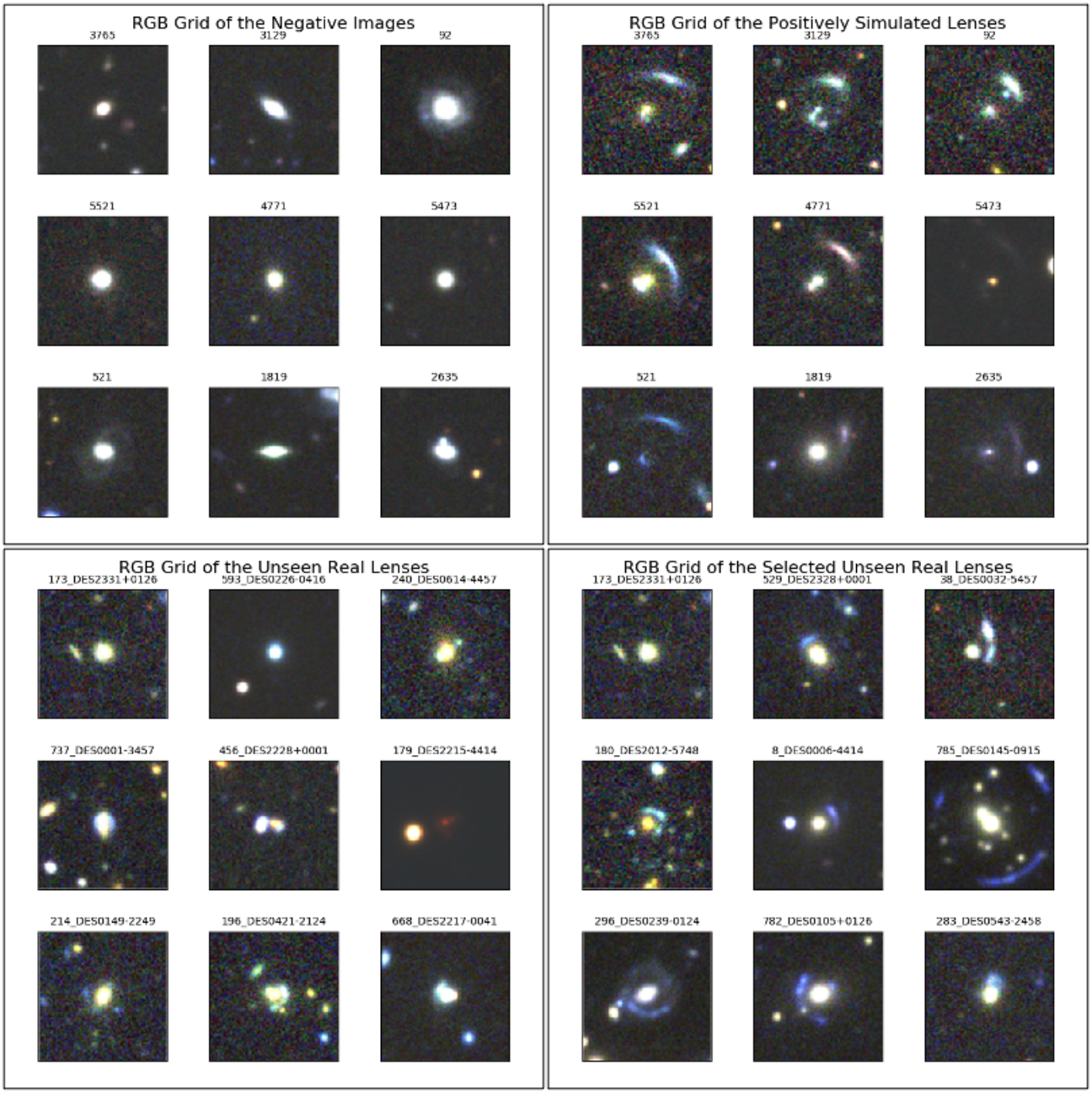

Figure 3.3: Samples of nine random RGB composite images of the negative data (*top left*), positive data (*top right*), unseen real lenses data (*bottom left*), and the selected unseen real lenses data (*bottom right*).

Looking at Figure 3.3, we are able to see that the data containing previously identified gravitational lensing systems (*bottom left*), contains multiple morphologies of lensing systems

and not just the arclets which we see in our positive simulated images (*top right*). Therefore, the CNN identifies these unseen real lenses as negative images, since the multiple lensed source images (*bottom left*) appear as point-like sources and not the bright arcs expected in the positive images (*top right*). To overcome this difference in training and real lensing datasets, 83 images were visually selected of the 389 unseen real lensing images that contain bright arc-like morphologies similar to those seen in the positive simulated images. A sample of nine random RGB composite images of the selected unseen real lenses can be seen in the *bottom right* of Figure 3.3.

Using the 83 selected unseen real lenses, we tested the CNNs performance again, obtaining the mean accuracy and loss values for the selected data, $15.42 \pm 5.67\%$ and 8.95 respectively. After evaluating the CNN model against the selected unseen real lensing data, the following confusion matrix was obtained,

$$\begin{bmatrix} TN = 0 & FP = 0 \\ FN = 74 & TP = 9 \end{bmatrix}.$$

From this one can seen that the CNN model correctly predicts 9lenses and incorrectly predicts 74 lenses, out of the 83 selected unseen real lenses. It appears that the CNN is not very effective at identifying the difference between images containing lenses and those that do not. The results of our CNN's performance against the unseen testing data, unseen real lenses and the selected unseen real lenses are summarised in Table 3.1.

Table 3.1: Table showing the results obtained from our CNN for the unseen testing dataset, the unseen real lenses dataset and the unseen selected lenses dataset.

Dataset	**Accuracy**	**Loss**	**True Negative**	**False Positive**	**False Negative**	**True Positive**
Unseen Testing	$99.73 \pm 0.07\%$	0.81	$6,013$	21	0	$6,034$
Unseen Real	$11.92 \pm 2.75\%$	10.49	0	0	355	34
Unseen Selected	$15.42 \pm 5.67\%$	8.95	0	0	74	9

We ran our CNN an additional four times, to ensure that there is continuity of the observed results. The respective loss and accuracies of each of these runs are presented in Table 3.2, with their respective graphs in Figures 3.4 and 3.5. These loss and accuracies show the CNNs performance against the unseen testing data, the unseen real lenses and selected lenses, across all the runs. The confusion matrices of each run including ours, is seen in Table 3.3, shows the performance of the CNN evaluated against the unseen real lensing data.

Table 3.2: Table showing the mean loss and accuracy scores obtained for the different datasets, during each of the five runs of our CNN.

Run	**Training Accuracy** (%)	**Training Loss**	**Validation Accuracy** (%)	**Validation Loss**	**Unseen Test Accuracy** (%)	**Unseen Test Loss**	**Unseen Real Lenses Accuracy** (%)	**Unseen Real Lenses Loss**	**Selected Unseen Real Lenses Accuracy** (%)	**Selected Unseen Real Lenses Loss**
Our CNN	97.15 ± 2.81	0.07	97.03 ± 1.55	0.08	99.73 ± 0.07	0.81	11.92 ± 2.75	10.49	15.42 ± 5.67	8.95
Run 1	97.12 ± 2.85	0.07	97.08 ± 1.71	0.08	99.76 ± 0.06	0.01	9.92 ± 2.27	11.50	10.60 ± 3.99	10.37
Run 2	97.07 ± 2.85	0.07	96.81 ± 2.00	0.08	99.73 ± 0.03	0.01	11.05 ± 2.89	11.23	9.64 ± 5.11	10.39
Run 3	97.19 ± 2.85	0.07	96.98 ± 1.65	0.08	99.80 ± 0.04	0.01	8.95 ± 2.13	11.31	7.47 ± 2.89	10.34
Run 4	97.24 ± 2.81	0.07	97.00 ± 1.69	0.08	99.76 ± 0.05	0.01	8.59 ± 1.35	11.43	8.92 ± 3.10	10.42

Table 3.3: Table showing the confusion matrices of our CNN and the additional 4 runs, when evaluating the unseen real lenses dataset

Run	**True Negative**	**False Positive**	**False Negative**	**True Positive**
Our CNN Run	0	0	352	37
Run 1	0	0	363	26
Run 2	0	0	362	27
Run 3	0	0	358	31
Run 4	0	0	356	33

A cumulative list was created of lenses that were correctly predicted by the CNN over the multiple runs, when evaluating against the unseen real lenses dataset, resulting in a total of 57 lenses. There were lenses that were correctly predicted by the CNN in all five runs, and these amounted to 18 lenses. The histograms of the 57 identified lenses in the g, r and i-band magnitudes can be seen in Figures C.1, C.2 and C.3 respectively in the Appendix. The images of these lensing systems are 100×100 pixels in size as described in Section 2.3. Looking at these histograms, we are able to deduce the magnitudes of the strong gravitational lenses identified by our CNN. In Figures C.1, C.2, and C.3, the identified lenses have magnitudes in the ranges of $17 < g < 24$, $16 < r < 24$, and $16 < i < 24$, this is in accordance with the magnitude cuts made for the positive data (Section 2.2), with a few identified lenses being fainter.

To ensure that the CNN architecture considered in this thesis is not overfitting, we perform standard 5-fold cross-validation (with random re-initialisation of initial weights of the CNN and different random seeds) and plotted the loss and accuracy dependencies for training and validation data (Figure 3.2). The obtained training/validation plots are typical for

the optimal training scenario (Figure 3.1 vs Figure 3.2). The results of the 5-fold cross-validation also produce consistent accuracies for the training and validation data (Table 3.2 and Figures 3.4 - 3.5) and all five CNNs identified the same 18 unseen lenses (from the unseen real lenses dataset). These indications clearly show that our CNN can correctly identify lenses from our dataset, which was created for the very narrow class of strong gravitational lenses with bright arcs. However, because of the absence of training data containing other types of gravitational lenses, the CNN performs poorly on the whole dataset of the unseen gravitational lenses.

For the unseen real lenses dataset, containing a set of known lenses, we extracted postage stamp images (100×100 pixels) at the coordinates retrieved from the various respective catalogs presented in Section 2.3. Note that some catalogues record the positions of multiple lensed source images rather than the lens itself. As such, in a few cases the postage stamps may not be centred on the lens. Thus, a sub-optimal performance of the CNN is detected, as it is trained on images centred on the lens galaxies. We have removed lensing systems that were duplicated in our unseen real lenses dataset to form a list of $\frac{252}{389}$ unique lensing systems. These duplicated lenses are a result of the same systems appearing in multiple catalogues or multiple images of a particular lens system being recorded. Without duplications, a total of $\frac{42}{252}$ (16.67%) lenses were correctly predicted across all five of our CNN runs, and the same $\frac{14}{42}$ (33.33%) were observed during every run. These 42 lens systems are presented in the Table B.1. The amount of times each lens system has been correctly predicted across all five runs (Our CNN run + four additional runs) is seen in the last column of Table B.1.

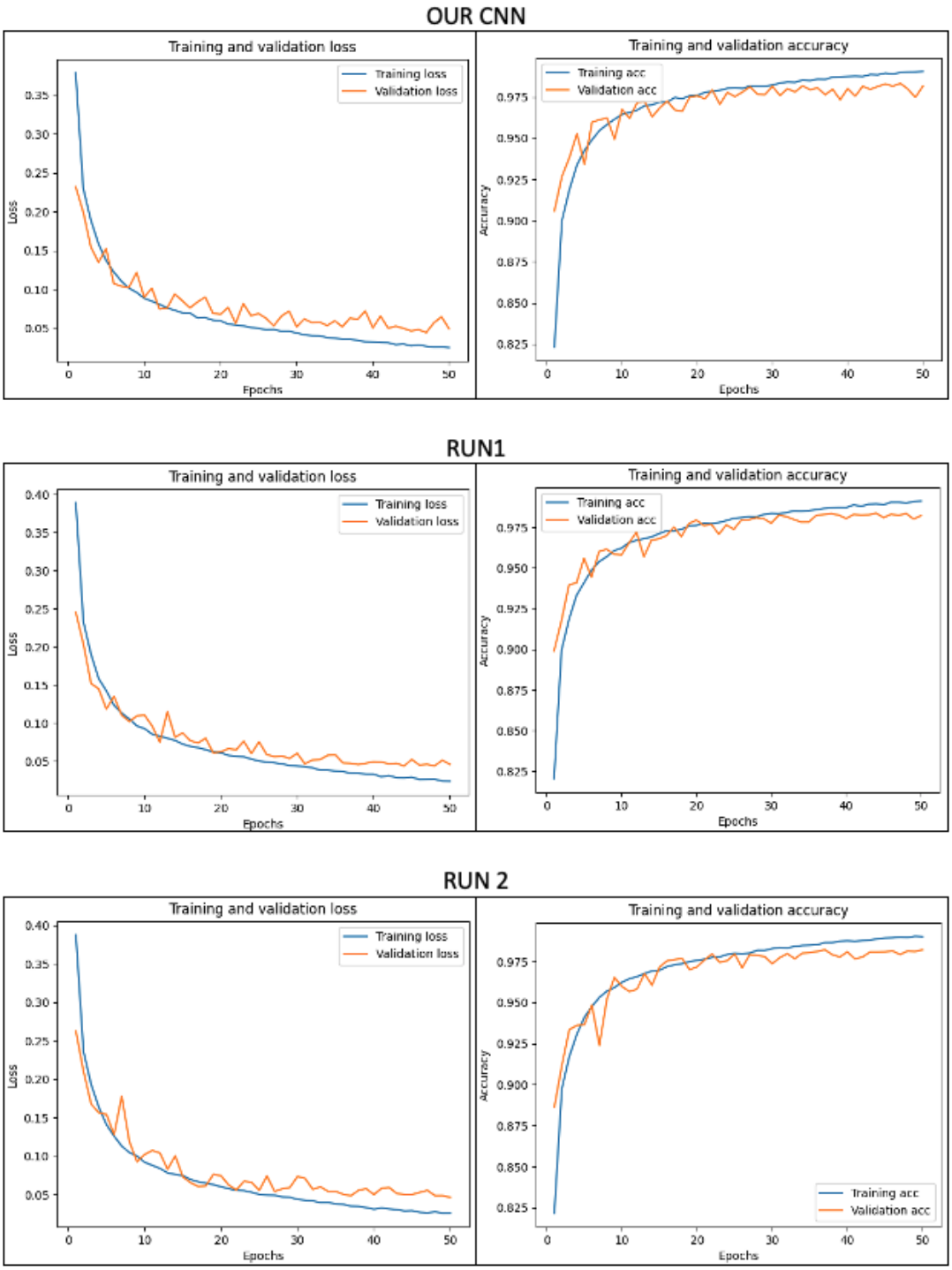

Figure 3.4: First page of loss and accuracy graphs of the training and validation data during our CNN run and an additional four runs of the same CNN.

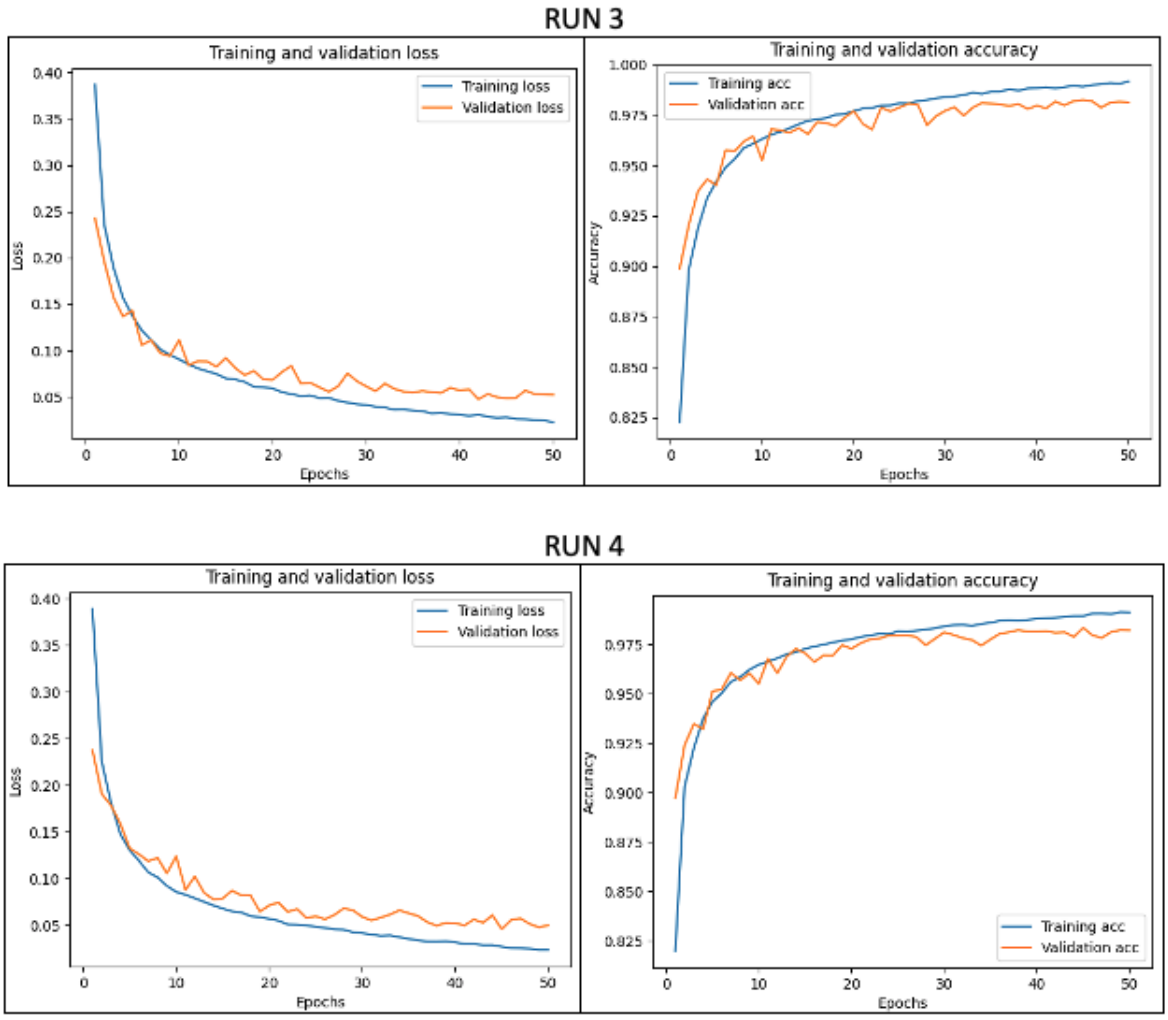

Figure 3.5: Second page of loss and accuracy graphs of the training and the validation data during our CNN run and an additional four runs of the same CNN.

Chapter 4

Discussion

This chapter discusses the results obtained in Chapter 3 and compare it to previous studies. We also present experimental CNNs made through trial and error in Section 4.1. These CNNs show some of the attempts taken to gain the optimal CNN which we use in this thesis. We also discuss several possible improvements in Section 4.2.

We trained and validated our CNN as described in Section 2.4. The training and validation datasets consist of galaxy images from DES and simulated gravitational lenses as discussed in Sections 2.1 and 2.2. Looking at the loss and accuracy graphs in Figure 3.2, we are able to determine that, our CNN is able to learn and predict whether or not a gravitational lens is present in the images. A training accuracy and loss of $97.15 \pm 2.81\%$ and 0.07 has been obtained, as well as a validation accuracy and loss of $97.03 \pm 1.55\%$ and 0.08. An accuracy score of $99.73 \pm 0.07\%$ and a loss, of 0.81 was achieved when the CNN was evaluated against the unseen testing data.

When evaluating the CNN against unseen real lenses (Section 2.3), it predicted $11.92 \pm 2.75\%$ correctly with a loss score of 10.49. This shows that the CNN struggled to predict real lenses

against the simulated ones. When looking at the nine random RGB composite images found in the various datasets, as seen in Figure 3.3, we can deduce that the positive simulated lensing systems are not very realistic, due to most multiple images being point source like. The reason for this is because the real lenses contain various morphologies of lensed source images and not just bright arcs, which form the basis of the simulated data.

In an effort to better understand the relatively low accuracy scores, we searched through the unseen real dataset for real lenses that appear to have similar bright arcs to those seen in the positive simulated dataset. The idea being that if the CNN is focusing on visible differences, this would hopefully improve the overall accuracies of the CNN when running against the limited unseen real dataset, called the selected unseen real lenses. After doing 5-fold cross validation, we obtained an accuracy of $15.42 \pm 5.67\%$ with a loss of 8.95. However, when comparing our results across all five runs, as seen in Table 3.2, we do not see a consistent improvement of the accuracy of the selected unseen real lenses, but rather that the results are similar to that of the unseen real lenses. We can deduce from this that the CNN is learning from both visible and underlying properties within the images that are present in the positive simulated dataset which are not present in the unseen real lenses dataset. This difference is the leading cause of the relatively low accuracies. This difference is further exaggerated by the limited availability of currently identified real lenses to form a training dataset as well as a dataset to test the overall accuracy of the trained CNN, hence the need for simulated lensing data as training data.

In a previous study which identified strong gravitational lensing systems using a CNN algorithm, an accuracy of $98.6 \pm 0.1\%$ was obtained when evaluating their CNN against an unseen testing data which is similar to the data used to train the CNN (Jacobs et al., 2019b). However, it is unclear what their accuracy was when evaluating their CNN against previously identified lenses. Their testing scores are similar to ours when training and validating their

CNN against images containing galaxies and positive simulated lenses.

In Section 1.4, we looked at various projects which used ML algorithms to find lensing systems, including various forms of CNNs. For the majority of the projects which documented their accuracies against real lenses, they achieved similarly low accuracies to those achieved by our CNN. In Pertillo et al. (2019), their CNN was able to correctly predict 20% of the strong gravitational lenses in a given set (Petrillo et al., 2019). In Metcalf et al. (2019), their best CNN produced a true positive rate of 22% for an unseen real lenses dataset, indicating that it only predicted 22% correctly (Metcalf et al., 2019). In Huang et al. (2019), their CNN correctly identified 40% of their unseen dataset. The majority of these papers mostly used a combination of simulated and real lensing data to train and test their models. They obtained very high accuracies when testing their ML algorithms against datasets that are similar to their training and validation data. However, there appears to be inconsistencies when evaluating against real lensing data.

Similarly, our CNN performed well during training and validation when trained with simulated data. As with others, our CNN struggled to find similarities between the positive simulated training data and the real lenses. Our CNN was able to identify $\frac{57}{389}$ (14.65%) real lenses correctly. After investigating these 57, we found that some of these lenses were duplicated, due to multiple papers recording the same lensing systems of multiple lensed source images. After removing these duplicates, a total of $\frac{42}{252}$ unique lensing systems were identified, these are presented in Table B.1. This results in a comparatively similar accuracy to those from previous studies when factoring in the sizes of the relative datasets and running times. There are millions of galaxies that exist and have been observed in various surveys, and only a few strong gravitational lensing systems have been discovered so far. Although, our CNN only finds $\sim 14.65\%$ of the lensing systems, it will still be able to finding strong gravitational lensing systems without using any real lenses as training dataset, reducing hu-

man effort.

4.1 Results Using Alternative CNN Architectures

During the course of this project we attempted various architectures of CNN models to gain the best possible results with the given constraints - time, resources, availability of data, etc.. The goal is to achieve the highest possible accuracies when passing real lenses through the CNN after training - with a perfect score being 100%.

To determine the optimal foundation of our CNN model, we started by creating two models, one with a *sci-kit learn* back-end and one with a *Keras* back-end. This is to determine which back-end is the most suitable when analysing features within images. We determined that the *Keras TensorFlow* back-end was the most successful, with multiple packages at our disposal and higher overall accuracies across the board.

We then created a CNN model and tried various options to obtain the best possible performance. We tried multiple variations of classifier architectures along with variations of early stopping, data augmentation, the number of k-fold iterations, adjusting the learning rate, batch sizes and epochs. Various experiments were performed on permutations of these factors. Only a few are presented here to show the attempts to obtain more desirable results.

In experiment 1, we used a CNN architecture similar to that of the Jacobs paper (Jacobs et al., 2019b) with $33,215,329$ trainable parameters. We trained this CNN model with 50 epochs and a learning rate of 0.0002, a data augmentation multiplier of 2, a batch size of

128, and used a k-fold cross validation where $k = 5$. No early stopping was used as we wanted to see the loss graphs over all epochs, and noted that in previous experiments where early stopping was used, the process of training the CNN was halted too aggressively. A training accuracy and loss of $99.08 \pm 1.91\%$ and 0.02 and a validation accuracy and loss of $97.77 \pm 1.01\%$ and 0.08 were obtained. We obtained an average test accuracy and loss of $99.75 \pm 0.08\%$ and 0.01 respectively when evaluating our CNN against the unseen testing data. We also obtained the average scores of the accuracy and loss when evaluating the CNN against real lenses of $6.27 \pm 1.91\%$ and 13.73. When evaluating the CNN against the selected real lenses, an average accuracy and loss of $6.02 \pm 1.70\%$ and 13.40 was obtained. The loss and accuracy graphs for the training and validation data of experiment 1 are seen in Figure 4.1.

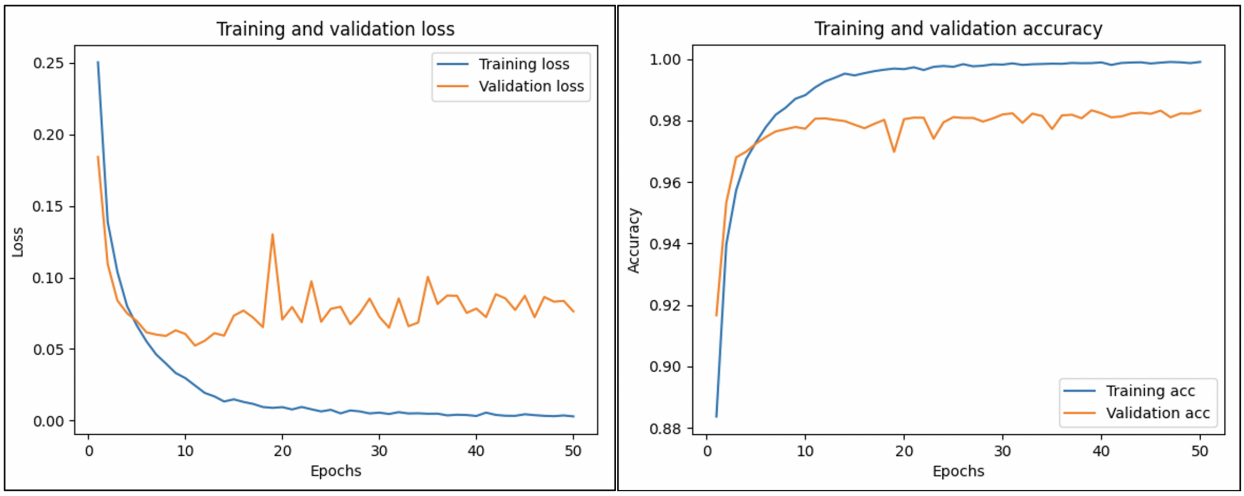

Figure 4.1: The loss (*left*) and accuracy(*right*) graphs of the training and validation data of experiment 1.

In Figure 4.1, the training loss line decreases very clearly following that of a good-fit as seen in Figure 3.1. However, the validation loss line is increasing from epoch 10 onwards.

This graph very clearly shows that the CNN is over-fitting and will generalise poorly and this can be seen in the training and validation accuracy graphs. This indicates that the CNN is learning the features of the training dataset too aggressively and is not learning a wide-spread feature set of the general image population, ensuring that it will not be able to generalise enough to distinguish between images it has not seen before.

In experiment 2, a different CNN model was used which contains multiple convolutional layers. This model has a total of $140,449$ trainable weights, and is trained using a batch size of 128, 50 epochs and a data augmentation multiplier of 2. We used an early stopping with a patience of 10, 5-fold cross validation, and a learning rate of 0.0002. We gained a training accuracy and loss of $96.55 \pm 3.19\%$ and 0.08 respectively. We also gained a validation accuracy of $95.21 \pm 2.15\%$ and loss of 0.12. The training and validation loss and accuracy graphs are seen in Figure 4.2. Experiment 2 obtained an average accuracy and loss when evaluating the CNN against the unseen testing data of $99.37 \pm 0.10\%$ and 0.03. When evaluating the CNN model against the real lenses, accuracy and loss of $9.77 \pm 1.58\%$ and 12.61 was obtained. The selected data gained an accuracy and loss of $11.57 \pm 4.61\%$ and 12.21.

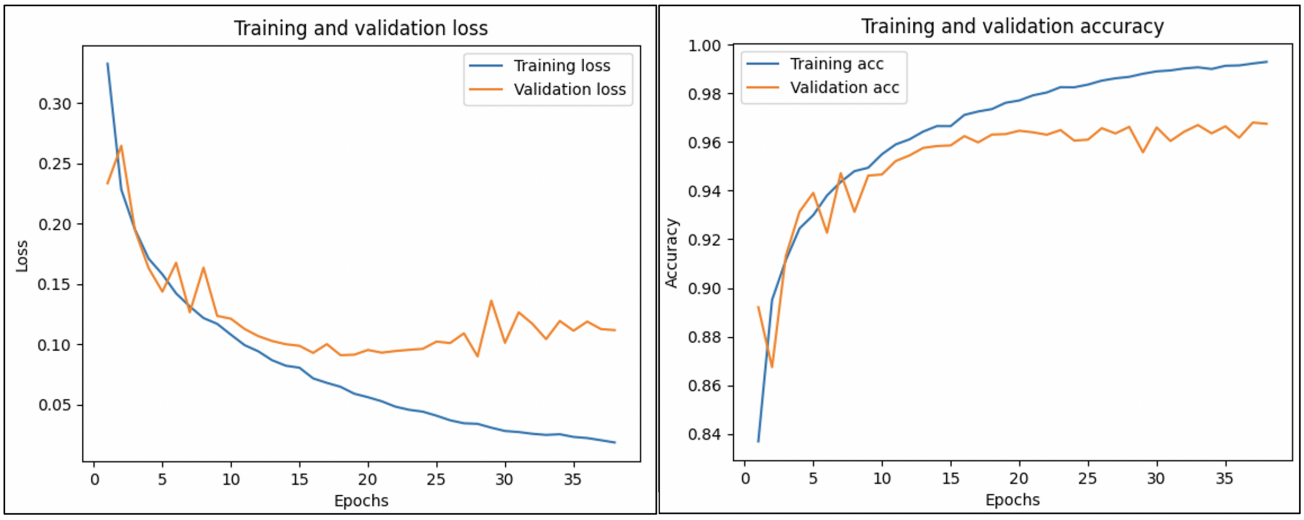

Figure 4.2: The loss (*left*) and accuracy (*right*) graphs of the training and validation data of experiment 2.

Figure 4.2, shows the loss and accuracy graphs during training and validation. The training loss graph shows a clear decrease to the end of training, indicating that there is a low learning rate present. The validation loss line flattens out at around $15 - 20$ epochs. However, the gap between the two loss lines increases, showing that the training dataset is unrepresentative of the validation dataset. This implies that the CNN is not able to learn well enough from the given training dataset to accurately differentiate between images that contain lensing systems and those that do not. Furthermore, any tests against unseen real data would yield similarly poor results.

After experimenting with various different types of architectures, the majority of which are not documented but are derived in a similar fashion to the above two experiments, we finally settled with the CNN architecture and datasets documented in Chapter 2, and the results of which are presented in Chapter 3.

4.2 Possible Improvements

Detailed below are potential improvements which could be applied to the CNN to yield better results. Had time and resources not been a factor in this Masters project, these avenues of improvement would have been investigated and applied where beneficial. In this section, the possible improvements would address the differences between our training and real datasets, and the potential flaws of our CNN algorithm, and provide possible enhancements thereof.

It is important to note that in our CNN, we only use simulations of lenses for the training data to increase the amount of strong gravitational lenses in the various g, r and i-bands. This is similar to the method of semi-supervised learning, where a simulator using real data can provide realistic synthetic images (Shrivastava et al., 2016). However, the problem with using only simulated images for training, is that these generated images are not real and may produce traits that only exist within the simulated images and does not produce the traits present in real gravitational lensing images. This may result in the CNN algorithm learning certain traits made through human or simulation errors rather than gravitational arcs (Shrivastava et al., 2016). This is seen in our CNN results, as the overall accuracies when running the CNN against the real unseen dataset is lower than we initially hoped for. To overcome this, we would include more real lensing data in the training data, particularly including more variation in the gravitational lensing morphologies.

Understanding that gravitational lensing systems have different morphologies is crucial to potentially improve the overall accuracies against the real lensing data. As shown in Figure 1.4, there exists multiple morphologies of the lensed source images, which we could observe. However, the simulated lenses only include bright gravitational arcs and none of the other morphologies. Therefore, our CNN model will not be able to learn the different lenses if only certain bright arcs are included in the training data. To improve the results, the training

data would need to include the various types of lensing morphologies at large volumes, the three seen in Figure 1.4 and the other morphologies of the lensed images are created with regards to the positions of the lens, the source and the observer.

Aside from fixing the training data, we could improve the results by increasing the volume of the training and validation data. We can gain more lensing images from the few gravitational lensing databases that exist, and more simulated lensing data from previous projects, along with various applications of our data augmentation multiplier. This increase in training and validation data should help our CNN model to learn the lensing features which it is not currently learning.

Various researchers who found gravitational lensing systems use well performing lens finding ML algorithms. The most well known lens finding algorithms that have proved to be highly successful includes: YattaLens (Sonnenfeld et al., 2017), Arcfinder (Alard, 2006; Seidel and Bartelmann, 2007; More et al., 2012), and the Ringfinder (Gavazzi et al., 2014). Using one of these or similar lens finding algorithms will improve results in conjunction with the output of our CNN to verify whether or not a given image contains a lensing system.

An improvement regarding the ML application of our selected CNN is to use different CNN architectures and methods of learning. We could use reinforcement learning which would ensure that the CNN model constantly adapts itself to obtain better results (Liu, 2017). In our current CNN, we make use of various max-pooling layers. While max-pooling layers have their advantages of reducing the number of pixels in a given image, thereby speeding up training. These layers also have the disadvantage of losing large amounts of information about the image as it is passed on to the various layers, potentially reducing the overall performance of the CNN (Rimal, 2018). A suitable way to overcome this is by using a capsule neural network. In a capsule neural network, a group of neurons (capsule) represents

different traits of the same image feature. These traits, or features, are processed and a probability of whether or not that specific feature is present in the image is calculated (Mensah et al., 2019). The capsule neural networks have not become mainstream yet, and needs to be tested in various scenarios to determine their strengths and weaknesses (Shah et al., 2018).

Finally, another improvement could be to change the way we label the lenses, by using a rating system, as seen in many papers that search for these lensing systems, rather than a boolean value of 0 or 1. This rating system can either be done visually or through a software program. Huang et al. (2020) has done this visually, where Jacobs et al. (2019) used the LensRater software (Jacobs et al., 2019a; Huang et al., 2020). The neural network will then be able to give a grade that relates to the probability that a lens has been perceived in a given image. Thus, we can determine how well our CNN is performing realistically by assessing the probability of an image containing a lens. Currently, our CNN identifies whether or not there is most definitely a lens visible in the given image, if there is any uncertainty, our CNN notes the image as not containing a lens. As a result, our CNN is not able to determine which images might have lenses in them, losing many possible or highly likely lensing candidates.

After looking at the different possible improvements raised above, one could wonder if using data from other sky surveys (such as Euclid or LSST) might improve our CNNs performance. Our CNN could be applied to the sky surveys as they contain similar optical data to that of DES. However, the CNN would have to be trained and evaluated using data from the same sky survey, as the telescopes used in the surveys contain different instrumentation which produces different images. One instrumental aspect that could influence the CNNs performance when using one sky surveys data to train the CNN and to evaluate it against another survey data, is the resolution of the images. Comparing the resolution, mirror size and whether or not the telescope is space based or ground based, shows that data collected

from the sky surveys may differ, as seen in Table 4.1. These parameters are not the only instrumentation variables that influence the observed images of the telescopes. The quality and resolution of telescopes in space is greater than ground-based telescopes because the atmosphere of earth distorts the light coming from distant objects. A telescope with poorer resolution may observe a spiral galaxy as a gravitational lens, which could negatively affect the training of a CNN, as it is learning the incorrect objects. If using various sky surveys, the affects of the instruments should be taken into account, to determine if there is any factors that may negatively affect the performance of a CNN.

Table 4.1: Table showing a few different instrumental aspects of DES, Euclid and LSST, comparing the resolution and primary mirror size.

Sky Survey	**Sky or Ground Based**	**Resolution**	**Primary Mirror Size**	**References**
DES	Ground Based	0.263 arcseconds per pixel	4m	Diehl et al. (2017)
Euclid	Space Based	0.1 arcseconds per pixel	1.2m	Laureijs et al. (2011)
LSST	Ground Based	0.2 arcseconds per pixel	8.4m	Ivezić et al. (2019)

Chapter 5

Conclusion

The applications of gravitational lensing are vast, and their value in astronomical observations enables us to look at very distant objects in the universe, allowing for the observation of some of the earliest galaxies. We were able to implement ML through the use of a CNN algorithm to automate the search for these lenses in DES. Our results are presented and discussed in Chapters 3 and 4 respectively. Due to the time constraints of this project, we were not able to implement the various improvements discussed in Section 4.2.

This project had multiple limitations which influenced its effectiveness. The main limitation of this project is that the CNN was unable to find all the known types of gravitational lensing systems as it was only trained on a select class of lenses with bright arcs. This can be improved by using various gravitational lensing systems to train and test our CNN. This data can be retrieved from the various databases and research projects that are constantly being updated. Another limitation is that some of the images in the unseen real lenses dataset (retrieved from various papers), were centred on the lensed source images rather than the lens itself, effectively producing duplicates of the same lensed system ($\frac{57}{389}$ lenses correctly predicted). Thus, the duplications should be removed to gain a more accurate representation

of the CNNs performance ($\frac{42}{252}$ unique lenses correctly predicted).

The aim of this project was to identify strong gravitational lenses within DES using an ML technique. In order to complete our aim, there were two main objectives that needed to be completed. The first objective was to create a CNN algorithm and train it with negative and positive data. The collection and creation of these two datasets is discussed in Sections 2.1 and 2.2. We tested the CNN against the unseen testing data to determine its performance at differentiating between images containing lenses and those that do not. Using a configurable CNN, we were able to manipulate it with little-to-no code changes. The benefit of this was quickly realised when testing various configurations to identify improvements, as discussed in 4.1. This allowed us to rapidly test various configurations without the laborious task of changing code in multiple places wherever applicable to start a run which can also be prone to human error.

The second objective was to test our CNN against the unseen real lenses to determine how well it performs with real lenses. The CNN was only trained using negative images from DES and positive simulated lenses, and was able to detect the presence of a gravitational lens in an image with an accuracy of $11.92 \pm 2.75\%$. Thus, the CNN (in one run) was able to find strong gravitational lensing systems, although not to the extent which we hoped for. Our CNN found 34 lenses in an unseen real dataset, and after inspecting the data, we realised that the images of our unseen real lensing data contain lenses with various morphologies and not just the bright arcs which we saw in the simulated data. We visually selected lenses which appear similar to our positive data and was able to detect 9 lenses using the CNN. Our CNN was able to correctly predict $15.42 \pm 5.67\%$ of the selected unseen lensing data. To improve our results, we would either have to use more realistic simulations or include more real data, or a combination of both, with reinforcement learning. More possible improvements have been addressed in Section 4.2.

We ran the CNN a total of five times in order to achieve repeatable results. The CNN was able to identify a total of $\frac{57}{389}$ ($\frac{42}{252}$ unique, as seen in Table B.1) lensing systems correctly, of which the same $\frac{18}{389}$ lenses were observed during all five runs and $\frac{14}{252}$ of these were unique. We were able to identify strong gravitational lensing systems to an accuracy of 14.65% after a total of five runs. This is not too far off from the best accuracy rates obtained by some of the most accurate CNNs that currently exist, which range from 20% to 40% when searching for lenses. Given the opportunity and resources to run the entire DES dataset through the CNN, it would have been able to detect potentially new gravitational lenses. It is clear that even with the modern technologies and the ML algorithms that currently exist, identifying gravitational lensing systems is difficult because of their various lensing morphologies and their rareness as a result of the particular conditions necessary to generate these systems. However, these results obtained provide a foundation for future investigations. By using more precise simulations and more real lenses to train the CNN, an improvement in performance of our CNN could be seen.

The known lenses that we have used in this project is only a small portion of the $\sim 1,000$ lenses that currently exist, and the lenses that we have used are not representative of the entire set. A next logical step for this project is to use the CNN on a large amount of data from DES to determine new lensing candidates. These candidates can then be visually inspected to determine if they are actually lenses or not. This process is similar to the inspection processes found in several previous studies when identifying new gravitational lensing systems (Jacobs et al., 2019b; Metcalf et al., 2019; Petrillo et al., 2019). Through the use of citizen science projects and the ever increasing amount of strong gravitational lensing systems found and added to databases such as the Master Lens Database [1], more real lenses can be used as training data, improving the CNN's overall performance as well. This method

[1] admin.masterlens.org

of using ML to detect objects can be applied to several other astronomical objects, enlarging the observance of those objects and expanding our knowledge of them.

Bibliography

Abbott, T. M. C., Abdalla, F. B., Allam, S., Amara, A., Annis, J., et al. (2018). The dark energy survey: Data release 1. *The Astrophysical Journal Supplement Series*, 239(2):18. doi: 10.3847/1538-4365/aae9f0.

Alard, C. (2006). Automated detection of gravitational arcs. *arXiv e-prints*. arXiv: astro-ph/0606757.

Arnold, L., Rebecchi, S., Chevallier, S., and Paugam-Moisy, H. (2011). An Introduction to Deep Learning. In *European Symposium on Artificial Neural Networks (ESANN)*, Proceedings of the European Symposium on Artificial Neural Networks (ESANN), Bruges, Belgium. HAL number: hal-01352061.

Atlas, L. E., Homma, T., and Marks, R. J. (1988). An artificial neural network for spatio-temporal bipolar patterns: Application to phoneme classification. In Anderson, D. Z., editor, *Neural Information Processing Systems*, pages 31–40. American Institute of Physics. Retrieved from http://papers.nips.cc/paper/20-an-artificial-neural-network-for-spatio-temporal-bipolar-patterns-application-to-phoneme-classification.pdf.

Bailer-Jones, C. A. L., Gupta, R., and Singh, H. P. (2002). *An Introduction to Artificial Neural Networks*, page 51. arXiv: astro-ph/0102224.

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. doi: 10.1142/S0218271810017160.

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

Bartelmann, M. (2010). TOPICAL REVIEW Gravitational lensing. *Classical and Quantum Gravity*, 27(23):233001. doi: 10.1088/0264-9381/27/23/233001.

Bartelmann, M. and Schneider, P. (2001). Weak gravitational lensing. , 340(4-5):291–472. doi: 10.1016/S0370-1573(00)00082-X.

Bell, J. (2020). *Machine Learning: Hands-On for Developers and Technical Professionals*. Wiley. ISBN: 9781119642190.

Bennett, C. L., Larson, D., Weiland, J. L., Jarosik, N., Hinshaw, G., et al. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results. , 208(2):20. doi: 10.1088/0067-0049/208/2/20.

Bisnovatyi-Kogan, G. and Tsupko, O. (2017). Gravitational Lensing in Presence of Plasma: Strong Lens Systems, Black Hole Lensing and Shadow. *Universe*, 3:57. doi : 10.3390/universe3030057.

Bolton, A. S., Koopmans, L. V. E., Barnabè, M., Treu, T., and Czoske, O. (2011). Two-dimensional kinematics of SLACS lenses – IV. The complete VLT–VIMOS data set *. *Monthly Notices of the Royal Astronomical Society*, 419(1):656–668. doi: 10.1111/j.1365-2966.2011.19726.x.

Bouwens, R. J., Stefanon, M., Oesch, P. A., Illingworth, G. D., Nanayakkara, T., et al. (2019). Newly Discovered Bright z $\sim$ 9-10 Galaxies and Improved Constraints on Their Prevalence Using the Full CANDELS Area. , 880(1):25. doi: 10.3847/1538-4357/ab24c5.

Brownlee, J. (2017). Gentle introduction to the adam optimization algorithm for deep learning. Retrieved from https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/.

Brownlee, J. (2019a). How do convolutional layers work in deep learning neural networks? Retrieved from https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/.

Brownlee, J. (2019b). How to use learning curves to diagnose machine learning model performance. Retrieved from https://machinelearningmastery.com/learning-curves-for-diagnosing-machine-learning-model-performance/.

Cervantes-Cota, J. L., Galindo-Uribarri, S., and Smoot, G. F. (2019). The legacy of einstein's eclipse, gravitational lensing. *Universe*, 6(1). doi: 0.3390/universe6010009.

Chollet, F. (2018). *Deep Learning with Python.* Manning Publications Co., America. ISBN: 9781617294433.

Chollet, F. et al. (2015). Keras. Retrieved from: https://keras.io.

Ciresan, D., Meier, U., Masci, J., Gambardella, L. M., and Schmidhuber, J. (2011). Flexible, high performance convolutional neural networks for image classification. *International Joint Conference on Artificial Intelligence IJCAI-2011*, pages 1237–1242. doi: 10.5591/978-1-57735-516-8/IJCAI11-210.

Collett, T. E. (2015). The population of galaxy–galaxy strong lenses in forthcoming optical imaging surveys. *The Astrophysical Journal*, 811(1):20. doi: 10.1088/0004-637x/811/1/20.

Collett, T. E., Oldham, L. J., Smith, R. J., Auger, M. W., Westfall, K. B., et al. (2018). A precise extragalactic test of General Relativity. *Science*, 360(6395):1342–1346. doi: 10.1126/science.aao2469.

Correspondent, A (1969). Artificial Intelligence: Making Machines Behave. *Nature*, 222(5198):1028. doi: 10.1038/2221028a0.

Courbin, F., Bonvin, V., Suyu, S. H., Marshall, P. J., Rusu, C. E., et al. (2016). H0licow – v. new cosmograil time delays of he 04351223:h0to 3.8 per cent precision from strong lensing in a flat cdm model. *Monthly Notices of the Royal Astronomical Society*, 465(4):4914–4930. doi: 10.1093/mnras/stw3006.

Davies, A., Serjeant, S., and Bromley, J. M. (2019). Using convolutional neural networks to identify gravitational lenses in astronomical images. *Monthly Notices of the Royal Astronomical Society*, 487(4):5263–5271. doi: 10.1093/mnras/stz1288.

de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., and Valentijn, E. A. (2012). The kilo-degree survey. *Experimental Astronomy*, 35(1-2):25–44. doi: 10.1007/s10686-012-9306-1.

Diaz Rivero, A. and Dvorkin, C. (2020). Direct detection of dark matter substructure in strong lens images with convolutional neural networks. , 101(2):023515. doi: 10.1103/PhysRevD.101.023515.

Diehl, H. T., Buckley-Geer, E. J., Lindgren, K. A., Nord, B., Gaitsch, H., et al. (2017). The DES Bright Arcs Survey: Hundreds of Candidate Strongly Lensed Galaxy Systems from the Dark Energy Survey Science Verification and Year 1 Observations. *The Astrophysical Journal Supplement Series*, 232(1):15. doi: 10.3847/1538-4365/aa8667.

Dyson, F. W., Eddington, A. S., and Davidson, C. (1920). Ix. a determination of the deflection of light by the sun's gravitational field, from observations made at the total eclipse of may 29, 1919. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 220(571-581):291–333. doi: 10.1098/rsta.1920.0009.

Eddington, A. S. (1919). The total eclipse of 1919 May 29 and the influence of gravitation on light. *The Observatory*, 42:119–122. Bibcode: 1919Obs....42..119E.

Einstein, A. (1911). Über den Einfluß der Schwerkraft auf die Ausbreitung des Lichtes. *Annalen der Physik*, 340(10):898–908. doi: 10.1002/andp.19113401005.

Einstein, A. (1936). Lens-Like Action of a Star by the Deviation of Light in the Gravitational Field. *Science*, 84:506–507. doi: 10.1126/science.84.2188.506.

Ellis, R. S. (2010). Gravitational lensing: a unique probe of dark matter and dark energy. *Philosophical Transactions of the Royal Society of London Series A*, 368(1914):967–987. doi: 10.1098/rsta.2009.0209.

Faure, C., Kneib, J.-P., Covone, G., Tasca, L., Leauthaud, A., et al. (2008). First catalog of strong lens candidates in the cosmos field. *The Astrophysical Journal Supplement Series*, 178. doi: 10.1086/526426.

Fraknoi, A., Morrison, D., and Wolff, S. (2016). *Astronomy (The Textbook)*. ISBN: 978-1-938168-28-4.

Gavazzi, R., Marshall, P. J., Treu, T., and Sonnenfeld, A. (2014). Ringfinder: Automated detection of galaxy-scale gravitational lenses in ground-based multi-filter imaging data. *The Astrophysical Journal*, 785(2):144. doi: 10.1088/0004-637x/785/2/144.

Hoshino, H., Leauthaud, A., Lackner, C., Hikage, C., Rozo, E., et al. (2015). Luminous red galaxies in clusters: central occupation, spatial distributions and miscentring. *Monthly Notices of the Royal Astronomical Society*, 452(1):998–1013. doi: 10.1093/mnras/stv1271.

Huang, X., Storfer, C., Ravi, V., Pilon, A., Domingo, M., et al. (2020). Finding Strong Gravitational Lenses in the DESI DECam Legacy Survey. *The Astrophysical Journal*, 894(1):78. doi: 10.3847/1538-4357/ab7ffb.

Hurwitz, J. and Kirsch, D. (2018). *Machine Learning for Dummies*. John Wiley Sons, USA. ISBN: 978-1-119-45495-3.

Ilbert, O., Capak, P., Salvato, M., Aussel, H., McCracken, H. J., et al. (2008). COSMOS Photometric redshifts with 30-bands for 2-deg2. *The Astrophysical Journal*, 690(2):1236–1249. doi: 10.1088/0004-637x/690/2/1236.

Iliashenko, O., Bikkulova, Z., and Dubgorn, A. (2019). Opportunities and challenges of artificial intelligence in healthcare. In *E3S Web of Conferences*, volume 110 of *E3S Web of Conferences*, page 02028. doi: 10.1051/e3sconf/201911002028.

Ivezić, Ž., Kahn, S. M., Tyson, J. A., Abel, B., Acosta, E., et al. (2019). LSST: From Science Drivers to Reference Design and Anticipated Data Products. *The Astrophysical Journal*, 873(2):111. doi: 10.3847/1538-4357/ab042c.

Jacobs, C., Collett, T., Glazebrook, K., Buckley-Geer, E., Diehl, H. T., et al. (2019a). An extended catalog of galaxy–galaxy strong gravitational lenses discovered in DES using convolutional neural networks. *The Astrophysical Journal Supplement Series*, 243(1):17. doi: 10.3847/1538-4365/ab26b6.

Jacobs, C., Collett, T., Glazebrook, K., McCarthy, C., Qin, A. K., et al. (2019b). Finding high-redshift strong lenses in DES using convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 484(4):5330–5349. doi: 10.1093/mnras/stz272.

Jaelani, A. T., More, A., Oguri, M., Sonnenfeld, A., Suyu, S. H., et al. (2020). Survey of Gravitationally lensed Objects in HSC Imaging (SuGOHI) - V. Group-to-cluster scale lens search from the HSC-SSP Survey. *Monthly Notices of the Royal Astronomical Society*, 495(1):1291–1310. doi: 10.1093/mnras/staa1062.

Jaki, S. L. (1978). Johann Georg von Soldner and the gravitational bending of light, with an English translation of his essay on it published in 1801. *Foundations of Physics*, 8(11):927–950. doi: 10.1007/BF00715064.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer Texts in Statistics. Springer New York. ISBN: 9781461471387.

Jeong, J. (2019). The most intuitive and easiest guide for convolutional neural network. Retrieved from: https://towardsdatascience.com/the-most-intuitive-and-easiest-guide-for-convolutional-neural-network-3607be47480:∼:text=Rectangular%20or%20cubic%20shapes%20can,a%20single%20long%20feature%20vector.

Jordan, M. I. and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260. doi: 10.1126/science.aaa8415.

Karpathy, A. (2020). Cs231n: Convolutional neural networks for visual recognition. Retrieved from: https://cs231n.github.io/neural-networks-3/.

Kelly, P. L., Diego, J. M., Rodney, S., Kaiser, N., Broadhurst, T., et al. (2018). Extreme magnification of an individual star at redshift 1.5 by a galaxy-cluster lens. *Nature Astronomy*, 2:334–342. doi: 10.1038/s41550-018-0430-3.

Kingma, D. P. and Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv e-prints*. arXiv: 1412.6980.

Kneib, J.-P. and Natarajan, P. (2012). Cluster lenses. *The Astronomy and Astrophysics Review*, 19. doi: 10.1007/s00159-011-0047-3.

Kochanek, C. S., Falco, E. E., Impey, C. D., Lehar, J., McLeod, B. A., et al. (1999). The Evolution of Gravitational Lens Galaxies. *arXiv e-prints*. arXiv: astro-ph/9910165.

Koekemoer, A. M., Mack, J., Lotz, J. M., Borncamp, D., Khandrika, H. G., et al. (2017). The Hubble Space Telescope Frontier Fields Program. Bibcode: 2017HEAD...1610523K.

Koopmans, L. V. E. and Treu, T. (2004). The lenses structure dynamics survey. *Multiwavelength Cosmology*, page 23–26. doi: 10.1007/0-306-48570-2_4.

Kurata, J. (2018). Deep learning with keras. Retrieved from: https://app.pluralsight.com/library/courses/keras-deep-learning/table-of-contents.

Lanusse, F., Ma, Q., Li, N., Collett, T. E., Li, C.-L., et al. (2017). Cmu deeplens: deep learning for automatic image-based galaxy–galaxy strong lens finding. *Monthly Notices of the Royal Astronomical Society*, 473(3):3895–3906. doi: 10.1093/mnras/stx1665.

Laureijs, R., Amiaux, J., Arduini, S., Augueres, J. L., Brinchmann, J., et al. (2011). Euclid Definition Study Report. *arXiv e-prints*. arXiv: 1110.3193.

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324. doi: 10.1109/5.726791.

Li, R., Napolitano, N. R., Tortora, C., Spiniello, C., Koopmans, L. V. E., et al. (2020). New High-quality Strong Lens Candidates with Deep Learning in the Kilo-Degree Survey. *The Astrophysical Journal*, 899(1):30. doi: 10.3847/1538-4357/ab9dfa.

Link, F. (1937). Sur les conséquences photométriques de la déviation d'Einstein. *Bulletin Astronomique*, 10:73–90. Bibcode: 1937BuAst..10...73L.

Liu, Q., Zhu, H., Liu, C., Jean, D., Huang, S., et al. (2020). Application of machine learning in drug development and regulation: Current status and future potential. *Clinical Pharmacology Therapeutics*, 107. doi: 10.1002/cpt.1771.

Liu, Y. H. (2017). *Python Machine Learning By Example*. Packt Publishing. ISBN: 9781783553112.

Marshall, P. J., Treu, T., Melbourne, J., Gavazzi, R., Bundy, K., et al. (2007). Superresolving distant galaxies with gravitational telescopes: Keck laser guide star adaptive optics andHubble space TelescopeImaging of the lens system SDSS j07373216. *The Astrophysical Journal*, 671(2):1196–1211. doi: 10.1086/523091.

Marshall, P. J., Verma, A., More, A., Davis, C. P., More, S., et al. (2016). SPACE WARPS - I. Crowdsourcing the discovery of gravitational lenses. , 455(2):1171–1190. doi: 10.1093/mnras/stv2009.

Meneghetti, M. (2016). *Introduction to Gravitational Lensing - Lecture scripts*. ISBN: 978-3-030-73582-1.

Mensah, K. P., Adebayo, F. A., Ayidzoe, A. M., and Baagyire, Y. E. (2019). Capsule networks – a survey. *Journal of King Saud University - Computer and Information Sciences*. doi: 10.1016/j.jksuci.2019.09.014.

Metcalf, R. B., Meneghetti, M., Avestruz, C., Bellagamba, F., Bom, C. R., et al. (2019). The strong gravitational lens finding challenge. *Astronomy & Astrophysics*, 625:A119. doi: 10.1051/0004-6361/201832797.

Mitchell, T. (1997). *Machine Learning*. McGraw-Hill Science/Engineering/Math, USA. ISBN: 0070428077.

Mohammed, M., Khan, M., and Bashier, E. (2016). *Machine Learning: Algorithms and Applications*. ISBN: 9781498705387.

Moolayil, J. (2018). *Learn Keras for Deep Neural Networks: A Fast-Track Approach to Modern Deep Learning with Python*. Apress. ISBN: 9781484242407.

More, A., Cabanac, R., More, S., Alard, C., Limousin, M., et al. (2012). The CFHTLS-Strong Lensing Legacy Survey (SL2S): Investigating the Group-scale Lenses with the SARCS Sample. *The Astrophysical Journal*, 749(1):38. doi: 10.1088/0004-637x/749/1/38.

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., et al. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1):1. doi: 10.1186/s40537-014-0007-7.

Niepert, M., Ahmed, M., and Kutzkov, K. (2016). Learning Convolutional Neural Networks for Graphs. *arXiv e-prints*. arXiv:1605.05273.

Oguri, M., Inada, N., Strauss, M. A., Kochanek, C. S., Kayo, I., et al. (2012). The Sloan Digital Sky Survey Quasar Lens Search. VI. Constraints on Dark Energy and the Evolution of Massive Galaxies. *The Astronomical Journal*, 143:120. doi: 10.1088/0004-6256/143/5/120.

O'Shea, K. and Nash, R. (2015). An Introduction to Convolutional Neural Networks. *arXiv e-prints*. arXiv: 1511.08458.

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., et al. (1999). Measurements of Ω and Λ from 42 High-Redshift Supernovae. , 517(2):565–586. doi: 10.1086/307221.

Petrillo, C. E., Tortora, C., Vernardos, G., Koopmans, L. V. E., Verdoes†Kleijn, G., et al. (2019). LinKS: discovering galaxy-scale strong lenses in the Kilo-Degree Survey using convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 484(3):3879–3896. doi: 10.1093/mnras/stz189.

Petrillo, E. (2019). *Discovering gravitational lenses with artificial intelligence*. PhD thesis, University of Groningen. doi: 10.33612/diss.100697045.

Planck Collaboration, Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., et al. (2020). Planck 2018 results. VI. Cosmological parameters. *Astronomy & Astrophysics*, 641:A6. doi: 10.1051/0004-6361/201833910.

Poberezhskiy, I., Luchik, T., Zhao, F., Frerking, M., Basinger, S., et al. (2021). Roman space telescope coronagraph: engineering design and operating concept. 11443:114431V. doi: 10.1117/12.2563480.

Pourrahmani, M., Nayyeri, H., and Cooray, A. (2018). LensFlow: A convolutional neural network in search of strong gravitational lenses. *The Astrophysical Journal*, 856(1):68. doi: 10.3847/1538-4357/aaae6a.

Quinn, P., Axelrod, T., Bird, I., Dodson, R., Szalay, A., et al. (2015). Delivering SKA Science. In *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 147. arXiv: 1501.05367.

Refregier, A. (2003). Weak Gravitational Lensing by Large-Scale Structure. *Annual Review of Astronomy and Astrophysics*, 41:645–668. doi: 10.1146/annurev.astro.41.111302.102207.

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., et al. (1998). Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. , 116(3):1009–1038. doi: 10.1086/300499.

Rimal, K. (2018). Retrieved from: https://software.intel.com/content/www/us/en/develop/articles/understanding-capsule-network-architecture.html?language=es.

Rosenblatt, F. (1957). *The Perceptron, a Perceiving and Recognizing Automaton Project Para.* Report: Cornell Aeronautical Laboratory, Report Number: 85-460-1. Cornell Aeronautical Laboratory. Retrieved from: https://books.google.co.za/books?id=P_XGPgAACAAJ.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3:210–229. doi: 10.1147/rd.33.0210.

Schaub, R. (2019). What are Neural Networks made of? *arXiv e-prints*. arXiv: 1909.09588.

Schechter, P. L. (2004). The hubble constant from gravitational lens time delays. *Proceedings of the International Astronomical Union*, 2004(IAUS225):281–296. doi: 10.1017/s1743921305002085.

Schneider, P. (2003). Gravitational lensing as a probe of structure. *arXiv e-prints*. arXiv: astro-ph/0306465.

Schneider, P., Ehlers, J., and Falco, E. E. (1992). Gravitational Lenses. In *Gravitational Lenses, XIV, 560 pp. 112 figs.. Springer-Verlag Berlin Heidelberg New York. Also Astronomy and Astrophysics Library*, page 112. doi: 10.1007/978-3-662-03758-4.

Seidel, G. and Bartelmann, M. (2007). Arcfinder: an algorithm for the automatic detection of gravitational arcs. *Astronomy & Astrophysics*, 472(1):341–352. doi: 10.1051/0004-6361:20066097.

Shah, P. L., Gupta, T. K., Dhakad, J. B., and D'silva, M. R. (2018). A review paper on understanding capsule networks. *International Journal of Engineering Development and Research (IJEDR)*, 6:58–65. Retrieved from: http://www.ijedr.org/papers/IJEDR1804013.pdf.

Shajib, A. J., Treu, T., Birrer, S., and Sonnenfeld, A. (2020). Massive elliptical galaxies at $z \sim 0.2$ are well described by stars and a Navarro-Frenk-White dark matter halo. *arXiv e-prints*. arXiv: 2008.11724.

Shalev-Shwartz, S. and Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, USA. ISBN: 1107057132.

Shrivastava, A., Pfister, T., Tuzel, O., Susskind, J., Wang, W., et al. (2016). Learning from Simulated and Unsupervised Images through Adversarial Training. *arXiv e-prints*. arXiv: 1612.07828.

Shu, Y., Brownstein, J. R., Bolton, A. S., Koopmans, L. V. E., Treu, T., et al. (2017). The Sloan Lens ACS Survey. XIII. Discovery of 40 New Galaxy-scale Strong Lenses. *The Astrophysical Journal*, 851(1):48. doi: 10.3847/1538-4357/aa9794.

Smith, S. (1936). The Mass of the Virgo Cluster. , 83:23. doi: 10.1086/143697.

Sonnenfeld, A., Chan, J. H. H., Shu, Y., More, A., Oguri, M., et al. (2017). Survey of Gravitationally-lensed Objects in HSC Imaging (SuGOHI). I. Automatic search for galaxy-scale strong lenses. *Publications of the Astronomical Society of Japan*, 70(SP1). doi :10.1093/pasj/psx062.

Sonnenfeld, A., Treu, T., Gavazzi, R., Suyu, S. H., Marshall, P. J., Auger, M. W., and Nipoti, C. (2013). The sl2s galaxy-scale lens sample. iv. the dependence of the total mass density

profile of early-type galaxies on redshift, stellar mass, and size. *The Astrophysical Journal*, 777(2):98. doi: 10.1088/0004-637x/777/2/98.

Sonnenfeld, A., Verma, A., More, A., Allen, C., Baeten, E., et al. (2020). Survey of gravitationally-lensed objects in hsc imaging (sugohi). vi. crowdsourced lens finding with space warps. *Astronomy & Astrophysics*. arXiv: 2004.00634.

Soucail, G., Fort, B., Mellier, Y., and Picat, J. P. (1987). A blue ring-like structure in the center of the A 370 cluster of galaxies. , 172:L14–L16. Bibcode: 1987AA...172L..14S.

Soucail, G., Mellier, Y., Fort, B., Mathez, G., and Cailloux, M. (1988). The giant arc in A 370 : spectroscopic evidence for gravitational lensing from a source at Z=0.724. , 191:L19–L21. Bibcode: 1988AA...191L..19S.

Stark, D., Auger, M., Belokurov, V., Jones, T., Robertson, B., et al. (2013). The cassowary spectroscopy survey: A new sample of gravitationally lensed galaxies in sdss. *Monthly Notices of the Royal Astronomical Society*, 436(2):1040–1056. doi: 10.1093/mnras/stt1624.

Treu, T. (2010). Strong Lensing by Galaxies. *Annual Review of Astronomy and Astrophysics*, 48:87–125. doi: 10.1146/annurev-astro-081309-130924.

Treu, T., Marshall, P. J., and Clowe, D. (2012). Resource letter gl-1: Gravitational lensing. *American Journal of Physics*, 80(9):753–763. doi: 10.1119/1.4726204.

Valls-Gabaud, D. (2006). *The conceptual origins of gravitational lensing*, volume 861 of *American Institute of Physics Conference Series*, pages 1163–1163. doi: 10.1063/1.2399715.

Valls-Gabaud, D. (2009). *A Conceptual and Cultural History of Gravitational Lensing*, volume 409 of *Astronomical Society of the Pacific Conference Series*, page 57. Bibcode: 2009ASPC..409...57V.

Wallach, I., Dzamba, M., and Heifets, A. (2015). AtomNet: A Deep Convolutional Neural

Network for Bioactivity Prediction in Structure-based Drug Discovery. *arXiv e-prints*. arXiv: 1510.02855.

Walsh, D., Carswell, R. F., and Weymann, R. J. (1979). 0957 + 561 A, B - Twin quasistellar objects or gravitational lens. *Nature*, 279:381–384. doi: 10.1038/279381a0.

Wambsganss, J. (1998). Gravitational lensing in astronomy. *Living Reviews in Relativity*, 1(1):12. doi: 10.12942/lrr-1998-12.

Wong, K. C., Sonnenfeld, A., Chan, J. H. H., Rusu, C. E., Tanaka, M., et al. (2018). Survey of Gravitationally Lensed Objects in HSC Imaging (SuGOHI). II. Environments and Line-of-Sight Structure of Strong Gravitational Lens Galaxies to z $\sim$ 0.8. *The Astrophysical Journal*, 867(2):107. doi: 10.3847/1538-4357/aae381.

Zalesky, L. and Ebeling, H. (2020). AStroLens: Automatic Strong-Lens Modeling of X-ray Selected Galaxy Clusters. *Monthly Notices of the Royal Astronomical Society*. doi: 10.1093/mnras/staa2180.

Zhang, A., Lipton, Z. C., Li, M., and Smola, A. J. (2020). *Dive into Deep Learning*. Retrieved from: https://d2l.ai.

Zhang, Y. and Zhao, Y. (2015). Astronomy in the Big Data Era. *Data Science Journal*, 14:11. doi: 10.5334/dsj-2015-011.

Zwicky, F. (1937). Nebulae as Gravitational Lenses. *Physical Review*, 51:290–290. doi: 10.1103/PhysRev.51.290.

Appendix A

Explanation of the Different Datasets

For clarification purposes an explanation is added to describe the different datasets that are used to either train, validate or test our CNN. This is seen in the following list, where the dataset's name is given in italics with a brief description:

Negative Data: This dataset contains 100×100 pixel images clipped from the $10,000 \times 10,000$ pixel images retrieved from DESDR1, as discussed in Chapter 2.1. This dataset of $24,192$ images contains no known gravitational lensing systems.

Positive Data: This is the simulated gravitationally lensed images with a size of 100×100 pixels, as discussed in Chapter 2.2. This dataset contains a total of $24,192$ simulated lensed images.

Total Dataset: Both the positive and negative datasets are multiplied by a factor of 2, through data augmentation resulting in $48,384$ images each. Combining these two datasets a total of $96,768$ useable data is provided.

Training Data: This is 80% of the total dataset, used to train and help our CNN learn the difference between the positive and negative data.

Validation Data: This is the remaining 20% of the total dataset. This is used to validate the performance of the CNN, after it has learnt from the training data.

Testing Data: This is the unseen data not used at all during the training or validation process of our CNN. This dataset contains two separate testing datasets, and each is functionally different, testing different aspects of our CNN. The formation of these datasets have been discussed in Chapter 2.3. These datasets are discussed below under the *Unseen Testing Data* and *Unseen Real Lensing Data*.

Unseen Testing Data: This refers to the $6,034$ images collected separately of positive and negative data each. These images have been generated using the same processes as those discussed in Chapters 2.1 and 2.2. These images also have a size of 100×100 pixels. Testing our CNN against this data, provides us with an insight into how well our CNN is able to learn and predict data that it has never seen before.

Unseen Real Lensing Data: This refers to 389 images of strong gravitational lensing systems identified in previous studies prior to this project. These images are clipped to 100×100 pixels. These images are used to evaluate our CNNs performance when identifying real lensing data that it has never seen before.

Selected Unseen Real Lensing Data: This refers to 89 visually selected images of the unseen real lensing data. These images contain similar bright arc features, similar to that of the positive simulated data. These images are used to determine if the CNN will perform better when evaluating images that look similar to that of the positive training data.

Appendix B

Table of Identified Strong Gravitational Lenses

Table B.1: Table showing the 42 unique strong gravitational lensing systems that our CNN was able to identify correctly across all five runs. The lensing system's name in DES (first column), along with its RA and Dec coordinates in degrees (second and third columns) are presented. The lenses have been obtained from previous studies and those respective references are provided, in the fourth column, as: [1] Jaelani et al. (2020). [2] Diehl et al. (2017). [3] Wong et al. (2018).[4] Stark et al. (2013). [5] Jacobs et al. (2019b).[6] More et al. (2012).[7] Faure et al. (2008). [8] Sonnenfeld et al. (2017). The last column shows the amount of times that each lens has been correctly predicted across all five runs.

DES Name	**RA**	**Dec**	**References**	**Detected Amount**
$DES0011-4623$	2.97298	-46.23866	2	5
$DES0023-4914$	5.93348	-49.39171	2	5

continued on next page

continued from previous page

DES Name	RA	Dec	References	Detected Amount
$DES0025-4123$	6.49127	−41.55583	2	5
$DES0035-5123$	8.84584	−51.50662	2	5
$DES0100-0707$	14.70392	−7.36575	4	5
$DES0212-0750$	33.35217	−7.73189	6	5
$DES0312-3623$	48.40600	−36.17770	5	5
$DES2208+0209$	332.24990	2.11520	1	5
$DES2214+0126$	333.57842	1.17607	6	5
$DES2234+0001$	338.33310	−0.32640	1	5
$DES2242+0001$	340.99900	−0.08030	1	5
$DES2242+0209$	340.50490	2.46960	3	5
$DES2305+0209$	346.74280	2.42860	1	5
$DES2319+0043$	349.97260	0.63690	1	5
$DES0000+0209$	0.00000	2.41528	7	4
$DES0006-4206$	1.51255	−42.13928	2	4
$DES0031+0043$	8.07330	1.01020	1	4
$DES0201-0416$	29.90930	−3.98290	1	4
$DES0212-0333$	32.81500	−3.72990	1	4
$DES0232-0333$	38.38370	−3.46710	1	4
$DES0002+0043$	0.80450	0.90690	1	3
$DES0212-2415$	33.28860	−24.22920	5	3
$DES0222-1041$	35.82638	−10.98013	6	3
$DES0030-4414$	7.77070	−44.04910	2	2
$DES0213-0541$	33.53363	−5.59233	6	2

continued on next page

continued from previous page

DES Name	**RA**	**Dec**	**References**	**Detected Amount**
$DES0216-1041$	34.41483	−10.55554	6	2
$DES0218-0458$	34.53060	−5.26010	1	2
$DES0223-0333$	36.04370	−3.60150	1	2
$DES0231-0207$	37.70060	−1.98410	1	2
$DES0239-0124$	39.97140	−1.58270	1	2
$DES0008-5457$	2.06884	−55.0659	2	1
$DES0020-5040$	5.45183	−50.47489	2	1
$DES0021-4040$	5.39120	−40.66790	2	1
$DES0032-5457$	8.35058	−54.76009	2	1
$DES0034+0209$	8.61730	2.42270	1	1
$DES0108+0126$	16.78860	1.29180	1	1
$DES0158-0416$	29.22650	−4.40710	1	1
$DES0203-0958$	30.95825	−9.71486	6	1
$DES0212-0416$	32.82704	−4.45811	6	1
$DES0219-1041$	35.17963	−10.87124	6	1
$DES2211+0001$	333.04760	−0.13890	1	1
$DES2214+0043$	333.45500	0.94750	1	1

Appendix C

Histograms of the Identified Lenses

Magnitudes

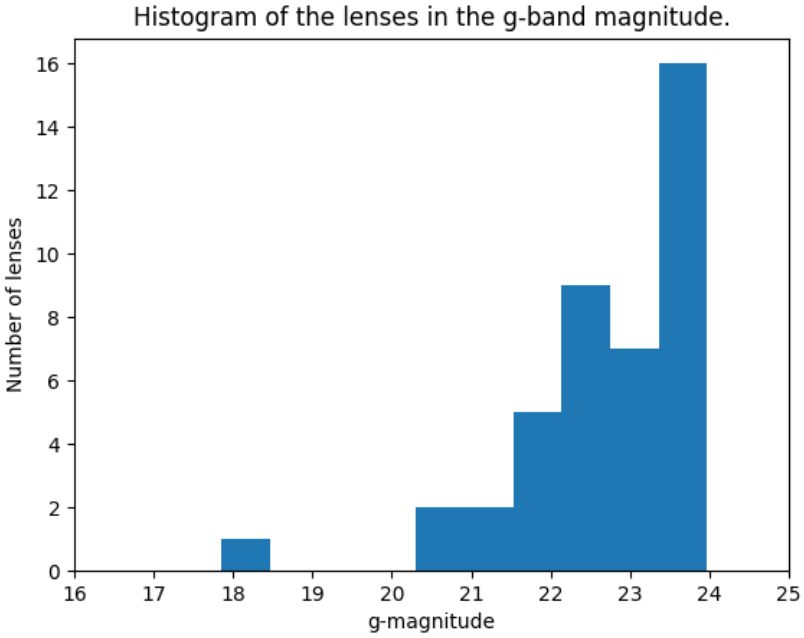

Figure C.1: Histogram of magnitudes of the 42 uniquely identified lenses of the unseen real set in the g-band.

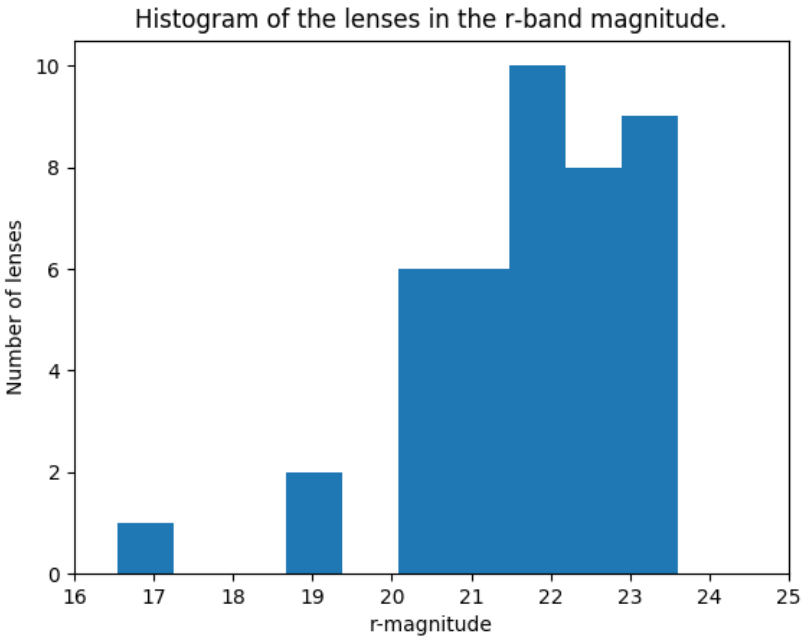

Figure C.2: Histogram of magnitudes of the 42 uniquely identified lenses of the unseen real set in the r-band.

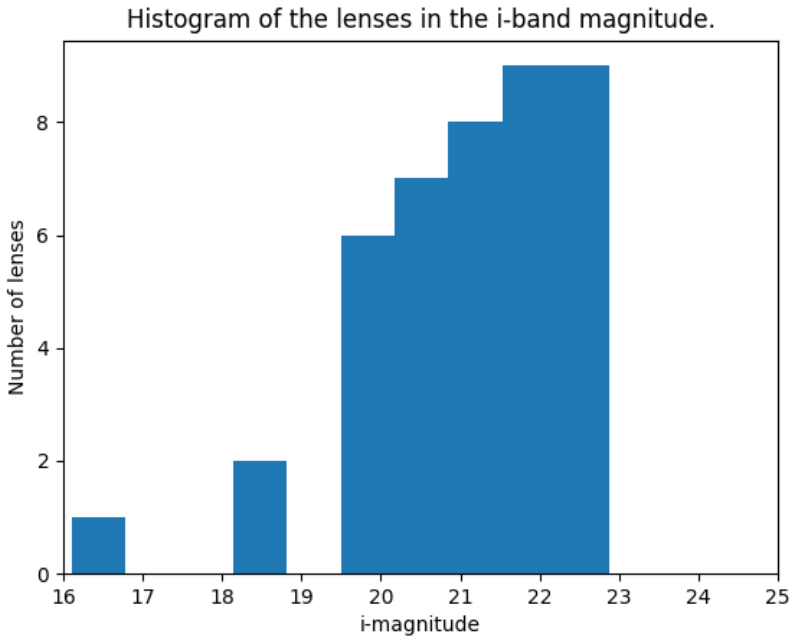

Figure C.3: Histogram of magnitudes of the 42 uniquely identified lenses of the unseen real set in the i-band.