Automating Strong Galaxy-Scale Gravitational Lens Modelling with Neural Networks

Christopher James Pearson
Student ID: 4309384

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

October 2021

‘Light thinks it travels faster than anything but it is wrong. No matter how fast light travels, it finds the darkness has always got there first, and is waiting for it.’

-Terry Pratchett

Supervisor: Dr Simon Dye

Examiners:

Submitted: 19 October 2021
Examined:
Final Version:

Abstract

Strong galaxy-scale gravitational lensing provides a powerful means of studying galaxy formation, constraining cosmology and understanding the evolution of large-scale structure. The presence of a foreground lensing galaxy deflects the light rays of a background source into multiple images, with their configuration dependent upon the lens mass distribution. Modelling of this distribution to reproduce the lensed images not only aids in measuring the dark matter content of the foreground galaxy, but allows for the reconstruction of the background source in order to study high-redshift galaxy populations. Modelling is typically performed by relatively slow parametric parameter-fitting techniques requiring manual inspection. However, upcoming large-scale surveys like the Legacy Survey of Space and Time (LSST) and *Euclid* will discover tens of thousands of strong lenses, with this vast quantity driving the growing use of automated machine learning for both identifying and modelling such systems extremely quickly. Convolutional neural networks (CNNs) can extract information from lensed images in order to predict parameters of the lens mass model, but require large simulated data sets for training and testing.

In this thesis, I explore the effectiveness of deep learning CNNs for estimating strong galaxy-scale lens mass model parameters when applied to upcoming wide-field survey data, investigating the practicalities they face and comparing and combining them with conventional modelling methods. I construct a CNN and train it on my own simulated lensing images with the imaging characteristics of the *Euclid* VIS band, LSST r-band, and LSST gri multiband. The CNN is trained to predict parameters of a smooth singular isothermal ellipsoid (SIE) projected mass profile that would best fit the observed image. Multiple aspects of this method are investigated, beginning with a comparison of its accuracy and reliability when applied to the survey data sets. The impact of multiband imaging is analysed, as well as the impact of lens light subtraction commonly performed by conventional modelling techniques. I show that for images including lens galaxy light, the CNN recovers the lens model parameters with an acceptable accuracy, with precision improved on average by 34 ± 5 per cent when lens light is subtracted. Additionally, the inclusion of multiband data improves performance regardless of lens light subtraction. While similar accuracies and precision are obtained for single epoch *Euclid* VIS and LSST r-band data sets, adding g- and i-band images to the latter increases precision by 24 ± 2 per cent without lens light and by 20 ± 2 per cent with lens light. I also examine the gains in performance through stacking images, and the impact of lens mass-light alignment. For the latter, when orientation and ellipticity of the lens light profile are allowed to differ from those of its mass profile, just as with real galaxies, the network performs most consistently when trained with a moderate amount of scatter between the two profiles.

I next seek ways of improving the method for application to real images, starting by implementing an existing technique to create a new Bayesian CNN that can predict both model parameters and their uncertainties. New data sets are then simulated for training and testing the network, allowing the network to either predict parameters of the SIE profile or predict those of the more general elliptical power law profile. To examine how the CNN performs at fitting these profiles to more complex mass models, these data sets feature a range of increasingly realistic lensing systems, from smooth parametric mass and light profiles to featuring real background sources, complex hydrodynamically-generated foreground mass distributions and line-of-sight structure.

In order to assess the suitability of the Bayesian CNN as a whole, I compare its performance when tested on these data sets to that of a conventional fitting method: the semilinear inversion technique PyAutoLens. In addition, I present a method for combining the network with such inversion methods where the CNN provides initial priors on the latter's parameters. Across the test sets, I find that the CNN achieves errors 19 ± 22 per cent lower than when applying PyAutoLens blindly. Compared to PyAutoLens alone, the initial centring of its priors on CNN-predicted parameters instead achieves 27 ± 11 per cent lower errors. If the prior widths are additionally initialised according to CNN-predicted Bayesian uncertainties, errors are reduced further to 37 ± 11 per cent as the uncertainties help it to avoid local minima in parameter space, with errors also 17 ± 21 per cent lower than the CNN by itself. While the CNN is undoubtedly the fastest modelling method, the combination of the two increases the speed of conventional fitting alone by a factor of 1.73 and 1.19 with and without CNN-predicted uncertainties, respectively. This, combined with greatly improved precision, highlights the benefits obtainable through combining neural networks with conventional techniques in order to achieve an efficient automated modelling approach. I finish off this thesis by discussing the scientific applications requiring such an approach, and examining how tightly these methods can constrain cosmological parameters through the modelling of double source plane lenses to aid in our understanding of the Universe.

Acknowledgements

First and foremost, I would like to say a massive thank you to my supervisor Simon. Starting all the way back as my undergraduate tutor, you have continuously guided and inspired me over these past eight years, for which I am eternally grateful. I cannot thank you enough for all your help, from assisting me in my Master's project and PhD research papers to providing me with so much advice over the course of writing this thesis. You have been an amazing tutor, supervisor and friend, and it is safe to say I would not be here without you, so once again, thank you.

Next, I would like to thank everyone in CAPT for being such a wonderful and inclusive group of boundless joy and enthusiasm, and notably those who allowed me to do this PhD in the first place. You all deserve heaps of praise, and I hope to see you at the Christmas Party, which hopefully will not be cancelled for 'fire' or 'global pandemic' reasons this time. To Nan and Jacob, thank you for all your aid and humour over the past few years: I could not have achieved this work without it, and I wish you guys all the best. Plus a special mention to Frazer, whose Scientific Computing module inspired me to pursue gravitational lensing in the first place.

Undoubtedly, another massive thank you goes to my family. You have all been beside me the whole way, inspiring me, improving my confidence, and encouraging me to take opportunities when they arrive. You have helped me revise for exams, practice talks, and prepare for interviews. You have been enthusiastic about astronomy and keen to learn about my research, and during stressful times you have been there to remind me to take a step back and have a break now and again. Thank you for all your help with this thesis and my other works, and for bringing so much humour and joy over the past few years. To Mum and Dad, where to begin? Your support and faith in me has been incredible, and not a day goes by that I am not grateful to have you as my parents. Thank you for encouraging me to pursue this path, and instilling in me a dedication to working hard to achieve my dreams.

To all those above, and to everyone I have most certainly forgotten to mention, this has been a lifelong journey and you have made it all worthwhile.

Thank you.

Publications

The work contained in this thesis is original research, and has not been submitted for any other degree or professional qualification. The work herein is my own except when otherwise explicitly stated.

This thesis presents material that has been published in the following papers:

1. Pearson, J., Li, N. and Dye, S., 2019. The use of convolutional neural networks for modelling large optically-selected strong galaxy-lens samples. Monthly Notices of the Royal Astronomical Society, 488(1), pp.991-1004.
2. Pearson, J., Maresca, J., Li, N. and Dye, S., 2021. Strong lens modelling: comparing and combining Bayesian neural networks and parametric profile fitting. Monthly Notices of the Royal Astronomical Society, 505(3), pp.4362–4382.

Specifically, Chapter 1 contains material from (2.). The work in Chapter 2 is based on (1.). The contents of Chapters 3 & 4 were published in (2.), with Figures 3.1 and 3.2 adapted from (2.). Chapter 5 features material from both (1.) and (2.).

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 General Background

The aims of extragalactic astronomy and cosmology have always been to understand galaxies and the larger scale structures of the Universe, and how they evolve over their lifetime since the Universe's inception. Ever since the discovery of objects beyond the Milky Way, astronomers and cosmologists alike have sought to delve further and further back through cosmological time to discover the building blocks of the Universe as it is today. A major step forward in this endeavour came with the discovery of discrepancies between spiral galaxy rotation curves and their observed stellar masses (Rubin and Ford Jr, 1970), supporting the existence of dark matter which plays a key role in the formation of structure on galactic and cosmological scales. Similarly impactful was the discovery of the accelerated expansion of the Universe in the 1990s, inferring the presence of dark energy through observations of Type Ia supernovae as standard candles (Riess et al., 1998).

Decades later, we now have a deeper understanding of galaxy evolution and structure formation, with density fluctuations in the early Universe leading to the formation of small dark matter haloes and filaments (Lacey and Cole, 1993; Schneider, 2006; Frenk and White, 2012; Conselice, 2014). The resulting hierarchical growth of these structures, along with the galaxies primarily within them, produces the large-scale structure we see today (Liddle and Lyth, 2000; Kraljic et al., 2018) as well as the dark matter substructures found within the mass density profiles of galaxies and galaxy clusters (Moore et al., 1999a; Ibarra et al., 2019; Haggar et al., 2021). However, many questions remain as yet unanswered regarding the exact mechanisms behind this, which are compounded by our lack of understanding of dark matter and dark energy.

While many possibilities have been proposed as to the nature of dark matter (Bertone et al., 2005; Peter, 2012), cold dark matter (CDM) has so far proved to be the most enduring. Being non-relativistic, these particles quickly coalesce under gravity into potential wells in which galaxies grow, allowing small-scale structure to form rapidly in the early Universe (Blumenthal et al., 1984). Weakly interacting massive particles (WIMPs) are the current primary candidates: non-relativistic theoretical particles predicted by extensions to the standard model that only interact through gravity (Roszkowski et al., 2018). Meanwhile, the simplest form of modelling dark energy is through the cosmological constant, Λ, of general relativity, representing the energy density of space (the ‘vacuum energy’). Together, the ΛCDM cosmological model successfully predicts a wide array of observations, from baryon acoustic oscillations (BAOs) and cosmic microwave background (CMB) polarisation to large-scale structure and accelerated expansion (Perivolaropoulos and Skara, 2021).

ΛCDM is also generally supported by simulations of the growth of structure over cosmic time, such as those produced by N-body simulations that model how N particles dynamically evolve (Bagla, 2005). While these can involve, for example, galaxies and gas clouds, many N-body simulations consider only gravity and aim to simulate the behaviour of dark matter. As such, they have been used to study the formation and topology of large-scale structure (Gott III et al., 1987; Blake et al., 2014), test various models of dark matter in order to compare with observations (Winther et al., 2019; Schwabe et al., 2020), investigate the clustering and shapes of dark matter haloes (Poole et al., 2016; Vega-Ferrero et al., 2017), and provide models of dark matter halo density profiles (Navarro, 1996; Mostoghiu et al., 2019). For galaxy evolution, hydrodynamical models are often chosen instead (Vogelsberger et al., 2020). These are based on N-body simulations but take into account physical processes such as the hydrodynamical motion of baryonic matter in order to study galaxy properties, such as those between galaxy morphology and colour (Correa et al., 2017), and compare to observations (Bignone et al., 2020) or other simulations (Haggar et al., 2021). They run over a large range of dynamical scales, making them computationally expensive, and so handle any physics below their resolution (such as star formation and AGN feedback) semi-analytically.

However, while ΛCDM simulations are in agreement with observations regarding large-scale structure, they still suffer from a number of problems. For example, CDM simulations used to overpredict the dark matter power spectrum on small scales, producing too high a number of dark matter subhaloes around galaxies compared to observed dwarf galaxy populations like those of the Milky Way (known as the 'missing satellites problem': see Klypin et al., 1999; Moore et al., 1999b). However, improved observations and simulations now yield too many observed subhaloes ('too many satellites problem': see Kim et al., 2018; Kelley et al., 2019; Kim and Peter, 2021), with such simulated satellites being too massive and too dense (known as the 'too-big-to-fail problem': see Boylan-Kolchin et al., 2011, 2012). Likewise, simulated CDM haloes around dwarf, irregular and low surface brightness galaxies all have more cuspy profiles than their observed flatter density profiles, especially concerning given these should all be dominated by dark matter (known as the 'core-cusp' or 'cuspy halo' problem: see Flores and Primack, 1994; Moore, 1994). Given these issues, and that experimental searches have yielded no convincing direct detections of WIMPs, there is now some doubt that these particles make up the dark matter inferred from observations (Bertone and Tait, 2018). Instead, some are looking to alternative particles such as axions (Nagano et al., 2019) or to other possibilities including primordial black holes (Jedamzik, 2020) and theories of modified gravity (Belgacem et al., 2019).

In addition to this, dark energy as the cosmological constant still faces major outstanding problems. Most notably, current quantum field theories predict a value of the constant 120 orders of magnitude larger than that needed to explain the Universe's observed accelerated expansion. This is known as the 'cosmological constant fine-tuning problem' (Weinberg, 1989; Del Popolo and Le Delliou, 2017), which could necessitate the need for an additional cosmological parameter to counteract this imbalance.

In order to constrain the nature of dark matter and dark energy, as well as study large-scale structure and galaxy evolution, many observational methods have been developed over the past few decades. For example, anisotropies in the CMB suggest a flat Universe, indicating its mass-energy density should equal the critical density. As measurements of the CMB angular power spectrum show that matter can only account for around 30 per cent of this (Ade et al., 2014), the presence of additional (dark) energy is required to explain these observations. Additionally, odd and even peaks in the CMB power spectrum correspond to regions of maximal and minimal density, respectively, and ratios of odd to even peaks provide measures of the relative densities of baryonic and non-baryonic matter in the Universe (Tegmark et al., 2004).

Meanwhile, the separations of galaxies highlight a preferred length scale for variations in the density of large-scale structure, arising from BAOs in the matter power spectrum (Bassett and Hlozek, 2010). Measurements of these separations can therefore be used to study dark matter through constraining cosmological parameters such as the matter density, Ω_m (Percival et al., 2007). Additionally, BAOs provide distance scales that can be used as standard rulers in order to measure cosmological distances through the expanding Universe. Likewise, the fixed absolute magnitudes of Type Ia supernovae allows them to be utilised as standard candles (Kowal, 1968) for this purpose. Combined with redshift measurements, these methods provide a distance-redshift relation that is used to infer cosmological models of dark energy (Huterer and Shafer, 2017).

Properties of dark energy can also be investigated through the constraining of cosmological parameters such as the Hubble constant, itself a measure of the expansion rate of the Universe. Many techniques have seen use in this area, from 'Early Universe' CMB anisotropy measurements (Aghanim et al., 2020) to 'Late Universe' distance ladders calibrated by standard candles to measure the expansion rate of the Universe (Riess et al., 2016). However, there remain significant discrepancies between measured values: this has become known as the 'Hubble tension', in which early and late time measurements disagree by at least 4σ (e.g. Knox and Millea, 2020; Di Valentino et al., 2021; Freedman, 2021). Gravitational lensing has also been used in this area, through the measuring of time delays between multiple images of distant quasars (see Section 1.1.2).

The properties of different dark matter candidates produce different distributions of dark matter at various scales (such as due to the low free-streaming length of CDM) that evolve differently with time. In order to shine a light on the properties of dark matter and its influence on galaxy and structure formation, the dark matter power spectrum and its evolution over cosmic time must therefore be well understood, from the lowest mass subhaloes to the large-scale structures of the cosmic web. As such, a number of techniques are used to measure and constrain dark matter, including those already discussed (for reviews of such methods, see e.g. Sumner, 2002; Pécontal et al., 2009; Baudis, 2018; Tao, 2020). For example, we have already seen that measurements of spiral galaxy rotation curves indicate the presence of dark matter haloes, and the same holds true for elliptical galaxy velocity dispersions. Additionally, x-ray emission from galaxy clusters indicate the presence of large quantities of hot gas that can only be explained should the clusters have a large dark matter component (Allen et al., 2002). Mass density profiles of galaxy clusters also provide powerful tools when accurately modelled through gravitational lensing, with measured cosmic shear providing constraints on the dark matter density and the power spectrum amplitude σ_8 (e.g. To et al., 2021). Additionally, strong gravitational lensing of individual galaxies can constrain their two-dimensional projected dark matter profiles, with their halo masses and substructures informing galaxy evolution models and dark matter simulations (see Section 1.1.2).

As gravitational lenses can help us understand the influence and properties of both dark matter and dark energy, we will explore these objects and their usage in more detail over the following sections.

1.1.1 Overview of Gravitational Lensing

In Einstein's general theory of relativity, the phenomenon of gravitational lensing arises as a consequence of the curvature of spacetime deflecting the paths of light rays. Light rays from a source such as a high-redshift galaxy travel along the shortest path through spacetime (a null geodesic, equivalent to a path of stationary traversal time in Fermat's principle), while massive objects curve spacetime around them according to their gravitational field. When such an object is situated between an observer and the light source, the paths of passing light rays become curved as they follow their geodesics through a now-curved spacetime. Each ray in a small circular bundle of light rays follows a slightly different curved path, distorting the bundle into an ellipse. This can result in the light source appearing magnified and distorted in shape from the point of view of the observer. Such distortions are generally similar to those produced by convex lenses in optics, and so these massive intervening objects are called gravitational lenses.

While the possibility of gravity deflecting light had been proposed prior to Einstein (e.g. Soldner, 1804), the impact of general relativity led to significantly larger deflections than had previously been predicted. Such deflections were later confirmed by Frank Dyson and Arthur Eddington for stellar light rays passing around the Sun during a total solar eclipse (Dyson et al., 1920), resulting in a change in the apparent positions of background stars. This is one example of gravitational microlensing, in which stars and even planets can act as lenses, changing the observed brightness or positions of background stars as they distort their shapes on unresolvable scales. Beyond the Solar System, microlensing requires near perfect alignment between the observer, lens and source, and can only be inferred from the monitoring of light curves as opposed to direct imaging.

While it was determined that the Sun and similar stars were not massive enough to produce multiple or distorted images, Zwicky (1937) proposed that galaxies acting as both lenses and sources would have the mass and spatial separation to yield an observable effect. On these larger scales, light from background galaxies can be distorted by foreground galaxies, groups and even galaxy clusters. The strengths of these distortions depend on the amount and distribution of mass in the foreground lens, as well as the position of the source and the distances between the source, lens and observer (see Section 1.2 for more details). Depending on these quantities, the lensing effect can produce anywhere from minor distortions in the apparent shapes of background sources to producing multiple images of the same source. Hence, we call the former weak gravitational lensing, while the latter provides the definition of strong gravitational lensing.

Weak lensing arises when only minor distortions are made to the apparent shapes of sources, elongating a circularly symmetric source to no more than would be observed for the most elliptical unlensed galaxies. The detection of weak lensing therefore relies on analysing the ellipticities, and hence alignments, of many of these sources in a statistical manner, with the resulting shear measurements used to obtain the convergence and hence determine the lens mass distribution (see

Section 1.2.4). Weak lensing is most commonly utilised when modelling galaxy clusters (Schrabback et al., 2018; Umetsu, 2020; Hong et al., 2021) and measuring the cosmic shear due to large-scale structures (Hikage et al., 2019; Heymans et al., 2021; DES Collaboration et al., 2021), due to their wide mass distributions distorting the shapes of many background galaxies simultaneously, although it has also been applied to galaxy- and group-scale lenses (e.g. Luo et al., 2018; Chen et al., 2019b; Schrabback et al., 2021).

This thesis instead focuses on strong gravitational lensing. Here, a galaxy, group or cluster acts as a lens to greatly magnify the light of a background source galaxy, resulting in multiple images of the same galaxy. Additionally, the appearance of these images can be distorted into arcs of light, and when the lens and source are perfectly aligned along the observer's line of sight (LOS) the resulting images can form a ring or four-pointed cross around the lens, known as an Einstein ring and Einstein cross, respectively. The first gravitational lens of this type was not discovered until 1979, consisting of two lensed images of a quasar either side of a foreground galaxy (Walsh et al., 1979). Even then, almost a decade went by before the first Einstein ring was detected by Hewitt et al. (1988) at radio wavelengths. In this work, I consider only strong galaxy-scale gravitational lensing, wherein the foreground lens is a single galaxy, as such systems offer the simplest laboratories for testing the problems in astrophysics and cosmology presented in the following section. The theory and equations behind this are detailed in Section 1.2.

1.1.2 Uses of Strong Galaxy-Scale Gravitational Lensing

Strong galaxy-scale lensing has proven itself to be a highly versatile tool for studying various physical properties of the Universe. Measurements of the observed distortion allow for modelling of the projected mass density profile of the foreground galaxy, which contains information on the dark matter content and substructure within the lens (Sonnenfeld et al., 2015; Shu et al., 2017; Küng et al., 2018).

Advancements have been made towards detecting this substructure (Vegetti and Koopmans, 2009; Vegetti et al., 2014; Hezaveh et al., 2016; Bayer et al., 2018; Brehmer et al., 2019; Ritondale et al., 2019), with such properties aiding in galaxy evolution models (e.g. Bolton et al., 2012) to help shed light on how galaxy morphologies change with time.

Lensing maintains the surface brightness of sources, but the resulting magnified images allow for the probing of high-redshift source populations, especially if their original surface brightness distributions can be reconstructed. With the addition of redshift measurements, this too can provide valuable information on galaxy evolution, and as such has received a recent surge in interest (e.g. Dye et al., 2018; Lemon et al., 2018; McGreer et al., 2018; Rubin et al., 2018; Salmon et al., 2018; Sharda et al., 2018; Shu et al., 2018; Sharon et al., 2019; Collett and Smith, 2020; Khullar et al., 2020; Inoue et al., 2020). Reconstructing the unlensed morphology of a source is possible if the mass profile of the lens is well constrained (Warren and Dye, 2003; Suyu et al., 2006; Nightingale et al., 2018; Powell et al., 2021), and allows for a more in-depth study of its properties, for example, its rotation curves (Dye et al., 2015; Geach et al., 2018).

The projected mass profiles of the foreground lenses, when combined with other methods such as galaxy rotation curves, may be used to obtain approximate 3D (deprojected) mass density profiles. Such profiles can be of use in testing general relativity (Collett et al., 2018) and cosmological models (Eales et al., 2015; Krywult et al., 2017; Rana et al., 2017; Davies et al., 2018; Yang et al., 2020; Giani and Frion, 2020; Birrer et al., 2020; Shajib et al., 2020; Wang et al., 2020; Colaço et al., 2021; Luo et al., 2021). General relativity informs us that light rays passing through the potential well of a foreground lens at different positions experience differing time dilation in addition to following separate curved geodesics. Gravitational time delays and geometric path differences between lensed images, paired with variable sources such as quasars, hence provide measurable time delays that can constrain the value of the Hubble constant (H_0) irrespective of the distances to or between the galaxies (e.g. Suyu et al., 2017; Bonvin et al., 2017; Chen et al., 2018; Birrer et al., 2019; Liao et al., 2019; Taubenberger et al., 2019; Birrer and Treu, 2020; Wong et al., 2020; Denzel et al., 2021; Li et al., 2021). Recently, there has been much work on extending this to gravitationally lensed supernovae, which may provide far tighter constraints on H_0 (Collett et al., 2019; Oguri, 2019; Foxley-Marrable et al., 2020; Bag et al., 2021; Bayer et al., 2021). This is because their standardisable absolute brightnesses and well-understood light curves are less stochastic than quasars, their light curves evolve over weeks rather than years, and the fading of their light curves allows for easier reconstruction of the background source (Goldstein et al., 2018). There remains significant tension in the predicted value of the Hubble constant between techniques focusing on early-universe and late-universe physics (Freedman, 2017; Mörtsell and Dhawan, 2018; Bengaly et al., 2020; Pandey et al., 2020; Vagnozzi, 2020), with the uncertainties in lensing estimates being partially reduced through the modelling of larger strong lens catalogues.

1.1.3 Upcoming Surveys

Obtaining large catalogues of lenses has proven challenging due to their rarity; the multiple imaging of strong lensing requires the foreground and background galaxies to be almost perfectly aligned along an observer's line of sight. Surveys involved in identifying strong lenses have to date observed only hundreds, with most lying at low redshift. These include the Sloan Lens ACS (SLACS) survey (Bolton et al., 2006), the CFHTLS Strong Lensing Legacy Survey (SL2S; Cabanac et al., 2007), the Sloan WFC Edge-on Late-type Lens Survey (SWELLS; Treu et al., 2011), the BOSS Emission-Line Lens Survey (BELLS; Brownstein et al., 2011) and lenses found in the Dark Energy Survey (Dark Energy Survey Collaboration et al., 2005).

Upcoming surveys are expected to remedy this, producing billions of galaxy images containing tens of thousands of strong lensing systems (Collett, 2015). These include the European Space Agency's *Euclid* telescope (Laureijs et al., 2011), and the Rubin Observatory Legacy Survey of Space and Time (LSST) which will use the ground-based Vera C. Rubin Observatory (formerly the Large Synoptic Survey Telescope (Ivezić et al., 2008)). *Euclid* is due to launch in 2023 with the primary aim of measuring the acceleration of the Universe up to a redshift of $z = 2$ to study dark matter and dark energy. Using its visible-wavelength camera (VIS) and near-infrared photometer (NISP-P) and spectrograph (NISP-S), *Euclid* will map the distribution of up to two billion galaxies as it studies the large-scale structure of the Universe, covering 15,000 deg^2 over its six year mission. Meanwhile, LSST will begin science operations in 2023, covering around 18,000 deg^2 in six bands (u, g, r, i, z, y) every few nights, and repeatedly doing so over the course of ten years. Among other goals, this will allow the survey to study dark matter and dark energy on an unprecedented scale.

Combined, these surveys will generate many tens of petabytes of data only processable through automation or crowdsourcing. While the latter may help when it comes to identification, modelling thousands of lenses in an efficient manner will undoubtedly require automated modelling methods much faster than conventional parametric techniques.

1.2 Lensing Theory

Having provided an overview of strong gravitational lenses and a discussion of their uses in the previous section, the following details their theoretical framework, including the two-dimensional projected mass profiles used in modelling them. Derivations in this section primarily follow the work of Narayan and Bartelmann (1996) and Meneghetti (2006), with further information on gravitational lensing available from a range of sources (e.g. Schneider, 1992; Schneider et al., 1992;

Mollerach and Roulet, 2002; Kochanek, 2004; Schneider et al., 2006; Treu, 2010).

1.2.1 Effective Refractive Index

In Minkowski spacetime, the deflection of light rays resulting from gravitational lensing is very much analogous to deflections in geometric optics as light passes through a medium with a refractive index n. For a lens whose Newtonian potential Φ is much smaller than c^2 (i.e. $\Phi/c^2 \ll 1$, which is valid for all galaxy-scale lenses), its effective refractive index is

$$n = c/c' = \frac{1}{1 + \frac{2}{c^2}\Phi} \approx 1 - \frac{2}{c^2}\Phi \tag{1.1}$$

where $c' = 1 + \frac{2}{c^2}\Phi$ is the reduced speed of light through the potential. The angle of deflection is then defined as the gradient of n perpendicular to the path of the light ray, integrated along this path,

$$\vec{\hat{\alpha}} = -\int \vec{\nabla}_{\perp} n \; dl = \frac{2}{c^2} \int \vec{\nabla}_{\perp} \Phi \; dl. \tag{1.2}$$

This deflection is a factor of two larger than that expected by Newtonian gravity, as the refractive index is analogous to taking into account the local curvature of spacetime around the lens. As the deflection angle is small, we can simplify the above by instead integrating along an unperturbed ray with the same impact parameter, rather than along the deflected ray itself: this is called the Born approximation.

1.2.2 Thin Lens Approximation

To aid in calculating the deflections, it is important to note that the spatial extent of a lens along the line of sight is far smaller than the distances between the observer, lens and source. The deflections therefore occur over a comparatively small region, and so the lens can be approximated as a thin mass sheet with its mass distribution projected onto it along the line of sight. This is known as the thin lens approximation, with the mass sheet referred to as the lens plane. For a lens of density ρ, its projected surface density is

$$\Sigma(\vec{\xi}) = \int \rho(\vec{\xi}, z)\ dz \tag{1.3}$$

for a two-dimensional lens plane vector $\vec{\xi}$, and a distance z along the unperturbed path from the point of closest approach.

If we consider a lens to be a point source with mass M and a light ray along a path with impact parameter b, the lens will have a gravitational potential

$$\Phi(b, z) = \frac{GM}{(b^2 + z^2)^{1/2}}, \tag{1.4}$$

which results in a deflection angle of

$$\hat{\alpha} = \frac{2}{c^2} \int \vec{\nabla}_\perp \Phi\ dz = \frac{4GM}{c^2 b}. \tag{1.5}$$

Likewise, a circularly symmetric lens is found to produce a deflection angle of

$$\hat{\alpha}(\xi) = \frac{4GM(\xi)}{c^2 \xi} \tag{1.6}$$

for a light ray at radius ξ, with the total mass contained within this radius given as

$$M(\xi) = 2\pi \int_0^\xi \Sigma(\xi')\xi'\ d\xi'. \tag{1.7}$$

Within this approximation, we can then calculate the deflection angle at a position $\vec{\xi}$ for a more complex lens, by considering the contributions from all the mass elements (treated as point sources) at positions $\vec{\xi}'$:

$$\vec{\hat{\alpha}}(\vec{\xi}) = \frac{4G}{c^2} \int \frac{(\vec{\xi} - \vec{\xi}')\Sigma(\vec{\xi}')}{|\vec{\xi} - \vec{\xi}'|^2} \, d^2\xi', \tag{1.8}$$

having defined $\vec{b} \equiv \vec{\xi} - \vec{\xi}'$ to be the impact parameter of each mass element.

1.2.3 The Lens Equation

When describing the effects of gravitational lensing, it is useful to consider a graphical representation of the deflection of a light ray such as the one shown in Figure 1.1. Here, light rays originating from a source pass by a foreground lens of mass M, whose distortion of spacetime changes their trajectory such that they reach a distant observer. To this observer, the rays appear as if they originated from another point in space, at the location of their lensed images. The distances between the observer and the lens, the observer and the source, and the lens and the source are given by D_L, D_S, and D_{LS}, respectively. These are given as angular diameter distances, noting that $D_{LS} \neq D_S - D_L$ in general for a curved spacetime.

In this diagram, the optic axis passes through the centre of the lens, with the source a distance $\vec{\eta}$ away from the axis on the source plane, and the lensed image a distance $\vec{\xi}$ away from the axis on the lens plane. We use $\vec{\theta} = \vec{\xi}/D_L$ and $\vec{\beta} = \vec{\eta}/D_S$ for the angular positions of the lens and source, respectively, and from geometric identities we obtain the relation,

$$\vec{\theta} D_S = \vec{\beta} D_S + \vec{\hat{\alpha}} D_{LS} \tag{1.9}$$

assuming the angles involved are small compared to the distances involved. Introducing the reduced deflection angle

$$\vec{\alpha}(\vec{\theta}) = \frac{D_{LS}}{D_S} \vec{\hat{\alpha}}(\vec{\theta}), \tag{1.10}$$

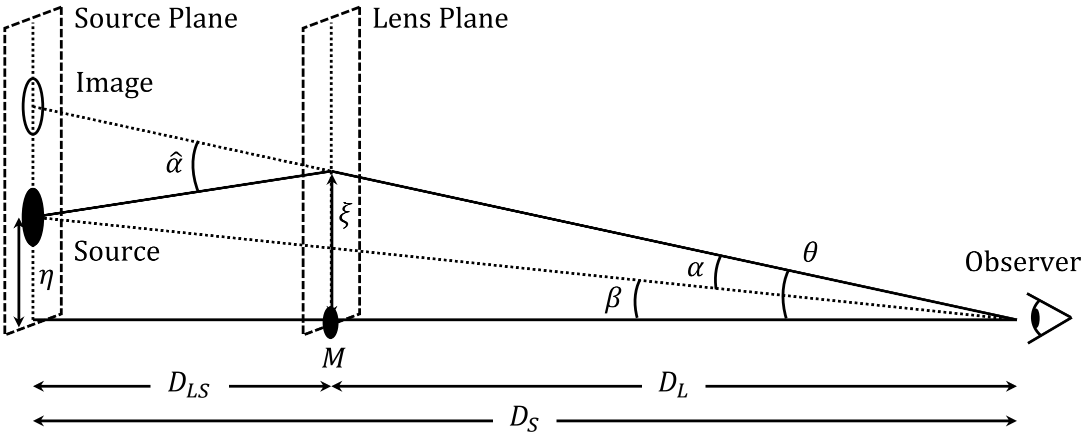

Figure 1.1: Ray diagram for gravitational lensing. With respect to the line of sight of a lensing object of mass M, rays (solid line) from a source at angle β change direction as they pass close to the lensing object. This results in the rays appearing to originate from an apparent source at angle θ.

we arrive at the *lens equation*

$$\vec{\beta} = \vec{\theta} - \vec{\alpha}(\vec{\theta}), \tag{1.11}$$

which relates the lens and source planes through the deflection arising from $\Sigma(\vec{\theta})$, and whose non-linearity allows for multiple lensed images of the source to form around the lens.

For a circularly symmetric lens, we can consider a special case where the source is directly behind the lens ($\beta = 0$). Combining Equations 1.6 and 1.11 results in an angular deflection in the lens plane of $\theta = \theta_E$, where

$$\theta_E = \sqrt{\frac{4GM(\theta_E)}{c^2}\frac{D_{LS}}{D_S D_L}} \tag{1.12}$$

is the angular Einstein radius. In this case, the source is lensed into a ring fully encircling the lens, with its radius depending on the total mass contained within the ring as well as all three angular diameter distances. The Einstein radius is also a useful quantity for general strong lensing, as the magnification of images typically peaks for images on the ring, and multiple images of a source are often separated by around $2\theta_E$.

As it can be easier to work in dimensionless coordinates than physical ones, especially when implementing lensing computationally, we can define a length scale ξ_0 on the lens plane (for example, image pixels). Combined with the corresponding scale $\eta_0 = \xi_0 D_S / D_L$ on the source plane, we obtain the dimensionless forms of the vectors for the lens and source plane, respectively,

$$\vec{x} \equiv \frac{\vec{\xi}}{\xi_0}, \quad \vec{y} \equiv \frac{\vec{\eta}}{\eta_0}. \tag{1.13}$$

We also define a scaled deflection angle

$$\vec{\alpha}(\vec{x}) = \frac{D_L D_{LS}}{\xi_0 D_S} \vec{\hat{\alpha}}(\xi_0 \vec{x}) = \frac{1}{\pi} \int \frac{(\vec{x} - \vec{x}')\kappa(\vec{x}')}{|\vec{x} - \vec{x}'|^2} d^2 x' \tag{1.14}$$

where $\kappa(\vec{x}) = \Sigma(\vec{x})/\Sigma_{cr}$ is the dimensionless surface density (or convergence), and

$$\Sigma_{cr} = \frac{c^2}{4\pi G} \frac{D_S}{D_L D_{LS}} \tag{1.15}$$

is the critical surface density. A lens which satisfies the condition $\Sigma(\vec{x}) > \Sigma_{cr}$ anywhere indicates that it is capable of producing multiple images, with such a condition being a requirement for typical strong lenses (but not strictly necessary should they possess certain unusual density profiles: see Subramanian and Cowling, 1986). Finally, we can hence rewrite the lens equation as

$$\vec{y} = \vec{x} - \vec{\alpha}(\vec{x}) \tag{1.16}$$

which allows computed deflection angles (see later) to map positions on the lens plane to those on the source plane. The mapping is performed in this direction as the non-linear lensing effect allows multiple lens plane positions to correspond to a single source plane position.

1.2.4 Magnification and Critical Curves

While lensing changes the size and shape of the source, it conserves its surface brightness, increasing the total source flux and leading to a magnified, if distorted, view of the source. The appearance of the lensed source depends on properties of the lens mass profile, the angular diameter distances involved, and the position of the source relative to the lens along the line of sight. For the latter, this is because the level of magnification varies according to the surface mass profile of the lens. By considering the limit of point sources, the local behaviour of the nonlinear mapping between source and image positions in Equation 1.11 can be linearised. As such, the distortion can be described by the Jacobian matrix,

$$A \equiv \frac{\partial \vec{\beta}}{\partial \vec{\theta}} = \left(\delta_{ij} - \frac{\partial \alpha_i(\vec{\theta})}{\partial \theta_j} \right) \tag{1.17}$$

for components (i, j) on the lens plane, where δ_{ij} is the Kronecker delta function. The determinant of the Jacobian gives the local solid-angle distortion, and hence we can quantify the magnification μ by its inverse,

$$\mu = \frac{\text{image area}}{\text{source area}} = \frac{\delta\theta^2}{\delta\beta^2} = \frac{1}{\det A} = \frac{1}{(1-\kappa)^2 - |\gamma|^2}, \tag{1.18}$$

which can be described by two quantities: the convergence κ and shear γ. These respectively dictate the isotropic re-scaling of the source size and the amount by which the lensed image is stretched (sheared) in a given direction (note that the observed magnification of extended sources will also depend on their surface brightness distribution).

From analysing these quantities, we find regions of infinite magnification on the lens plane, defined by two lines: the tangential and radial critical curves. Images forming along a tangential or radial critical curve are greatly distorted tangentially or radially to this line, respectively. Such images are formed when sources lie along corresponding lines in the source plane, known as tangential and radial

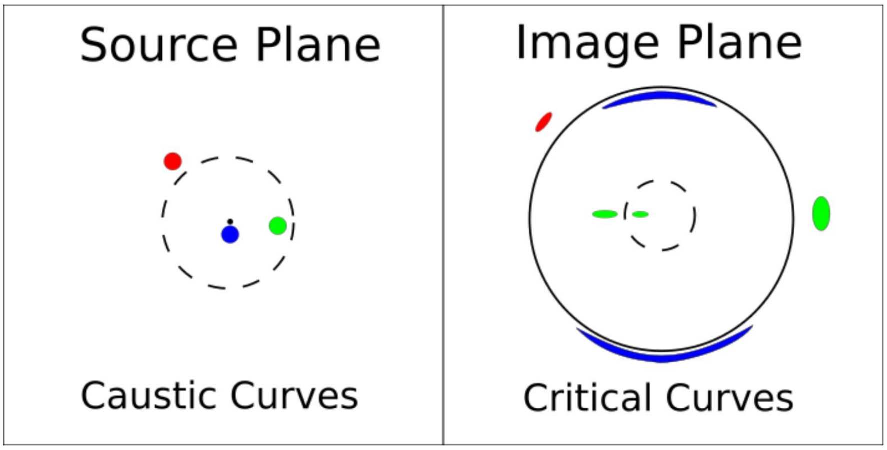

Figure 1.2: Source and image configurations for a circularly symmetrical non-singular lens. The left diagram shows the sources located around caustics, while the right diagram shows the resulting lensed images positioned around critical curves. Each coloured source produces the corresponding coloured images. (Figure credit: Schäfer et al., 2020)

caustics. The number of images increases by two each time the source enters the region of a caustic, although the innermost image is usually too faint to be observed. Circularly symmetric lenses have point-like tangential caustics, while those of elliptical lenses take the form of an astroid with four cusps. Examples of these lines and possible image configurations are shown in Figures 1.2 and 1.3.

As a source approaches a foreground circularly symmetrical lens, it experiences only minor changes to its magnification and shape (weak lensing) but to an observer appears to remain outside the tangential critical curve. Crossing the outer (radial) caustic results in two more images of the source; these initially form together as a single radially-elongated image on the opposite side of the lens at the radial critical curve before separating. When the source nears the inner (tangential) caustic, the innermost image demagnifies while the two outer images experience strong tangential magnification. These eventually form an Einstein ring along the tangential critical curve once the source is aligned directly behind the lens.

While much of this remains the same for elliptical lenses, some changes occur due the elongation of the critical curves and the presence of the cuspy tangential

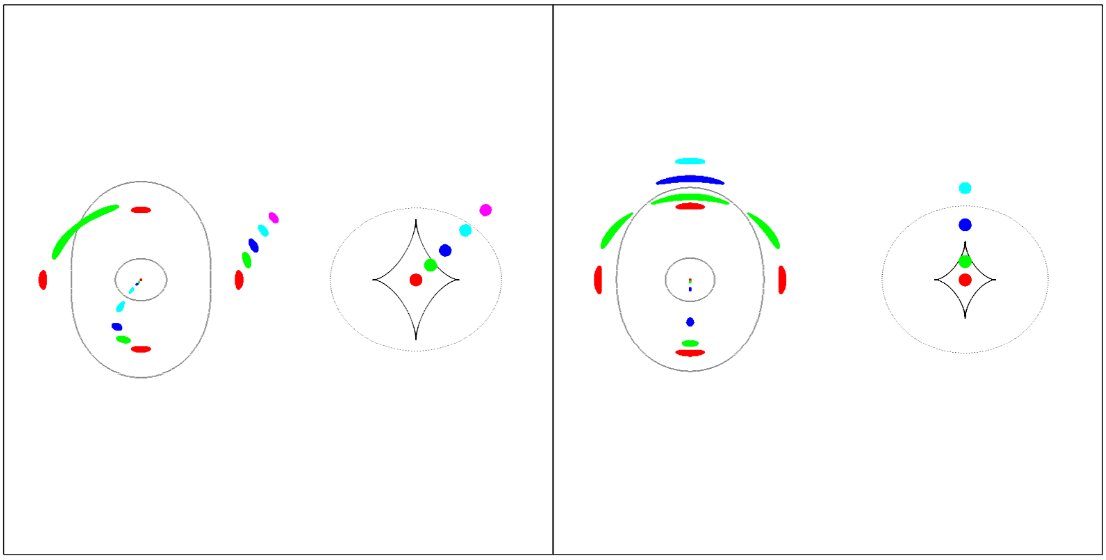

Figure 1.3: Source and image configurations for an elliptical lens. Left panel: source crossing a fold caustic. Right panel: source crossing a cusp caustic. Each panel contains two diagrams: the left diagram shows the lensed images positioned around critical curves, while the right diagram shows the sources located around caustics. Each coloured source produces the corresponding coloured images. (Figure credit: Narayan and Bartelmann, 1995)

caustic. After the source passes the outer caustic, the positions of the two new images can vary depending on the source's angle of approach. As the source reaches and crosses the inner caustic, a new image appears at the tangential critical curve, strongly magnified along it before again splitting into two. Once the source aligns with the lens, the four outer images now form an Einstein cross, with two images positioned inside the tangential critical curve and two outside.

Clearly, the mass distribution of the lens changes the appearance of the source images, with parameters like its ellipticity capable of producing different image configurations. This relationship between mass distributions and observations is hence something that can be learned by a machine, as we shall explore later in this thesis.

1.2.5 Lens Mass Profiles

As we have seen, the magnification and deflection angle depend upon the two-dimensional mass distribution of the foreground lens. If we assume that matter within galaxies behaves like an ideal gas in thermal and hydrostatic equilibrium, a simple solution can be obtained in which the density at a given radius r from the centre is given by

$$\rho(r) = \frac{\sigma_v^2}{2\pi G r^2}, \tag{1.19}$$

where σ_v is the one-dimensional velocity dispersion which is constant across the galaxy (Narayan and Bartelmann, 1996). This is known as the singular isothermal sphere (SIS) model, and corresponds to a projected surface density of

$$\Sigma(\xi) = \frac{\sigma_v^2}{2G\xi}. \tag{1.20}$$

Such a profile lacks a core radius and instead contains a singularity at its centre, causing the innermost (demagnified) lensed image to vanish. Despite this singularity, it is capable of reproducing the flat rotation curves of spiral galaxies (Keeton, 2001; Meneghetti, 2006) and continues to provide a good fit to the mass profiles of early-type galaxies that make up the majority of strong galaxy-scale lenses (e.g. Koopmans et al., 2006, 2009; Auger et al., 2010a). This, combined with the profile's simplicity (its lens equation is simply $\vec{y} = \vec{x} - \vec{x}/|\vec{x}|$) makes it a common choice for modelling such lenses.

As galaxies are frequently non-spherical in shape, a model also popular for strong galaxy-scale lensing is the singular isothermal ellipsoid (SIE; e.g. Kormann et al., 1994). This expands on the SIS model by introducing ellipticity, $e = 1 - q$ (where q is the semi-minor to semi-major axis ratio), along a preferred orientation, and is the model used most extensively in this thesis. I provide here the forms of isothermal ellipsoid deflection angles presented in Keeton (2001), which have been separated into horizontal and vertical components (α_1, α_2) along with positions on the lens plane (x_1, x_2):

$$\alpha_1(x_1, x_2) = \frac{b_n q}{\sqrt{1-q^2}} \tan^{-1} \frac{\sqrt{1-q^2} x_1}{\psi + q^2 s}, \tag{1.21}$$

$$\alpha_2(x_1, x_2) = \frac{b_n q}{\sqrt{1-q^2}} \tanh^{-1} \frac{\sqrt{1-q^2} x_2}{\psi + q^2 s}, \tag{1.22}$$

where $\psi^2 = q^2(s^2 + x_1^2) + x_2^2$, and s is the core radius, set to zero for SIE profiles[1]. The normalisation factor

$$b_n = 4\pi \left(\frac{\sigma_v}{c}\right)^2 \frac{D_{LS}}{D_S} \tag{1.23}$$

is equal to the angular Einstein radius in the limit of a singular, spherical model. For lenses rotated by an angle ϕ, the ellipticity be separated into two components of complex ellipticity,

$$e_1 = \frac{1-q^2}{1+q^2} \cos 2\phi\,, \quad e_2 = \frac{1-q^2}{1+q^2} \sin 2\phi\,. \tag{1.24}$$

Such deflections were utilised in this thesis to simulate images containing galaxy-scale lenses. More details on this simulation process are given in Section 2.1.2.

In addition to the SIE model, I also utilise a more generalised form beginning in Chapter 3: the power law profile. The density distribution of this model follows $\rho \propto r^{-n}$, having introduced an extra lens model parameter, the power law slope n. This model covers most of the simple, physically interesting density profiles, with the velocity dispersion now a function of radius. Spherical power law models produce the deflection angles and convergences

$$\alpha(\theta) = b_n \left(\frac{\theta}{b_n}\right)^{2-n}, \quad \kappa(\theta) = \frac{3-n}{2} \left(\frac{\theta}{b_n}\right)^{1-n}, \tag{1.25}$$

---

[1] A limit of defining such elliptical mass profiles is that if $q = 1$ then $(\alpha_1, \alpha_2) \to \infty$. However, as $q \to 1$, the arctan and arctanh functions aid in preventing the deflections from growing to infinity, and so in practice, one can set values of q arbitrarily close to 1 while maintaining finite deflections.

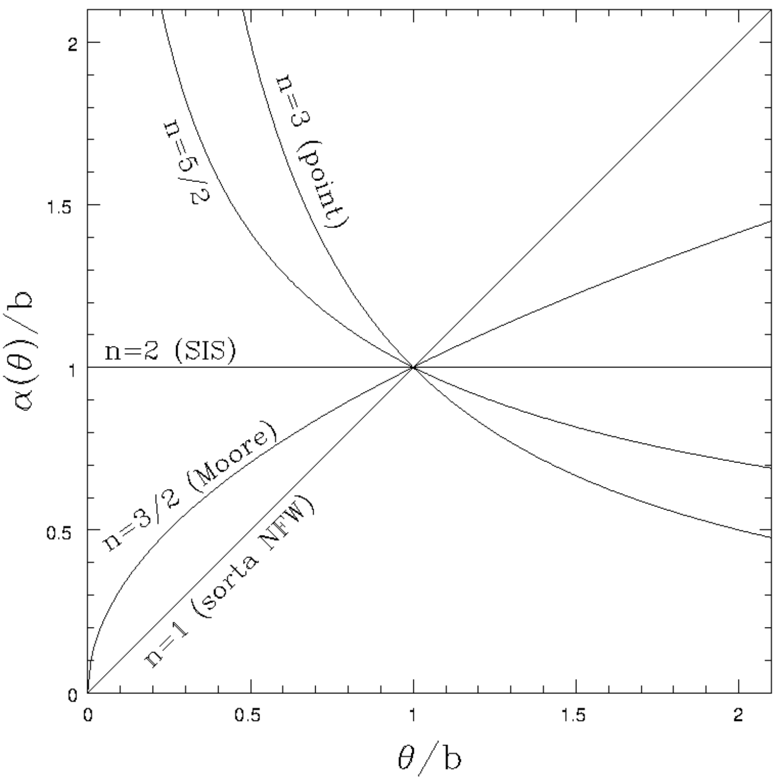

Figure 1.4: Deflection angles of power law mass profiles as functions of angular radius on the lens plane (both in units of the normalisation b_n), for various power law slopes n ranging from 1 to 3. Compared to the SIS model, deflections of more extended profiles ($n < 2$) fall to zero at the lens centre while those of more concentrated profiles ($n > 2$) diverge. (Figure credit: Kochanek, 2004)

where the definition of the normalisation b_n depends on the chosen slope (Schneider et al., 2006). A value of $n = 2$ corresponds to an SIS model, which approaches a point mass lens as $n \to 3$ (and hence $\kappa = 0$ with a central singularity). If instead $n \to 1$, the profile instead becomes similar to the inner region of an NFW model (Navarro, 1996) commonly used for dark matter haloes, although here κ tends to a constant unlike the central cusp of the NFW model (for which $\kappa \propto \ln \theta$). Examples of how the deflections change with slope are given in Figure 1.4. A variable slope can allow such profiles to provide better fits to many galaxy lenses; studies have shown that while close to isothermal, the total mass density profiles of massive elliptical galaxies tend to have slightly higher power law slopes (Dutton and Treu, 2014; Cappellari et al., 2015; Jee et al., 2015). For example, Auger et al. (2010a) obtained values of $\langle n \rangle = 2.08 \pm 0.03$ and scatter $\sigma_n = 0.16 \pm 0.02$ for 73 early-type galaxies at $z < 0.5$.

Modelling lenses with power law profiles can also aid in galaxy evolution models. For example, while Koopmans et al. (2009) and Barnabè et al. (2011) found no significant evolution of slope with redshift up to $z = 0.5$ and $z = 0.3$, respectively, Ruff et al. (2011) analysed 11 early-type galaxy strong lenses and found a mild negative trend between power law slope and redshift for $z = 0.2 - 0.9$. This was supported by Sonnenfeld et al. (2013), who investigated this evolution of power law slope for tens of massive early-type galaxy lenses at $z = 0.2 - 0.8$ and found that for a fixed stellar mass and effective radius the slope became steeper at lower redshifts. While no evidence for a redshift-dependent slope was found by Cui et al. (2017) for 118 gravitational lenses combined with other observations, mild evolutions were once again suggested in work by Holanda et al. (2017) and Chen et al. (2019a), with the latter studying 161 strong galaxy-scale lenses. These trends hence suggest that dissipative processes may contribute to the evolution of early-type galaxies since $z = 1$, such as through dissipative mergers causing baryons to cool and fall to the galaxy centre (Ruff et al., 2011).

Using the formalism of Tessore and Metcalf (2015) for the elliptical power law model used in this thesis, we rewrite the convergence as

$$\kappa(R) = \frac{3-n}{2}\left(\frac{b_n}{R}\right)^{n-1}, \tag{1.26}$$

where $R = \sqrt{q^2 x_1^2 + x_2^2}$ is the elliptical radius. We also define the elliptical angle $\varphi = \arctan(q x_1, x_2)$ such that the coordinate transformation is

$$x_1 = R/q \cos\varphi\,, \quad x_2 = R \sin\varphi\,. \tag{1.27}$$

Following the complex formulation of lensing used in Tessore and Metcalf (2015), which will not be detailed here, we obtain the complex deflection angle

$$\alpha(R,\varphi) = \frac{2b_n}{1+q}\left(\frac{b_n}{R}\right)^{n-2} \mathrm{e}^{\mathrm{i}\varphi}\,{}_2F_1\left(1, \tfrac{n-1}{2}; \tfrac{5-n}{2}; -\tfrac{1-q}{1+q}\,\mathrm{e}^{\mathrm{i}2\varphi}\right), \tag{1.28}$$

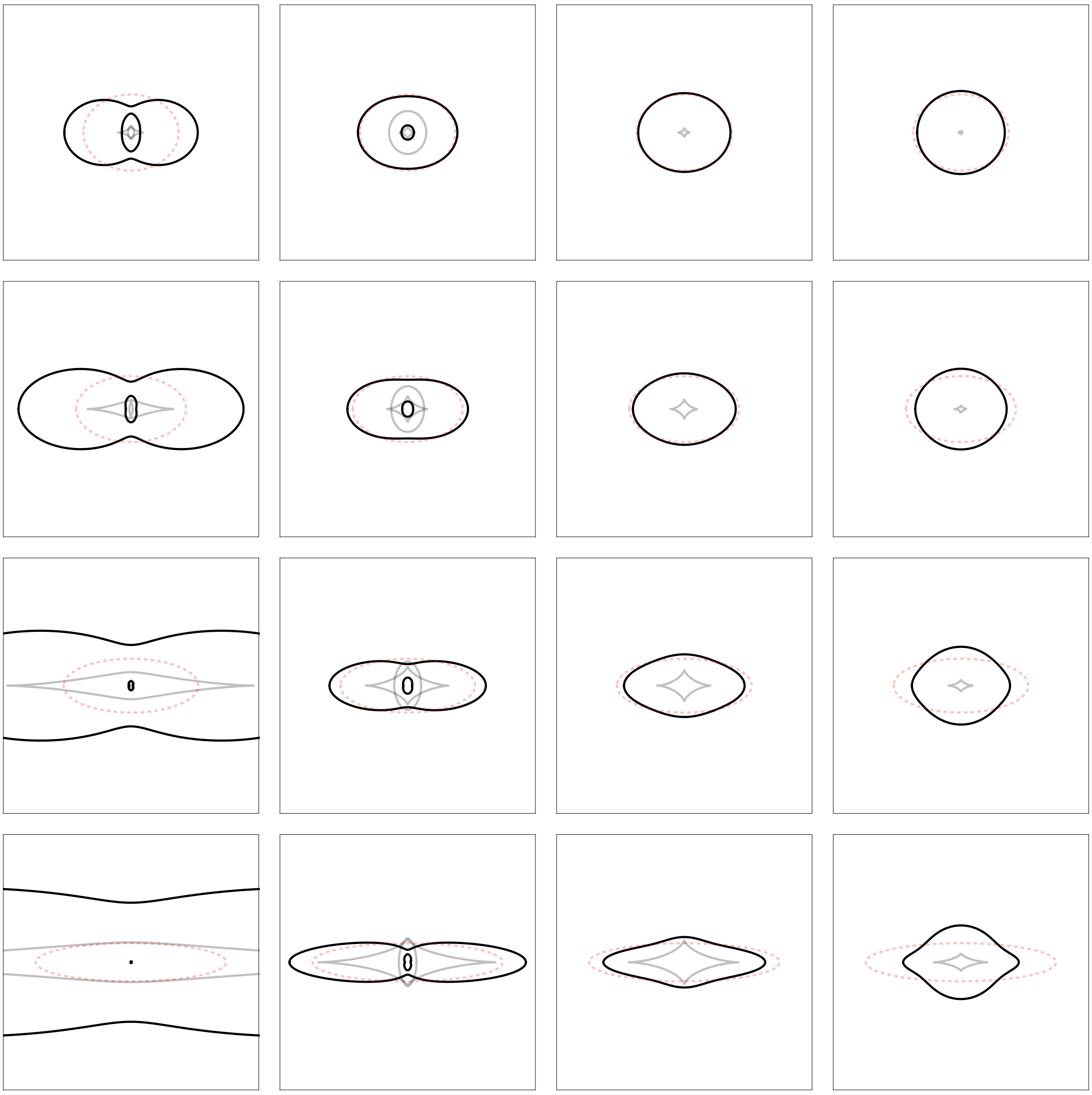

Figure 1.5: Critical lines (black) and caustics (grey) for elliptical power law profiles. Dotted red ellipses represent the axis ratio, and have semi-minor axes equal to b which, for the purpose of illustration, decreases as $q^{1/2}$. Left to right: power law slope n increasing from 1.25 to 2.75 in steps of 0.5. Top to bottom: axis ratio decreasing from 0.8 to 0.2 in steps of 0.2. (Figure credit: Tessore and Metcalf, 2015)

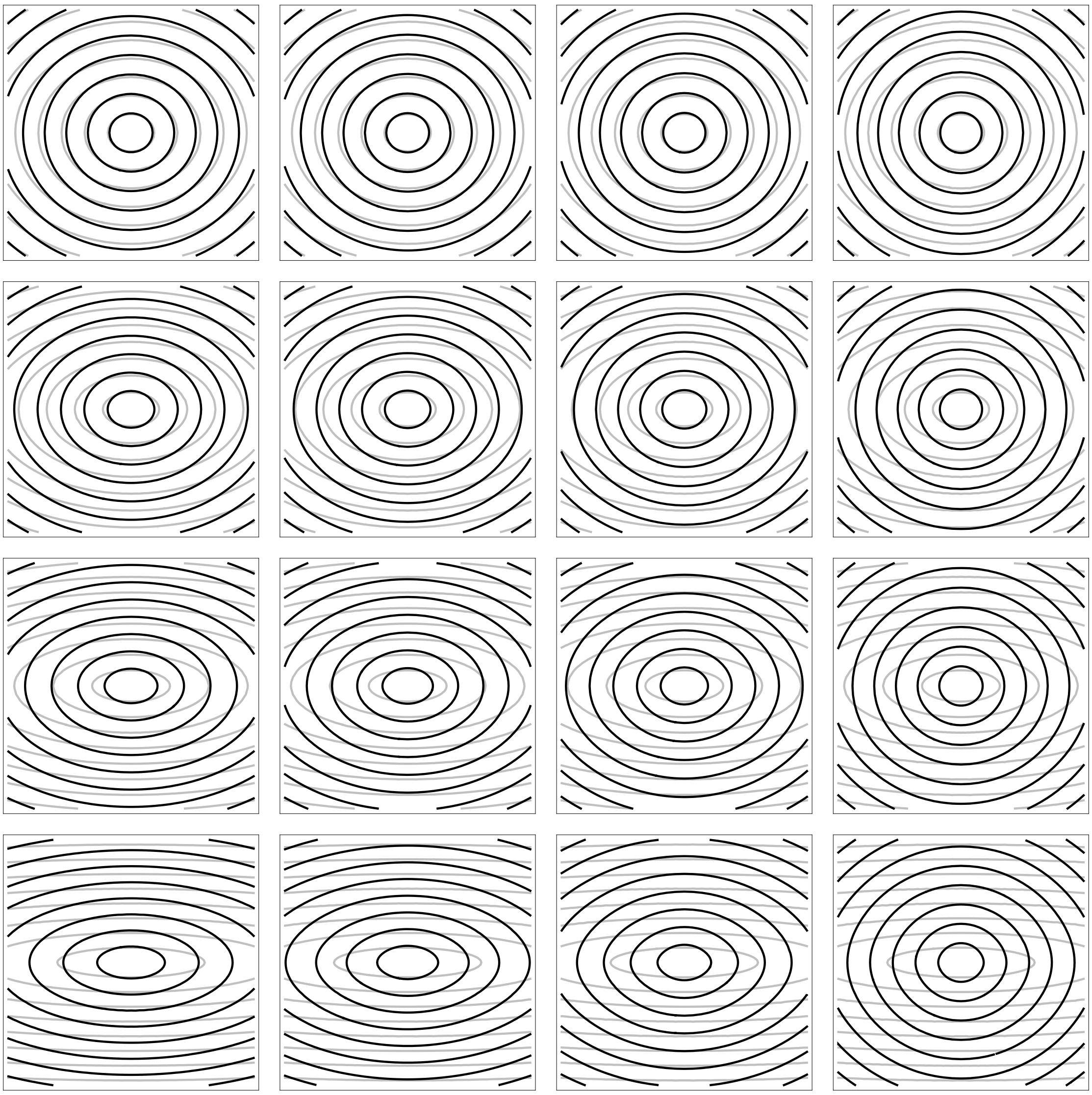

Figure 1.6: Isocontours of the elliptical power law profile's deflection potential (black) and surface mass density (grey). The power law slope varies from $n = 1.25$ (left) to $n = 2.75$ (right) in steps of 0.5, while the axis ratio varies from $q = 0.8$ (top) to $q = 0.2$ (bottom) in steps of 0.2. (Figure credit: Tessore and Metcalf, 2015)

where ${}_2F_1(a, b; c; z)$ is the Gaussian hypergeometric function, and assign $\alpha_1 = \alpha_{real}$ and $\alpha_2 = \alpha_{imag}$ when simulating the lens. Examples of the resulting critical lines and caustics for elliptical power law profiles of different axis ratios and slopes are presented in Figure 1.5, while examples of their corresponding deflection potentials and surface mass densities are given in Figure 1.6.

A major limitation of this profile is that the hypergeometric function must be evaluated numerically, and so its presence is the main contributor in slowing the computations of the deflections. For this reason I revert to using Equations 1.21 and 1.22 when simulating SIE models as opposed to setting $n = 2$ and using the above form. Additionally, the slope parameter proves challenging when fitting profiles to real galaxy-scale lenses, significantly slowing the modelling process for conventional techniques. We will see in the next section how machine learning has been applied to this area to help quicken both the identification and modelling of these lenses.

1.3 Existing Techniques for Lens Detection and Modelling

With the advent of large scale surveys, vast data sets of images will become available over the next decade. In order to cope with so much information, multiple automated methods have been developed over the years for rapidly and accurately identifying strong gravitational lenses. These have included geometrical quantification techniques (Seidel and Bartelmann, 2007; Bom et al., 2017), the analysis of colour bands (Gavazzi et al., 2014; Maturi et al., 2014), spectroscopic analysis (Baron and Poznanski, 2016; Ostrovski et al., 2017; Talbot et al., 2021), and the Histogram of Oriented Gradients (HOG) feature extraction method (Avestruz et al., 2017). Convolutional Neural Networks (CNNs) have seen considerable use in this area (Jacobs et al., 2017; Lanusse et al., 2017; Petrillo et al., 2017; Schäfer et al., 2018; Davies et al., 2019; Jacobs et al., 2019; Metcalf et al., 2019; Canameras et al., 2020; He et al., 2020; Li et al., 2020; Huang et al., 2021; Magro et al., 2021; Gentile et al., 2022; Jacobs et al., 2022; Morgan et al., 2022; Wilde et al., 2022), along with other machine learning methods (Cheng et al., 2020), as these require neither the spectroscopic data nor arbitrary geometric measurements often employed by the other techniques.

CNNs are a subset of deep neural networks that have in recent years become popular for handling large amounts of data. They have been extensively used for rapid feature extraction and classification of images, and have seen a wide range of applications in astronomy, from discovering exoplanets to classifying gravitational wave signals to measuring photometric redshifts and the structures of high-redshift galaxies (e.g. Fluke and Jacobs, 2020; Paillassa et al., 2020; Schuldt et al., 2020; Tohill et al., 2020; Wu and Peek, 2020). Such networks have been shown to be very effective at correctly identifying thousands of lenses purely from images, and are able to do so extremely quickly. However, this first requires the CNN to be trained on many tens of thousands to hundreds of thousands of images; with so few real images of gravitational lenses, these training sets must be simulated instead. While lens detection has also seen the application of citizen science (More et al., 2015; Sonnenfeld et al., 2020), there may be little overlap between lenses identified by citizen science and those by machine learning (Knabel et al., 2020), highlighting the need for multiple approaches in order to obtain the most complete sample.

Following identification of a lens system, mass modelling is typically performed using parametric techniques. These obtain the lens mass model parameters that best fit the observed image. Different techniques have been developed for this purpose, for example those that involve reconstructing sources parametrically or on pixellated grids (Warren and Dye, 2003; Vegetti and Koopmans, 2009; Nightingale et al., 2018; Etherington et al., 2022) or the use of shapelets (Birrer et al., 2015;

Birrer and Amara, 2018). Modelling can often require time-consuming processes including point spread function estimation, deblending and subtraction of the lens galaxy light through repeated modelling, and image masking prior to modelling. For example, Rojas et al. (2021) deblends the lens light from the source based primarily on their colours, Bolton et al. (2008a) and Shu et al. (2016) fit the lens light with a smooth radial B-spline model (detailed in Bolton et al., 2006), and Nightingale et al. (2018) and Birrer and Amara (2018) repeatedly model the lens light as one or more superpositions of parametric profiles. To circumvent some of this initial processing, current modelling techniques now simultaneously fit both the lens galaxy light and mass profile but at the expense of an even slower modelling speed.

With CNNs successfully used for detecting lenses, they have since been shown to provide a promising alternative method of lens modelling. Hezaveh et al. (2017) demonstrated this for the first time, training a combination of networks to predict lens mass model parameters and applying them to simulated and real Hubble Space Telescope (HST) images. A method of obtaining uncertainties on these predictions was presented by Levasseur et al. (2017), using a CNN as an approximate Bayesian neural network (BNN; see Section 1.4). While the initial training took multiple days on a GPU machine, when applied to test images they reported an increase in lens modelling speed of several orders of magnitude compared to parametric methods, demonstrating the potential application of CNNs for this purpose. This was later extended for application to interferometric observations (Morningstar et al., 2018), along with the demonstration of machine learning to additionally reconstruct the background source from CNN-predicted parameters (Morningstar et al., 2019).

Since then, CNNs and similar networks have seen much use in lens modelling, including source reconstruction (Chianese, 2019; Chianese et al., 2020), redshift and lens velocity dispersion estimation (Bom et al., 2019) and the detection and modelling of dark matter substructure (Brehmer et al., 2019; Alexander et al., 2020; Lin et al., 2020; Ostdiek et al., 2020; Rivero and Dvorkin, 2020; Varma et al., 2020; Vernardos et al., 2020; Alexander et al., 2021; Ostdiek et al., 2022a,b). CNNs have also seen use in general galaxy deblending and segmentation (Burke et al., 2019; Hausen and Robertson, 2020), which could have application to strong gravitational lenses. Recent work by Maresca et al. (2021) showed how CNNs could easily identify unphysical source reconstructions outputted by semi-analytic parametric modelling methods, allowing for an automated approach to dealing with incorrect models. With regard to using CNNs to obtain parametric lens mass profile parameters, Schuldt et al. (2021) focused on ground-based imaging, leaving in foreground lens light and making use of four filters to distinguish sources composed of Hubble Ultra-Deep Field (HUDF; Beckwith et al., 2006) galaxies before examining how well the CNN-predicted models translated into predicting time delays and image positions. Park et al. (2021) applied an approximate BNN to the modelling of time-delay lenses consisting of lensed AGN, combining results with simulated time delays for H_0 inference across hundreds of such lenses. Wagner-Carena et al. (2021) presented a hierarchical inference framework for such BNN lens modelling to avoid biases introduced by differences between training and test data sets, while an alternative approach was proposed by Legin et al. (2021) which used a second network to perform a density estimation technique in order to improve BNN predictions. And Madireddy et al. (2019) presented a pipeline for both lens detection and modelling, including denoising and deblending (removing lens light), using multiple different deep neural networks in a modular fashion that provided informative latent spaces and uncertainties at each stage.

The CNN has become prevalent within the field of gravitational lensing as an important classification and modelling method, and is the main focus of this thesis. In the next section we will explore how exactly this form of machine learning works, and how it has become such an effective tool in an astronomer's arsenal.

1.4 Machine Learning

While the theory of artificial neural networks (ANNs) has been around since the 1940s, their popularity has increased dramatically only over the last decade thanks to advances in deep learning and computing power. With developments to Nvidia's graphics processing units (GPUs) in 2009, ANNs could be trained in a fraction of the time and allowed for deeper, more complex architectures. Soon after, networks began outperforming traditional methods and paved the way for the 'deep learning revolution'. Since then, a range of different network architectures have been applied to various problems including pattern, image & speech recognition, object detection and classification, clustering, time series data analysis and natural language processing. They have seen use across many fields, from astronomy to medical imaging to commercial applications like Google Translate. In this section I will present an overview of neural network architecture and functionality with a focus on convolutional neural networks, and discuss the mechanisms of how they can be trained to produce desired results.

1.4.1 Neurons & Neural Networks

The concepts of ANN architecture and learning were first devised by McCulloch and Pitts (1943) and arise directly from those of the brain, with Hebb (1949) presenting the idea of pathways in the brain strengthening with each use. Just as the brain is made up of interconnected neurons, so too are neural networks, although for the latter these artificial neurons, or nodes, are grouped into layers. Nodes in a given layer are not connected to each other but are instead connected between layers, conventionally to those adjacent to that layer. Mathematically, each of these nodes is a computation whose output y is the dot product of its inputs x (the outputs of nodes from the previous layer) and the strengths of their connections to the node w (weights), plus the strength of the node itself b (bias), which is then passed through some non-linear 'activation' function f:

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{1.29}$$

This non-linearity allows the network to learn non-linear relationships in parameter space, present in most real-world data.

Originally these nodes were perceptrons (Rosenblatt, 1958), capable of only a binary output; networks consisting of one or multiple layers of these nodes were hence single-layer and multi-layer perceptrons, respectively. As this binary approach was limited, multiple continuous activation functions have since been developed, suitable for a much broader range of applications. The activation function chosen throughout this work is the Rectified Linear Unit (ReLU; Nair and Hinton, 2010) function

$$f(x) = \max(0, x) \tag{1.30}$$

which acts non-linearly on the nodes such that any negative values are set to zero. This activation function was chosen as it is both computationally efficient and helpful in avoiding the vanishing gradient problem, which can prevent network weights from changing during training (details of training are presented in Section 1.4.3), and hence this form is commonly used by deep neural networks in place of sigmoidal activation functions.

The data for the the network to analyse acts as the input to the first layer (the 'input layer'). Computations at each node in that layer are performed with the outputs passed on to the next layer, and this forward pass of information continues until the final values are outputted in the 'output layer'. Deep neural networks are a type of ANN that contain at least one 'hidden' layer between the input and output layers, which allow for more complex relationships to be learned.

The simplest type of layers are fully connected, i.e. each node in the layer is connected to every node in the preceding layer; if this is the case for all layers then the network is itself fully connected. While the simple structure of fully connected networks makes them applicable to a wide range of problems, they are rarely used to handle images and other complex data as they can be less effective and more computationally expensive than other more specialised networks. Instead, deep neural networks often incorporate other types of layers.

1.4.2 Convolutional Neural Networks

As the main focus of this thesis is on the application of convolutional neural networks (CNNs) to strong gravitational lens modelling, these networks are described in more detail here. CNNs are a subset of neural networks that have become popular over the past few years for their ability to deal with the large amounts of information contained in images, such as for edge detection and face recognition. Inspired by research on the visual cortex by David Hubel and Torsten Wiesel in the 1950s and 1960s, Fukushima (1980) introduced the precursor 'neocognitron' neural network, which would later become the more modern CNN developed by LeCun et al. (1989).

CNNs contain grid-like convolutional layers that apply convolutional filters (also known as kernels) to their inputs in order to extract information, in a similar manner to image processing techniques (e.g. Villa et al., 2016). Each kernel consists of a small grid of weights, with an additional bias term dictating the overall strength of that kernel. During a convolution, the kernel scans across the image, taking the dot product of its weights with each section of the image (called the kernel's receptive field).

For a given convolutional layer, the number and dimensions of the kernels are set by hyperparameters, as are the horizontal and vertical strides (the distances in pixels that a kernel moves per step across the image). When an image is fed into the layer as an input, each kernel is convolved with it to produce a two-dimensional feature map; convolutions take place across the image depth, such that the output has a depth of one. Hence the output of the layer is a stack of feature maps, with a depth equal to the number of kernels applied. While all of these kernels share the same dimensions, their values are individually learned as the network is trained.

CNNs also usually contain other layer types in addition to convolutional layers. Pooling layers decrease the image size to both speed up computation and allow later convolutional layers to learn more abstract features from the lower resolution image. These are often chosen to be max-pooling, outputting the maximum values of each kernel-sized region of the image, due to such maximal values being the most informative for spatially-encoded information than, for example, average values (Chollett, 2017). For this reason, max-pooling layers are used in this work. Additionally, batch normalisation can be implemented after each convolutional layer, which normalises their output in order to increase the stability of the network. After feature extraction, a ‘flatten’ layer converts the stacks of images from the last convolutional layer into a one-dimensional vector. This is then passed to one or more fully connected layers, where each node in the layer is individually connected to all nodes in the preceding layer, in order to identify relationships in the data that the convolutional layers cannot.

1.4.3 Training a Network

In order to use a neural network, it must be trained on data in order to optimise its weights and biases and reduce its error (loss). In this regard, networks are usually classified as supervised or unsupervised, the former requiring its training data to be labelled with the correct values. This helps supervised networks to learn what to output for a given input, but requires us to have prior knowledge of the details of the training set. Unsupervised networks do not require this labelling but as a result are mainly used for clustering and so are not the focus of this work.

Neural networks are usually used for either classification or regression tasks. Classification networks output values corresponding to the probability of an input belonging to a given class or category; these can be binary or a decimal value between 0 and 1. Such networks have therefore been employed in gravitational lens finding, classifying for example whether an image contains a lens or not. However, this thesis focuses on modelling lenses through obtaining values for lens mass profile parameters, and as such uses a regression network where the output contains a series of numbers that can take any value, each representing one of the parameters.

Networks also contain hyperparameters which are set before training and often fine-tuned through repeated training attempts, whether by hand or through an automated process such as Bayesian hyperparameter optimisation. These include the number and types of layers, the number of training iterations (epochs) and the step size taken to reduce the loss after each iteration (learning rate). As part of constructing a network, a suitable cost function must be selected. Cost functions evaluate the performance of the network during training, which for a supervised regression network compares the predictions y' to the true values y in some way. A suitable cost function must be differentiable with respect to the weights and biases of a network, can be written as an average, and must not depend on any activation values besides those of the output layer. In the following chapter, the chosen cost function is the mean squared error (MSE) calculated for n input images,

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i')^2 \tag{1.31}$$

which is reduced as the network trains. As the cost function dictates how the network learns, changing it inevitably impacts network performance; MSE is commonly used as a cost function due to its simple implementation and ability to heavily penalise outliers leading to more reliable predictions. An optimiser algorithm is also required to inform the network how it should update the weights and biases after each iteration to reduce its overall loss. In some cases the optimiser also updates the learning rate, reducing it from initially large steps that help the network explore parameter space down to small steps that allow it to converge on precise values.

Training supervised convolutional neural networks is performed over a number of epochs, typically requiring tens to hundreds of thousands of training images. For a given epoch, these images pass through the network in batches (also called mini-batches), with the loss for each batch calculated via the cost function. This loss is then used to update the weights and biases of the network using processes known as mini-batch gradient descent and backpropagation. In gradient descent, the gradient of the cost function is calculated with respect to the weights and biases. The weights and biases are then updated by taking small steps (proportional to the learning rate) in the direction of the negative gradient in order to minimise the loss. Backpropagation, derived in the 1960s, first saw use in neural networks by Rumelhart et al. (1986), with its modern form popularised by LeCun et al. (1989). In backpropagation, gradient descent begins at the output layer with gradients propagated backwards through the network so that all weights and biases are updated despite the network's complexity. How exactly the gradients and learning rate are used in this manner depend on the choice of optimiser, which can itself introduce additional hyperparameters. Throughout this thesis, the Nadam optimiser (Dozat, 2015) is used for this purpose, which is a combination of methods based on the gradient descent algorithm.

Once all batches of images have passed through the network and a final loss stored, but before the next epoch begins, a separate validation set of images is passed through the network to see how well the network has generalised to images beyond those it has trained on. While the validation loss has no impact on training, it provides a measure of whether the network is under- or over-fitting to the training data, in which case a modification to the training set or hyperparameters is required. Should the training and validation losses be similar and low enough for the given task, no further updates or modifications are made, and the now trained network is then tested on images from a completely separate test set. This acts as a last check of its generalisation in case the above modifications result in overfitting to the validation set, and marks the last step of training. The trained network is then ready for application to new data.

1.5 Thesis Overview

As discussed earlier in this chapter, gravitational lensing benefits astronomy and cosmology in many ways. This includes providing a means of studying the properties of distant galaxies by modelling their mass profiles and reconstructing background sources. It also aids in our understanding of the evolution of large scale structure through dark matter measurements and constraining cosmology.

In order to cope with future surveys that are set to discover many thousands of lenses, automated approaches are required. Additionally, these surveys will revisit the same lenses multiple times, inevitably leading to the repeated remodelling of any given lens as the data improves. In preparation for this, machine learning methods have thus seen use in modelling strong lenses quickly and efficiently, including the prevalent use of CNNs for this task. However, the functionality and limitations of such networks have not yet been fully explored, nor has a comprehensive analysis been performed of their accuracy and efficiency compared to conventional modelling methods.

In this thesis, I seek to examine the use of deep learning convolutional neural networks as a competitive strong galaxy-scale gravitational lens modelling method when applied to upcoming wide-field survey data. In Chapter 2, I construct and train a CNN in order to investigate the practicalities faced by such networks for lens modelling, analysing their accuracy and reliability when applied to strong galaxy-scale lens images with the imaging characteristics of the LSST and *Euclid* surveys. The network architecture is presented along with details of my simulated data sets used for training and testing. CNN results for images with foreground lens light subtracted are compared to those with lens light included, testing on both single band and multiband imaging. I also analyse how network performance varies with the mass model parameters themselves, the impact of mass-light profile alignment, as well as the gain in accuracy through stacking images.

In Chapter 3, the theory of Bayesian probability and variational inference are presented before these are utilised to create a Bayesian neural network capable of predicting its own uncertainties, for use in the rest of the thesis. After a hyperparameter fine-tuning process, I provide the architecture and training process of this Bayesian CNN. This is followed by details of the more realistic simulated data sets used for its training and testing, with the multiple test sets featuring increasingly complex lensing systems from smooth mass profiles to those with hydrodynamically-simulated substructure, line-of-sight structure and real source galaxies.

In Chapter 4, comparisons are made between the Bayesian CNN and parametric fitting methods when tested on the above data sets, including both SIE and elliptical power law profile modelling. The semilinear inversion technique PyAutoLens is chosen as the parametric fitting method for this comparison, and I additionally explore a way of combining these methods together through utilising CNN predictions as priors for PyAutoLens. An analysis of their modelling speeds is also presented.

In Chapter 5, a summary and discussion of the work in this thesis are presented, covering the key findings, limitations and difficulties experienced. Final conclusions are presented, and the chapter finishes with an overview of the potential future work and applications of both the Bayesian CNN and its combination with conventional modelling methods.

In Chapter 6, these applications are discussed in more detail, with a focus on those requiring large samples of modelled strong lenses. This is followed by an examination of one such application to constrain cosmological parameters through the modelling of double source plane lens systems, in which I propagate the results of the earlier chapters to obtain the precision that can be expected for these cosmological parameters in light of future surveys.

Chapter 2

CNNs for Strong Lens Modelling

With the need for efficient lens modelling techniques to cope with future large-scale survey data, the next chapter covers the initial work carried out during this PhD on using machine learning for this purpose, investigating the practicalities faced when applied to data sets with the imaging characteristics of these surveys. Specifically, I focus on *Euclid* and LSST for the simulated images, with the latter consisting of either one or three bands. To begin, the methodology of this work is discussed in Section 2.1, including the chosen architecture and training process for my CNN along with the simulation process used to generate its training and testing image data sets of lenses. The results of my investigation into multiple aspects of using such networks for lens modelling are then presented in Section 2.2. First, I compare LSST and *Euclid* results along with quantifying the benefits from using LSST multiband imaging, and study how errors vary with the lens parameters themselves. In addition, I examine how the CNN copes when trained to model the lens mass profile with and without prior subtraction of foreground lens light. This is followed by an assessment of the impact of assumed mass and light alignment during training when lens light is not subtracted. Lastly, the performance gain from stacking images and improved signal-to-noise is presented. These results are then followed by a summary of this chapter in Section 2.3, and are further discussed in Chapter 5. This chapter features work from my publication *'The use of convolutional neural networks for modelling large optically-selected strong galaxy-lens samples'* (Pearson et al., 2019).

2.1 Methodology

In Section 1.3 we saw how convolutional neural networks have been used to speed up lens modelling by several orders of magnitude once trained, and Section 1.4 detailed their theory, operation and general structure. Now, I begin by describing the chosen CNN architecture and training process used for the rest of this chapter, including the mass model parameters it will predict. I follow this by providing details of the lensed image simulations used to train the network, in preparation for the investigations presented in the next section.

2.1.1 CNN Architecture

The modelling of gravitational lens mass profiles is a regression task that requires the use of a supervised CNN; see Sections 1.4.2 and 1.4.3 for details of their architecture and operation. To select the network architecture for this work, a grid search was performed over network hyperparameters, including the number and ordering of each type of layer, layer widths (i.e. the number of kernels applied in a given layer), and convolutional kernel dimensions. This was done by training and testing each architecture on 10,000 and 1000 images, respectively, to obtain the highest accuracy with minimal overfitting while also seeking to limit training times (and hence computational cost). The work by Hezaveh et al. (2017) was also taken into account, which demonstrated that a network with relatively few layers can perform just as well as far more complex network architectures. Figure 2.1 shows the chosen network architecture, consisting of six convolutional

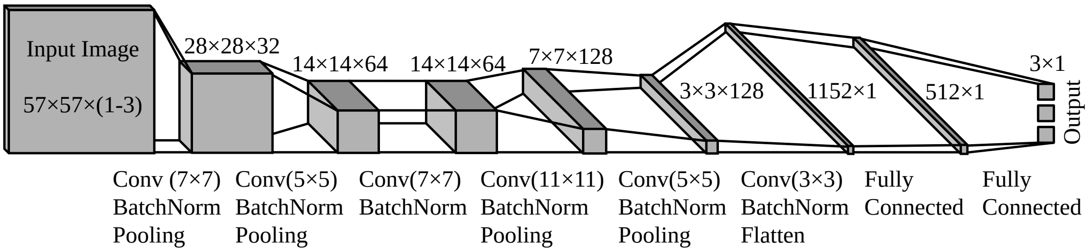

Figure 2.1: Structure of the CNN used in this work, showing the input image and the output shape of each block of layers. The types of layers in each block are given underneath, along with the kernel sizes (in pixels) used. In total there are six convolutional layers, each with batch normalisation, four max-pooling layers and two fully connected layers. A ‘flatten’ layer is also included to connect the multidimensional data to the 1D fully connected layer, and ReLU activation is used throughout. Numbers given above or beside each output are the output dimensions. Further details can be found at the end of Section 2.1.1.

layers, four pooling layers, and a hidden fully connected layer before the final output layer (which itself is fully connected). Six convolutional layers provided the best efficiency for this work in terms of accuracy and training time, and batch normalisation was performed after each one. A kernel stride of one was chosen and zero-padding was incorporated at the image edges to maintain the image size through the convolution. The pooling layers were chosen to be max-pooling (see Section 1.4.2), and ReLU activation functions were used throughout (see Equation 1.30 and Section 1.4.1).

The mass model parameters that we wish to obtain are those of an SIE lens: the Einstein radius θ_E, orientation ϕ and ellipticity e of the elliptical projected mass profile (see Section 1.2). Such profiles therefore range from circular ($e = 0$) to highly elliptical ($e \to 1$). However, the network was trained to predict the Einstein radius and two components of complex ellipticity (see Equation 1.24) later converted to orientation and ellipticity, as this gave more reliable values than directly predicting these parameters.

As well as refining hyperparameters, the network’s performance was improved further by pre-processing the inputs, with images intensity scaled so the counts in each pixel lay in the range 0-1. The mass model parameters were also scaled to the range 0-10 before being passed to the network and later re-scaled back when determining model performance. Typically, outputs of a neural network are scaled to the range 0-1, however the range 0-10 achieved higher accuracy and precision for this work.

In terms of the weights and biases, each layer has the following:

- Convolutional layer: For input height x_1, width x_2 and depth (number of bands) D, the input is an (x_1, x_2, D) matrix, with the input to the first layer being the original image of depth $D = 1$, or $D = 3$ for multiband images. The output is an (x_1, x_2, N) matrix, where N is the number of kernels applied in that layer. The bias and weights of each kernel can change as the network trains, but remain fixed as the kernel is applied across the image. Each kernel of size (k_1, k_2) has a depth equal to that of the input, and has an associated bias, resulting in a total of $k_1 \times k_2 \times D \times N$ weights and N biases per layer. The kernel sizes used are given in Figure 2.1.
- Batch normalisation layer: When used after a convolutional layer, a batch normalisation layer normalises across each feature map of the input using two trainable parameters. Hence for input depth D, this results in $2 \times D$ trainable weights.
- Max-pooling layer: Pooling uses a 2×2 kernel with a stride of two (i.e. the kernel shifts by 2 pixels between each application as it moves across an image), hence for input dimensions (x_1, x_2, D) the output has dimensions (floor(x_1/2), floor(x_2/2), D).
- First fully connected layer: The input is a flattened 1152-node array [the preceding layer has output dimensions (3, 3, 128)], and the output is a 512-node array, hence there are 1152 × 512 weights and 512 biases.
- Final layer: Input is a 512-node array, output is a 3-node array (one output per parameter to estimate), hence there are 512 × 3 weights and 3 biases.

- Total: 1 band = 2,395,267 trainable parameters.
- Total: 3 bands = 2,398,403 trainable parameters. This difference comes from the first (convolutional) layer having deeper kernels to match the depth of images with multiple bands (i.e. $D = 3$ above): see Section 1.4.2.

Multiple versions of the network were trained, using the single-band and multi-band imaging characteristics of the LSST as well as those of the single-band *Euclid* VIS instrument, with and without prior subtraction of foreground lens light. Each network was trained on 50,000 simulated images over 150 epochs, with identical architectures and hyperparameters. During training, the mean squared error was used as the network's cost function to determine its overall error, and the best results for network optimisation were obtained using the Nadam optimiser (see Section 1.4) and a learning rate of 1e-3. The CNN was run on a GPU machine (featuring an Nvidia GeForce GTX 1080 Ti GPU), allowing for much faster processing; for example, training over 150 epochs on 50,000 single-band images takes less than an hour, and 3-band images in less than 1.5 hours, while testing on such data sets takes only a few seconds.

2.1.2 Image Data Set Simulation

This work utilises a supervised approach to machine learning, so requires training the network on a large data set of simulated strongly-lensed images labelled by their true lens model parameters, as discussed in Section 1.4. The procedures for generating these images are described in the following section.

To make the training and test sets, I created a strong gravitational lens simulator: an overview of its simulation process is given below. The simulator involved multiple detection criteria that, if at any point were not met, would restart the generation process for a given image. A more in-depth description of many aspects of the simulator, including the detection criteria, are described after this overview:

1. Set initial hyperparameters:
    (a) Select the survey (LSST or Euclid) and set the corresponding telescope properties, including exposure time, number of exposures, telescope gain, image size, image resolution, filter bandwidths, point spread function (PSF) full width at half maximum (FWHM), readout noise, dark current, sky background surface brightness, and AB magnitude zeropoints.
    (b) Choose the number of images to simulate, and the amount by which to temporarily upscale the resolution during the simulation process (the resolution was doubled in this work, such that one pixel in the final image was simulated as four pixels).
    (c) Define the cosmology, the lens mass model, the lens and source light profiles, the criteria required for a sufficient strong lens detection, and whether to sample the lens mass model parameters from uniform or realistic distributions.
2. Load galaxy spectral energy distributions (SEDs), and define image arrays.
3. For each image:
    (a) Define the following parameters: orientations and axis ratios of mass and light profiles for the lens and source, Sersic indices of light profiles, the position of the lens in the image, and the ratio of the Einstein radius to the lens light effective radius.
    (b) Select redshifts of the lens and source, and calculate angular diameter distances and Mpc length scales of the lens and source plane.
    (c) For the lens, select the velocity dispersion (if required) and determine the Einstein radius (and effective radius using the above ratio). Combine with the fundamental plane (Hyde and Bernardi, 2009) to obtain the r-band absolute magnitude.

(d) For the source, select the r-band absolute magnitude from its luminosity function, combine this with the Faber-Jackson relation (Nigoche-Netro et al., 2010) to obtain velocity dispersion, and use the fundamental plane to obtain the effective radius.

  Detection Criteria: check that image and counter image are resolved in at least one band, using the Einstein radius, the effective radius of the source, and the telescope seeing.

(e) Lens flux: begin by selecting an SED for the lens given its redshift. This now needs to scaled according to the chosen r-band absolute magnitude in order to obtain apparent magnitudes and flux in each filter:

  i. For each filter, integrate over the SED to obtain the relative magnitudes in each band. Repeat for a redshifted version of the SED and compare to obtain the K-corrections for each filter.

  ii. Scale the non-redshifted relative magnitudes according to the r-band absolute magnitude to obtain absolute magnitudes for each filter.

  iii. Applying the K-corrections, convert absolute magnitudes to apparent magnitudes, and convert these first to photon flux, then to electron flux, and finally to analogue-to-digital unit (ADU) counts following Ivezic et al. (2010).

(f) Source flux: select an SED for the source, and repeat the above.

(g) Create the lens plane at the upscaled resolution, and apply the lens equation using the deflection angles of the chosen mass model to obtain the corresponding positions on the source plane.

  Detection Criteria: select the position for the centre of the source based on the Einstein radius of the lens, to help ensure multiple imaging of the source.

(h) Create an (unlensed) source image and a lensed source image using the lens and source plane position mapping (at the upscaled resolution) and the light profile parameters. Downscale these images back to the desired resolution, scale their intensities according to the counts determined for each filter, and convolve with the telescope PSF.

Detection Criteria: compare source image and lensed source image to check that magnification and tangential shearing is sufficient.

(i) Create the lens light profile at the upscaled resolution, downscale to the chosen resolution, rescale by the expected counts, and convolve with the PSF.

(j) Noise: apply shot noise to the lens, the lensed source, and the background sky counts, and add these together into one image. Subtract the pre-noise lens light profile to leave shot noise residuals, and add readout noise and dark current.

Detection Criteria: determine the signal-to-noise ratio (see Section 2.2.4) and check it meets the required threshold.

(k) Store parameters in arrays, store the final image in a matrix (for saving as a batch of images), and optionally save the image individually.

4. Save arrays of stored parameters and matrices containing generated images, to be loaded later as training and test data.

As the majority of lenses are early-type galaxies, the lens mass profile adopted was the SIE model (see Section 1.2), commonly used as a good fit for strong lens profiles. For the same reason a Sérsic profile was used to model the light profile of the source and lens, with the Sérsic index randomly drawn from a normal distribution with a mean of 4 (i.e. that of a de Vaucouleurs profile). For sources, r-band absolute magnitudes were sampled from the double-Schechter luminosity function of Kelvin et al. (2014), while for lenses these were instead obtained using the fundamental plane relation from Hyde and Bernardi (2009). Redshifts for the lens and source were selected from uniform distributions with upper limits of $z_{\mathrm{lens}} = 2$ and $z_{\mathrm{source}} = 6$, based on trial data sets and work by Collett (2015), and used to obtain r-band apparent magnitudes. Spectral energy distributions (SEDs) of appropriate ages were then selected uniformly from LSST simulated object SEDs (Connolly et al., 2014, available from the LSST GitHub repository), redshifted accordingly, and combined with the r-band magnitudes to obtain the flux in each band. These SEDs use the Bruzual and Charlot models (Bruzual and Charlot, 2003) with a Chabrier (2003) initial mass function (IMF), and either an exponential decline or an instantaneous burst of star formation.

The data sets used for training and testing the CNN contained tens of thousands of single or multiband images (see Section 2.2 for further details). The multiband images were generated using three filters of the LSST with RGB=(i, r, g), using the observatory's CCD filter response curve (Connolly et al., 2014, available from the LSST GitHub repository), and likewise the *Euclid* VIS filter response curve. These curves are presented in Figure 2.2. I adopted the native pixel scale of the LSST (0.2 arcsec pixel^{-1}; Ivezić et al., 2008; Abell et al., 2009) and *Euclid* (0.1 arcsec pixel^{-1}; Racca et al., 2016), and from inspecting the data sets and the distribution of known Einstein radii the postage stamp images were fixed at 57×57 pixels, as this maintained the visibility of lensed arcs while minimising both simulation and training times to handle the many data sets used in this work. Image size was fixed despite differences in resolution between LSST and Euclid resulting in lensed arcs appearing smaller in the images of the former compared to the latter, as this ensured that the CNN architecture outlined in the previous section remained constant between LSST and Euclid images (except for the depth of multiband LSST images). Even so, the extra space around the lensing system in the LSST images would not impede a suitably-trained network due to its ability to focus on regions of interest, except for perhaps longer training and testing times than strictly necessary. Typical exposures were used for each visit, with LSST images consisting of two 15s exposures (Abell et al., 2009) and *Euclid* VIS images

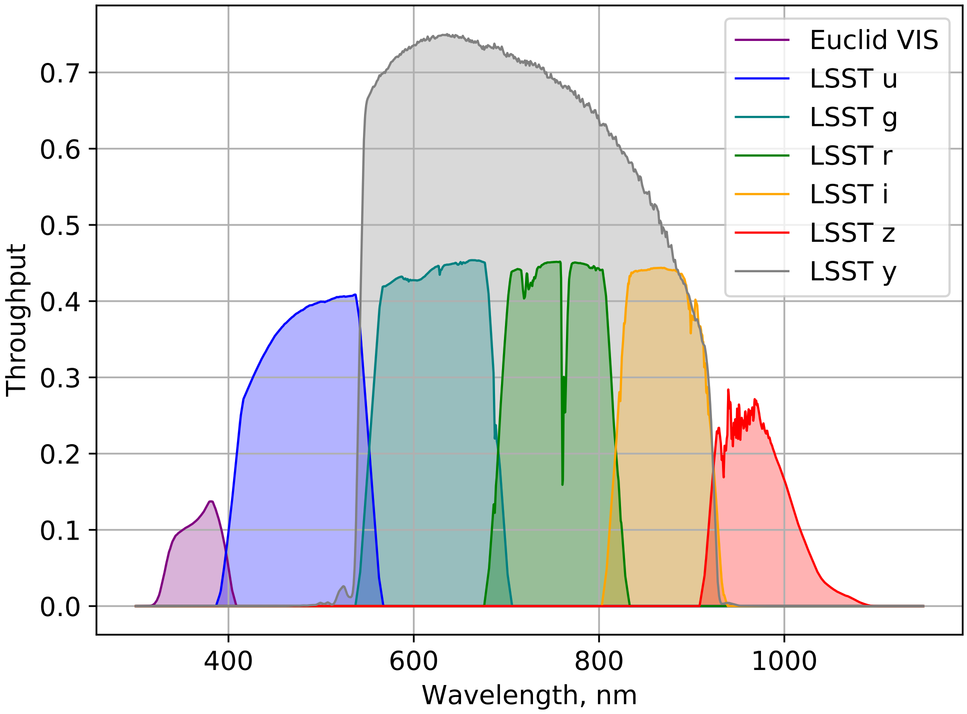

Figure 2.2: Filter response curves for *Euclid* VIS and the six (u, g, r, i, z, y) filters of LSST. For LSST image simulations in this work, I focus on the g, r and i bands, which together cover approximately the same wavelength range as *Euclid* VIS.

consisting of a single 565s exposure (Cropper et al., 2018), with AB magnitude zeropoints of 27.09, 28.58, 28.50, 28.34, 27.95, 27.18 for LSST (Ivezic et al., 2010) and 25.5 for *Euclid* (Collett, 2015).

After generating the lensed image of the background source, light from the lens was added and the image then convolved with a Gaussian point spread function with a full width at half-maximum of 0.7 and 0.17 arcseconds for LSST and *Euclid*, respectively. The sky background was added (Ivezic et al. (2010) for LSST, Niemi (2015) and Collett (2015) for *Euclid* VIS) along with shot noise, the expected read noise (five electrons per readout) and dark current (two electrons per pixel per second; Radeka et al., 2009). For this work, multiple data sets required the subtraction of the foreground lens light. This was implemented by subtracting the true light profile convolved with the point spread function, leaving only shot noise residuals. It should be noted that this subtraction of the true light profile is idealistic, and hence worse residuals are expected to be present in real data; this

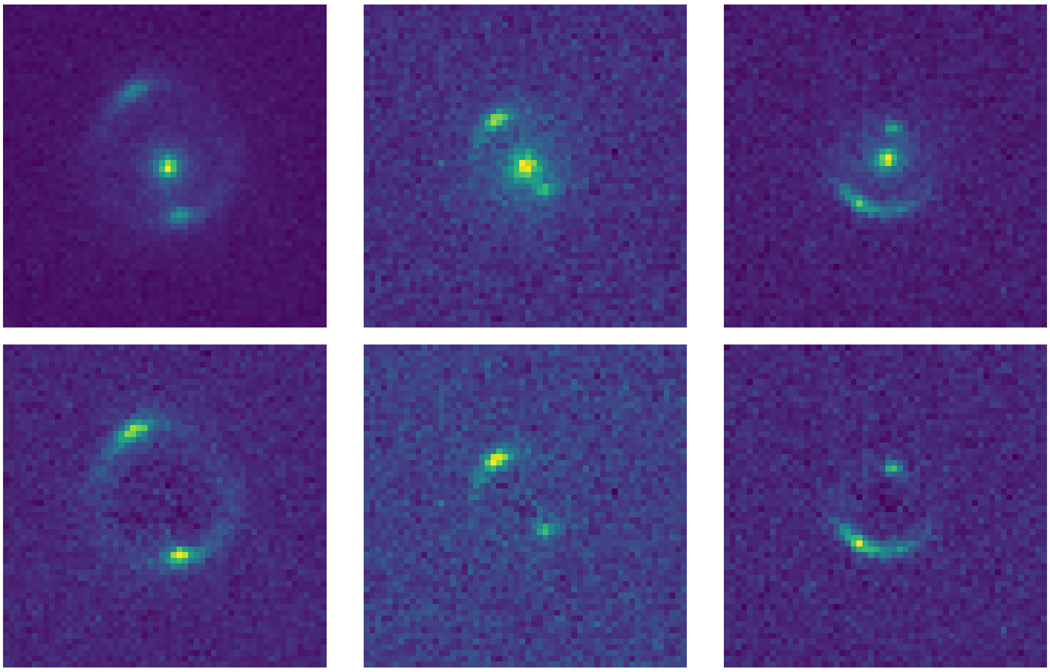

Figure 2.3: Example *Euclid* VIS images demonstrating foreground lens light subtraction, produced from the gravitational lens simulator. Top images have their lens light included, while bottom images are for the same systems once lens light has been subtracted.

is discussed further in Section 5.4. Examples of lens light subtraction are shown in Figure 2.3 for simulated *Euclid* images.

For these postage stamp images the lens centre positions followed a normal distribution about the image centre with a standard deviation of one pixel, and for all galaxies the offset between centre of the mass and centre of the light was taken to be zero. To ensure sufficient training across parameter space, the data sets used for training the network consisted of lenses whose Einstein radii, orientations and ellipticity were uniformly distributed in the ranges 0.2-3.2 arcseconds, 0-180°, and 0.0-0.6, respectively.

For both training and test data sets, the ellipticity and orientation of the lens light profile was scattered relative to the mass profile according to the following distributions measured from real lens samples. Light profile axis ratios were set to those of the mass profile divided by a factor taken from a normal distribution of mean 1.02 and standard deviation of 0.12 (Bolton et al., 2008b). The mass profile orientation was sampled from a normal distribution centred on the chosen light profile orientation with standard deviation of 10° (Bolton et al., 2008b). The impact of these mass-light correlations on network performance was also investigated: see Section 2.2.3.

The data sets used for testing the network were simulated to have distributions of parameters in line with the expected observations of the LSST and *Euclid*. Orientation of the light profiles were uniformly distributed, while mass profile axis ratios were selected from a normal distribution of mean 0.78 and standard deviation of 0.12 (Koopmans et al., 2006). Likewise, velocity dispersions were selected from a normal distribution with mean 250 and standard deviation of 44 km s^{-1} based on results from SLACS papers (e.g. Shu et al., 2017), and used to determine lens Einstein radii R_{Ein} (Equation 1.23). Lens effective radii R_e were then drawn from the observed distribution of the ratio R_{Ein}/R_e from Koopmans et al. (2006) (a normal distribution of mean 0.52 and standard deviation of 0.17) truncated to the range 0.1-10, and together with Einstein radii were in agreement with known distributions (Koopmans et al., 2006; Gavazzi et al., 2007; Bolton et al., 2008a). For these generated test data sets, the distributions of Einstein radii, velocity dispersions and redshifts are given in Figures 2.4, 2.5, and 2.6.

The lens and source could be offset from one another to produce arcs rather than full rings, with all source centroid positions determined randomly from a uniform distribution within the Einstein radius of the lens in order to ensure a lensed image. Multiple detection criteria were implemented to ensure strong gravitational lenses were produced, based on the criteria in Collett (2015):

1. Centre of the source must be multiply-imaged (must lie within the Einstein radius θ_E for this work),
2. Image and counter image must be resolved in at least one band (image centres are separated by $\sim 2\theta_{\mathrm{E}}$, so $(2\theta_{\mathrm{E}})^2 > (2R_{\mathrm{e,source}})^2 + \mathrm{seeing}^2$),

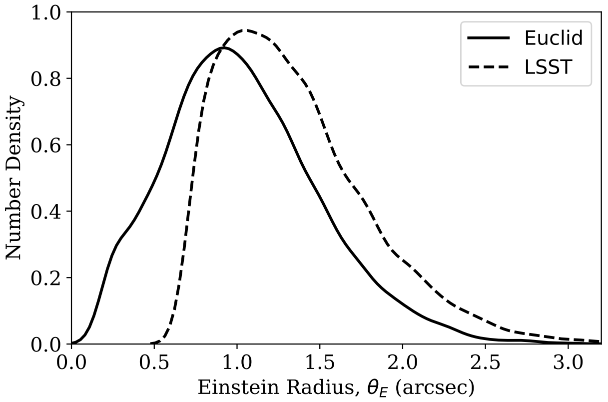

Figure 2.4: Distribution of Einstein radii from the gravitational lens simulator for *Euclid* and LSST test data sets. *Euclid* has mean 1.04 and width (standard deviation) 0.46 arcsec, and the LSST has mean 1.35 and width 0.45 arcsec.

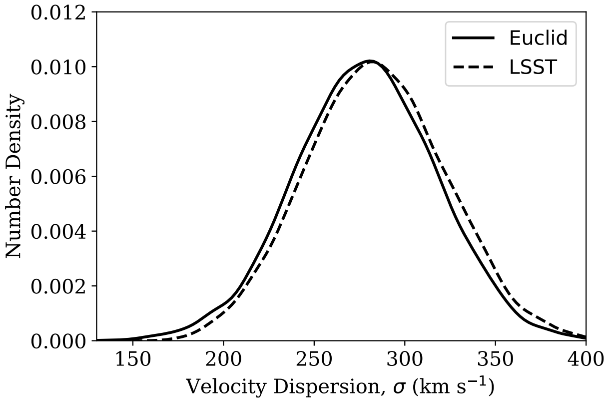

Figure 2.5: Distribution of velocity dispersions from the gravitational lens simulator for *Euclid* and LSST test data sets. *Euclid* has mean 278 and width (standard deviation) 39 km s^{-1}, and LSST has mean 284 and width 38 km s^{-1}.

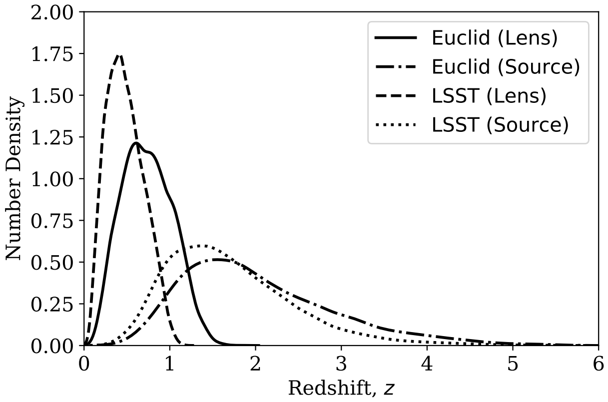

Figure 2.6: Distribution of lens and source redshifts from the gravitational lens simulator for *Euclid* and LSST test data sets. *Euclid* (lens) has mean 0.74 with width (standard deviation) 0.29, *Euclid* (source) has mean 2.04 with width 0.89, LSST (lens) has mean 0.48 with width 0.21 and LSST (source) has mean 1.71 with width 0.74.

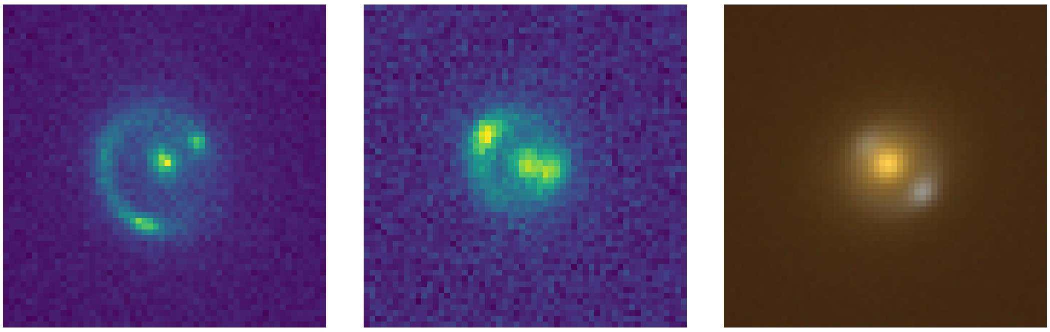

Figure 2.7: Example images produced from the gravitational lens simulator. From left to right: simulated *Euclid*, r-band LSST and colour LSST using RGB=(i, r, g). Images have a pixel scale of 0.1 arcsec pixel^{-1} for *Euclid* and 0.2 arcsec pixel^{-1} for LSST.

3. Sufficient magnification ($\mu_{\mathrm{TOT}} > 3$) and tangential shearing ($\mu_{\mathrm{TOT}} R_{\mathrm{e,source}} >$ seeing) in at least one band,
4. Sufficient signal-to-noise ratio (SNR $>$ 20) in at least one band.

Some examples of the outputs of the simulation are shown in Figure 2.7, for *Euclid*, single-band LSST and multiband LSST (RGB = i, r, g).

2.2 Investigating the Practicalities of CNN-based Lens Modelling

Having simulated data sets of lensed images with the imaging characteristics of the LSST single-band, LSST multi-band and *Euclid* VIS, the CNN was trained and tested on each of these separately, using the architecture outlined in Section 2.1.1. CNNs with identical architectures were trained on these data sets. Now, we examine the performance of these networks when applied to a variety of different test sets. In the following section, I discuss how network performance varies a) between subtracting and including lens light in the simulated images, b) with differing assumptions about how mass follows light in both training and test images, and c) with signal-to-noise and the stacking of images.

Throughout this chapter, errors on network-predicted lens model parameters are computed from the distributions of differences between the true and predicted parameter values across the test image set. As these distributions can be non-Gaussian in nature, we quantify these using the 68 per cent confidence interval either side of the mean and refer to it hereafter as a parameter's 'error'.

2.2.1 Investigation I - Subtracting Lens Light

To begin, the CNN was trained and tested on each of the simulated data sets with foreground lens light subtracted. Lens light subtraction is often performed by lens modelling software so as to allow for easier modelling of the background source light (see Section 1.3). The subtraction used for this work, described in the previous section, makes use of the known lens light profile parameters and so over-simplifies this process. CNN performance may vary slightly with different lens subtraction techniques, but testing these is not the focus of this work; I instead assume the case of ideal subtraction leaving only shot noise residuals. The three data sets had the imaging characteristics of LSST r-band, LSST multiband (RGB=i, r, g) and *Euclid* VIS, each consisting of 10,000 images. The network predicted the Einstein radii and two components of complex ellipticity of the foreground lens mass profiles, with the two components later converted to orientation and ellipticity.

Tables 2.1 and 2.2 respectively list the resulting CNN errors and median biases for each of the parameters and data sets, with the distributions of parameter errors presented in Figures 2.8 and 2.9. The large scatter in ellipticity results in Figure 2.9 compared to the other parameters is expected as ellipticity has a smaller effect on lensed source images and hence making it harder to predict. From the figure and Table 2.2, it is clear that there is very little biasing of results towards over- or under-predicting parameters when lens light is subtracted, with Einstein radius and ellipticity biases not exceeding 0.6 per cent and 1.6 per cent, respectively.

Table 2.1: Errors (68 per cent confidence intervals) on CNN-predicted parameters for the three survey data sets. Results are presented both for images which have had the foreground lens light subtracted and for images with the lens light present. Errors are computed from the distributions of differences between true and predicted parameter values across 10,000 test images (see beginning of Section 2.1).

Mass model parameter	*Euclid* VIS	LSST r-band	LSST (gri)
Lens light subtracted			
Einstein radius, θ_E (arcsec)	0.043	0.054	0.042
e_1	0.066	0.062	0.047
e_2	0.066	0.062	0.048
Ellipticity, e	0.057	0.053	0.041
Orientation, ϕ (radians)	0.164	0.155	0.114
Lens light included			
Einstein radius, θ_E (arcsec)	0.064	0.091	0.071
e_1	0.089	0.090	0.072
e_2	0.092	0.089	0.072
Ellipticity, e	0.079	0.077	0.064
Orientation, ϕ (radians)	0.225	0.220	0.172

Biases for all parameters are much smaller than the errors presented in Table 2.1, and do not appear to be affected by the choice of survey.

Single-band LSST and *Euclid* share similar results for orientation and ellipticity, but *Euclid* achieves higher precision for Einstein radii owing to the telescope's higher image resolution and narrower point spread function. Taking the mean average and standard deviation over Einstein radius, orientation and ellipticity, switching from a single band to three bands reduces parameter errors by 24 ± 2 per cent, as expected due to the increased information available to the network. This in turn results in the LSST gri results reaching the lowest errors for all parameters.

To investigate in what areas the network struggles to predict parameters, Figure 2.10 presents how the errors on these parameters vary with the value of each parameter. The results for Einstein radius θ_E are given as fractional errors, (i.e. 68 per cent confidence interval of $(\theta_{\mathrm{E}}^{\mathrm{predicted}} - \theta_{\mathrm{E}}^{\mathrm{true}})/\theta_{\mathrm{E}}^{\mathrm{true}}$), with the error bars in the figure corresponding to the standard errors. From the top row we see that errors increase for smaller Einstein radii, as expected; the limited image resolution

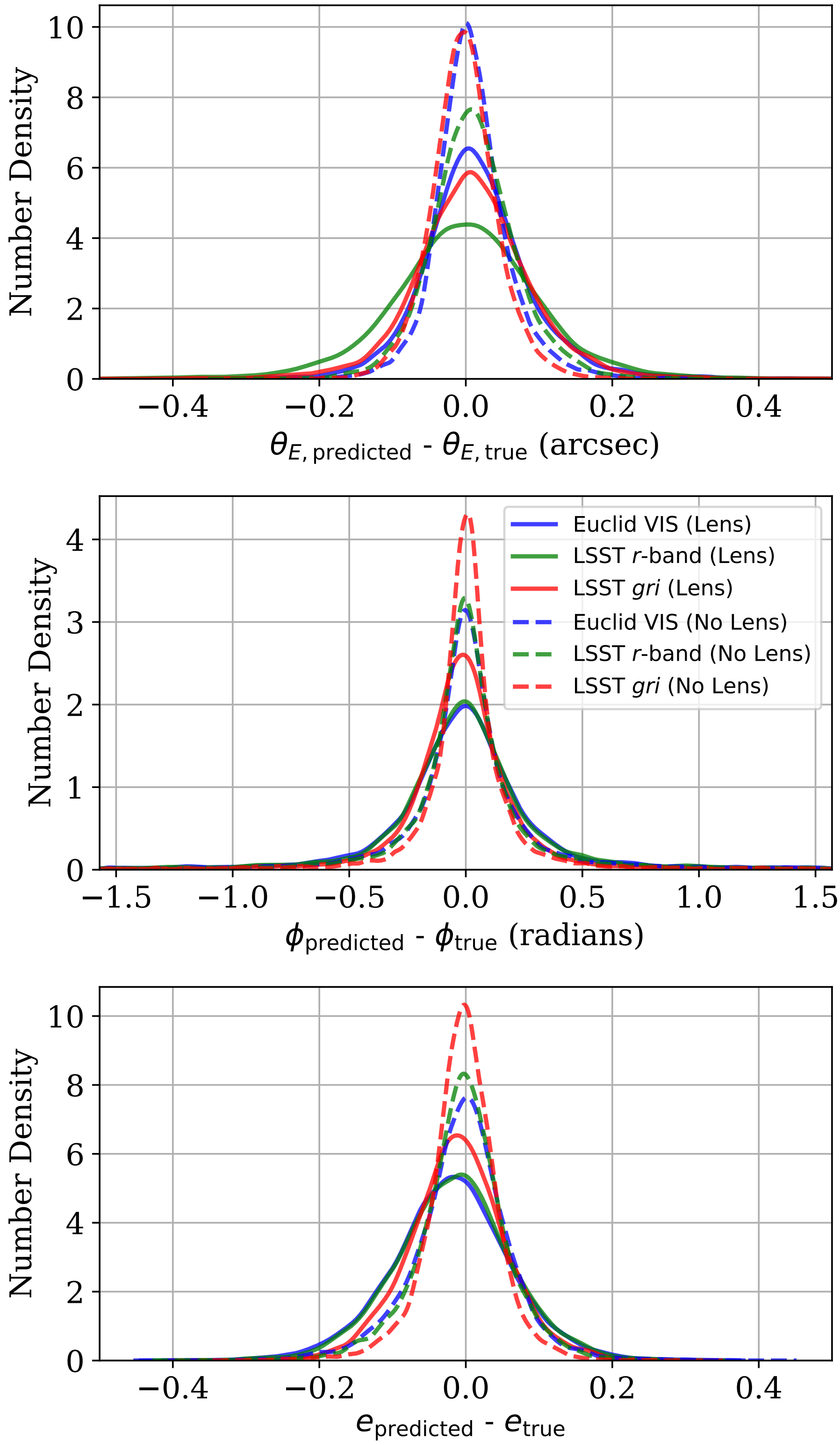

Figure 2.8: Distribution of the differences between the network's predicted parameters and their true values for test data sets of 10,000 images. From top to bottom: Einstein radius, orientation and ellipticity of the lens mass profile. These distributions correspond to single-visit *Euclid* VIS (blue), LSST r-band (green) and LSST *gri* data sets (red), both with the lens light included and subtracted (solid and dashed lines, respectively).

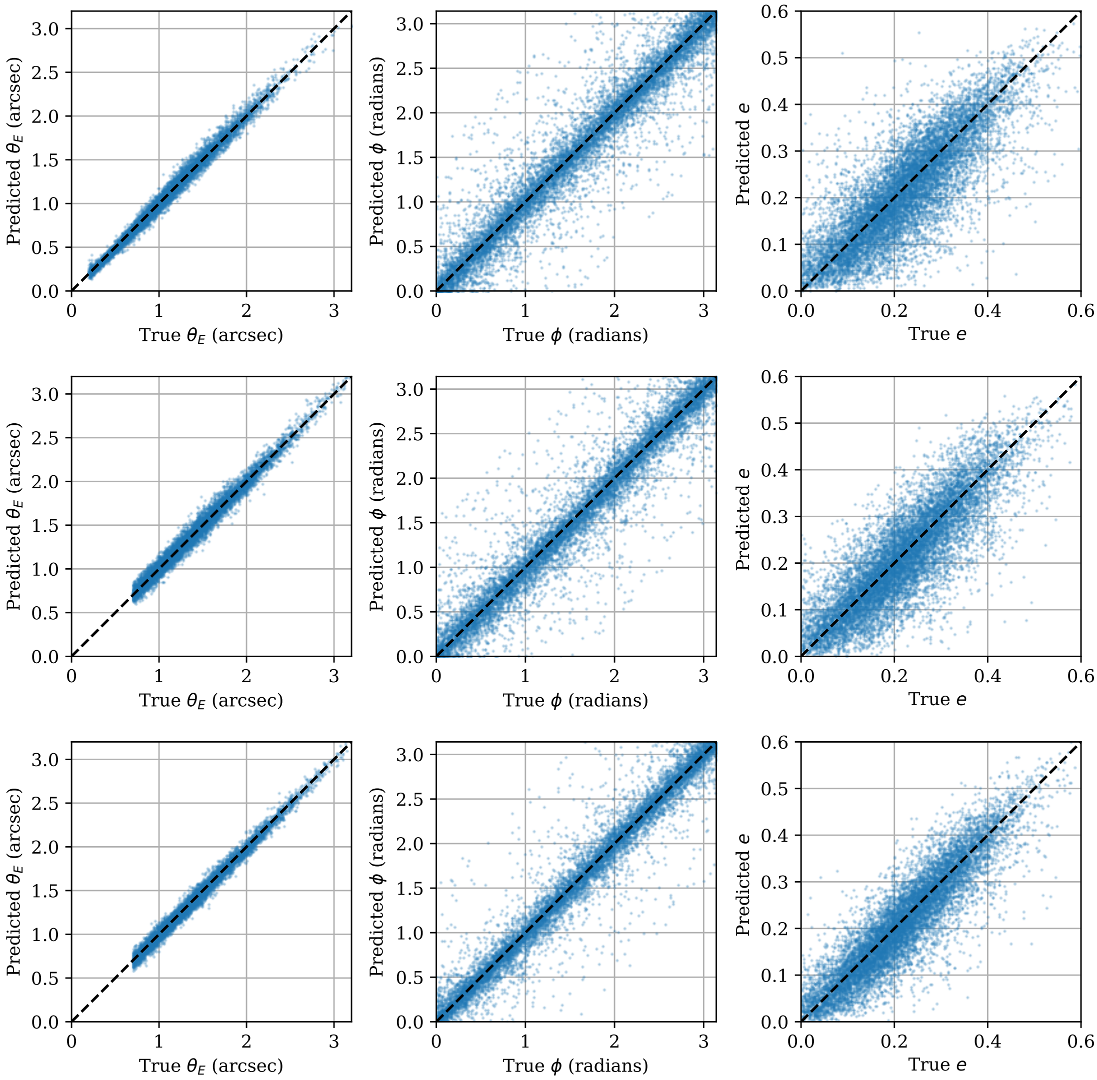

Figure 2.9: Comparison of network-predicted lens parameters with the true values for images with lens light subtracted. From left to right: Einstein radius, orientation and ellipticity of the lens mass profile. From top to bottom: *Euclid* VIS, LSST r-band and LSST gri.

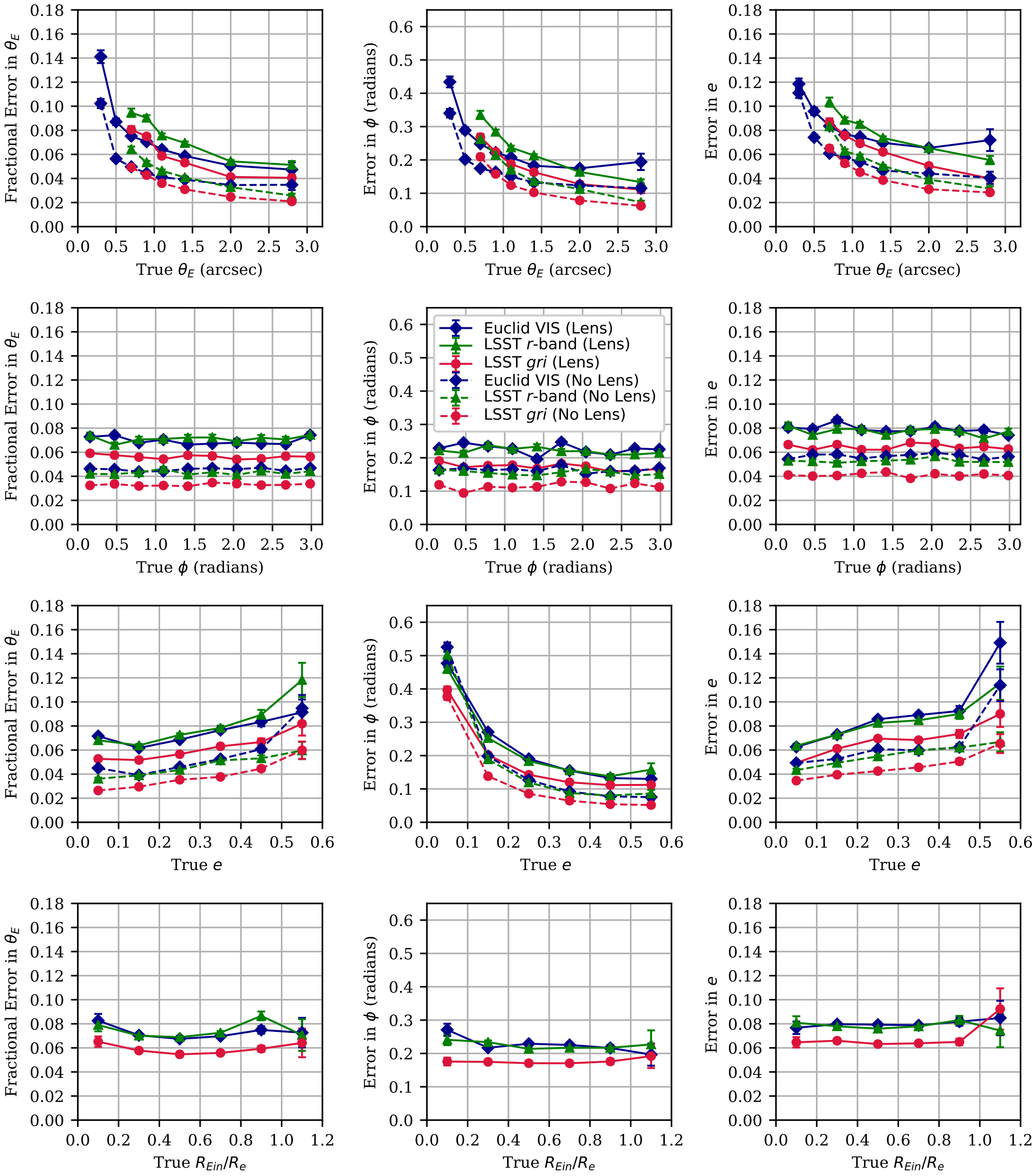

Figure 2.10: The variation of network-predicted lens model parameter errors with each parameter value. The solid and dashed lines correspond to images with lens light included and subtracted, respectively. The blue diamonds, green triangles, and red circles correspond to *Euclid* VIS, LSST r-band and LSST gri images, respectively. From left to right: Einstein radius, orientation and ellipticity of the lens mass profile. From top to bottom: Einstein radius, orientation, ellipticity and $R_{\mathrm{Ein}}/R_{\mathrm{e}}$, the ratio between Einstein radius and effective radius. The error in Einstein radius is given as the fractional error, and the error bars are the corresponding standard errors.

Table 2.2: Biases of CNN-predicted parameters for the three survey data sets, computed from taking the median average across 10,000 test images of the predicted parameter values minus their true values. Percentage biases are given next to the absolute biases for Einstein radius and ellipticity, with negative values indicating under-predictions compared to the true values.

Mass model parameter	*Euclid* VIS		LSST r-band		LSST (gri)	
Lens light subtracted						
Einstein radius, θ_E (arcsec)	0.005	(0.5%)	0.008	(0.6%)	-0.003	(-0.2%)
e_1	0.001		0.001		0.004	
e_2	-0.002		0.005		0.004	
Ellipticity, e	-0.003	(-1.3%)	-0.003	(-1.3%)	-0.003	(-1.6%)
Orientation, ϕ (radians)	0.001		-0.003		0.003	
Lens light included						
Einstein radius, θ_E (arcsec)	0.010	(1.1%)	0.000	(0.0%)	0.010	(0.6%)
e_1	-0.001		-0.002		0.000	
e_2	0.007		0.003		0.005	
Ellipticity, e	-0.019	(-9.1%)	-0.016	(-7.7%)	-0.012	(-5.4%)
Orientation, ϕ (radians)	-0.009		-0.010		-0.017	

results in the network having reduced information on the shapes of smaller, more pixellated arcs. However, no correlation is observed between parameter errors and orientation on the second row, with errors clearly unaffected as the network should not show a preference for certain orientations.

From the third row, we see that all parameter errors are affected by ellipticity. More elliptical lenses result in lensed arcs less closely following the shape of the Einstein ring, producing larger errors in Einstein radius. Yet this makes it easier to identify the direction of the elongation, reducing orientation error as expected in the central graph. However, in the third graph we observe that an increased ellipticity also increases the error on ellipticity predictions. This can be explained as higher ellipticities result in larger caustics within which the source galaxy can reside, and hence produce more varied lensed arcs that the network may struggle to learn. As such, this trend vanishes when presented as the fractional error in ellipticity.

2.2.2 Investigation II - Including Lens Light

Having analysed results when training and testing the CNN on images with lens light subtracted, we now examine how this performance changes when instead training and testing on images containing lens light. Leaving in lens light may make it more difficult for lens modelling software to distinguish the source light, but avoids issues of modelling residuals left behind from unsuccessful lens subtraction. An alternative to subtraction is masking of lens light, but such masks are usually applied manually; automatic masking may result in greatly varying results and is left to other work.

With the presence of lens light in these images, it was necessary to ensure realistic differences between the light and mass profiles. Motivated by observations (Koopmans et al., 2006; Bolton et al., 2008b), the ellipticity and orientation of the mass profiles were scattered about those of the light profiles with standard deviations of 0.12 and 10°, respectively, for both training and testing of the network. This is investigated in more detail in Section 2.2.3.

The test results, showing how the new CNN errors change with each parameter, are again shown in Figure 2.10. These closely follow the same trends as images without lens light, although errors are higher on average. As lens light can interfere with modelling by overlapping source light, I also investigated how the errors changed with the ratio between the Einstein radius and the lens light profile's effective radius, $R_{\mathrm{Ein}}/R_{\mathrm{e}}$, shown on the bottom row. We observe that the network is mostly unaffected by this ratio, at least over the range tested, which is an encouraging indication that lens model parameters can still be reliably estimated even when lens light begins to dominate the lensed source image.

Test results are again presented in Tables 2.1 and 2.2 and in Figure 2.8, with predicted parameters against their true values shown in Figure 2.11. Regarding biases, those for Einstein radius and the complex ellipticity components remain

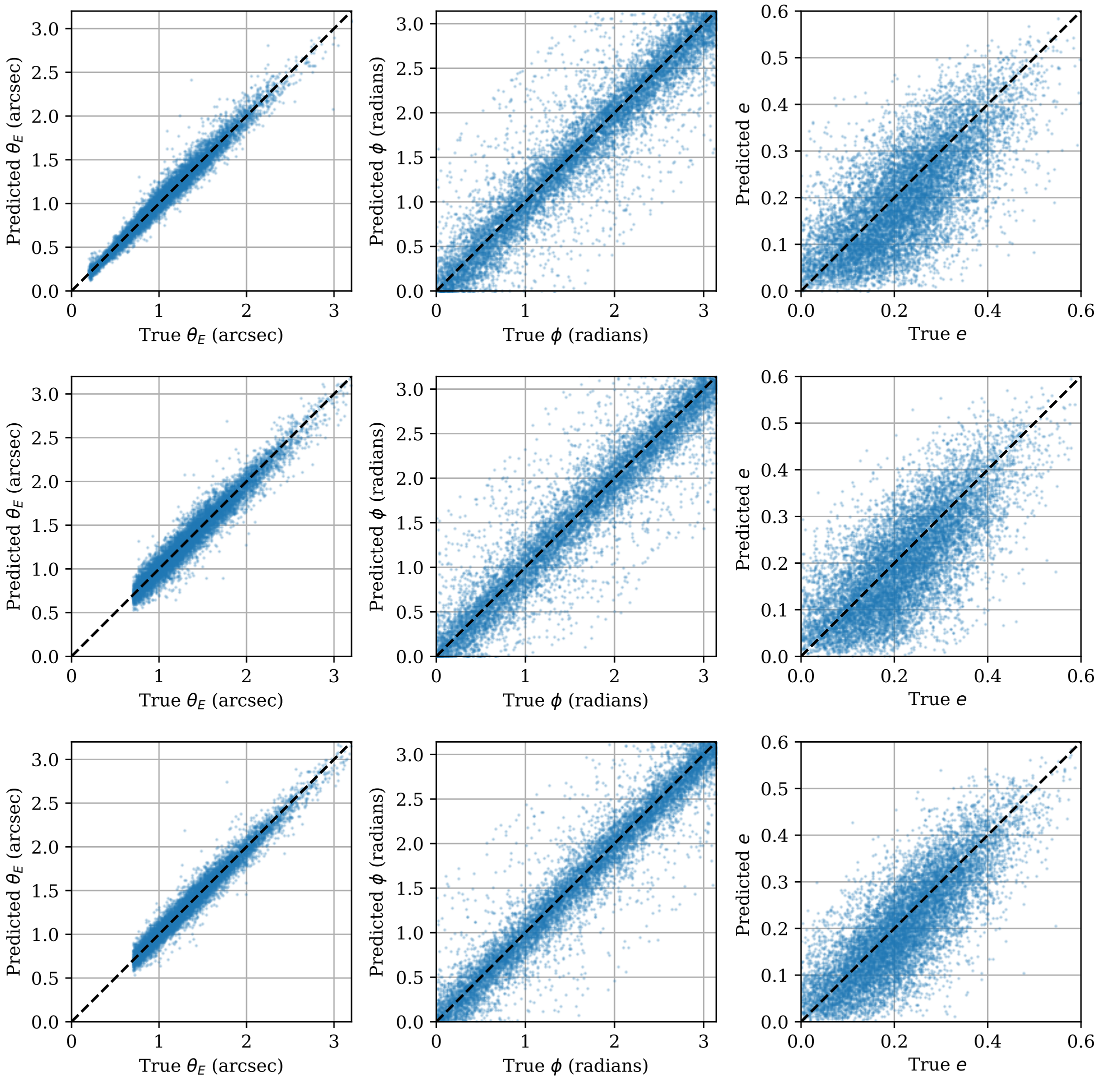

Figure 2.11: Comparison of network-predicted lens parameters with the true values for images including lens light. From left to right: Einstein radius, orientation and ellipticity of the lens mass profile. From top to bottom: *Euclid* VIS, LSST r-band and LSST gri.

overall similar to those for lens light-subtracted images, while those for orientation have increased in magnitude. However, the largest change is seen in ellipticity, with biases now reaching over 9 per cent and values generally under-predicted. This is because the CNN struggles to model lenses with higher ellipticity so resorts to following more closely to the lens light, and, due to the upper bound when randomly selecting ellipticity, highly elliptical simulated galaxies are more likely to have light profiles with lower ellipticities than their mass profiles. Despite this, biases for the parameters are still small relative to the errors in Table 2.1 and continue to show little correlation with the choice of survey.

Just as for images without lens light, the CNN obtains the smallest errors for orientation and ellipticity using the three-band LSST data set. This is to be expected as the addition of lens light to all three data sets would do little to impact their relative performance, with these parameter errors reduced by increased available information. However, the lowest errors for Einstein radius arise from the *Euclid* data set, as this parameter is now primarily affected by image resolution due to the increased difficulty in separating the source and lens light. With the lens light present the colour information can be used to distinguish between the foreground lens and background source. Even if the colouration of the two were similar, the multiple bands still act to improve the signal-to-noise. However, the change from using a single LSST band to three bands now has a reduced impact compared to images with lens light subtracted, with errors only decreasing on average by 20 ± 2 per cent likely due to the worse network performance overall.

Comparing the results with those in Section 2.2.1, it is clear that the inclusion of lens light degrades the results for all data sets, including the multiband LSST results. Instead, subtracting lens light from three-band LSST images decreases errors by 41, 36 and 34 per cent for Einstein radius, ellipticity and orientation, respectively. This is compared to decreases of 34, 28 and 27 per cent for *Euclid* VIS and 40, 32 and 30 per cent for LSST r-band, resulting in an average decrease of 34 ± 5 per cent across the parameters and data sets. Orientation error changes the least since this is dominated by the larger errors associated with predicting orientation for rounder lenses, seen in Figure 2.10.

Training and testing on multiband images containing lens light produces higher errors than those of single-band images with lens light subtracted. This demonstrates that successful subtraction of the foreground lens light leads to a more significant reduction in errors than incorporating multiband data. However, this lens subtraction is for the ideal case where light profile parameters are known, so this result may be lessened depending on the choice of subtraction technique.

Overall, the network achieves the greatest performance for LSST *gri* images in which lens light has been subtracted, having the lowest errors for all parameters. Clearly networks trained on images without lens light still benefit significantly from the use of multiple colour bands, which increases the precision of all predicted parameters.

2.2.3 Investigation III - Light & Mass Profile Alignment

The presence of lens light in the images used to train and test the network clearly has a significant influence on its performance. The data sets used to investigate this in the previous section contained ellipticities and orientations that could differ between the mass and light profiles. Such a scatter is observed in the real Universe (e.g. Koopmans et al., 2006; Bolton et al., 2008b), and can offset the positions of lensed sources (Harvey et al., 2016). In this section, a series of tests are conducted to assess how well the network performs with varying mismatch between mass-light alignment in the training and testing images, which was used to inform the data sets used in the previous section. For each image, the parameters of the mass profile were first selected (as described in Section 2.1.1), followed by those of the light profile. For the latter, the ellipticity e_{light} and orientation ϕ_{light} were drawn from two normal distributions with means set by the corresponding mass profile values, and with variable standard deviations in order to control the scatter in alignment between the profiles, such that

$$\phi_{\text{light}} = \mathcal{N}(\mu = \phi_{\text{SIE}},\, \sigma_{\phi})\,, \tag{2.1}$$

$$e_{\text{light}} = 1 - q_{\text{light}} = 1 - \frac{q_{\text{SIE}}}{\mathcal{N}(\mu = 1,\, \sigma_{e})}\,, \tag{2.2}$$

where $q = b/a$ is the axis ratio, $\mathcal{N}(\mu,\sigma)$ is a normal distribution of mean μ and standard deviation σ, and the standard deviations of the orientation offset and ellipticity offset are σ_{ϕ} and σ_{e}, respectively. Hence, larger values of σ_{e} and σ_{ϕ} correspond to less correlation between the light and mass profiles.

Three tests were performed, in each case training on 50,000 images and testing on 10,000 images, with the imaging characteristics of the *Euclid* telescope. For the first two tests, the σ for one parameter was varied in the training and test data while the other was kept at $\sigma = 0$ in both data sets. For the third test both σ_{e} and σ_{ϕ} were varied by the same multiples of their SLACS values (see Section 2.1.2) in the training and test data. For these tests, we used more highly magnified sources by restricting source positions to within the lens caustic, rather than within the Einstein radius as before, to gain a clearer view of the impact of the mass-light alignments of orientation and ellipticity on CNN performance and parameter errors.

In Figure 2.12, in most cases the parameter errors remain relatively insensitive to the offset between mass and light orientation regardless of how the CNN is trained. For example, the figure indicates how a network trained on a scatter of $\sigma_{\phi} = 60°$, akin to almost no correlation between the profiles, will vary little with the scatter of the test data set, as it is clearly not associating the two profiles. However, we observe notable trends for all three parameters when trained on images with no

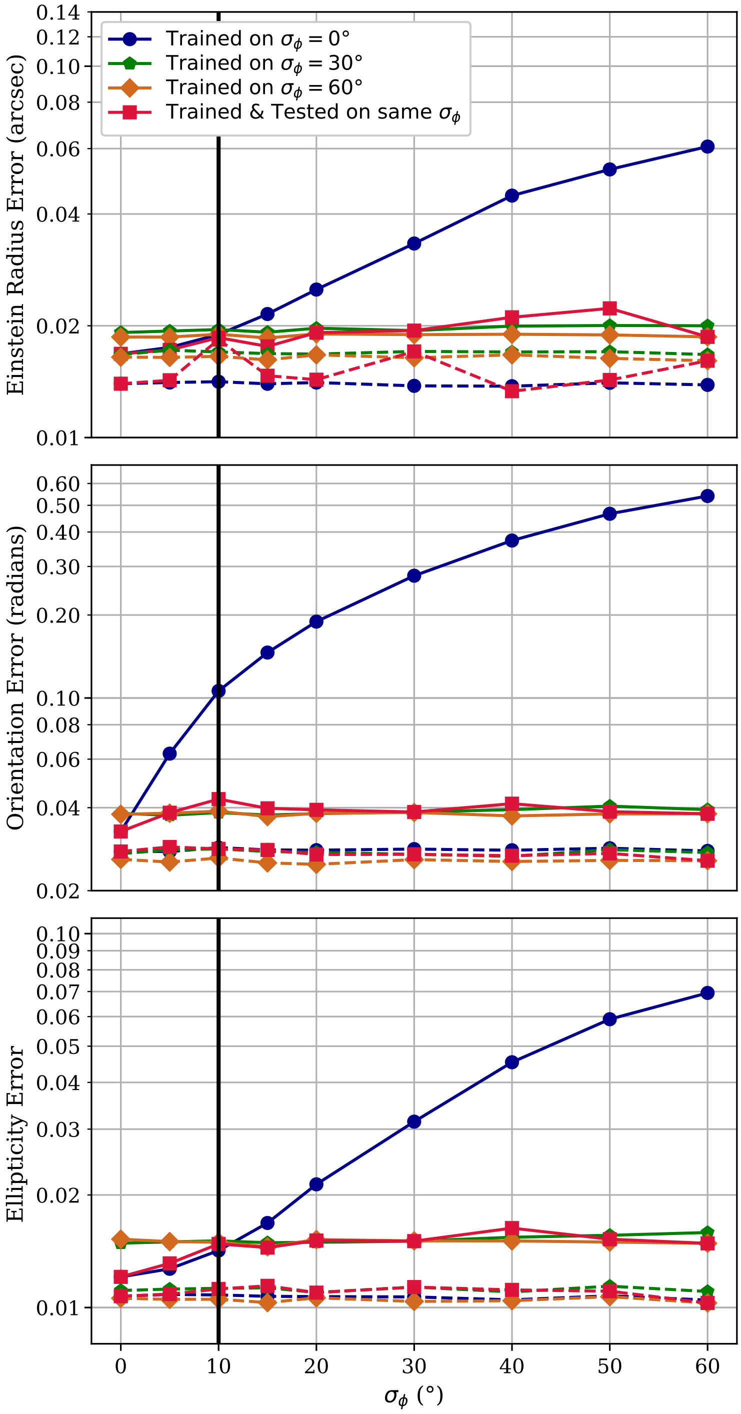

Figure 2.12: Error on mass model parameter predictions as a function of the scatter between light and mass orientation in the test data. For these results, the ellipticity of the light profile follows that of the mass with zero scatter (i.e. $\sigma_e = 0$), for both training and test data. From top to bottom: Results for Einstein radius, orientation and ellipticity. The dashed lines indicate the results for data sets with lens light subtracted, which hence act as a control. The vertical black line indicates the standard deviation used elsewhere in this work, $\sigma_\phi = 10°$, based on SLACS results.

offset (given by the blue circles). In this case, the CNN more easily determines lens parameters when both mass and light profiles in the test set are more closely oriented.

When trained with $\sigma_\phi = 0°$ the network associates the foreground lens light with the mass profile parameters, relying on the lens light to determine them, and so its predictions are less reliable when tested with greater scatter in mass-light orientation. This can clearly lead to large errors, especially for orientation, when the network is applied to such misaligned systems. Einstein radii are obtained from studying the lensed arcs of the source, and for elliptical lenses different lens orientations and ellipticities for the same source position will result in different arc configurations. Hence, the trend seen in the upper plot of Figure 2.12 is less pronounced than the other parameters as to obtain the Einstein radius the network must still rely more on the lensed source light. For ellipticity error as a function of σ_ϕ, shown in the lower plot of Figure 2.12, the ellipticity of the light and mass profiles are kept identical, so if the network only learned from the lens light then it should perform equally well regardless of σ_ϕ. The fact that this is not observed thus indicates that the network is still learning from the lensed source light to an extent when predicting mass profile ellipticity.

Errors in parameter predictions are also presented in Figure 2.13, now when varying the scatter between the light and mass profile ellipticity, σ_e, in the test set. When trained with $\sigma_e = 0$, similar trends are again observed for Einstein radius and ellipticity, whose errors are more affected by this scatter in ellipticity than the previous scatter in orientation. This is likely because from Figure 2.11 we know that ellipticity is already generally hard for the CNN to predict, so it may rely more heavily on that of the light profile which in turn affects the prediction of Einstein radius: Should the ellipticity expected from the lensed source light be different to that of the light profile, the CNN would handle this discrepancy through (incorrectly) changing its prediction of Einstein radius. However, for ori-

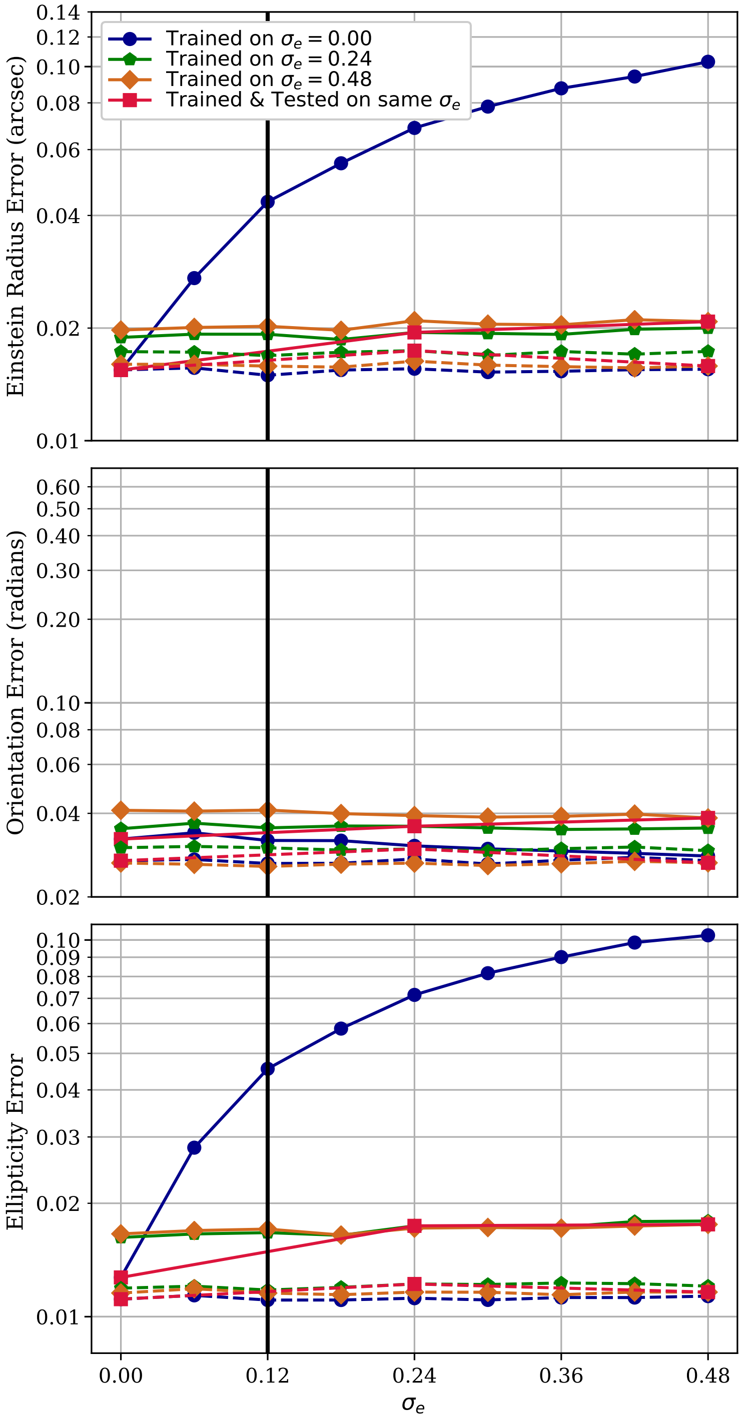

Figure 2.13: Error on mass model parameter predictions as a function of the scatter between light and mass ellipticity in the test data. For these results, the orientation of the light profile follows that of the mass with zero scatter (i.e. $\sigma_\phi = 0$), for both training and test data. From top to bottom: Results for Einstein radius, orientation and ellipticity. The dashed lines indicate the results for data sets with lens light subtracted, which hence act as a control. The vertical black line indicates the standard deviation used elsewhere in this work, $\sigma_e = 0.12$, based on SLACS results.

entation error the $\sigma_e = 0$ trend we observed in Figure 2.12 has disappeared, as the CNN can just use the lens light to obtain the direction of ellipticity. But what if the lens light appears circular? As discussed when varying σ_ϕ, the network utilises the lensed source light to measure ellipticity even with the presence of lens light, and so can use the direction of this distortion to measure orientation.

It should be noted that in the middle panel of Figure 2.13 the $\sigma_e = 0$ trend for orientation error has not only flattened, but actually reversed somewhat. This would suggest that a larger scatter between the test set's light and mass profile ellipticities can result in lower orientation prediction errors. However, this is because the mass profiles' axis ratios in the test sets are sampled from a normal distribution with a mean of 0.78 that is closer to the upper limit of 1 (an ellipticity of zero) than the lower limit of 0.4 (an ellipticity of 0.6). As a result, the tails of the distribution are unevenly cut off, biasing the sampling of light profile ellipticities to slightly higher values on average for larger σ_e, with these larger values allowing for more precise orientation predictions, as per Figure 2.10.

Figure 2.14 shows the results of simultaneously scattering both ellipticity and orientation between the light and mass profiles of the test data. With the inclusion of lens light, the lowest average error is achieved when training and testing on lenses with zero scatter between the profiles, with values for $\{\theta_E(\text{arcsec}),\ \phi(\text{rad}),\ e\}$ of $\{0.015,\ 0.032,\ 0.013\}$. However, this is still an average of 11 ± 8 per cent larger than that of images without lens light, for which errors of $\{0.015,\ 0.027,\ 0.011\}$ were obtained by the network. We also observe similar trends to before; training with no scatter produces high prediction errors when mass does not closely follow light. Training instead with a moderate amount of scatter, i.e. somewhere between a $(\sigma_e,\ \sigma_\phi)$ of (0.12, 10) and (0.24, 20), gives consistent network performance across a wide range of mismatch between the mass and light profiles. Across the tests, no systematic biases in predicted parameter values were identified except if trained on zero scatter. In this case the network tends to over-estimate the Einstein radius

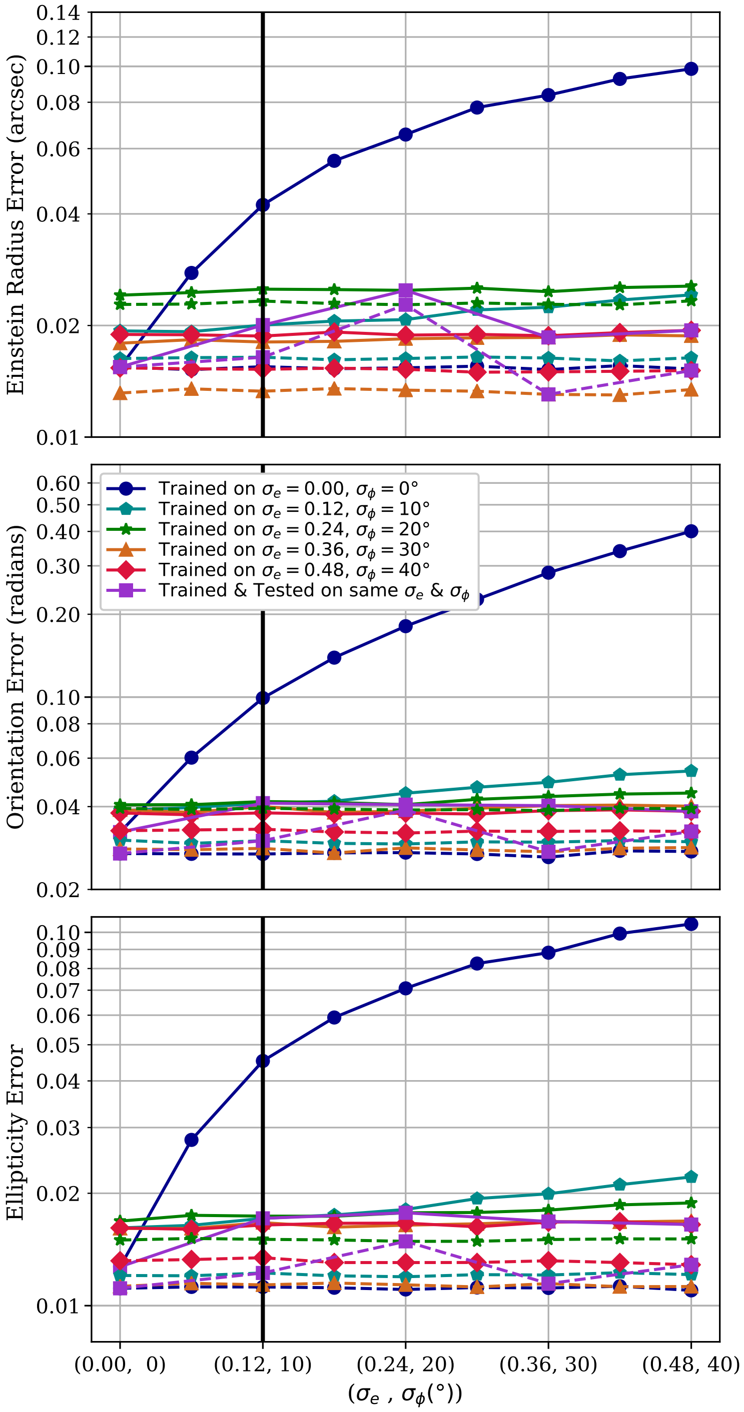

Figure 2.14: Error on mass model parameter predictions as a function of the light-mass scatter of both ellipticity and orientation in the test data. From top to bottom: Results for Einstein radius, orientation and ellipticity. The dashed lines indicate the results for data sets with lens light subtracted, which hence act as a control. The vertical black line indicates the standard deviations used elsewhere in this work, ($\sigma_e = 0.12$, $\sigma_\phi = 10°$), based on SLACS results.

and ellipticity by a small but significant amount of up to 2 per cent, again because of the previous discussion of biased sampling of light profile ellipticities and the effect on Einstein radii predictions when such ellipticities are different to those that would be predicted based on the lensed source light.

From these results, it is clear that despite predicting mass profile parameters the CNN will attempt to learn from the lens light if it is present in the images. This can in turn result in poor accuracy should a realistic difference in mass-light alignment not be taken into account during training. As such, the training and test data sets used elsewhere in this work contained moderate scatter between the light and mass profiles, with $\sigma_e = 0.12$ and $\sigma_\phi = 10^\circ$, to both align with observed SLACS distributions and to ensure the network did not rely too heavily on the lens light.

2.2.4 Investigation IV - Signal-to-Noise & Image Stacking

So far, we have investigated the performance of the network with regard to lens parameters, of both the mass and light profiles, as well as the difference between imaging characteristics of the LSST and *Euclid*. Another major factor that influences modelling performance is the signal-to-noise ratio (SNR) of the lensed images. With these surveys set to repeatedly cover the sky over several years, many images of gravitational lenses can be stacked over time to improve the signal. In this section, I show how network performance depends on the SNR of the images used for training and testing. I limit the discussion here to images with lens light subtracted since the previous trends and relative errors discussed when comparing inclusion and subtraction of lens light remain the same here. Two tests are conducted: the first investigates how CNN errors vary with the SNR for both LSST and *Euclid* images. The second puts this into the context of surveys, showing how CNN errors depend on the number of stacked survey visits and how this changes with the number of stacked survey visits assumed during network training. For this second test, while only results for *Euclid* images are presented, we investigate stacks of up to 100 visits to estimate the expected relative performance gain available for LSST.

To begin, the SNR was determined as follows. The 5×5 pixel regions in each corner of the image were used to obtain the mean and standard deviation of the background noise. The mean background was subtracted from the image, and the N pixels making up the lensed source were then chosen as those above 3-sigma of the background noise. Finally, the SNR was defined for the total signal S of the lensed source as

$$\mathrm{SNR} = \frac{S}{\sqrt{S + N\sigma^2}} \tag{2.3}$$

where $\sqrt{S}$ is the shot noise of the lensed source and σ is the observed standard deviation of the background noise, encompassing all other sources of noise (sky background, dark current and readout noise)[1]. In the first test, I generated a training set and test set containing 50,000 and 10,000 single-exposure images. Their mass-light orientation and ellipticity offsets were scattered by $\sigma_\phi = 10°$ and $\sigma_e = 0.12$ as discussed in Section 2.2.3, and lens light then subtracted following the procedure in Section 2.1.2. The randomly drawn lens model parameters give rise to a distribution in SNR within the training and test data. To assess how the network performance depends on SNR, I binned the test images by SNR and measured the errors of predicted lens model parameters in each bin.

Results for *Euclid* VIS, LSST r-band and LSST gri images are respectively plotted in the upper, centre and lower panels of Figure 2.15. These errors initially decrease rapidly with higher signal-to-noise, however there is little decrease in error past approximately SNR=100. There is also some scatter in the higher signal-to-noise images due to the network having fewer of these brighter, lower redshift galaxies to train on. The results quantify how the network performs significantly better

---

[1] From testing, the use of $\sqrt{N}\sigma$, obtained from the four 5×5 pixel regions, as a measure of the background noise in Equation 2.3 is found to be accurate to within 2 per cent of that obtained from combining readout noise with shot noise from the sky background and dark current.

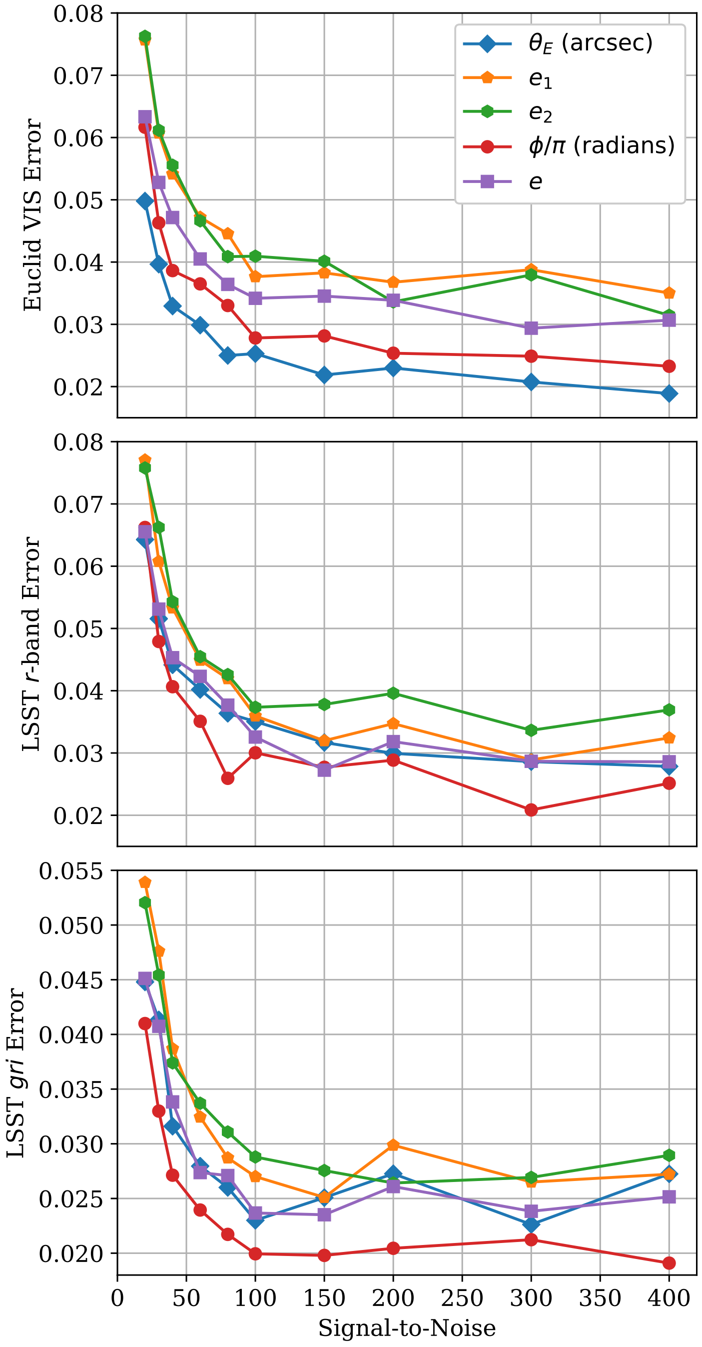

Figure 2.15: Errors for the CNN's predicted lens model parameters as functions of the signal-to-noise ratio (SNR) of the source in the test images. Both training and test images had their lens light subtracted to leave shot noise residuals. From top to bottom: simulated images for the *Euclid* telescope VIS filter, LSST r-band and LSST gri. Data points lie at the lower SNR bin boundaries, with the last point encompassing all images with SNR$\geq$400. The blue diamonds, orange pentagons, green hexagons, red circles and purple squares correspond respectively to Einstein Radius, the first and second components of complex ellipticity, orientation and ellipticity of the lens. The values of orientation ϕ are divided by pi to more easily compare the parameters.

on higher signal-to-noise images, with prediction errors for SNR$\geq$400 images a factor of 1.6-2.7 (mean of 2.2 ± 0.3) times smaller compared to that of SNR=20-30 images. Hence, parameter predictions for higher SNR images are much more likely to be closer to their true values, with little change in results beyond around SNR$\geq$100.

For the second test, I analysed how errors vary with the stacking of training and test images, and investigated how the resulting SNR of the training data set affects the performance of the CNN. Until now we have only considered single-visit images when training and testing the CNN, however *Euclid* is expected to undertake four exposures of each object (Laureijs et al., 2011; Racca et al., 2016; Cropper et al., 2018), and likewise LSST will undertake tens to hundreds of visits of each object (Abell et al., 2009). As such, training sets of 50,000 simulated *Euclid* images were produced for different noise levels corresponding to one, four, 10 and 100 visits. The four networks that resulted from training on these data sets were then applied to four test sets of 10,000 images generated with the same range of noise characteristics. The test sets all shared the same lens and source parameters, as did the training sets, in order to avoid the changing background noise influencing parameter selection that could otherwise bias lower SNR data sets towards higher magnification images. While ideally dithering would be performed to accurately replicate observations, to ease image generation no dithering was performed here as the images contained no artefacts (such as hot pixels, fixed pattern noise, cosmic rays, or satellite trails) that would necessitate this additional processing. With each test set containing a different number of stacked visits, each contained a different distribution of SNR. While these could extend to high SNR values, the distributions were all skewed towards lower values. For single-visit and stacked 4-, 10-, and 100-visit data sets, peaks in SNR occurred at 30, 40, 50, and 80, respectively. All combinations of test and training SNR were performed to investigate the impact of insufficient CNN training.

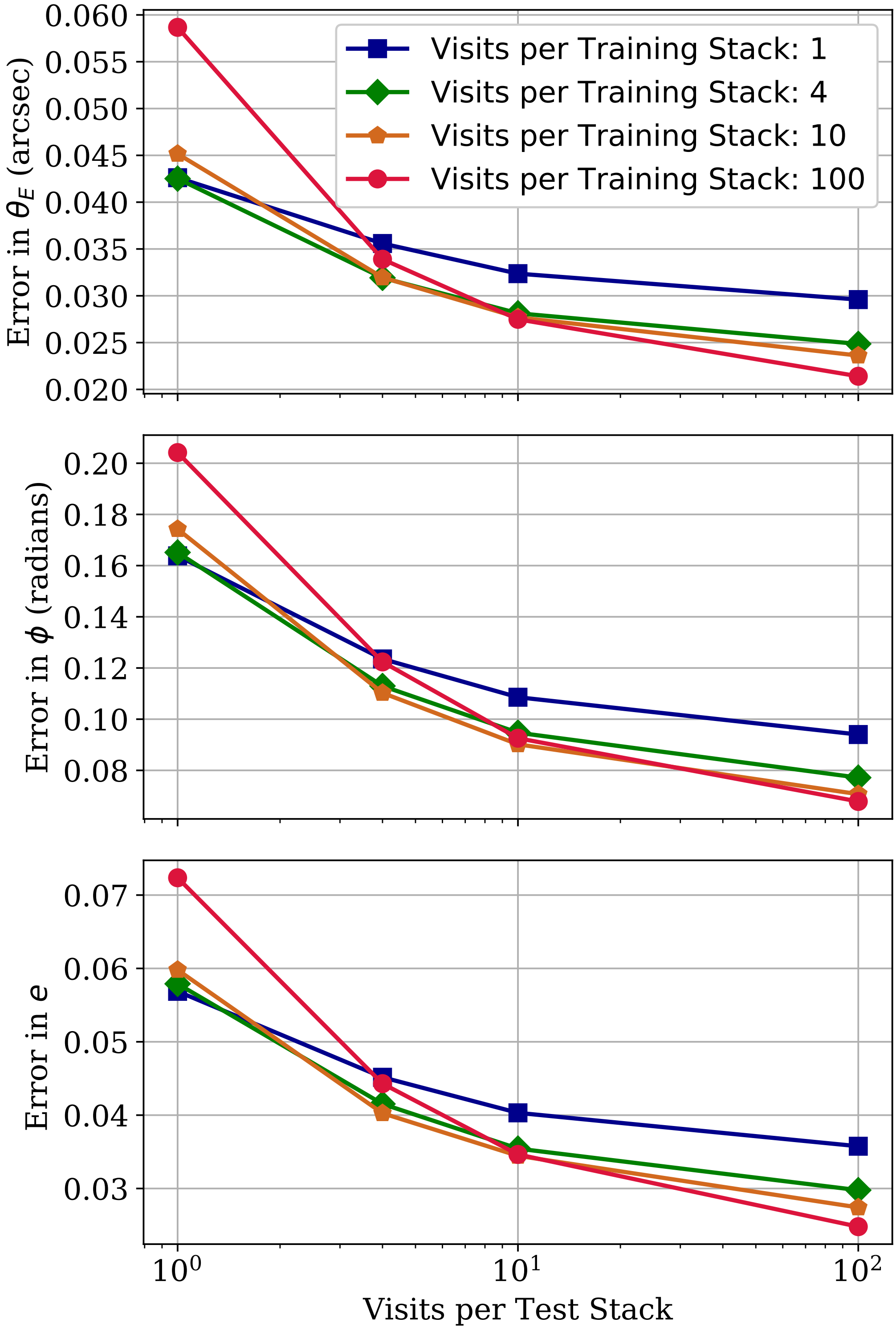

Figure 2.16: Errors for the CNN's predicted lens model parameters as functions of the number of visits making up each stacked image in the test set. From top to bottom are the Einstein radius, orientation and ellipticity of the lens mass profile. In each case, blue squares, green diamonds, orange pentagons and red circles correspond to where the network was trained on stacked images made up of 1, 4, 10 and 100 visits, respectively.

Figure 2.16 shows the errors for the predicted mass model parameters as functions of the number of visits making up stacked images in the test data set. It is clear that, over the range tested, network performance always improves when applied to more deeply stacked test images, as expected. For example, when trained on single-visit images (blue squares), testing on stacked images of 100 visits rather than single-visit images results in an average decrease in CNN error of 37 ± 5 per cent across the parameters $\{\theta_E(\text{arcsec}), \phi(\text{rad}), e\}$, from values of {0.04, 0.16, 0.06} to {0.03, 0.11, 0.04}. Likewise, performing this comparison when training on images containing 4, 10 and 100 visits results in average decreases in CNN error of 48 ± 5, 54 ± 5 and 65 ± 1 per cent, respectively. Additionally, when averaged across the four trained networks, errors are reduced by 31 ± 8, 41 ± 9, and 51 ± 11 per cent compared to testing on single-visit images when instead testing on 4-, 10-, and 100-visit stacked images, respectively.

Furthermore, there is a notable increase in accuracy for deeper stacked test images when the network is retrained on multi-visit stacks. When testing on stacks of 10 or more visits, a network trained on any number of stacked images will perform notably and consistently better than one trained only on single-visit images. The less noisy training data from even a few stacked visits clearly provide more information for the CNN to learn from. Compared to training on single-visit images, training on stacked images of 100 visits reduces average errors by 29 ± 1 per cent when both networks are tested on such 100-visit stacks. Conversely, when presented with fewer visits making up the test stacks, the network generally performs better when trained on fewer-visit stacks, as expected: a network trained on single-visit images achieves errors that are 23 ± 3 per cent smaller than a network trained on 100-visit images when tested on single-visit data sets. However, even when training and testing on images with the same number of visits the CNN does not achieve the same errors. Compared to training and testing on single-visit images, training and testing on 4-, 10-, and 100-visit stacked images decreases errors on average by 28 ± 2, 40 ± 4, and 55 ± 4 per cent, respectively, with the latter resulting in values of {0.02, 0.07, 0.02}.

These results highlight that for sufficient CNN training one must incorporate a large range of signal-to-noise ratios into the training set, and that there should be an emphasis on training on images with lower SNRs both in order to reduce their larger errors presented here, and because images with these ratios will make up the majority of early data releases. One could train a single CNN on a range of stacked visits, requiring a much larger training set, or train multiple CNNs to each handle a set number of stacked images (or even a select range of SNR). Regardless, there are significant improvements to CNN performance when utilising the stacking of images obtained by *Euclid* and LSST, whether that involves a few visits or hundreds. However, throughout the rest of this thesis I focus on training and testing on single-visit images; these contain the highest fraction of low SNR images so provide a lower bound for the CNN's accuracy and are the first images to become available from upcoming surveys.

2.3 Summary

In this chapter, I presented my work in investigating the practicalities faced when using CNNs for lens modelling, with a focus on application to the upcoming LSST and *Euclid* surveys. To begin, a convolutional neural network was constructed to estimate the mass model parameters of lenses from images, detailed in Section 2.1.1. Data sets of images were simulated for network training in Section 2.1.2, utilising SLACS lens data and SIE mass models, resulting in strong galaxy-galaxy lens images characteristic of those produced by the *Euclid* VIS filter and LSST (for both r-band and multiband gri imaging). The simulations allowed for the fast generation of guaranteed lenses, and benefited from multiple SEDs for various galaxy ages and metallicities, along with realistic distributions for parameters such as axis ratios and mass-light profile scatter. The CNN could then be trained on such data sets in order to predict the values of three lens mass model parameters: Einstein radius, and two components of complex ellipticity later converted to ellipticity and orientation.

Results of the investigations were presented in Section 2.2. The first involved training and testing the CNN on images with foreground lens light subtracted, and again with this lens light included. Results were compared for the different parameters and surveys, in terms of both overall errors and as functions of the lens parameters themselves. It was found that the multiple bands of LSST *gri* images resulted in the lowest CNN errors, and that the subtraction of lens light decreased errors by 27-41 per cent (an average of 34 ± 5 per cent) across the data sets and parameters. All parameter errors increased for smaller Einstein radii, and the same was seen in all but orientation error for increasing ellipticities. For the case where lens light is not subtracted, the next investigation found that differences in the mass and light profile alignment during training and testing greatly impacts network performance. I then examined how network performance scales with signal-to-noise ratio of the lensed source, observing that increasing SNR past 100 had little impact on the results. Finally, this was put in the context of future surveys by studying the improvement gained through stacking images within both the training and test sets. Together, these investigations can act as a guide for future training, and highlight the promising use of CNNs for lens modelling and their potential for further development in this area.

Chapter 3

From CNN to BNN: Enhancements and Improvements

We have now seen that convolutional neural networks can be applied effectively to lens modelling for data sets with the imaging characteristics of the upcoming *Euclid* survey and LSST. We have explored how network accuracy varies with lens mass and light as well as the impact of noise, however one obvious question arises: how does the network compare to conventional lens modelling methods? Before we can perform this comparison, it is worth noting that the images used for the above investigations have remained fairly simple. Additionally, until now the CNN has only been able to predict single values for the lens parameters, without any notion of uncertainty on individual predictions. Improvements to both of these areas would increase the network's adaptability and functionality, respectively, enabling it to handle more realistic data whilst producing additional information on its predictions. Hence, in this chapter I will begin by using the now-established method of incorporating Bayesian inference into CNNs in order to produce a Bayesian neural network (BNN) capable of predicting its own uncertainties. This is followed by details of the new network's architecture and training and testing process, before finishing with the enhancements made to the training and testing images in order to produce more complex data sets.

This chapter begins with Section 3.1 outlining the current and previous research performed with regard to Bayesian neural networks, including their initial application to strong gravitational lensing. The theoretical description of this probability framework and the established method of its incorporation into neural networks is summarised in Section 3.2. As part of the integration into neural networks, a fine-tuning process is required to balance an introduced hyperparameter and this is detailed in Section 3.3. Section 3.4 details the Bayesian CNN's architecture and training and testing process used for the remainder of this thesis, while the content of the training sets, including an overview of power law mass models, are given in Section 3.5. Section 3.6 covers the multiple data sets that will be used to test lens modelling methods in Chapter 4, followed by a short summary in Section 3.7. This chapter features work contained in my publication *'Strong lens modelling: comparing and combining Bayesian neural networks and parametric profile fitting'* (Pearson et al., 2021).

3.1 Background

Standard CNNs for regression problems such as those of lens modelling can be trained to predict parameter values, yet there is no way for these networks to reliably quantify uncertainties on such values. For classification networks, Denker and LeCun (1990) proposed that Bayesian statistics could be incorporated should network weights be assigned probability distributions rather than single values. Hence such networks would output meaningful probability distributions instead of point estimates, such as through repeated sampling of these weight probabilities; see Section 3.2.1. These Bayesian neural networks were studied further (e.g. MacKay, 1992; Neal, 1995; Graves, 2011; Hoffman et al., 2013), including the incorporation of variational inference (Jordan et al., 1999); see Section 3.2. However, the methodologies of these works are both challenging to implement and computationally expensive, especially with the rise of deep learning models.

More recently, we have seen the incorporation of an approximate Bayesian framework that can be applied to both classification and regression networks (Gal and Ghahramani, 2016a,b; Kendall and Gal, 2017). This is achieved primarily through the application of dropout (randomly selecting nodes and setting them to zero: see Srivastava et al., 2014) before each weight layer as a form of variational inference; this and dropout are detailed in Section 3.2.2. When applied to regression, this computationally inexpensive approach doubles the number of outputs because the network now predicts both values and their associated uncertainties. Levasseur et al. (2017) have since applied this to create an approximate Bayesian CNN framework for strong lens modelling, and in the following sections I present an implementation of this framework applied to my own work.

3.2 Bayesian Neural Networks

3.2.1 Bayesian Probability

In a Bayesian framework, uncertainties are quantified through replacing point estimates with full probability distributions. We first seek to obtain such distributions for the model parameters $\boldsymbol{\omega}$ (neural network weights and biases; not to be confused with mass model parameters) before moving on to output predictions. These posterior distributions are computed using Bayes' Theorem,

$$p(\boldsymbol{\omega}|\boldsymbol{D}) = \frac{p(\boldsymbol{D}|\boldsymbol{\omega})p(\boldsymbol{\omega})}{p(\boldsymbol{D})} = \frac{p(\boldsymbol{D}|\boldsymbol{\omega})p(\boldsymbol{\omega})}{\int p(\boldsymbol{D}|\boldsymbol{\omega})p(\boldsymbol{\omega})d\boldsymbol{\omega}} \tag{3.1}$$

where our initial beliefs about model parameters are contained within the prior distribution $p(\boldsymbol{\omega})$. Given some (training) data $\boldsymbol{D}$, this is updated through the likelihood $p(\boldsymbol{D}|\boldsymbol{\omega})$, and normalised using the Bayesian evidence $p(\boldsymbol{D})$ which involves marginalising over all possible parameters.

When new (test) data are introduced, we can determine the probability of an output $\boldsymbol{y}$ given an input $\boldsymbol{x}$ by calculating the posterior predictive distribution,

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{D}) = \int p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\omega}) p(\boldsymbol{\omega}|\boldsymbol{D}) d\boldsymbol{\omega}. \tag{3.2}$$

This Bayesian model averaging (BMA) involves taking a weighted average over all parameters by combining each parameter's posterior probability (akin to neural network training) with the probability $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\omega})$ of an output given that parameter and the input data (akin to neural network testing with a suitable probabilistic cost function).

However, both the posterior and the posterior predictive distribution require marginalising over the entire parameter space, which is computationally expensive and often intractable due to its complexity. This is especially true for deep neural networks as they can contain millions of parameters in the form of weights and biases, and their non-linearity ensures that no analytical solution can exist.

Instead, we seek ways of approximating this marginalising integral. While there are Markov Chain Monte-Carlo (MCMC) sampling-based methods such as Hamiltonian Monte Carlo and the Metropolis-Hastings algorithm, they can prove time-consuming to run and challenging to implement into neural networks (Charnock et al., 2022). Hence, for this work we focus on an alternative approach that can closely approximate a Bayesian framework, that of variational inference.

3.2.2 Variational Inference

Variational methods aim to treat inference as an optimisation problem, producing an approximation of the otherwise intractable posterior probability distribution for the parameters and latent variables. This involves optimising over a set of tractable *variational* distributions to select the one closest to the true posterior probability, then sampling from that distribution to obtain an approximate solution. Such methods are frequently better suited for stochastic gradient optimisation and usage with GPU acceleration than standard sampling methods.

With regard to neural networks, we have training data $(\boldsymbol{X}, \boldsymbol{Y})$ consisting of input images $\boldsymbol{X}$ and their corresponding true values $\boldsymbol{Y}$, and the set of unobserved weights and biases $\boldsymbol{\omega}$, which together produce the posterior $p(\boldsymbol{\omega}|\boldsymbol{X}, \boldsymbol{Y})$. Hence, for test data $(\boldsymbol{x}, \boldsymbol{y})$, the probability distribution of predicted values (the posterior predictive distribution) is obtained through marginalising over the weights,

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{X}, \boldsymbol{Y}) = \int p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\omega}) p(\boldsymbol{\omega}|\boldsymbol{X}, \boldsymbol{Y}) d\boldsymbol{\omega}. \tag{3.3}$$

We now seek a (tractable) variational distribution $q(\boldsymbol{\omega})$ that closely matches the posterior, i.e. $p(\boldsymbol{\omega}|\boldsymbol{X}, \boldsymbol{Y}) \approx q(\boldsymbol{\omega})$. This is achieved through minimising the Kullback-Leibler (KL) divergence between the two,

$$D_{\mathrm{KL}}(q(\boldsymbol{\omega})||p(\boldsymbol{\omega}|\boldsymbol{X}, \boldsymbol{Y})) = \int q(\boldsymbol{\omega}) \log \frac{q(\boldsymbol{\omega})}{p(\boldsymbol{\omega}|\boldsymbol{X}, \boldsymbol{Y})} d\boldsymbol{\omega} \tag{3.4}$$

which returns the expectation of the (logarithmic) difference between the distributions. Identical distributions would produce a value of $D_{\mathrm{KL}} = 0$ while deviations from one another cause the logarithm term to move away from zero, and weighting this by $q(\boldsymbol{\omega})$ ensures regions of higher probability are prioritised. Through Bayes' Theorem, minimising the KL divergence is found to be equivalent to maximising the log-likelihood of the network's prediction for a given input, however the theory and derivations behind this will not be explored further here.

For application to neural networks, a form of $q(\boldsymbol{\omega})$ is chosen such that the set of variational parameters to be optimised (i.e. weights) in a given layer are multiplied by a set of Bernoulli-distributed random variables with a given probability. Sampling from $q(\boldsymbol{\omega})$ is hence equivalent to performing dropout over network weights, incorporating dropout before each weight layer. Introduced by Srivastava et al. (2014), dropout is the process of randomly 'dropping' nodes (setting them to zero) temporarily for each forward pass through the network during training. This stops the network from becoming too reliant on certain pathways in order to prevent it from overfitting to its training data.

After training, performing inference using the approximated posterior predictive distribution

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{X}, \boldsymbol{Y}) \approx \int p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\omega}) q(\boldsymbol{\omega}) d\boldsymbol{\omega} \tag{3.5}$$

can be done by approximating this integral with a Monte Carlo (MC) integral, through repeatedly passing each test image through the network and using dropout during *testing* (known as MC dropout) to sample from the approximate parameter posterior. This sampling thus gives a measure of the *epistemic* uncertainty, i.e. how well the network has been trained.

The other type of uncertainty to be obtained is the *aleatoric* uncertainty, which describes errors coming from the input data itself such as the noise level of images. It is found that the earlier minimising of the KL divergence is equivalent to maximising the log-likelihood of the training set output parameters. Hence, obtaining aleatoric uncertainties can be achieved through modifying the network's cost function, choosing its form to be the negative of a Gaussian log-likelihood

$$-\mathcal{L} = \sum_k \frac{1}{2} \left\| y_{\mathrm{pred},k} - y_{\mathrm{true},k} \right\|^2 \exp\left(-s_k\right) + \frac{1}{2} s_k . \tag{3.6}$$

Here, $y_{\mathrm{pred},k}$ and $y_{\mathrm{true},k}$ are the predicted and true values of parameter k for a given training image, and $s_k = \log \sigma_k^2$ is the log-variance. σ_k represents the aleatoric uncertainty of each parameter, and due to its presence in both terms it is optimised through training and can be outputted for each image fed through the network. The log-variance is implemented rather than σ_k itself in order to avoid potential division by zero and to improve numerical stability.

This computationally inexpensive approach of variational inference produces an approximate Bayesian neural network whose weights, even after training, can change according to a probability distribution rather than remaining fixed. Sampling of the resulting posterior is achieved through repeated testing (taking the standard deviation as the epistemic uncertainty), with the network now outputting both predictions of the parameter values and predictions of their associated aleatoric uncertainties. The latter are then added in quadrature with the epistemic uncertainties to give the overall measure of the CNN's uncertainty for a given image. The median of the predicted values is taken to be the final result, with such averaging additionally serving to increase network accuracy over that of standard CNNs.

3.3 BNN Fine-Tuning

As part of the implementation of variational inference with dropout, we require the introduction of an additional hyperparameter, the dropout rate (or its reverse, the dropout keep rate). This gives the fraction of neurons in a given layer that are dropped, i.e. set to zero, at random during each forward pass through the network. Like other hyperparameters, the dropout rate must be fine-tuned in order for the CNN to reliably predict uncertainties, and this was carried out in the form of a grid search as follows.

To begin, multiple instances of the network were created with dropout applied throughout, except before the final output layer as this otherwise drastically decreased performance. They were each given a different dropout rate between 0.01 and 0.2, hence dropping between 1 and 20 per cent of neurons, but otherwise featured the same architecture and training process (see Section 3.4). They trained on 100,000 images, and the fine-tuning process was performed for both SIE and

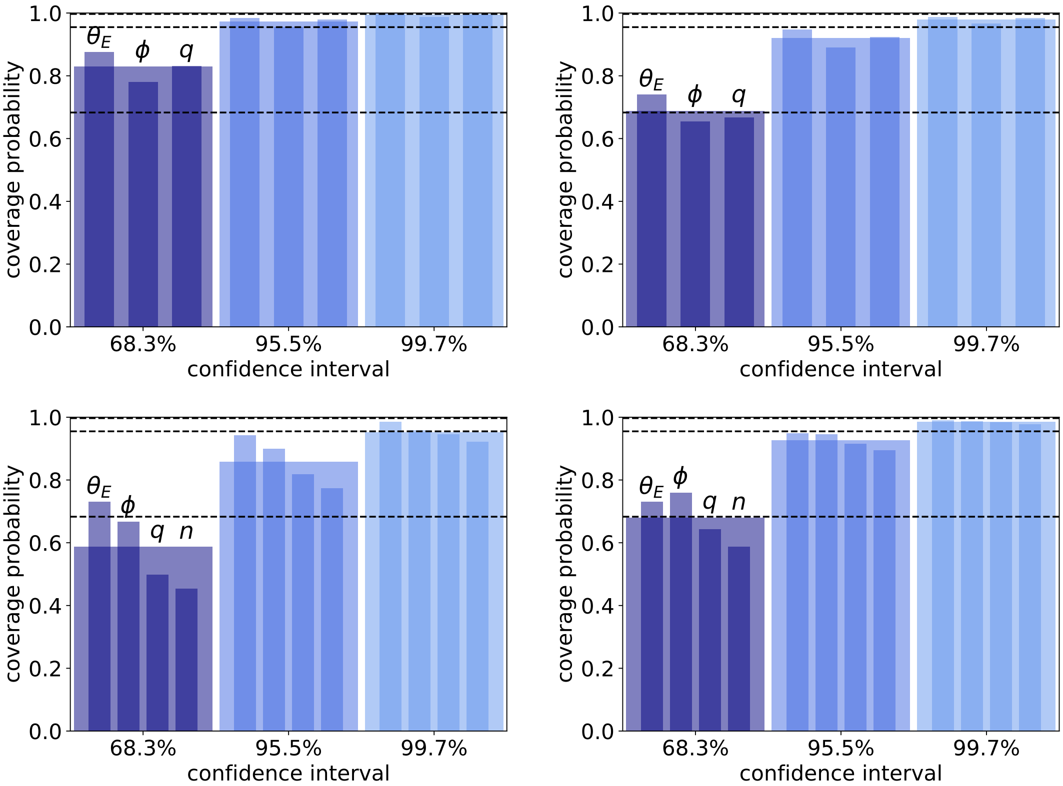

Figure 3.1: Coverage probabilities when fine-tuning dropout during CNN training to ensure appropriately-sized predicted uncertainties. Top and bottom rows show results for the CNN trained to predict SIE parameters (θ_E, ϕ, q) and power law parameters (θ_E, ϕ, q, n), respectively, indicated by the smaller bars. Mean coverage probabilities are given as the wider bars, and for the ideal case would reach each dashed line representing 68.3, 95.5 and 99.7 per cent coverage. Left: example results from using incorrect dropout rates (0.2 and 0.01 for SIE and power law, respectively). Right: results from using fine-tuned dropout rates.

power law lenses (see Section 3.5). The CNNs were then tested on 10,000 images covering parameter space with the same imaging characteristics as the training set, and predict parameter values and their associated uncertainties. Fine-tuning involves determining how accurate these predicted uncertainties are compared to the true scatter in predicted values across the test set (and hence across parameter space).

As such, coverage probabilities were obtained for each parameter; these give the fraction of the test set whose predicted value lies within a given confidence interval of the true value. These confidence intervals are determined based on the predicted 1σ uncertainties, additionally being scaled up to 2σ and 3σ intervals. Hence, the ideal case would show 68.3 per cent of test set predictions lying within the 1σ confidence interval, and so on. Some example coverage probabilities are given in Figure 3.1. After training and testing were performed, a dropout rate of 5 per cent (i.e. a keep rate of 95 per cent) was found to generate the most accurate uncertainties and so was implemented as the dropout rate for the rest of this thesis.

It is worth noting that even though the average coverage probabilities matched their desired confidence intervals, this was not true for each individual parameter, as shown in Figure 3.1. However, a dropout rate of 5 per cent also minimised the deviation between such individual parameter coverage probabilities. Nevertheless, this is an inherent limitation of the approximate Bayesian inference method that can lead to under- and over-estimated parameter uncertainties: to obtain correct coverage probabilities for each parameter would require a separate network for each parameter, which would then require significantly longer training and would lose out on understanding the complex relationships between parameters. As such, I continued to pursue a single network approach, with the development of a more complete Bayesian formalism for neural networks left to other work. The requirement of fine-tuning the dropout rate also presents a problem in its reliance on the training set; this is discussed in Section 5.3.

3.4 BNN Improvements

Through the incorporation of an approximate Bayesian formalism, we now have a Bayesian CNN capable of predicting its own uncertainties as well as parameter values. In the following section, I provide details of the network used hereafter in this thesis, in terms of the improvements to its design and operation over those in Chapter 2. The network architecture is given in Section 3.4.1 while Section 3.4.2 discusses the processes by which this network is trained and tested.

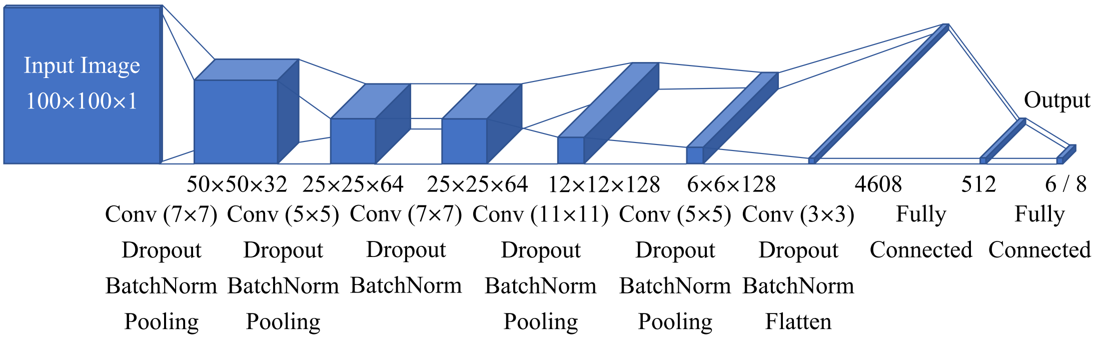

Figure 3.2: Structure of the neural network used in the remainder of this thesis, showing the input image and the output of each block of layers, in the same format as Figure 2.1. The network shares the same ordering of layers as before, however output dimensions have increased and dropout is now performed after each convolutional layer. Further details can be found at the end of Section 3.4.1.

3.4.1 Architecture

With the incorporation of Bayesian uncertainties and the increase in data set image size (see section 3.4.2), the neural network architecture now contains over 4 million trainable weights and biases. As shown in Figure 3.2, it features the six convolutional layers, four pooling layers and two fully connected layers that were used in Chapter 2. However, it now includes six dropout layers, with each applied after a convolutional layer and set to a dropout rate of 5 per cent. Additionally, following the work in Section 3.2.2, the cost function used in Chapter 2 has been replaced with a negative log-likelihood function, but otherwise the training process remains the same.

In order to ensure that the network performs optimally, the code GPyOpt (The GPyOpt authors, 2016) was used to try and optimise neural network hyperparameters that would otherwise require manual fine-tuning. These included the learning rate, convolutional kernel dimensions, layer widths (i.e. the number of outputs from a given layer), and the number of each type of layer. This avoids purely trial-and-error optimisation by utilising a Bayesian framework to update hyperparameters over a series of model iterations. Several of these optimisation runs were performed, from varying all hyperparameters simultaneously to only a few at a time, each run being allowed up to 100 iterations of training and testing to minimise its error. However, none were found to outperform the existing network architecture following its above changes, and so no further modifications were deemed necessary.

In Chapter 4, the Bayesian CNN is made to fit lenses with a power law profile (see Section 3.5) and so the slope parameter n of this profile, along with its associated uncertainty, is predicted in addition to the existing three parameters. Apart from these two extra outputs, this power law network features the same architecture and hyperparameters as those of the regular SIE-trained network.

3.4.2 Training and Testing

To train the network, 100,000 images were fed into the network in batches of 100, along with a validation set of 10,000 images. These 100×100-pixel 'postage stamp' cutouts of lenses featured a range of source types (see Section 3.5) in order for the network to cope with the wide range of realistic sources. Two Bayesian CNNs were trained: one to predict parameters for SIE lenses and a second for power law lenses. In both cases, the same pre-processing of images and true values were used as those employed in Section 2.1.1, including the predicting of complex ellipticity. However, to ease interpretations and comparisons of the results with other methods, going forward I provide orientation values in degrees and use axis ratios q in place of ellipticities e. For the latter change, the relation between the two is simply $e = 1-q$, and so the magnitudes of the results and their uncertainties remain unchanged.

Running on the same GPU machine as used in Chapter 2, training over 150 epochs on 100,000 images took less than two hours, while thousands of images were tested in the order of a few seconds to minutes. During testing, median averages for the predicted values and predicted uncertainties were obtained through repeated testing of each image, following the work in Section 3.2.2. This repeated testing samples the posterior, resulting in both reliable predicted uncertainties and more accurate predicted values. Hence the time taken depended on the number of repeats performed; for the results presented in this work each image was tested 100 times to provide sufficient accuracy in a manageable amount of time.

3.5 Training Data Set Simulation

This section details the gravitational lens image simulations used for training the Bayesian CNN in the remainder of this thesis. We focus now only on data sets with the imaging characteristics of the *Euclid* telescope's VIS filter to avoid repeat results, with the image size increased to 100×100 pixels to ensure they contain all light from lensed arcs. As with the previous CNN, the training set contained parametric lens mass models and parametric source light profiles, with central positions again following a normal distribution about the image centre with a standard deviation of one pixel.

As well as training to predict parameters of SIE mass models, I also trained separately to predict those of the power law ellipsoid (e.g. Tessore and Metcalf, 2015). The density profile

$$\rho(r) \propto r^{-n} \tag{3.7}$$

of this more general model changes with radius r according to an additional parameter, the power law slope n, that allows for increased flexibility but increases the complexity of the parameter space when modelling (see Section 1.2 for more details, including the use of power law deflection angles given by 1.28). As such, the training set was generated twice, once for each mass model, in order to train one CNN to predict three SIE parameters and another for the four power law parameters. The parameters of these training sets were uniformly distributed within the following ranges: 0.2-3.2 arcsec, 0-180°, 0.4-1.0, and 1.5-2.5 for Einstein radii,

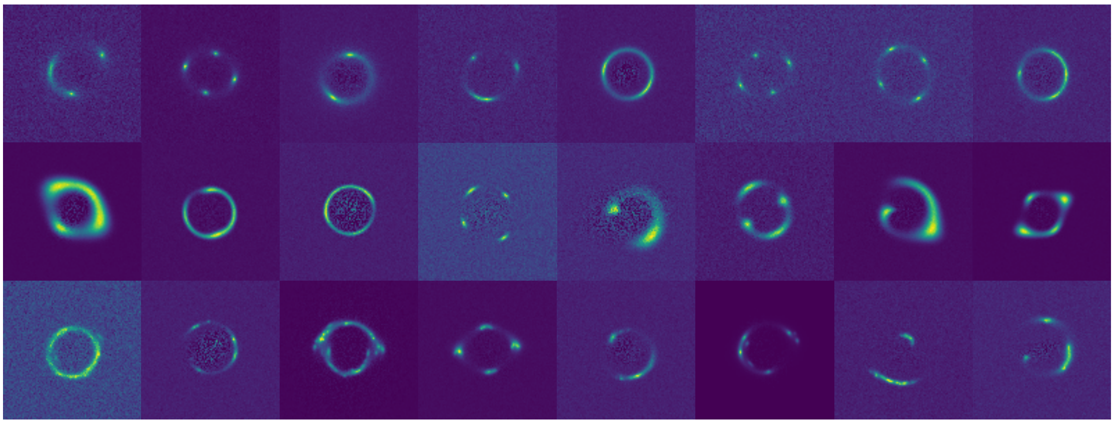

Figure 3.3: Examples of the simulated images used to train the Bayesian CNN. Top and middle rows contain Sérsic and Gaussian profile sources, respectively. The bottom row contains complex sources made up of multiple Sérsic profiles. The data set was simulated to have the expected imaging characteristics of the *Euclid* telescope's VIS filter.

orientations, axis ratios and slopes, respectively.

In addition to the sources with parametric Sérsic profiles used in Chapter 2, the training data set contained sources with Gaussian profiles as well as complex sources made up of multiple Sérsic profiles, in order to cover a wider range of morphologies. These Sérsic or Gaussian profiles were generated with variable Sérsic indices and variances respectively, each making up 25 per cent of the training set while the remaining 50 per cent contained multiple Sérsic profiles. Foreground lens light was once again subtracted after convolving with the point spread function to leave only shot noise residuals. Examples of the images used to train the Bayesian CNN are shown in Figure 3.3. With the above changes and improvements, the simulations otherwise followed the methodology detailed in Section 2.1.2.

3.6 Realistic Image Testing

Having discussed the Bayesian CNN's training data sets, we now detail those used to test the network and other modelling methods in the next chapter. These modelling methods were tested on a range of increasingly complex lensing systems, beginning with smooth SIE lenses and parametric sources before incorporating real Hubble Ultra-Deep Field (HUDF; Beckwith et al., 2006) source galaxies and CosmoDC2 (Korytov et al., 2019) line-of-sight (LOS) structure. Additionally, the foreground lenses were later replaced with more complex power law mass profiles or galaxies from the Evolution and Assembly of Galaxies and their Environments (EAGLE) cosmological hydrodynamical simulation suite (Crain et al., 2015; Schaye et al., 2015). As these data sets contained characteristics that differed from those of the training data, they provided a measure of the network's adaptability when applied to unseen data. These data sets are presented in more detail below.

The 1000-image data sets used for testing the lens modelling methods were generated with the Pipeline for Images of Cosmological Strong Lensing (PICS; Li et al., 2016a), through private communication with the author as the code has not been made publicly available. These data sets made use of their selection of HUDF galaxies for high-resolution, high-redshift sources, although I did not attempt to deconvolve these sources for this work. These single-band test images had the same size and imaging characteristics as the training set, but did not contain any shot noise residuals of the lenses as the errors introduced by lens removal depend on the technique used and are not the focus of this work. The magnitudes and angular sizes of the HUDF source galaxies were incorporated into the images, including the data set containing parametric (Sérsic and Gaussian) light profiles. As such, lens and source redshifts had the sole effect of changing the Einstein radii, so for the purposes of the simulation redshifts were taken to be $z_{lens} = 0.5$ and $z_{source} = 2.0$, respectively, with Einstein radii set manually.

PICS also allows for the inclusion of LOS structure through full ray-tracing of small light cones covering lens caustics, with light cones from the semi-analytic model CosmoDC2 based on the Outer Rim cosmological N-body simulation (Heitmann et al., 2019). One thousand of these small light cones were selected from within the field of view of the semi-analytic catalogue, each containing a primary lens

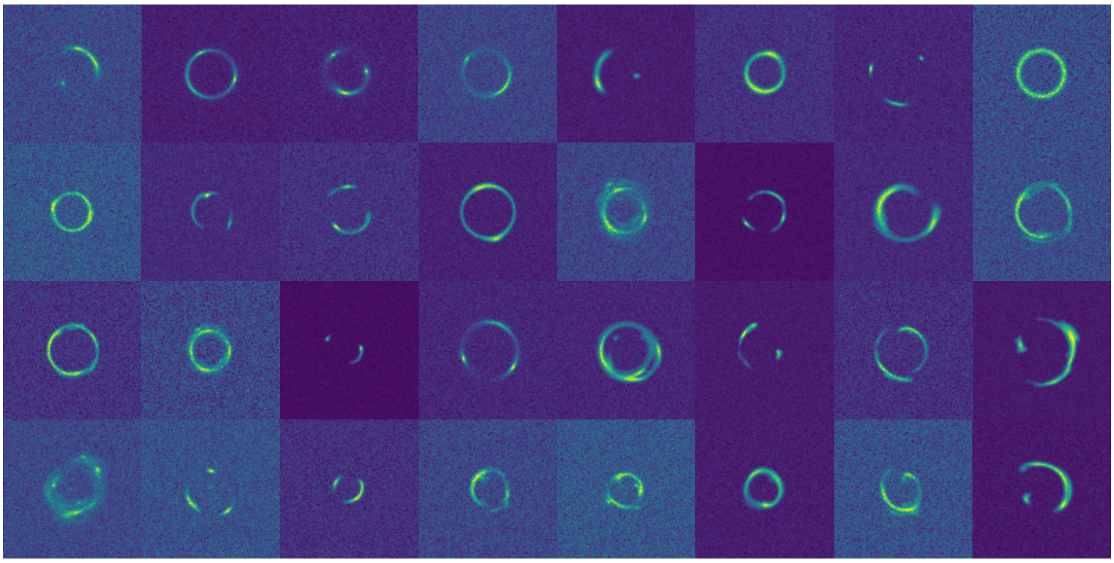

Figure 3.4: Examples of the simulated images produced from PICS software, used to test the different mass modelling methods. Top row: examples containing SIE lenses with parametric source profiles, with four Sérsic sources followed by four Gaussian sources. Second row: examples containing SIE lenses with HUDF sources, with the last four containing LOS structure. Third row: examples containing power law lenses with HUDF sources, with the last four containing LOS structure. Bottom row: examples containing EAGLE galaxy lenses with HUDF sources, with the last four containing LOS structure. The data sets were simulated to have the expected imaging characteristics of the *Euclid* telescope's VIS filter.

often residing in a galaxy group, with up to a hundred LOS galaxies between the observer and source. Test data sets were produced for a range of increasingly complex lensing systems, examples of which are shown in Figure 3.4:

- SIE lenses + parametric sources
- SIE lenses + HUDF sources, with and without LOS structure
- Power law lenses + HUDF sources, with and without LOS structure
- EAGLE galaxy lenses + HUDF sources, with and without LOS structure

For the SIE lens mass profiles, their parameter distributions followed those determined by PICS to be the main haloes in each CosmoDC2 light cone, with resulting Einstein radii, orientations and axis ratios distributed within the ranges 1.3-2.6 arcsec, 0-180° and 0.4-1.0, respectively. The power law lenses used the same distribution of lens parameters, with the addition of slope values randomly selected from a uniform distribution in the range 1.5-2.5, in line with the neural network training data. These main halo lenses were on average more massive, and therefore produced larger Einstein radii, than the majority of those expected by the *Euclid* survey. However, for this work these masses were not changed from the CosmoDC2 values as, where possible, we wished to maintain the halo masses and positions along the LOS for realistic ray-tracing through the light cones.

The EAGLE galaxy lenses were selected from the EAGLE galaxy catalogue by calculating the Einstein radius of each galaxy from their halo mass (assuming redshifts of $z_{lens} = 0.5$ and $z_{source} = 2.0$) and selecting one thousand of those with Einstein radii within the range 0.5-5.0 arcsec. This resulted in a sample of galaxies whose Einstein radii were naturally distributed between 0.5 and 3.2 arcsec. The parameters used as the true values for the results in Section 4.3.4 therefore followed different distributions than those of the CosmoDC2 haloes, although orientations and axis ratios were limited to the same ranges. They were obtained using PICS by performing a least squares fit to their known convergence maps (computed from the EAGLE particle data; see later), weighted by the square root of these maps, and supplying SIE mass profiles as the curve-fitting function. The logarithmic forms of the convergence maps were used as these eased the fitting process, and the fitting was performed five times with the median average taken for each parameter. The maps were also masked such that only pixels within a 3 arcsec radius of the lens centre were fitted, as masks of this radius provided the closest parametric fit to the maps with the lowest uncertainties in the fitted parameters. The same process was also used to obtain parameters for the data sets containing LOS structure, fitting either an SIE or power law mass profile. A discussion of the limitations of this fitting process is presented in Section 5.5.2.

To obtain a convergence map for a given EAGLE galaxy, PICS was supplied with its EAGLE particle data containing the three-dimensional positions and particle masses of each type of particle (dark matter, stars, and gas). A box size of

20×20 arcsec with a width of 512 pixels was found to be sufficient for this work, in that higher resolution boxes produced no discernible difference in the true values obtained from fitting convergence maps, nor in the resulting lensed images, and instead only led to convergence maps taking significantly longer to fit. Taking $z_{lens} = 0.5$, PICS then converted each type of particle data into a (pixellated) two-dimensional surface density profile. These were added together and finally converted to a convergence map by dividing by the critical surface density (Equation 1.15) assuming $z_{source} = 2.0$.

While the modelling methods only examine the observed images, the inclusion of LOS structure increases the observed Einstein radius compared to that of a solitary foreground lens. To eliminate such biases in the results, rather than using the Einstein radii of the foreground lenses as the true values, we instead measured the observed Einstein radii of each lensing system using the same convergence map fitting process as discussed previously.

3.7 Summary

In this chapter, I began by establishing the need for obtaining neural network uncertainties and more complex images in order to reliably compare its modelling to conventional techniques. I described the process by which the CNN in Chapter 2 was replaced by a Bayesian neural network through the implementation of the approximate Bayesian formalism detailed in Section 3.2 and the dropout fine-tuning process in Section 3.3. Details of the new network architecture, training and testing process were given in Section 3.4, while Section 3.5 presented the improvements made to the neural network's training set, including a description of the power law profile that the network will attempt to model. Lastly, Section 3.6 provided descriptions of the multiple new data sets of varying complexity for the network to be tested on. Summaries of both the training and test data sets

Table 3.1: Summary of the training and test data sets simulated in this chapter. Both Sérsic and Gaussian profiles are used for data sets containing parametric source light profiles.

Lens mass profile	Source light profile	LOS structure
Training		
SIE	Parametric	None
Elliptical power law	Parametric	None
Testing		
SIE	Parametric	None
SIE	HUDF	None
SIE	HUDF	CosmoDC2
Elliptical power law	HUDF	None
Elliptical power law	HUDF	CosmoDC2
EAGLE galaxy	HUDF	None
EAGLE galaxy	HUDF	CosmoDC2

are presented in Table 3.1.

With these in mind, we can now train the Bayesian CNN to handle more complex images, provide uncertainty measurements on its predictions rather than single values, and test on more realistic data sets. In the next chapter, we will explore how the Bayesian CNN compares to conventional modelling when applied to such images to identify the strengths and weaknesses of each, as well as investigating the potential combination of the two modelling methods.

Chapter 4

Combining CNNs with parametric modelling methods

Given the growing prevalence of neural networks both inside and outside of astronomy, it is well established that while trained neural networks are much faster at modelling, they are limited by the quality of their training data; mismatches between these and real observations can make neural networks less reliable than their conventional parameter-fitting counterparts. In this chapter, we seek to answer the following: how well can we expect convolutional neural networks (now approximate Bayesian neural networks) to obtain lens mass model parameters for increasingly realistic systems in future large-scale survey images? How do they compare to conventional parameter-fitting techniques? What benefits arise from a combination of machine learning and parametric modelling methods and can the disadvantages of one be counteracted by the other?

To answer these questions, methodologies of the previous chapter are combined in order to compare the modelling efficiency of approximate Bayesian CNNs (see Section 3.2) to that of a conventional parametric profile fitting approach, for the range of increasingly complex lensing systems detailed in Section 3.6. In this work, we choose the semi-linear inversion method of the open source code PyAutoLens (Nightingale et al., 2018) for this conventional fitting approach. In addition, we explore a new combination of the CNN with PyAutoLens, using the predictions of the former to inform the prior distributions of the latter. As discussed in Section 3.5, here we focus only on the imaging characteristics of the *Euclid* telescope to avoid an excess of results; the performance when applied to LSST imaging characteristics can be extrapolated following the work in Chapter 2.

This chapter starts by covering the parameter-fitting methods used in this work before presenting their test results alongside those of the CNN. An overview of PyAutoLens, along with the details of its configuration used in this work, are given in Section 4.1, with those of the combination method given in Section 4.2. Test results for each method are presented in Section 4.3, from application to images with simple parametric lenses and sources to those with hydrodynamical EAGLE galaxy lenses with real HUDF sources and line-of-sight structure. These are followed by an analysis of modelling speeds in Section 4.4, and a summary of this chapter in Section 4.5. The content of this chapter features work from my publication *'Strong lens modelling: comparing and combining Bayesian neural networks and parametric profile fitting'* (Pearson et al., 2021). The work presented is mine apart from the parametric modelling using PyAutoLens (see Section 4.1) which was conducted by Jacob Maresca, who coauthored the paper. In such modelling, I supplied Jacob Maresca with the image test sets (images only; no true parameter values), along with Bayesian CNN's predicted values and uncertainties where necessary (see Section 4.2). He then modelled the images using PyAutoLens and returned to me its output predictions, from which I then analysed the results.

4.1 PyAutoLens

The software PyAutoLens[1] (Nightingale et al., 2021) is used for modelling strong gravitational lenses, employing the semi-linear inversion method initially developed by Warren and Dye (2003). It simultaneously models the foreground lens and background source for strongly-lensed systems in a Bayesian framework, and can do so using a range of configuration options including modelling on a square or adaptive Voronoi grid, regularisation, and applying different source plane weightings such as weighting by source brightness. An example of PyAutoLens strong lens modelling is shown in Figure 4.1.

The result of modelling is a χ^2 statistic of the differences between the observed and model images, used to calculate the Bayesian evidence. By repeatedly updating the lens model parameters, this evidence is then maximised over a series of modelling iterations in an attempt to obtain the set of parameter values that correspond to the global maximum. For this work, the background source was reconstructed on a Voronoi grid that adapted to the lens magnification, with gradient regularisation applied uniformly across the source plane and allowed to vary in strength according to a regularisation prior. The inversion process was initialised using the input priors on the mass profile (see below), the regularisation prior, and a prior controlling the number of pixels in the source reconstruction.

Each pixel in the image plane was sub-gridded into 4 sub-pixels for ray tracing calculations: given the image resolution, further sub-gridding gave no noticeable improvement and only served to increase modelling time. Images were masked to increase modelling efficacy; this required fitting an annulus to significant image pixels in an automated fashion in order to cope with the many large test sets. To create the masks, minimum cross-entropy thresholding (Li and Lee, 1993) was applied to each image, producing a thresholded image whose calculated moments

---

[1] `https://github.com/Jammy2211/PyAutoLens`

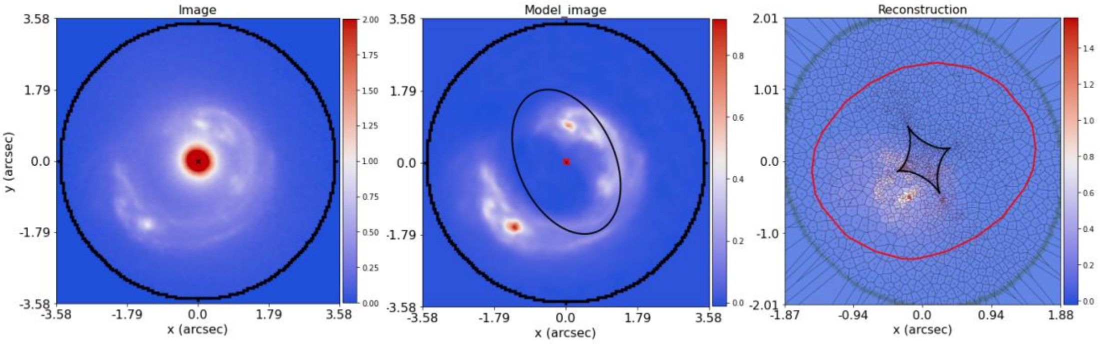

Figure 4.1: PyAutoLens modelling of a Hubble Space Telescope image containing the strong lens SLACSJ1430+1405. From left to right: original image, a fit to the lensed source galaxy, and the corresponding unlensed source reconstructed on a Voronoi grid. Critical lines and caustics are shown in the middle and right panels, respectively. (Figure credit: Nightingale et al., 2021)

were used to obtain its centroid. Then centred on this centroid was a circular annular mask, whose inner and outer radii were selected based on the innermost and outermost non-zero pixels in the thresholded image, to produce the thinnest annulus containing all such non-zero pixels. To avoid masking fainter source light, the width of the annular mask was slightly increased, with these inner and outer radii scaled by 90 per cent and 110 per cent, respectively. For an optimiser, PyAutoLens used the (Py)MultiNest Bayesian inference tool (Feroz et al., 2009; Buchner et al., 2014) for Markov chain Monte Carlo (MCMC) sampling of parameter space, run in constant efficiency mode with a sampling efficiency of 0.4.

PyAutoLens requires priors for the mass model parameters which are normally set manually by human inspection of the lensed image. The widths of the bounds of the priors are set relatively arbitrarily, and modelling can be reinitialised with modified priors should it converge on a poor fit. However, in this work we sought to compare PyAutoLens to the CNN though testing on thousands of lensed images, and so a method of automating PyAutoLens was required. As such, PyAutoLens was initialised using priors set either blindly or using CNN-predicted values (see Section 4.2). For the former, before modelling, Einstein radii were approximated by fitting a circle to image pixels above a 3-sigma threshold, by performing a least squares fit[2] using uniform weights. The radius of this circle, with uniform ±0.2 arcsec bounds, was then used as the prior for Einstein radius. This eased the fitting process while still leaving it fully-automated, so remained a fair comparison to the other automated modelling methods.

4.2 PyAutoLens + CNN

The parametric fitting of PyAutoLens can model the lens and source to a high degree of accuracy and in such a way that any poor fits are easily identifiable from their reconstructed images. But to do so requires a large amount of time and computing power as well as the manual effort of setting priors on the parameters. These are factors we would like to reduce, as they currently limit its applicability to the quantity of strong galaxy-galaxy lenses that future surveys will discover. PyAutoLens can be crudely automated by setting priors to cover a large region of parameter space (‘blind’ modelling) rather than basing them on human inspection, however this results in a drop in accuracy and further increased modelling time. In addition, the widths of the priors can present problems - if the bounds are too narrow, the software may not converge to the correct solution, while too large bounds result in many more iterations before a solution is obtained. Even then, wide bounds on the priors can result in the software converging on parameters that represent overmagnified or undermagnified solutions rather than recovering the true parameters (Maresca et al., 2021). While the inversion process can be initialised using a fit with a parametric source to aid in this endeavour, this can confuse modelling when presented with images of a complex source.

On the other hand, a trained CNN provides a much faster, computationally inexpensive and automated alternative, but requires a sufficiently large and realistic

---

[2] This used the `skimage.measure.CircleModel` function from the open source Python library scikit-image (van der Walt et al., 2014).

training set in order to avoid underfitting or overfitting to the data. While the CNN may not always be able to reach the high accuracy that PyAutoLens can achieve with human inspection (e.g. 0.1 per cent error on Einstein radius with negligible bias, for mock lenses modelled by PyAutoLens in Cao et al. (2022); see Section 4.3.1), compared to the automated blind modelling that would be required by PyAutoLens to cope with future surveys the CNN can potentially provide a more efficient and accurate method. Additionally, CNN predictions can act as priors for PyAutoLens, with such a combination of methods removing the need for human inspection and allowing for slower, yet still automated, accurate modelling to deal with more complex lenses.

PyAutoLens initialises lens model parameters by randomly sampling from a series of user-defined prior distributions which can be uniform or Gaussian. When setting up, these are typically manually selected based on a visual inspection of the lensed image. Automation of this process would require either generous uniform priors to ensure that no parameters are disallowed during fitting, or a means of making an estimation of a narrower set of priors. Since the CNN can now predict values for mass model parameters extremely quickly, this opens up the possibility of using these to inform the priors for PyAutoLens. Not only this, but the uncertainties that the CNN predicts can be used as the 1σ bounds, now using a Gaussian prior distribution (centred on the CNN's predicted parameters) rather than a uniform one.

Such a combination of CNNs and conventional parametric modelling can help alleviate the shortfalls that each method faces individually: CNN accuracy is limited by the quality of the training set, which ideally would feature non-parametric source and lens profiles for many hundreds of thousands of images, and without these the network may struggle with more complex lensing systems. Meanwhile, conventional parametric modelling does not require training, and hence offers a more flexible means of modifying the lensing model, such as changing the mass profile and the easy addition of components such as external shear. However, this modelling can take a large number of iterations before converging on a solution, and may require repeat modelling attempts.

Together, CNNs can vastly simplify the search over parameter space for parametric modelling methods and can additionally prevent methods using semilinear inversion, like PyAutoLens, from falling into local minima, especially when using CNN uncertainties to reduce the size of prior parameter space. Hence this combination may lead to a fully automated pipeline that can obtain accurate results over short time-scales for application to large samples of lenses.

4.3 Modelling Results for Increasingly Complex Lensing Systems

We now compare the performance of the approximate Bayesian CNN created in Section 3.4 to that of PyAutoLens modelling blindly ('PyAL (blind)' hereafter) for the range of increasingly complex test cases presented in Section 3.6. These techniques are also compared to combinations of the two, in which the CNN predictions are used as priors for PyAutoLens. The first ('PyAL + CNN') uses uniform priors of arbitrary width that are centred on the CNN-predicted model parameters. The second ('PyAL + CNN (1σ)') uses Gaussian priors centred on these parameters with the CNN-predicted 1-sigma uncertainties acting as the prior widths; see Section 4.2. We also tried setting the prior widths to twice the 1-sigma uncertainties reported by the CNN in case the uncertainties were underpredicted for the more complex test sets. However, overall the results were worse than passing the unmodified uncertainties, with distributions of errors (e.g. Figure 4.2) very close to those of PyAL + CNN, and so were not investigated further.

As discussed in Section 3.4.2, orientation values are now presented in degrees to ease interpretation, and likewise axis ratios q are used in place of ellipticities e, following the relation $e = 1 - q$. The uniform prior distributions of PyAL (blind) covered parameter space for all parameters except Einstein radius, which instead used ±0.2 arcsec bounds centred on the Einstein radii estimated by fitting a circular function directly to the lensed images (see Section 4.1). For the PyAL + CNN method, the bounds were ±0.2 arcsec, ±40°, ±0.1, and ±0.2 for Einstein radius, orientation, axis ratio and slope, respectively.

The modelling techniques were applied to simulated images containing SIE lenses, with parametric (Sérsic and Gaussian) and real HUDF sources, the results of which are presented in Sections 4.3.1 and 4.3.2, respectively. In the case of HUDF sources, results were compared for lenses with and without LOS structure, and this was repeated for the case of power law lenses (Section 4.3.3). Section 4.3.4 details how the techniques were instead applied to simulated images with complex lenses in the form of galaxies obtained from the EAGLE hydrodynamical simulation, again with HUDF sources and for those with and without LOS structure. These results, and the results of the thesis as a whole, are also discussed further in Chapter 5.

As in Chapter 2, throughout the remainder of this chapter we again quantify the error on predicted lens model parameters by the 68 per cent confidence interval computed from the distributions of differences between true and predicted parameter values across the test image set. This is separate from the 1σ uncertainty predicted by the CNN for a given predicted parameter value, obtained using the methods described in Chapter 3.

4.3.1 SIE Lenses + Parametric Sources

We began by testing for any inherent biases or differences between the modelling methods. Sets of 1000 images containing SIE lenses and parametric sources, with sources evenly split between Sérsic profiles and Gaussian profiles, were modelled using each method. The results are presented in Figure 4.2 as distributions of the differences between predicted values and true values for each mass model parameter. Fractional differences are used for Einstein radius and axis ratio. Tables 4.1 and 4.2 respectively give the errors and median biases for these distributions, where biases are taken to be the predicted parameter value minus the true value. Scatter plots of predictions against true values are given in Figure 4.3, including the CNN's predicted 1σ uncertainties. It is clear from these results that the CNN's predicted uncertainties are of an appropriate size and accurately reflect its errors; the large uncertainties in some orientation values are expected as they occur for round lenses with no well-defined orientation. It is worth noting that the CNN errors are significantly lower than those presented in Chapter 2; this is primarily due to doubling the size of the training set, which now also contains a greater variety of lensing systems. The incorporation of dropout and uncertainty prediction into the network architecture, along with testing on lenses with generally larger Einstein radii (and hence more available information), also serve to increase network performance.

The CNN by itself achieves a smaller error than that of PyAutoLens for axis ratio, but larger errors for the other parameters. Despite the significantly lower peaks for Einstein radius and orientation, the CNN has far fewer outliers, as shown in Figure 4.3, resulting in similar overall errors. This is because the CNN training prevents the CNN from producing the anomalous results that PyAutoLens can when it encounters local minima far from the correct model, while that very ability of PyAutoLens to explore parameter space enables it to provide highly precise results when it encounters the global minimum. However, for all mass model parameters the combination of the two methods does significantly better than PyAutoLens modelling blindly, with PyAL + CNN (1σ) achieving 36-44 per cent lower errors than PyAL (blind). There is also a notable benefit to including the CNN's predicted 1σ uncertainties in the priors of PyAutoLens, with 12-17

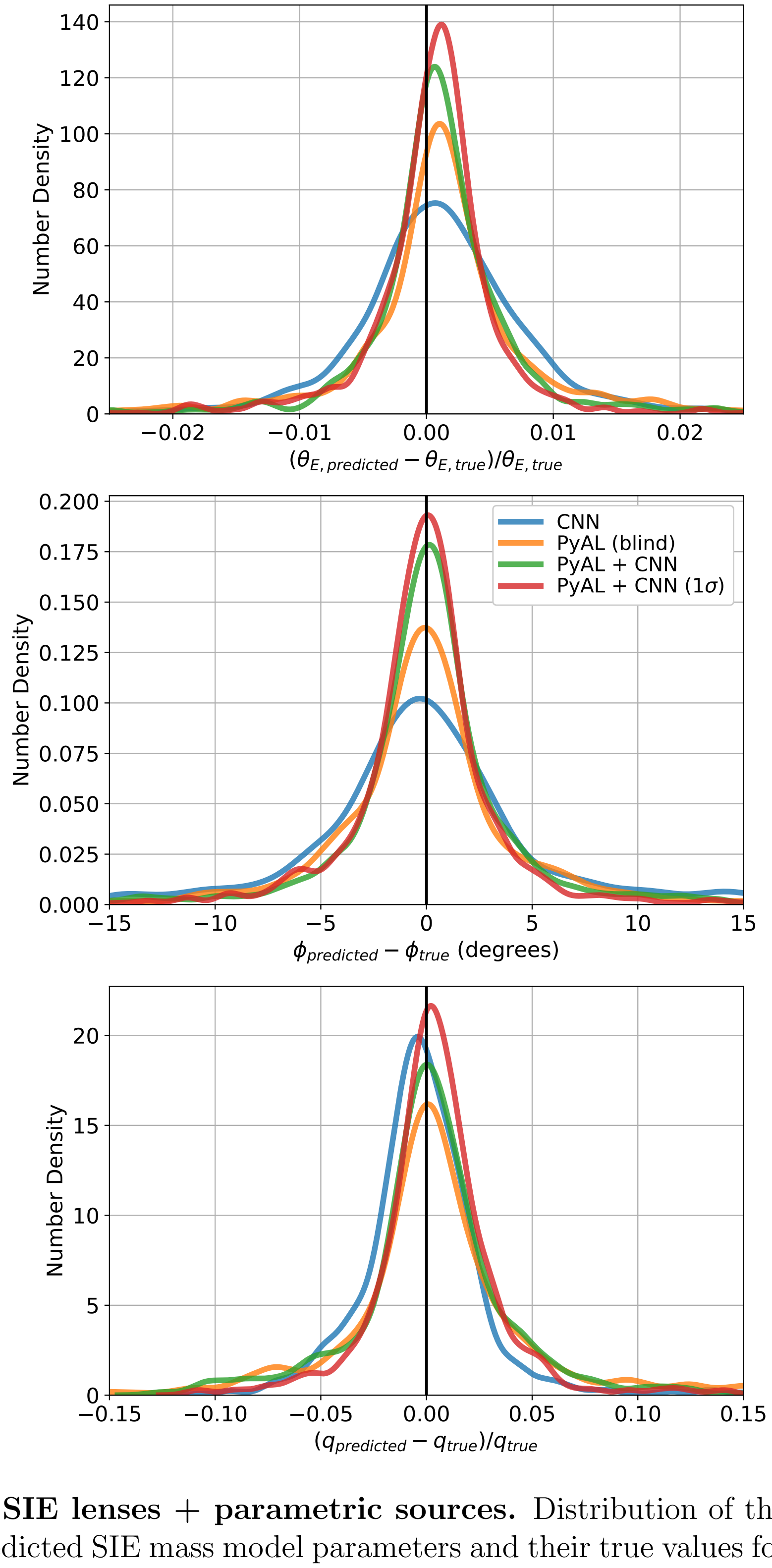

Figure 4.2: **SIE lenses + parametric sources.** Distribution of the differences between predicted SIE mass model parameters and their true values for a test data set of 1000 images containing parametric (Sérsic and Gaussian) sources. From top to bottom: Einstein radius, orientation and axis ratio of the lens mass profile. Einstein radius and axis ratio results presented as fractional differences. The distributions shown are those for the CNN (blue), PyAutoLens modelling blindly (orange), PyAutoLens using CNN predictions as priors (green) and PyAutoLens using CNN predictions and 1-sigma uncertainties as priors (red).

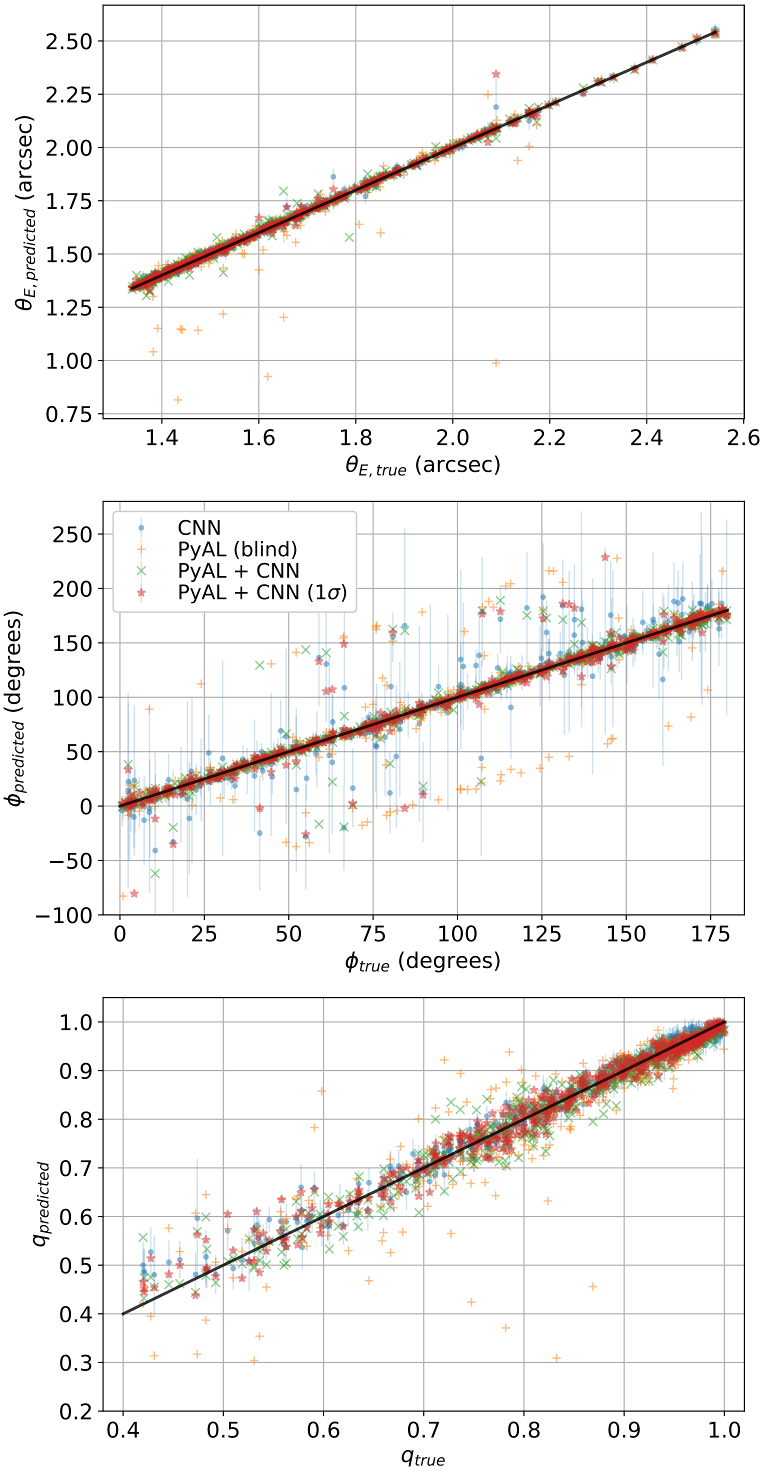

Figure 4.3: **SIE lenses + parametric sources.** Comparison of predicted SIE lens parameters with the true values for test data sets of 1000 images containing SIE lenses and parametric (Sérsic and Gaussian) sources. From top to bottom: Einstein radius, orientation and axis ratio of the lens mass profile.

Table 4.1: **SIE lenses + parametric sources.** The 68 per cent confidence intervals on predicted parameters for each modelling method, computed from the distributions of differences between true and predicted parameter values across 1000 test images.

Method	θ_E (arcsec)	ϕ (°)	q
CNN	0.0090	4.92	0.018
PyAL (blind)	0.0089	4.44	0.027
PyAL + CNN	0.0065	2.81	0.021
PyAL + CNN (1σ)	0.0055	2.47	0.017

per cent lower errors for PyAL + CNN (1σ) compared to PyAL + CNN. As we will see, this benefit from including 1σ uncertainties generally continues for the other test sets, and arises from replacing the uniform prior of PyAL (blind) with a Gaussian prior with a suitably estimated centre and width. This prior takes the place of human inspection through providing better starting points and constraints on PyAutoLens' parameter space, allowing it to converge on the correct model while avoiding local minima that it may otherwise become trapped in.

From Table 4.2, we see that the modelling methods experience little biasing of results, with no clear distinction between methods when comparing median biases for all parameters. For example, biases for Einstein radii and axis ratios do not exceed magnitudes of 0.1 per cent and 0.4 per cent, respectively. All biases are significantly smaller than the corresponding errors given in Table 4.1, and have negligible impact on modelling given their magnitudes. These results are to be expected given the simple parametric profiles used both in this test set and in the training of the network.

As previously discussed, PyAutoLens is typically initialised using human inspection of images, rather than the blind modelling performed here. While this would not be practical for application to future large-scale data sets, it is still worth considering how the errors of the automated methods presented here compare to PyAutoLens with human oversight, to gauge the impact of automating the modelling process. A comparison of this nature was not carried out for the data

Table 4.2: **SIE lenses + parametric sources.** The bias in the predicted parameters for each modelling method, computed from taking the median average across 1000 test images of the predicted parameter values minus their true values. Percentage biases are given next to the absolute biases for Einstein radius and axis ratio, with negative values indicating under-predictions compared to the true values.

Method	θ_E (arcsec)		ϕ (°)	q	
CNN	0.0016	(0.1%)	-0.18	-0.002	(-0.3%)
PyAL (blind)	0.0015	(0.1%)	-0.15	0.001	(0.2%)
PyAL + CNN	0.0013	(0.1%)	0.09	0.002	(0.2%)
PyAL + CNN (1σ)	0.0015	(0.1%)	-0.04	0.003	(0.4%)

sets in this thesis, but other works have provided examples of how well PyAutoLens can perform with human inspection. For example, Cao et al. (2022) tested PyAutoLens on 50 mock HST-resolution lenses and recovered power law model parameters to high accuracy: ±0.1 per cent for Einstein radius with negligible bias, and ±3.4 per cent for slope with 2.5 per cent systematic underprediction. However, Etherington et al. (2022) fitted power law models to 59 SLACS and BELLS GALaxy-Lyα EmitteR sYstems (GALLERY) lenses observed by HST utilising the now fully Bayesian PyAutoLens and from the posterior probability distribution function calculated a mean 1 per cent measurement precision on its modelled Einstein radii.

Additionally, Maresca et al. (2022) modelled uv-plane visibilities of Atacama Large Millimetre/sub-millimetre Array (ALMA) strong gravitational lenses, at an image resolution akin to that of Euclid. Across four galaxy-galaxy lenses modelled with an elliptical power law mass model with external shear, PyAutoLens produced uncertainties of ±0.01 arcsec, ±1-4°, ±0.01-0.05, and ±0.01-0.04 for Einstein radius, orientation, axis ratio, and power law slope, respectively. These uncertainties agree closely with the errors presented in Table 4.1 for SIE parameters, and we shall see in the next section that the same is true for modelling SIE lenses with HUDF sources. However, the rest of this chapter will show how introduction of LOS structure, EAGLE galaxy lenses or power law mass models raise errors of the automated methods significantly higher than may be produced by PyAutoLens with human inspection.

4.3.2 SIE Lenses + HUDF Sources

For the next set of testing, the impact of using real HUDF sources in place of simple parametric ones was examined, as well as the incorporation of additional structure along the LOS. Results for images without LOS structure are presented in Figures 4.4 and 4.6, and those with LOS structure in Figures 4.5 and 4.7. They follow the same format as before, and their associated errors and biases are given in Tables 4.3 and 4.4, respectively.

We first examine the results for images without LOS structure. In general, the accuracies for all modelling have dropped compared to testing on images with parametric sources, as expected due to the sources' complex non-parametric morphologies. As with the previous test set containing parametric sources, the CNN errors are higher for Einstein radius and orientation and lower for axis ratio when compared to PyAutoLens blind modelling. Likewise, we again see the combination of the two techniques performing significantly better than either of them separately, with PyAL + CNN (1σ) errors 31-39 per cent lower than PyAL (blind). In Figure 4.6, there are a notable number of axis ratio values that PyAL (blind) significantly underpredicts that the other methods do not. These scattered outliers generally correspond to the likewise underpredicted Einstein radii, and appear in later test results as well. They do not correlate to any specific lensing configurations, but are instead likely due to poor initialisations in parameter space, and as such do not appear for the CNN and CNN-assisted methods.

From Table 4.4, we see that the median biases for Einstein radius and axis ratio increase somewhat in magnitude compared to the previous test set: Einstein radius is over-predicted by up to 2 per cent while axis ratio is underpredicted by up to 1.4 per cent. However, these continue to remain notably smaller than the errors in

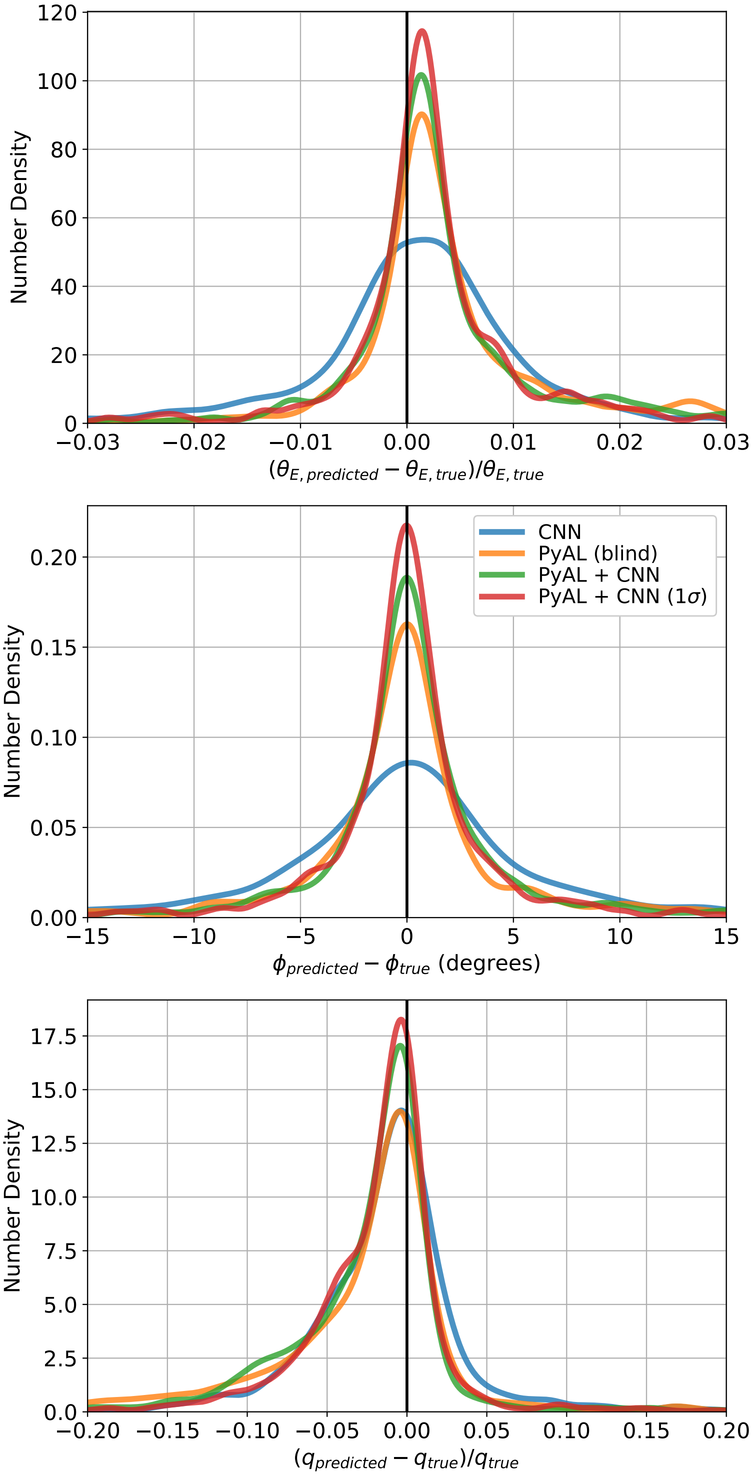

Figure 4.4: **SIE lenses + HUDF sources.** Distribution of the differences between predicted SIE mass model parameters and their true values for test data sets of 1000 images containing HUDF sources, without LOS structure. The distributions follow the same format as Figure 4.2.

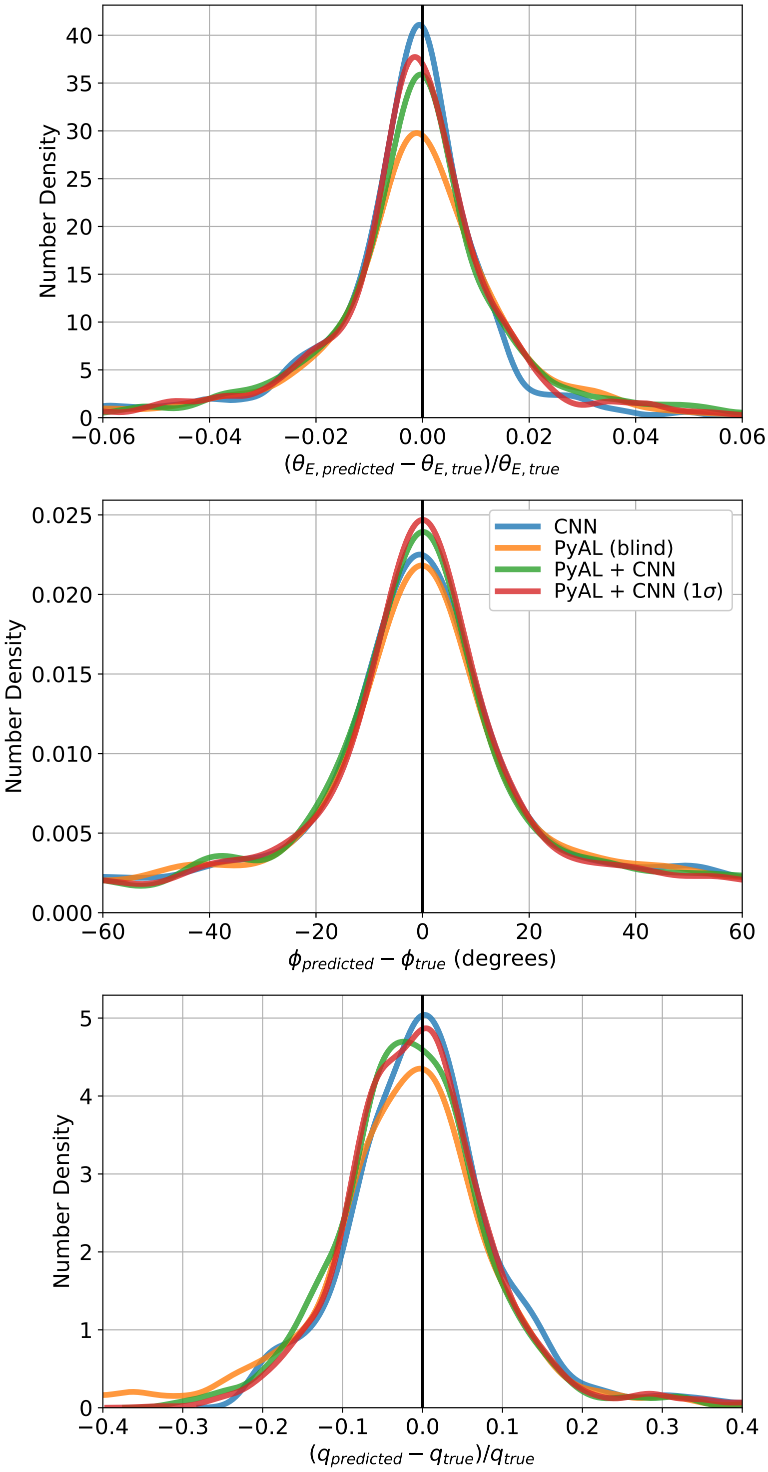

Figure 4.5: **SIE lenses + HUDF sources + LOS structure.** Distribution of the differences between predicted SIE mass model parameters and their true values for test data sets of 1000 images containing HUDF sources, with LOS structure. The distributions follow the same format as Figure 4.2.

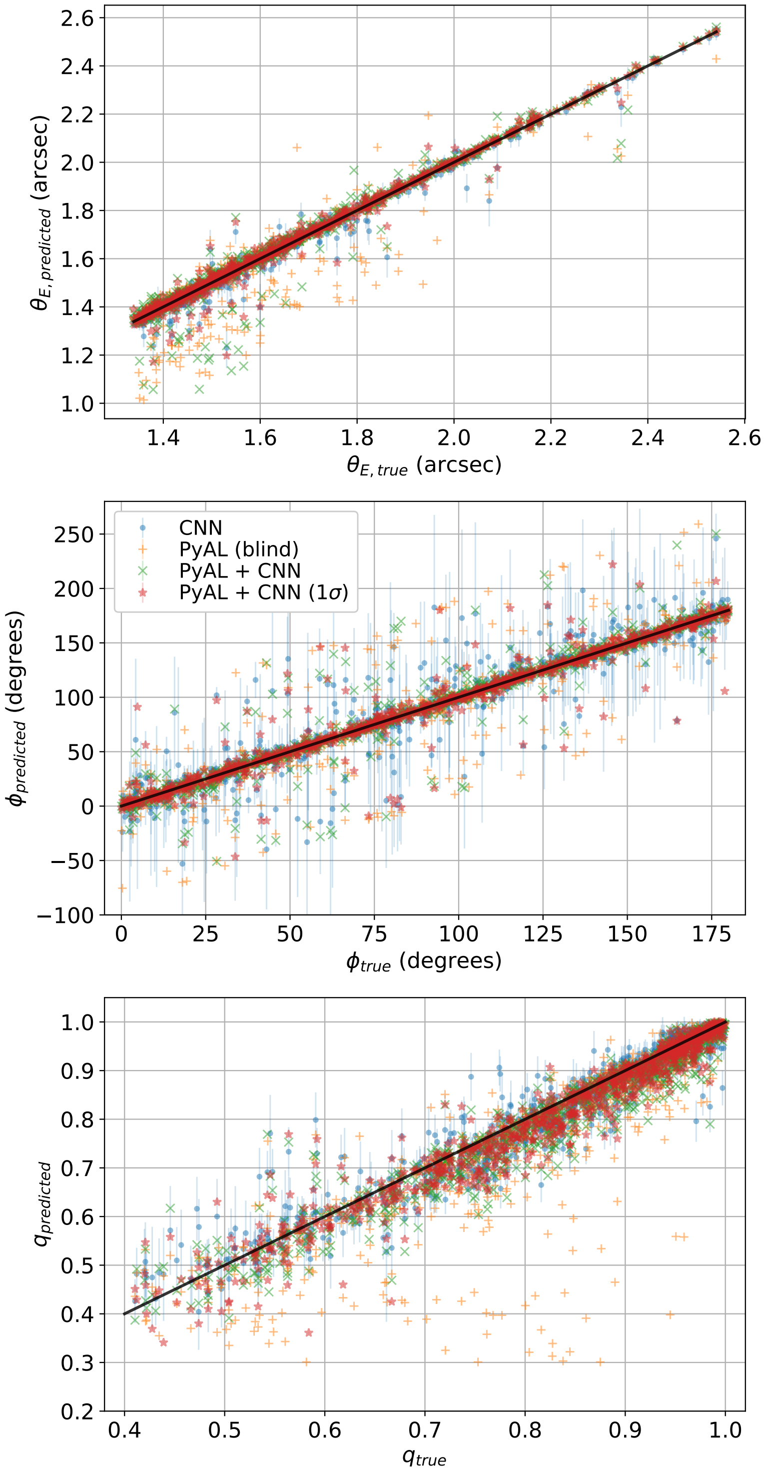

Figure 4.6: **SIE lenses + HUDF sources.** Comparison of predicted SIE lens parameters with the true values for test data sets of 1000 images containing SIE lenses and HUDF sources, without LOS structure. From top to bottom: Einstein radius, orientation and axis ratio of the lens mass profile.

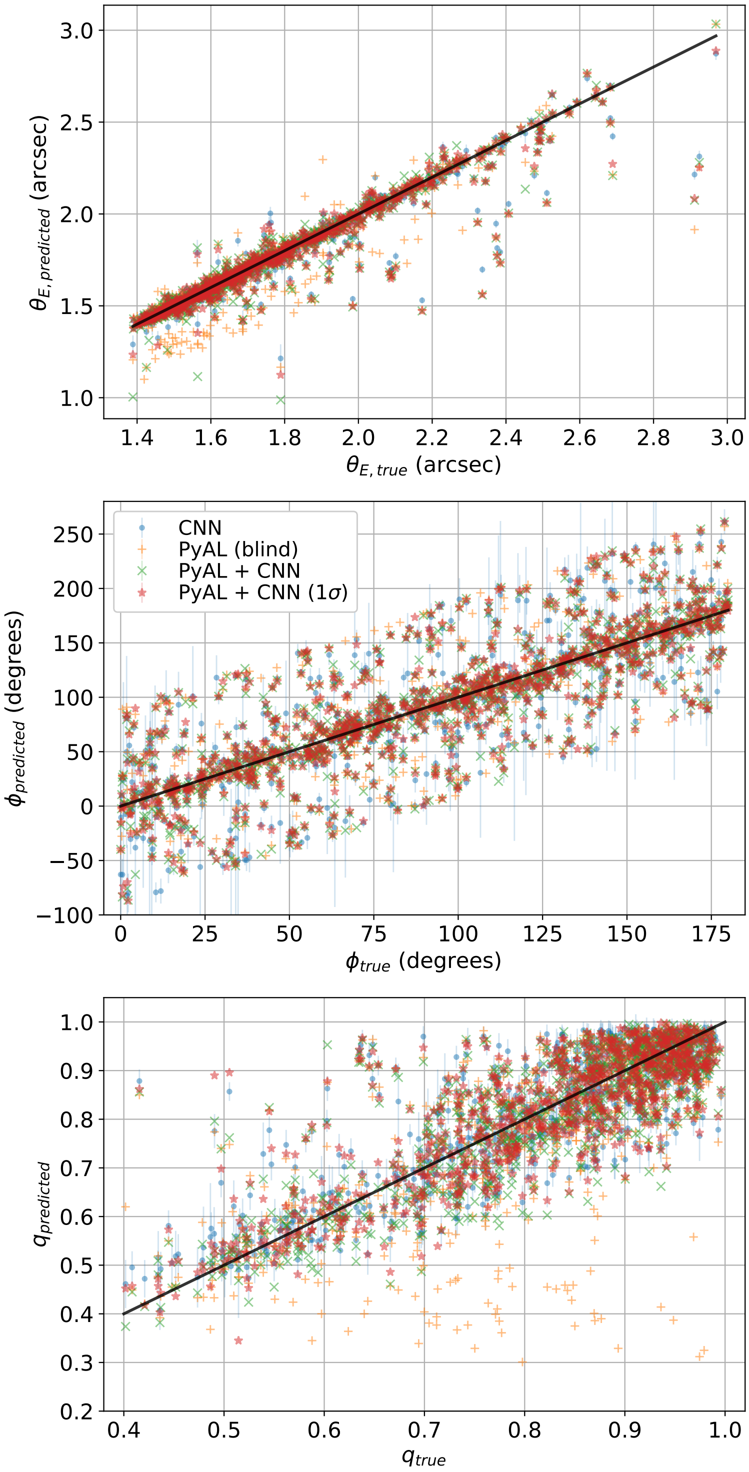

Figure 4.7: **SIE lenses + HUDF sources + LOS structure.** Comparison of predicted SIE lens parameters with the true values for test data sets of 1000 images containing SIE lenses and HUDF sources, with LOS structure. From top to bottom: Einstein radius, orientation and axis ratio of the lens mass profile.

Table 4.3: **SIE lenses + HUDF sources.** The 68 per cent confidence intervals on predicted parameters for each modelling method, computed from the distributions of differences between true and predicted parameter values across 1000 test images.

Method	θ_E (arcsec)	ϕ (°)	q
Without LOS structure			
CNN	0.013	6.10	0.029
PyAL (blind)	0.013	4.22	0.038
PyAL + CNN	0.010	3.02	0.030
PyAL + CNN (1σ)	0.009	2.57	0.026
With LOS structure			
CNN	0.021	30.7	0.067
PyAL (blind)	0.029	33.8	0.077
PyAL + CNN	0.026	28.0	0.068
PyAL + CNN (1σ)	0.023	27.1	0.065

Table 4.3, and there is little change in orientation biases, as expected. Regarding these biases, PyAL (blind) and PyAL + CNN perform very similar, while the CNN achieves the lowest for Einstein radii and axis ratio.

For images with LOS structure, results are similar to those above except for a general decrease in performance for all results, with errors increasing by factors of 2.2, 8.2 and 2.3 on average for Einstein radius, orientation and axis ratio, respectively. This is to be expected as the inclusion of extra mass along the LOS serves generally to increase the complexity of the lensing system compared to smooth parametric profiles, making the resulting images more difficult to model. As detailed in Section 3.6, to obtain the 'true' parameters for these images we fitted an SIE model convergence to the convergence of the combined SIE profile and LOS structure. It is worth noting however that the CNN was not trained on images containing HUDF sources or LOS structure, so while the drop in its accuracy is expected it still performs sufficiently well as an automated modelling method.

The CNN now achieves the highest accuracy for Einstein radii, with errors 10 per cent lower than PyAL + CNN (1σ), but the latter method continues to give the best results for orientation and axis ratio. For these parameters, the differences

Table 4.4: **SIE lenses + HUDF sources.** The bias in the predicted parameters for each modelling method, computed from taking the median average across 1000 test images of the predicted parameter values minus their true values. Percentage biases are given next to the absolute biases for Einstein radius and axis ratio, with negative values indicating under-predictions compared to the true values.

Method	θ_E (arcsec)		ϕ (°)	q	
Without LOS structure					
CNN	0.002	(0.1%)	0.10	-0.008	(-0.9%)
PyAL (blind)	0.003	(0.2%)	-0.07	-0.011	(-1.4%)
PyAL + CNN	0.003	(0.2%)	0.09	-0.011	(-1.3%)
PyAL + CNN (1σ)	0.003	(0.2%)	0.04	-0.009	(-1.0%)
With LOS structure					
CNN	-0.003	(-0.2%)	-0.50	-0.001	(-0.1%)
PyAL (blind)	-0.004	(-0.2%)	0.43	-0.013	(-1.6%)
PyAL + CNN	-0.002	(-0.1%)	-0.14	-0.010	(-1.3%)
PyAL + CNN (1σ)	-0.003	(-0.2%)	0.20	-0.003	(-0.5%)

between the distributions are reduced compared to the results for images without LOS structure. As such, compared to PyAL (blind) the errors are reduced by 9-28 per cent, 11-17 per cent, and 16-21 per cent for the CNN, PyAL + CNN, and PyAL + CNN (1σ), respectively. For the latter method, incorporating the CNN's predicted 1σ uncertainties in the priors of PyAutoLens now reduces errors by 3-10 per cent over PyAL + CNN, compared with 13-15 per cent for images without LOS structure. For both images with and without LOS structure, CNN-predicted uncertainties seen in Figures 4.6 and 4.7 continue to accurately reflect the CNN's errors for the vast majority of lenses, but for those containing LOS structure there are significantly more with greatly underpredicted uncertainties, due to not being represented in the CNN's training set.

With the addition of LOS structure, biases in orientation increase in magnitude for all modelling methods, reflecting the increased errors shown in Table 4.3. In addition, Einstein radii are now underpredicted, but their magnitudes, like those for axis ratio, do not appear to be affected. PyAL (blind) now experiences the largest biases for Einstein radius and axis ratio, respectively reaching magnitudes of 0.2 per cent and 1.6 per cent, but once again the biases of all methods remain significantly smaller than the errors in Table 4.3. Axis ratios are repeatedly underpredicted on average, due to the distortions to the lensed arcs brought about by the HUDF sources and LOS structure causing the lenses to appear more elliptical (lower axis ratios) than they actually are. Additionally, in part due to overpredicting the ellipticity, the modelling also generally predicts the aforementioned smaller Einstein radii to compensate for the observed distorted lensed images, with more elliptical lenses experiencing greater underpredictions of Einstein radii.

4.3.3 Power Law Lenses + HUDF Sources

Next, the same tests as Section 4.3.2 were repeated for power law lenses. For this, the same CNN architecture and training process was used but now training and testing on images containing lenses with power law mass profiles instead of SIE profiles. As such, the power law slope parameter n (see Equation 3.7) was also predicted by the network along with its associated 1σ uncertainty, but otherwise the training and testing was identical to that of Section 4.3.2. Figures 4.8 and 4.10 present results for images without LOS structure, while Figures 4.9 and 4.11 present those containing LOS structure, following the same format as before. Likewise, their associated errors and biases are respectively given in Tables 4.5 and 4.6.

For power law mass profiles without LOS structure, Einstein radius continues to be overpredicted while axis ratio is underpredicted, yet orientation appears unaffected in both the magnitude and direction of its biasing compared to those of SIE modelling. Biases for both Einstein radius and axis ratio increase compared to those of SIE modelling, now respectively reaching magnitudes of 0.8 per cent and 4.1 per cent. This is due to the difficulty in modelling power law slopes, discussed below, with all modelling methods significantly overpredicting values on average by up to 6.9 per cent.

The difficulty of predicting slope values is highlighted in Figures 4.10 and 4.11,

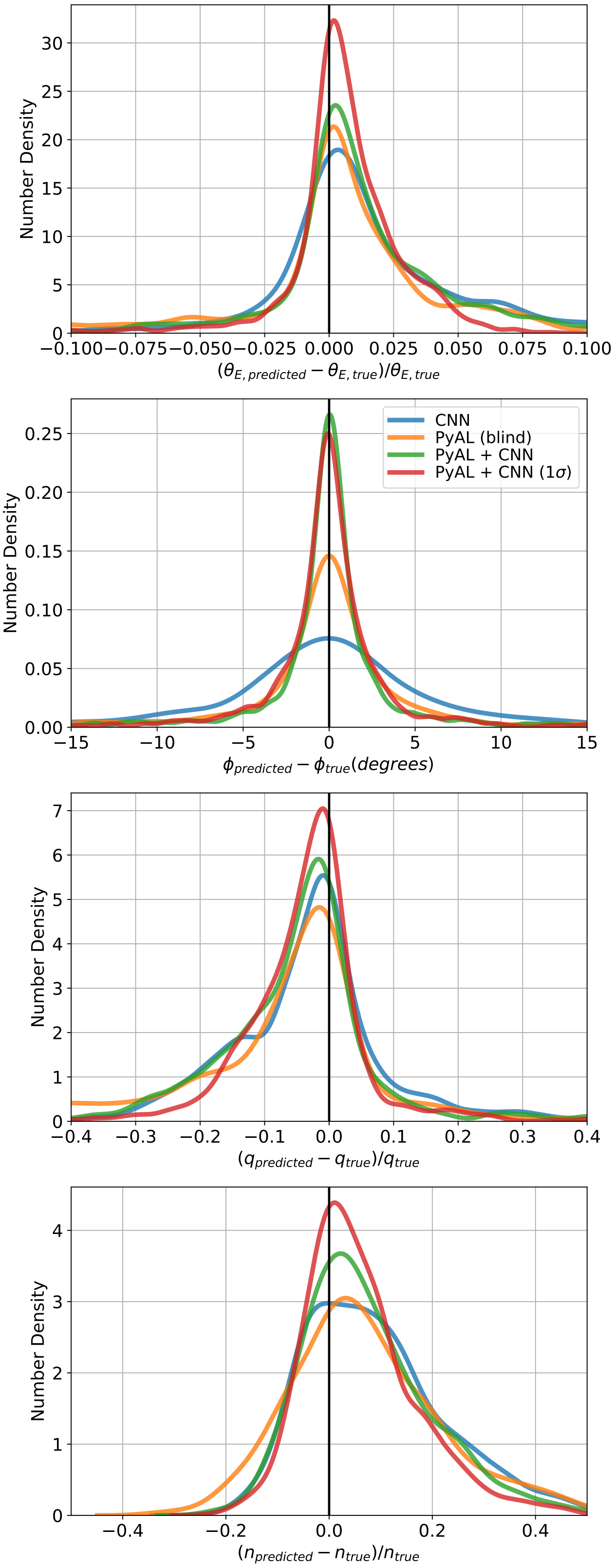

Figure 4.8: **Power law lenses + HUDF sources.** Distribution of the differences between predicted power law mass model parameters and their true values for test data sets of 1000 images containing HUDF sources, without LOS structure. The distributions follow the same format as Figure 4.2.

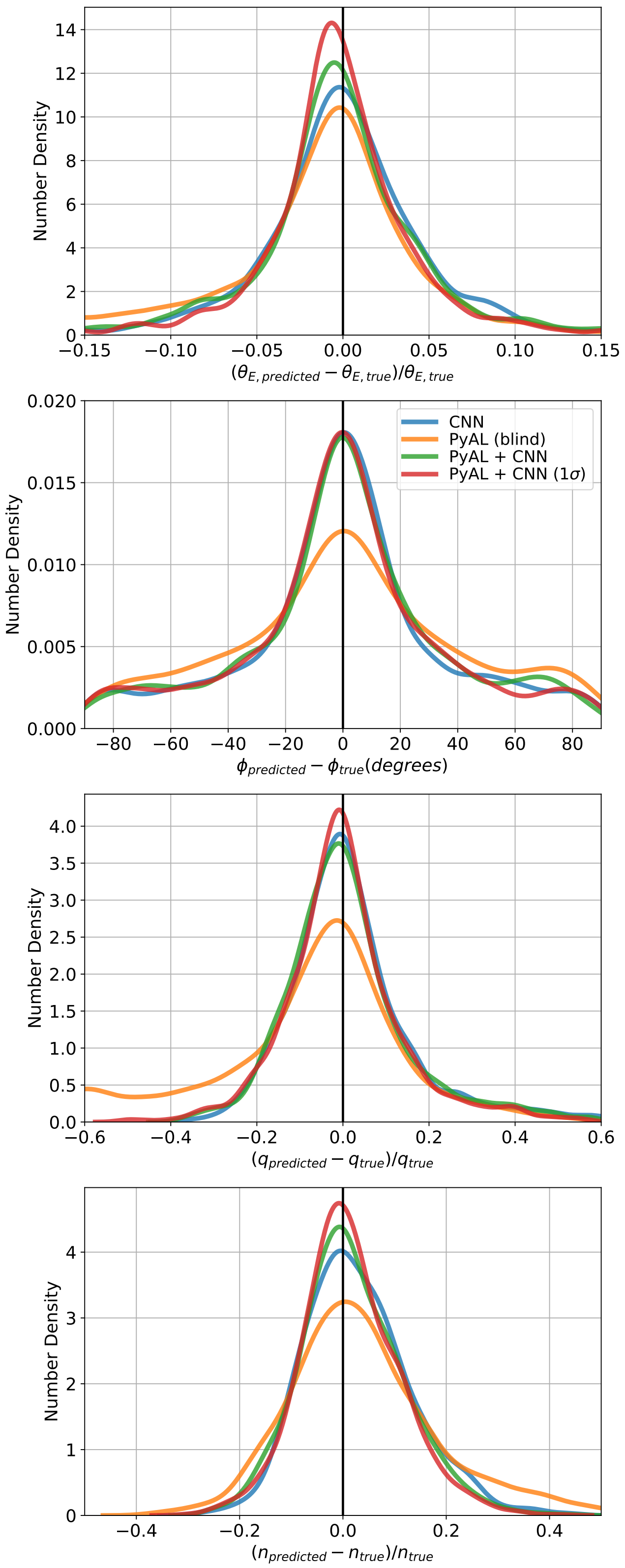

Figure 4.9: **Power law lenses + HUDF sources + LOS structure.** Distribution of the differences between predicted power law mass model parameters and their true values for test data sets of 1000 images containing HUDF sources, with LOS structure. The distributions follow the same format as Figure 4.2.

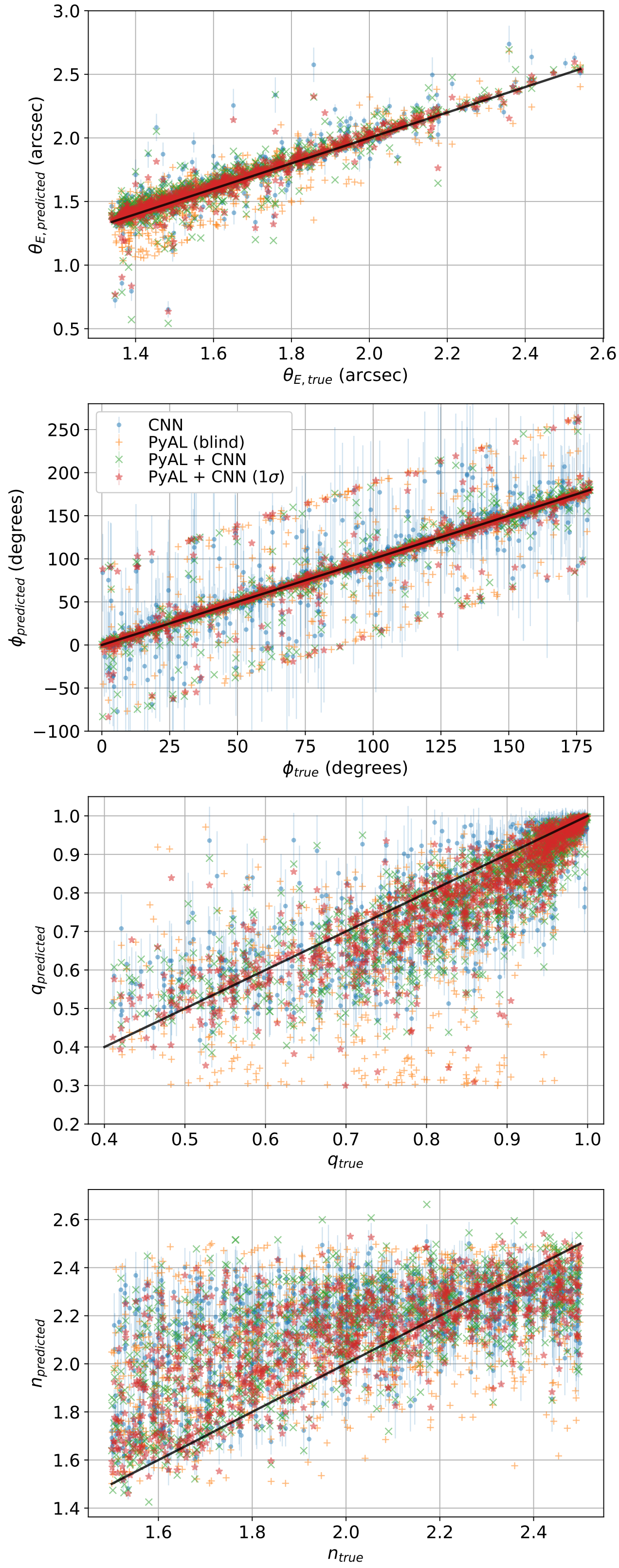

Figure 4.10: **Power law lenses + HUDF sources.** Comparison of predicted power law lens parameters with the true values for test data sets of 1000 images containing power law lenses and HUDF sources, without LOS structure. From top to bottom: Einstein radius, orientation and axis ratio of the lens mass profile.

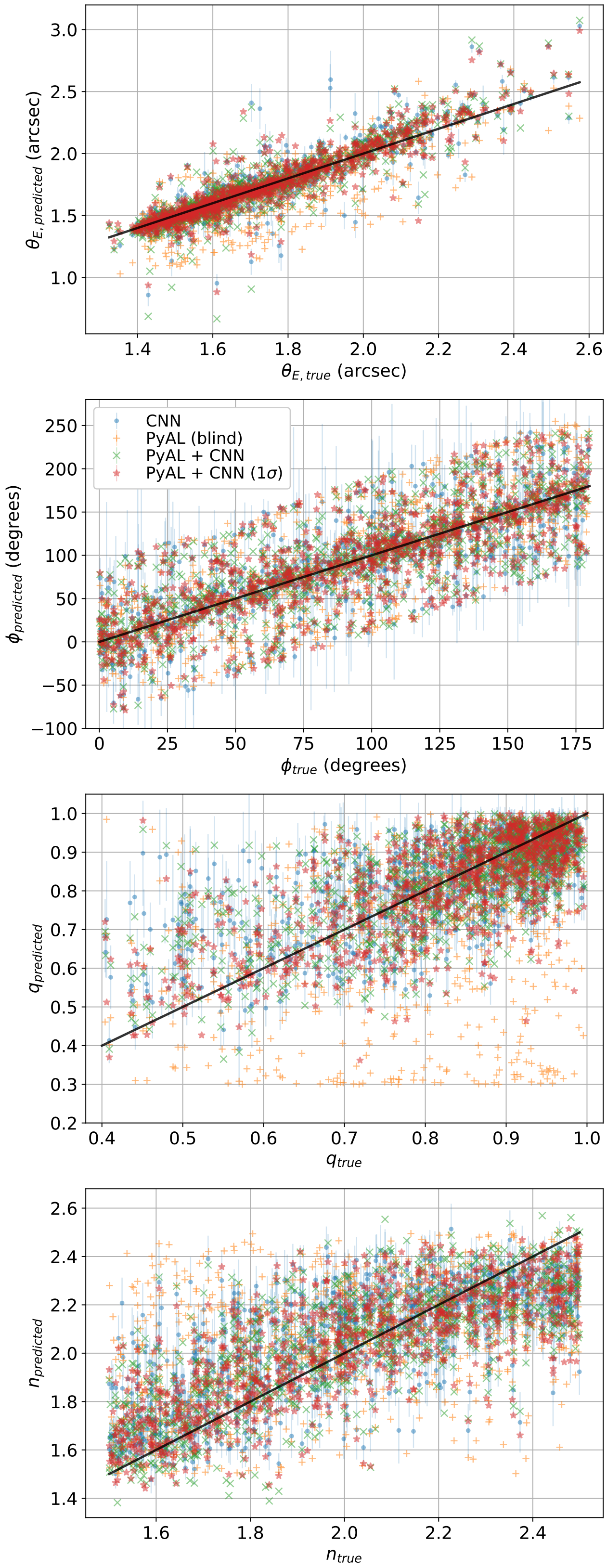

Figure 4.11: **Power law lenses + HUDF sources + LOS structure.** Comparison of predicted power law lens parameters with the true values for test data sets of 1000 images containing power law lenses and HUDF sources, with LOS structure. From top to bottom: Einstein radius, orientation and axis ratio of the lens mass profile.

with the highest accuracy achieved using PyAL + CNN (1σ) despite significant scatter remaining in both plots of slope values. Because of this complication, the results for other parameters are generally worse than for images containing SIE lenses, with errors increasing by factors of 1.2-2.9 and 1.1-4.2 for test sets with and without LOS structure, respectively. Einstein radii and axis ratios are predominantly affected, especially when modelling with PyAL (blind) whose results contain many more greatly underpredicted axis ratios regardless of the presence of LOS structure. This is a reflection of the overall degeneracy observed between slope and axis ratio (which in turn affects Einstein radius predictions); for increasing true slope values, predicted slopes in general go from overpredicted to underpredicted, while the opposite occurs for axis ratio predictions. These trends are reversed for increasing true axis ratios (rounder lenses). As such, the degeneracy is due to how deflection potentials seen in Figure 1.6 produced from increasing a lens' ellipticity (moving downwards in the figure) can appear similar to those produced by a lens with both lower ellipticity and power law slope (instead moving left or left and up in the figure), even if the surface mass density appears significantly different. Hence, both cases produce more elongated critical curves and caustics as shown in Figure 1.5, even if they differ in shape.

While the scatter in power law results is larger than that of SIE lenses, the uncertainties predicted by the CNN remain in general an accurate reflection of its errors for both lenses with and without LOS structure. These uncertainties appear to be slightly underpredicted for axis ratios, primarily for the large scatter seen at lower axis ratio values, and generally underpredicted for slope values. These underpredictions are due to the above difficulty in modelling slope leading to insufficient CNN training for more elliptical lenses, and would require further training in this area to cope with the wider variety of lensed image configurations caused by varying the power law slope.

For the test set without LOS structure, comparisons between modelling methods

Table 4.5: **Power law lenses + HUDF sources.** The 68 per cent confidence intervals on predicted parameters for each modelling method, computed from the distributions of differences between true and predicted parameter values across 1000 test images. The final column contains the confidence intervals for the power law slope parameter.

Method	θ_E (arcsec)	ϕ (°)	q	n
Without LOS structure				
CNN	0.050	7.37	0.083	0.27
PyAL (blind)	0.065	6.53	0.106	0.29
PyAL + CNN	0.049	2.63	0.080	0.25
PyAL + CNN (1σ)	0.029	2.50	0.062	0.20
With LOS structure				
CNN	0.070	32.2	0.093	0.19
PyAL (blind)	0.091	44.1	0.143	0.26
PyAL + CNN	0.068	34.1	0.096	0.19
PyAL + CNN (1σ)	0.060	32.1	0.091	0.18

remain similar to before, with errors of PyAL + CNN (1σ) 5-39 per cent lower than those of PyAL + CNN, and 32-62 per cent lower than those of PyAL (blind). Meanwhile, the CNN by itself achieves errors comparable to PyAL (blind), from 19 per cent lower to 13 per cent larger, due to the increasingly complex parameter space making it difficult for PyAL (blind) to avoid local minima. With regard to the orientation results seen in Figure 4.8, the CNN predictions are generally more scattered compared to PyAL (blind). However, while the results of the latter are for the most part centrally concentrated close to the true values, some are clearly offset by around ±90° due to the modelling falling into local minima in parameter space where perpendicular lenses can produce similar, if offset, lensed image configurations. For PyAL + CNN and PyAL + CNN (1σ) some of the outlying predictions remain, but these combinations of methods shift the majority of them back towards the true values, decreasing overall errors.

With LOS structure, the CNN and PyAL + CNN (1σ) methods respectively reduce errors by 23-35 per cent and 27-36 per cent compared to PyAL (blind). Similar to the test set containing SIE lenses and LOS structure, the difference between the CNN and PyAL + CNN (1σ) methods is smaller than for images without LOS structure, with the latter method obtaining only 0.3-14 per cent

Table 4.6: **Power law lenses + HUDF sources.** The bias in the predicted parameters for each modelling method, computed from taking the median average across 1000 test images of the predicted parameter values minus their true values. Percentage biases are given next to the absolute biases for Einstein radius, axis ratio and power law slope, with negative values indicating under-predictions compared to the true values.

Method	θ_E (arcsec)		ϕ (°)	q		n	
Without LOS structure							
CNN	0.012	(0.8%)	-0.04	-0.020	(-2.2%)	0.13	(6.9%)
PyAL (blind)	0.004	(0.3%)	0.02	-0.036	(-4.1%)	0.10	(5.1%)
PyAL + CNN	0.011	(0.7%)	0.00	-0.030	(-3.5%)	0.10	(5.1%)
PyAL + CNN (1σ)	0.008	(0.5%)	-0.05	-0.025	(-2.8%)	0.08	(4.2%)
With LOS structure							
CNN	-0.001	(-0.0%)	0.64	0.000	(0.0%)	0.03	(1.8%)
PyAL (blind)	-0.014	(-0.8%)	1.02	-0.034	(-4.0%)	0.04	(1.9%)
PyAL + CNN	-0.005	(-0.3%)	0.44	-0.005	(-0.6%)	0.01	(0.8%)
PyAL + CNN (1σ)	-0.005	(-0.3%)	0.04	-0.006	(-0.7%)	0.01	(0.7%)

lower errors over the former compared to the 28-66 per cent improvement obtained for images without LOS structure. Likewise, the incorporation of 1σ uncertainties now only reduces errors by 5-12 per cent over modelling with PyAL + CNN. The differences in results for images with LOS structure compared to those without are less prominent than those for SIE lenses due to larger errors overall: now, errors increase by factors of 1.6, 9.2, and 1.3 for Einstein radius, orientation and axis ratio, respectively, while errors in the slope actually decrease by 18 per cent. The modelling methods appear to generally overpredict slope values (i.e. more point-like lenses) in Figure 4.10, so the addition of LOS structure away from the central lens may allow the modelling to predict sufficiently lower slope values so as to slightly decrease the overall error in these results. The overpredicting of slope values is unlikely to have arisen from differences in their training and testing distributions as both were uniform and sampled between the same bounds.

Similar to the SIE results, the addition of LOS structure to the power law data set causes biases in Einstein radius to become underpredicted on average and increase the magnitude of orientation biases. However, the magnitudes of biases for Einstein radius, axis ratio, and power law slope are all reduced on average across the modelling methods, with more accurate predictions of the slope discussed above allowing for better modelling overall. For this data set, PyAL (blind) produces the largest median biases, likely due to the increased difficulty of parameter space causing it to converge on local minima.

4.3.4 EAGLE Lenses + HUDF Sources

The next stage involved testing on simulated images that contained more complex lenses, i.e. those that do not follow an analytical profile. A set of 1000 test images was simulated containing lensing galaxies taken from the EAGLE hydrodynamical simulations along with HUDF source galaxies (see Section 3.6). An identical test set was also simulated with the addition of LOS structure as applied previously for the parametric lenses. The known convergence maps of these lenses were fitted with SIE mass profiles in order to obtain the best-fit ‘true’ parameter values, before modelling the lensed images with an SIE mass profile using the previous methods. Figures 4.12 and 4.13 show the resulting accuracies of the different modelling methods, with corresponding scatter plots in Figures 4.14 and 4.15. Errors and biases are respectively presented in Tables 4.7 and 4.8, in the same format as previous results.

It should be noted that the EAGLE galaxy lenses modelled in this section have smaller Einstein radii on average than in previous sections (see Section 3.6), producing more pixelated arcs. The modelling methods are therefore expected to produce higher errors for all parameters as a result, on top of the difficulty in modelling more complex mass profiles; for example; for example, CNN error as a function of Einstein radius was shown in Figure 2.10. However, examining the results for lenses with Einstein radii above 1.3 arcsec, in line with previous test sets, such lenses continue to exhibit much larger errors than those seen in Section 4.3.2, more similar in magnitude to the rest of the EAGLE galaxy data set. As such, it appears the dominant source of error in these results arises from the complex

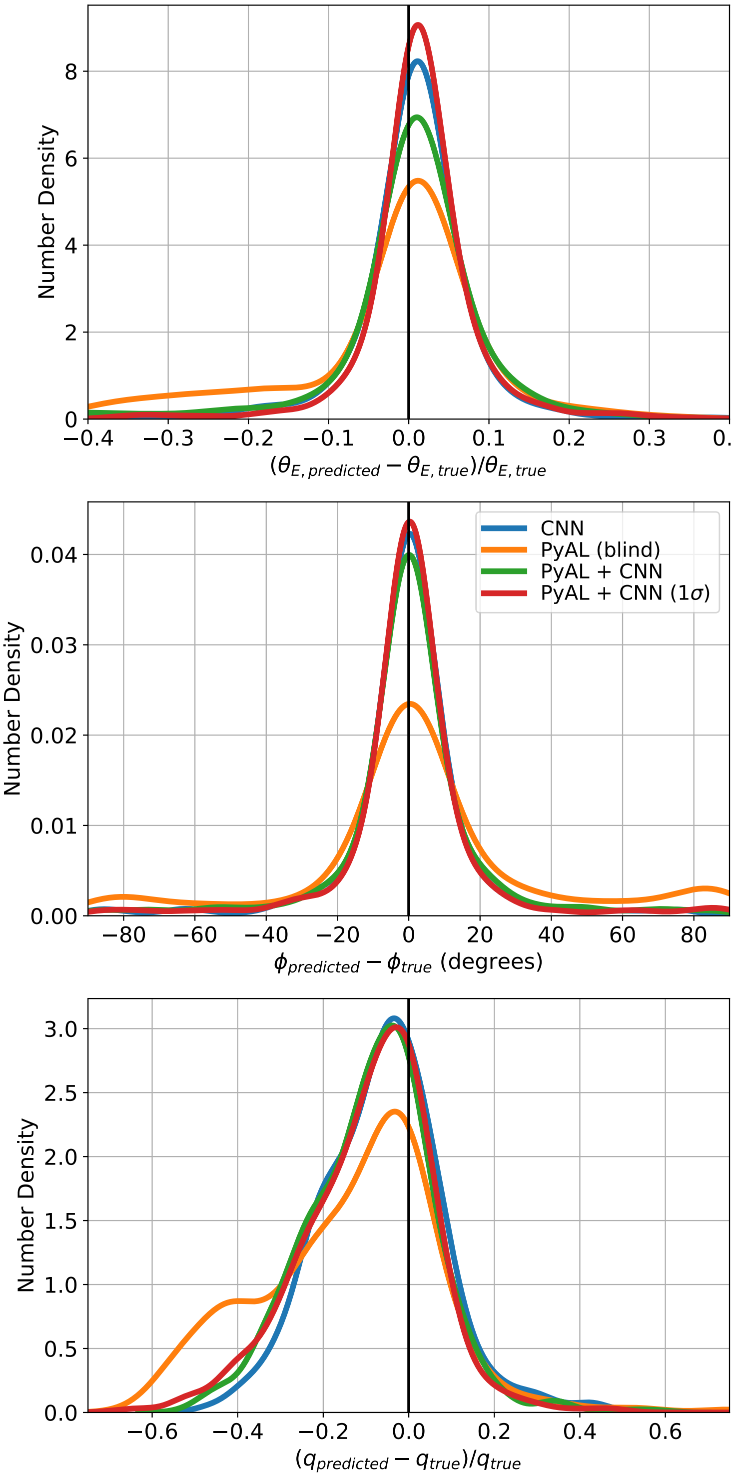

Figure 4.12: **EAGLE galaxy lenses + HUDF sources.** Distribution of the differences between predicted SIE mass model parameters and their true values for test data sets of 1000 images containing EAGLE galaxy lenses with HUDF sources, without LOS structure. The distributions follow the same format as Figure 4.2.

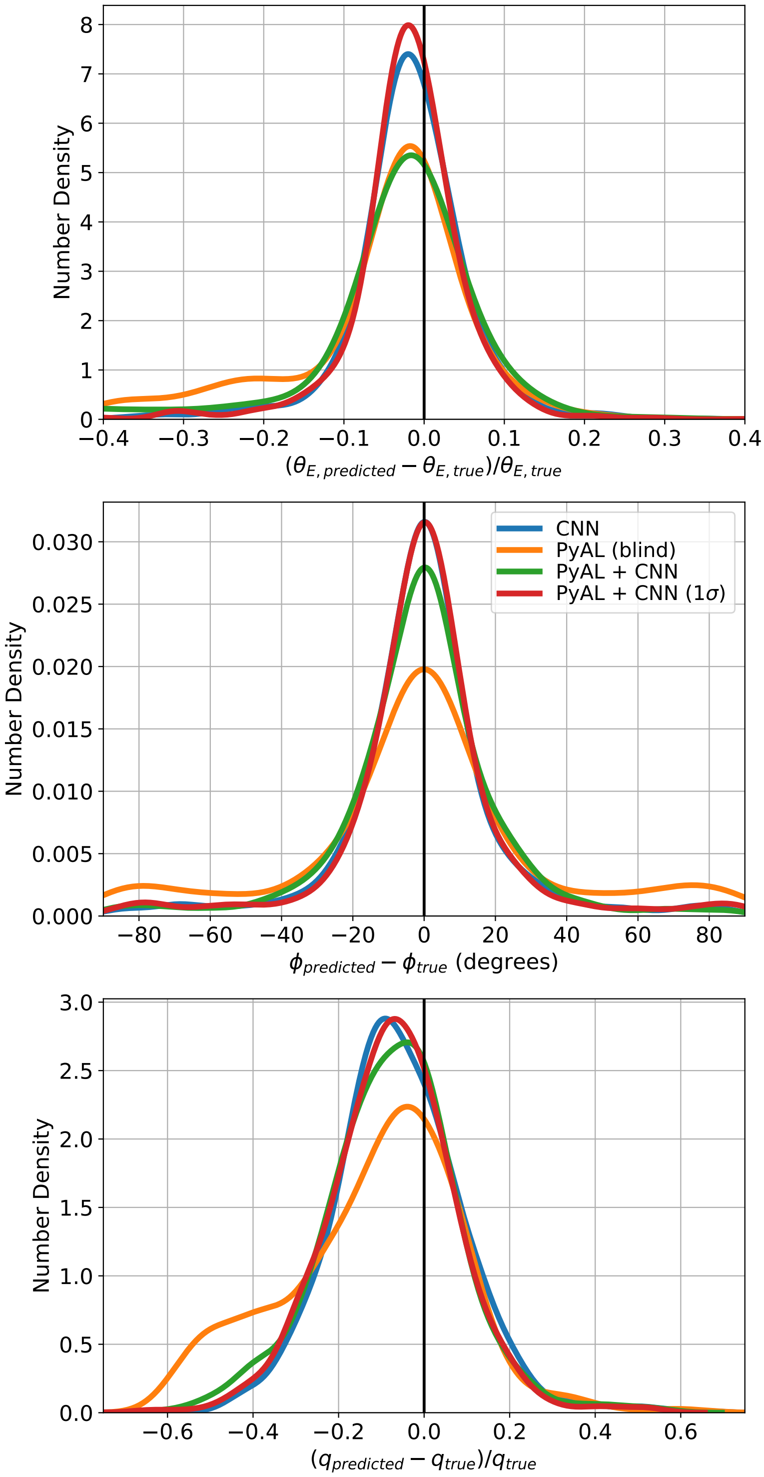

Figure 4.13: **EAGLE galaxy lenses + HUDF sources + LOS structure.** Distribution of the differences between predicted SIE mass model parameters and their true values for test data sets of 1000 images containing EAGLE galaxy lenses with HUDF sources, with LOS structure. The distributions follow the same format as Figure 4.2.

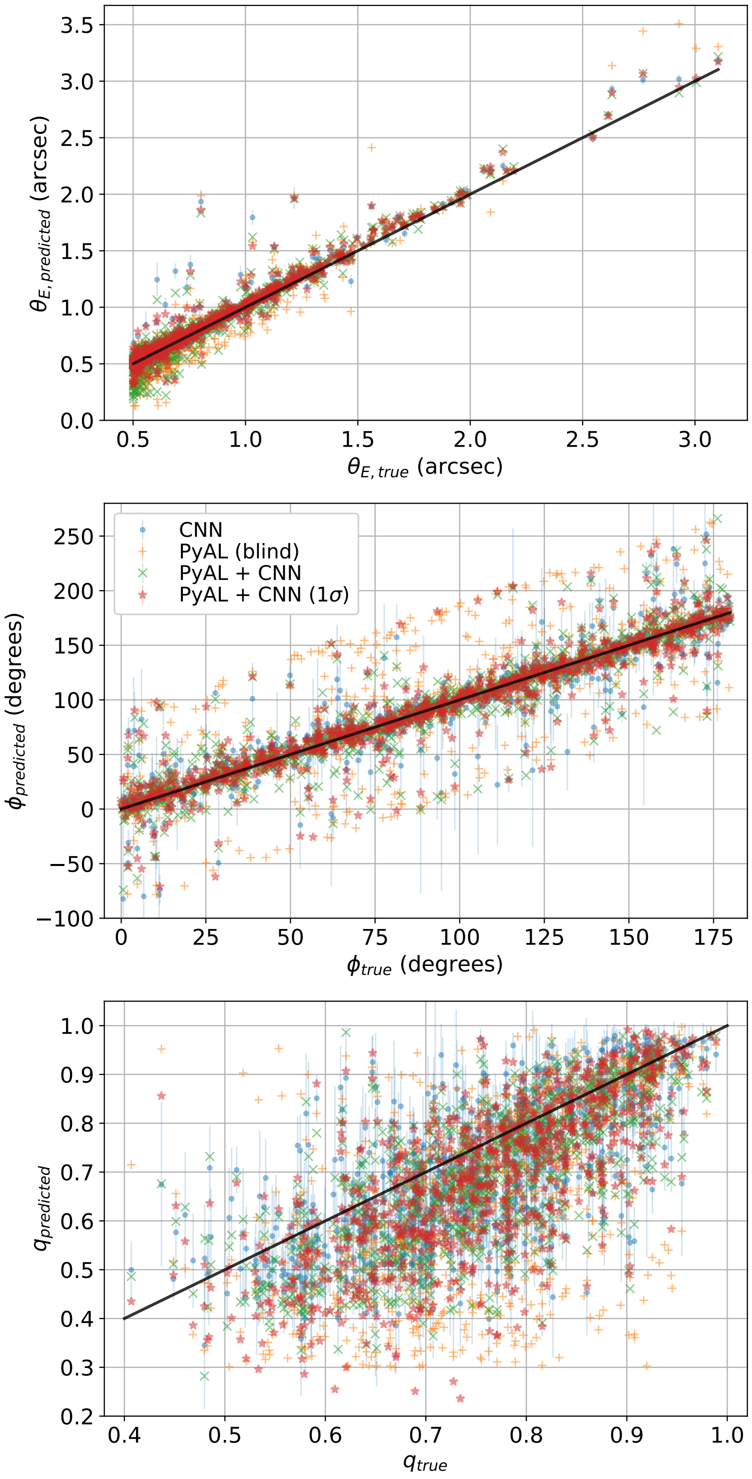

Figure 4.14: **EAGLE galaxy lenses + HUDF sources.** Comparison of predicted SIE lens parameters with the true values for test data sets of 1000 images containing EAGLE galaxy lenses and HUDF sources, without LOS structure. From top to bottom: Einstein radius, orientation and axis ratio of the lens mass profile.

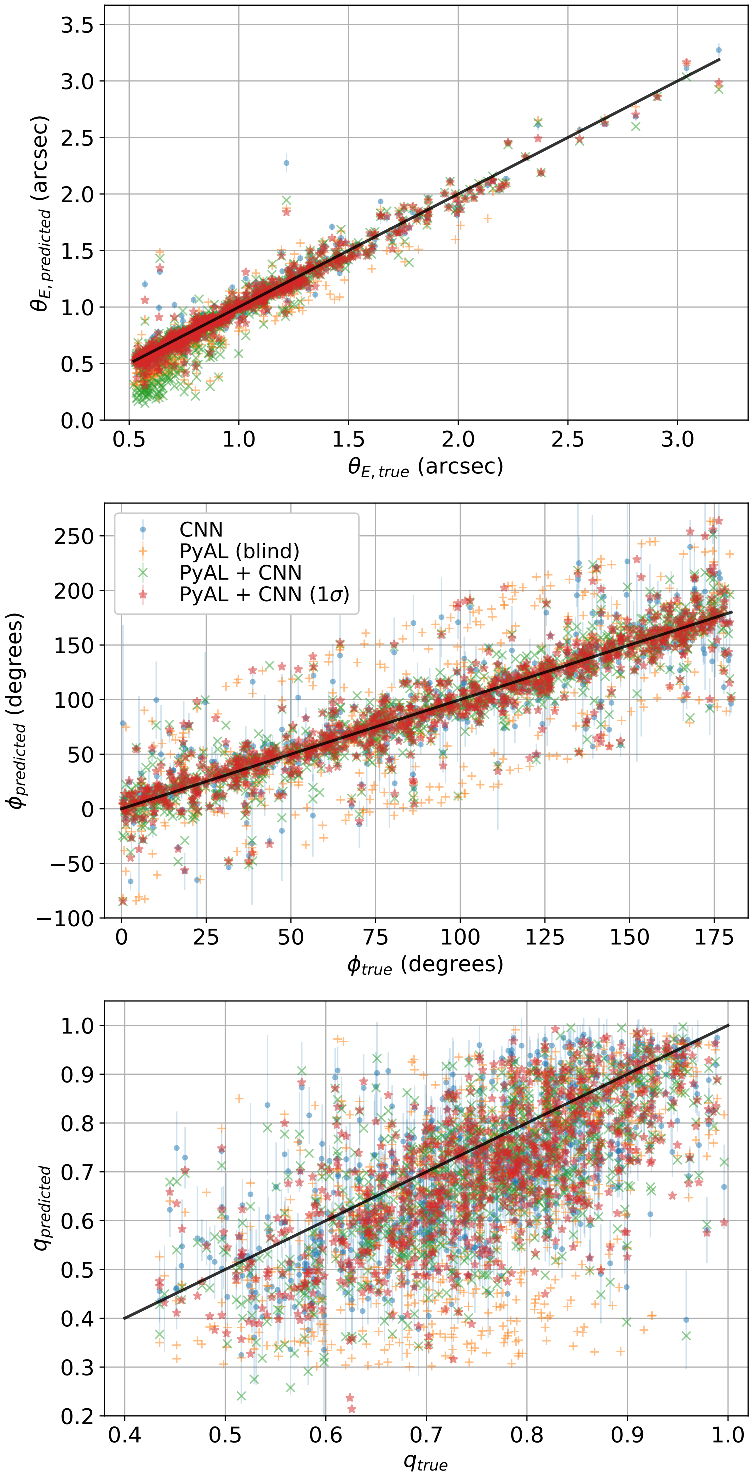

Figure 4.15: **EAGLE galaxy lenses + HUDF sources + LOS structure.** Comparison of predicted SIE lens parameters with the true values for test data sets of 1000 images containing EAGLE galaxy lenses and HUDF sources, with LOS structure. From top to bottom: Einstein radius, orientation and axis ratio of the lens mass profile.

Table 4.7: **EAGLE galaxy lenses + HUDF sources.** The 68 per cent confidence intervals on predicted parameters for each modelling method, computed from the distributions of differences between true and predicted parameter values across 1000 test images.

Method	θ_E (arcsec)	ϕ (°)	q
Without LOS structure			
CNN	0.035	8.70	0.108
PyAL (blind)	0.064	19.56	0.170
PyAL + CNN	0.044	9.12	0.116
PyAL + CNN (1σ)	0.034	8.12	0.119
With LOS structure			
CNN	0.046	12.6	0.114
PyAL (blind)	0.070	23.8	0.157
PyAL + CNN	0.060	15.4	0.121
PyAL + CNN (1σ)	0.043	12.4	0.114

foreground galaxies rather than the smaller Einstein radii.

Without LOS structure, the CNN achieves 36-55 per cent lower errors than PyAL (blind) despite the complex lenses in this test set, due to the complexity of parameter space causing the randomised automated initialisations of PyAL (blind) to frequently result in it converging on local minima. Providing more accurate priors enables PyAutoLens to converge on the correct parameters, and hence the PyAL + CNN (1σ) method gives similar results to the CNN, with 30-58 per cent lower errors than PyAL (blind). However, this method no longer outperforms the other CNN methods for all parameters: For Einstein radius and orientation, PyAL + CNN (1σ) achieves 11-22 per cent and 2-7 per cent lower errors compared to PyAL + CNN and the CNN, respectively, but for axis ratio the PyAL + CNN and CNN methods instead achieve 2.7 and 9.6 per cent lower errors than PyAL + CNN (1σ). However, this may simply be due to the large scatter observed in Figure 4.14, especially for axis ratio which is often underpredicted by the modelling methods. This is likely caused by two factors: a) the complex foreground galaxies produce fewer smooth, well-defined arcs than those trained on by the CNN, and b) the axis ratios of these galaxies can change as a function of radius, with the modelling methods only given final lensed images and hence probe only a part

Table 4.8: **EAGLE galaxy lenses + HUDF sources.** The bias in the predicted parameters for each modelling method, computed from taking the median average across 1000 test images of the predicted parameter values minus their true values. Percentage biases are given next to the absolute biases for Einstein radius and axis ratio, with negative values indicating under-predictions compared to the true values.

Method	θ_E (arcsec)		ϕ (°)	q	
Without LOS structure					
CNN	0.008	(1.1%)	0.13	-0.042	(-5.6%)
PyAL (blind)	0.004	(0.5%)	0.89	-0.077	(-10.1%)
PyAL + CNN	0.006	(0.8%)	0.28	-0.057	(-7.2%)
PyAL + CNN (1σ)	0.009	(1.2%)	0.15	-0.056	(-7.4%)
With LOS structure					
CNN	-0.014	(-1.8%)	-0.29	-0.050	(-6.9%)
PyAL (blind)	-0.022	(-2.6%)	0.29	-0.065	(-8.9%)
PyAL + CNN	-0.019	(-2.3%)	-0.02	-0.057	(-7.3%)
PyAL + CNN (1σ)	-0.015	(-1.8%)	-0.03	-0.054	(-7.4%)

of the whole convergence map used by the separate fitting method to obtain the 'true' values. Regardless of the presence of LOS structure, both Figures 4.14 and 4.15 show that for these complex EAGLE galaxy lenses the CNN underpredicts its uncertainties for all three parameters, with uncertainties barely visible for Einstein radii and clearly not representing the full scatter of the results. As such, were the network to be trained on similarly complex images, perhaps requiring a deeper architecture to do so effectively, not only would the CNN's predicted parameters become more accurate, but it would produce uncertainties that better represent the true uncertainty and hence allow PyAL + CNN and PyAL + CNN (1σ) to also achieve higher accuracies.

From Table 4.8, we see that replacing parametric mass profiles with those of EAGLE galaxies (without LOS structure) increases the relative biases of Einstein radii, now up to 1.2 per cent, increases magnitudes of orientation biases to values similar to those of images containing LOS structure, and significantly increases the axis ratio biases compared to those of SIE lenses to magnitudes reaching 10.1 per cent. These biases would have a major impact on such things as the determination of the Hubble constant, and so future work should be done to ensure they are reduced. It is possible that the biases here arise in part due to the fitting procedure used to obtain the true values, the difficulty of which is discussed in Section 5.5.2. For this data set, PyAL (blind) produces the largest bias for axis ratio but the smallest for Einstein radius, while the reverse is true for the CNN. When LOS structure is added to the images, as before the Einstein radii become under-predicted with magnitudes increasing up to 2.6 per cent. Meanwhile, biases for orientation and axis ratio remain relatively unchanged across the modelling methods, with biases larger than those for SIE or power law lenses with LOS structure as expected. The CNN now produces the smallest biases for both Einstein radius and axis ratio, and PyAL (blind) produces the largest biases.

Finally, for the test set containing LOS structure, PyAL + CNN (1σ) achieves the lowest errors for all parameters, but there is again little difference between its and the CNN's results, with PyAL + CNN (1σ) errors only 0.4-7 per cent lower than the CNN. As such their improvements over PyAL (blind) are almost identical, with PyAL + CNN (1σ) and CNN errors 28-48 per cent and 28-47 lower than PyAL (blind), respectively. There is still a benefit to incorporating the CNN's predicted 1σ uncertainties, which reduce errors by 6-28 per cent over those of PyAL + CNN. Due to the large amount of scatter in these results caused by modelling the EAGLE galaxy lenses, the presence of LOS structure has only a minor impact on the results, with errors that are factors of 1.3, 1.5 and 1.0 larger than for images without LOS structure for Einstein radius, orientation and axis ratio, respectively. With LOS structure, errors for the EAGLE galaxy lenses are factors of 2.2, 0.5 and 1.8 larger than those of SIE lenses with LOS structure for Einstein radius, orientation and axis ratio, respectively, compared to factors of 3.9, 3.1, and 4.2 when both test sets do not contain this extra structure.

4.4 Modelling Speeds

So far we have concerned ourselves with examining and comparing the overall accuracy and precision of each modelling method. However, if these methods are to cope with the vast quantities of lenses to be discovered by upcoming surveys then the time taken to model such lenses is also of significant importance. As such, we now consider the speed by which the different modelling methods can obtain the lens parameters seen in the previous results. Distributions of the time taken to model individual lenses using each modelling method are shown in Figure 4.16, for three of the test data sets: SIE lenses with HUDF sources, SIE lenses with HUDF sources and LOS structure, and EAGLE galaxy lenses with HUDF sources and LOS structure. Distributions for the CNN by itself are not shown as for a single lens the time taken is almost instant.

It is clear from the figure that the incorporation of the CNN helps quicken PyAutoLens' modelling for all three test data sets, although nowhere near as fast as the CNN by itself. Compared to PyAL (blind), PyAL + CNN increases modelling speeds by mean factors of 1.34, 1.08, and 1.14 for the three test data sets (1.19 on average), while PyAL + CNN (1σ) increases modelling speeds by mean factors of 1.56, 1.37 and 2.25 (1.73 on average). A significant improvement is obtained by incorporating the CNN's predicted uncertainties, especially for the EAGLE galaxy lenses; while all methods show some cases of lenses taking a long time to model, PyAL + CNN (1σ) has far fewer of these outliers, which themselves are generally much shorter than those of the other methods. This improvement arises from applying better priors on parameter space which allow PyAutoLens to converge faster and prevents it from falling into local minima that would give inaccurate parameter values. The use of (Py)MultiNest by PyAutoLens means that the closer it gets to the correct solution, the greater the number of samples that reach its acceptance threshold. This in turn quickens the Markov chain Monte Carlo (MCMC) sampling of (Py)MultiNest, further increasing the speed at

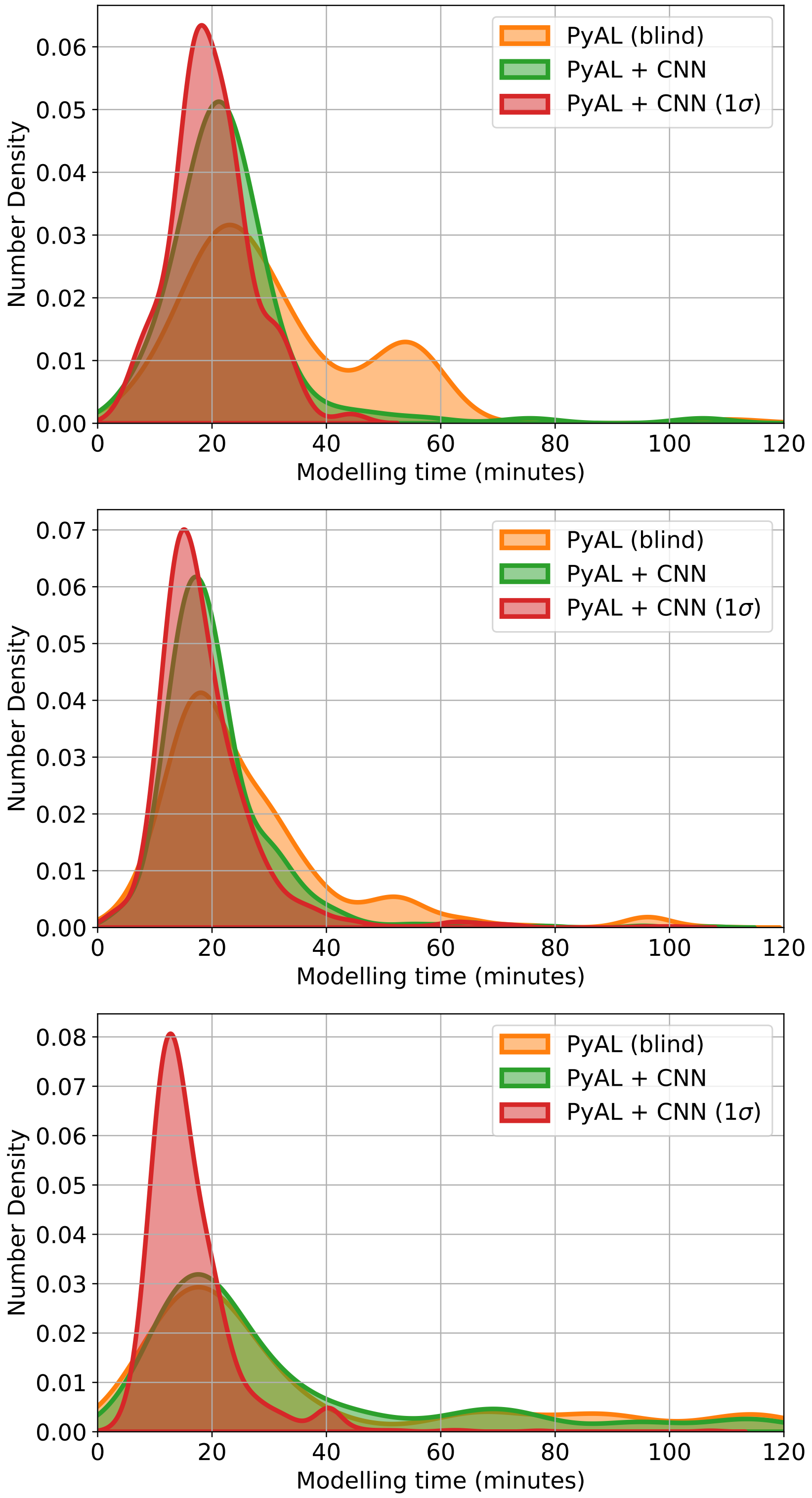

Figure 4.16: Distributions of the time taken for PyAutoLens to model each of the lenses making up a given test set when modelling blindly, with the CNN predictions as priors, and with the CNN predictions and uncertainties as priors. As the trained CNN can model lenses by itself almost instantly, its modelling times are not included here. From top to bottom: test sets consist of SIE lenses with HUDF sources, SIE lenses with HUDF sources and LOS structure, and EAGLE galaxy lenses with HUDF sources and LOS structure.

which PyAutoLens can reach the solution. Hence, starting in more optimal regions of parameter space can greatly quicken modelling. It is worth noting that the CNN-predicted uncertainties for the EAGLE galaxy lenses were generally underpredicted; while such priors for PyAutoLens may quicken modelling when centred on local or global minima, they are more likely to result in incorrect predictions. Additionally, those not centred on minima would greatly slow the modelling process, and so should more suitable priors be obtained for the EAGLE galaxy lenses, both the accuracy and modelling speed of PyAL + CNN (1σ) may be increased.

For the automated approach used in this work, only a single run of PyAutoLens is performed for each image, whereas conventionally not only would each image be inspected by eye beforehand, but some modelling may converge on incorrect values and hence require re-initialising with modified priors, further increasing modelling time and the need for human inspection. The CNN alone provides a much more rapid automated modelling than PyAutoLens or the combination method, making it the most suitable to handle upcoming large data sets extremely quickly. However, the fully automated PyAL + CNN (1σ) approach is still significantly faster and more accurate than PyAutoLens, so is well-suited to model lenses for which large uncertainties are predicted by the CNN.

4.5 Summary

With upcoming surveys like *Euclid* and LSST set to identify tens of thousands of gravitational lenses, conventional parametric parameter-fitting techniques like PyAutoLens will struggle to cope due to their relatively slow modelling speed and the need for manual inspection. Instead, faster automated approaches are required to either replace or enhance existing techniques.

In this chapter, we built upon the earlier work in this thesis by training an approximate Bayesian CNN to predict strong gravitational lens mass model parameters and comparing its performance to that of an automated version of the PyAutoLens semilinear inversion technique. These were applied to the range of increasingly complex lensing systems detailed in Chapter 3, including parametric SIE and power law lens mass profiles, EAGLE galaxy lenses, and LOS structure. In addition, we also explored the combination of these modelling methods, with CNN predicted values and uncertainties acting as priors for PyAutoLens.

Regarding the biases exhibited across the data sets in Section 4.3, axis ratios are generally underpredicted throughout due to distortions to the lensed arcs as previously discussed. While all modelling methods could exceed a 1 per cent median bias in Einstein radius, this was only seen for the EAGLE galaxy lenses: generally, more complex images result in larger average biases for all parameters, with magnitudes of those for Einstein radius spanning anywhere from zero up to 2.6 per cent. Without LOS structure, Einstein radii are over-predicted on average, and instead become under-predicted for images with LOS structure. This additional structure otherwise has little effect on the biases, except for increasing the magnitude of those for EAGLE galaxy lens Einstein radii. Over all the test data sets, PyAL (blind) is the only method that ever produced median orientation biases larger than 1°, or larger than 7.5 per cent for axis ratio (to a maximum of 10.1 per cent). Despite this, given the variety of biases exhibited by the modelling methods across the parameters and test sets, no one method appears to perform significantly better or worse in this regard. The biases observed in the CNN results also highlight how a single CNN can be biased even with the incorporation of dropout; this biasing could be reduced through combining the predictions of multiple networks each trained on a different sample of images.

Across the data sets and lens parameters in Section 4.3, the CNN errors are 19 ± 22 per cent lower than PyAL (blind) on average[3]. Hence, for the majority

---

[3] The impact of the relative performances of the modelling methods is also discussed further

of cases, the CNN results are on par with or exceed those of PyAutoLens, whose modelling is not typically automated, instead requiring manual inspection of images to obtain suitable priors. The combination of these two methods when using just CNN predicted values as priors, PyAL + CNN, improves upon PyAL (blind), achieving 27 ± 11 per cent lower errors across the parameters. However, it is frequently matched or exceeded by the CNN, with average errors only 4 ± 23 per cent lower than the network. In Section 4.4 it was found that while the CNN is by far the quickest modelling approach, this combination enhances the modelling speed of PyAutoLens, increasing it by a factor of 1.19 over PyAL (blind).

A substantial improvement is obtained when the CNN-predicted uncertainties are also incorporated into PyAutoLens' priors, PyAL + CNN (1σ). These uncertainties act to better constrain the prior distributions which help PyAutoLens to avoid local minima and converge faster on the correct solution. With these included, this combination of the two methods produces significantly lower errors than either one separately: PyAL + CNN (1σ) reduces errors on average by 37 ± 11 per cent across the parameters compared to PyAL (blind), 17 ± 21 per cent compared to the CNN, and 13 ± 9 per cent compared to PyAL + CNN. Additionally, PyAL + CNN (1σ) outperforms PyAL (blind) in all tests, as well as the CNN in the majority of cases. It also gives the highest PyAutoLens modelling speeds, increasing upon PyAL (blind) by a factor of 1.73 on average across the test sets. It should be noted that for the more complex lensing systems these uncertainties are often underpredicted, so should these be corrected an even greater improvement may be achieved.

We have seen in this chapter how CNNs can vastly simplify the search over parameter space for parametric modelling methods and how they can prevent methods using semi-linear inversion, like PyAutoLens, from falling into local min-

---

in Chapter 6, in which the errors are propagated to examine how well the methods can constrain cosmological parameters.

ima (the overmagnified and undermagnified solutions). Likewise, the combination method allows PyAutoLens to further refine CNN predictions in an automated fashion. While the CNN is undoubtedly the fastest modelling method, it suffers from a number of limitations. These, along with analyses of the results presented throughout this thesis, are discussed in the next chapter.

Chapter 5

Discussion & Conclusions

5.1 Overview

Strong galaxy-scale gravitational lensing is the deflection of the paths of light rays from a background source galaxy due to the presence of a foreground lensing galaxy, resulting in multiple images of the source that are often magnified and distorted. Such lensing provides a powerful tool for astronomy, seeing usage in a range of applications from studying galaxy evolution to constraining cosmological models. The complicated process of strong lens modelling has previously necessitated the use of techniques such as the parametric parameter-fitting of PyAutoLens, which are often relatively slow and require manual inspection. With the advent of large-scale surveys like LSST and *Euclid* soon to discover tens of thousands of strong lenses, automated techniques will be required to model these lenses quickly and efficiently. This has driven the use of machine learning in this area, with convolutional neural networks trained to recover lens mass model parameters extremely quickly. These networks apply convolutional kernels to images of lenses in order to extract information, but their training process requires a large data set of simulated images.

In this thesis, I have sought to investigate the use of convolutional neural networks in strong galaxy-scale lens modelling, examining the practicalities they face for this task, their accuracy and reliability when applied to upcoming surveys, how these compare to those of conventional modelling, and how CNNs and conventional modelling can be combined to gain the benefits of both. This work necessitated constructing a CNN and simulating my own data sets for training and testing. For later comparisons, these simulations contained various complexities, including HUDF source galaxies, power law and EAGLE galaxy lensing galaxies, and line-of-sight structure. The CNN was also redeveloped over the course of this thesis, through the creation and fine-tuning of a Bayesian CNN capable of predicting its own uncertainties.

5.2 Investigating Lens Modelling CNNs

In Chapter 2, I developed a CNN for strong gravitational lens modelling. To examine if such a technique would be applicable to images with the imaging characteristics of both LSST and the *Euclid* survey, I simulated data sets with such properties to both train and test the network on. These included both r-band and multiband gri imaging for LSST, and details of the CNN and simulations were presented in Section 2.1. Training on 50,000 'postage stamp' images of lenses allowed the CNN to predict parameters of the SIE mass model: Einstein radius, orientation, and ellipticity. Investigations into its application as a suitable modelling method were presented in Section 2.2. They included examining how accuracy varies between parameters and as functions of the parameter values themselves, as well as providing comparisons between *Euclid* and LSST, and single-band and multiband data. Additionally, the impacts of foreground lens light subtraction, lens mass-light alignment, and increasing signal-to-noise were assessed. Together, these investigations have enabled a means of evaluating the performance of the network when applied to upcoming wide-field surveys, and can act as a guide for future training.

For this work, the lens population consisted of isolated elliptical galaxies, adequately described by an SIE mass density profile without external shear. Additionally, the lens light centroid was assumed to be exactly aligned with the centre of mass (although the lens need not have been exactly centred in each image). However, in reality, lenses are rarely completely isolated, instead being subjected to perturbative effects from environmental structures, substructure in the lens itself and mass along the line of sight. In the test data sets, the velocity dispersions of the foreground lens population seen in Figure 2.5 may also be higher on average compared to the distribution of real lenses (Collett, 2015). This would correspond to the real population containing fewer lenses with large Einstein radii, resulting in a slight increase in overall CNN error, as per the top row of Figure 2.10. These simplifications enabled a more efficient investigation of the practicalities of using CNNs for lens modelling, but further work was required to analyse how the performance of the network would change when applied to more realistic data, as was explored in the later chapters.

5.2.1 Analysis of Results

For the inclusion of foreground lens light, the CNN is found to produce relatively large errors in its predictions of mass model parameters compared to those typically obtained by modelling methods (e.g. Maresca et al., 2022, and the results in Chapter 4 of this thesis), with such errors equal to or below 0.09 arcsec, 0.23 radians (13°) and 0.08 for Einstein radius, orientation and ellipticity, respectively. However, when trained and tested on images with prior lens light subtraction these errors are decreased by an average of 34 ± 5 per cent across the predicted parameters and data sets, with the greatest improvements seen for Einstein radius and LSST *gri* images. These results therefore support the use of lens light subtraction during image pre-processing for CNN lens modelling, as typically performed by conventional fitting methods. Significant improvements to the performance of the network are obtained when using multiple bands, with the use of three bands reducing parameter errors by 24 ± 2 per cent and 20 ± 2 per cent for images with and without lens light subtraction. However, the inclusion of colour information does not appear to be as effective as lens light subtraction, although such removal was for an idealised case so in practice these two may result in similar improvements. Overall, their combination proves most successful; training and testing on LSST *gri* images with lens light subtracted produces the lowest errors for all parameters. This is because any residual shot noise from the lens light will likely contrast with the background source light for colour images, and even if such noise is negligible the use of multiple bands can in essence raise the signal-to-noise of the input.

Analysing how parameter errors change with the parameter values themselves, we observe that all errors increase as fractional Einstein radius decreases, due to fewer lensed source pixels reducing the available information. Inclusion of lens light has little effect on the shape of the trends other than changing the absolute error values; parameter errors are unaffected by the ratio of Einstein radius to lens light effective radius, indicating that network performance is not impacted by the size of the lens light profile compared to that of the source. Orientation is found to have no effect on predictions, with a sufficient training set ensuring a rotation-invariant network as required for application to astronomical images. However, increasing ellipticity affects all three parameter errors, with those of ellipticity and fractional Einstein radius increasing while those of orientation decreasing. More elliptical lenses make determining orientation easier but also produce more varied and non-circular lensed sources, making it harder to predict the other parameters.

For data sets containing foreground lens light, differences between the light and mass profile ellipticity and orientation have a significant impact on the performance of the network. This is especially true when the network is trained on data sets with no or little difference between the two, due to the network being more influenced by the lens light as opposed to the lensed source light. In addition, even when trained and tested on images with zero scatter between light and mass profiles, errors are an average of 11 per cent larger than those for images with their lens light subtracted. No systematic biases are observed if the network is trained with some degree of mass-light scatter, providing reassurance that, notwithstanding environmental effects, measurements of H_0 from gravitational time delays can be improved simply through increasing the sample size of time-delay lenses. As a general rule, it is better to train a network assuming a moderate amount of scatter in the mass-light orientation and ellipticity offset than to train assuming that mass closely follows light. These results were also used to set the mass-light scatter in the data sets used throughout the rest of this work.

Errors for predicted parameters decrease approximately exponentially with increasing test image SNR up to SNR=100 after which there is little improvement. This indicates that the modelling of any detected lensed image measured to have SNR$\geq$100 will gain no benefit from further stacking, even if more observations of that system become available from surveys such as LSST. In general, testing on more deeply stacked images results in higher CNN accuracy, with networks trained on stacked images performing the best in this case. Hence, to deal with a large range of signal-to-noise ratios and stacked images a larger training set encompassing these ranges is required, with a focus on low SNR images, either used to train a single CNN or separate CNNs for different noise bands.

5.3 BNN Architecture

The next work in this thesis involved generating more complex training and testing data sets, and updating the neural network to better handle the resulting wider array of lensing images. In Chapter 3, I constructed my own Bayesian neural network for use in the remainder of this thesis, as such networks can both predict uncertainty measurements on their parameters and increase network accuracy through repeat testing. The method utilises dropout in both training and testing, requiring a fine-tuning process of the dropout rate to ensure uncertainties are correctly predicted. This process was performed in Section 3.3, with a dropout rate of 5 per cent resulting in the most accurate predictions of uncertainties. However, this dropout rate was selected from testing on simulated images and hence, as with general CNN training, the resulting uncertainties may not proves as accurate when applied to real data. If so, these uncertainties would likely be underpredicted, as observed when LOS structure and EAGLE galaxy lenses are introduced in Section 4.5. For real data, this reinforces the need for a comprehensive training set that covers all ranges of parameter space, noise levels, and image artefacts, in addition to the complex mass models presented in this work.

It is worth noting that the architectures of the neural networks in this thesis (see Figures 2.1 and 3.2) consisted of relatively few layers compared to other CNNs. For example, the method used by Hezaveh et al. (2017) to predict mass model parameters involved training a combination of four networks: Inception-v4 (Szegedy et al., 2017), AlexNet (Krizhevsky et al., 2012), Overfeat (Sermanet et al., 2013) and their own network. The larger number of layers used in these networks was required to train on higher resolution images which incorporated more complex features such as hot pixels, cosmic rays, pixel masking and occasional background objects. The images used in this thesis have a lower resolution and do not contain these features, hence the simpler network architecture does not appear to be detrimental, with the accuracy instead limited by image resolution. Simpler architectures contain far fewer neurons (and hence weights and biases), helping to quicken training and prevent overfitting to the training data set, but too few neurons results in insufficient training and poor accuracy. This tuning of network architectures is a well known issue, and as there are no standardised ways of building a network due to the variety of their applications, there is a degeneracy in that multiple network architectures can potentially achieve the same accuracy. I chose my architecture following a hyperparameter grid search while trying to minimise its complexity (see Section 2.1.1) and later checking its suitability with a Bayesian hyperparameter optimisation method (Section 3.4.1).

5.4 Realistic Data Sets

Following a description of the Bayesian CNN's architecture and training and testing processes, Section 3.5 detailed the updated simulations used for its training. Now focusing only on simulated *Euclid* images to avoid repeat results, the network was applied to larger cutouts featuring multiple source types enabling it to handle more realistic lensed images. Two data sets were created, doubled in size compared to previous training, so that the network could be trained to predict parameters for either the SIE model or the more challenging power law ellipsoid model. Additionally, a series of increasingly complex data sets of images were generated for use in testing both the network and the conventional fitting method PyAutoLens, as well as testing combinations of the two. Section 3.6 described these test sets, which ranged from simple parametric lenses and sources to EAGLE galaxy lenses with both real HUDF sources and CosmoDC2 line-of-sight structure.

As discussed in Section 3.6, the simulated lenses used in Sections 4.3.1 to 4.3.3 have a higher average Einstein radius than that expected from lenses detected by *Euclid*. The impact on CNN performance for smaller Einstein radii was observed in Figure 2.10, with less available information increasing errors for all parameters. From examining the results in the test sets as functions of Einstein radii we found no significant deviation in the relative accuracies between the modelling methods. Additionally, in Section 4.3.4 the increase in error for the EAGLE galaxy lenses was found to be dominated by these complex foreground galaxies rather than the smaller average Einstein radii. Hence, parametric lens mass profiles with smaller image separations are expected to produce slightly higher errors than for the data sets in Sections 4.3.1 to 4.3.3, but the relative errors between modelling methods would remain consistent.

It is also worth considering here the source-lens alignment, i.e. the distribution of doubles and non-doubles (quads, rings) within the data sets. While doubles are expected to dominate the catalogues of future surveys, such images are in general more difficult to model. In this work, a lower limit on magnification was used to ensure lensed images were being generated; this focused on sources in higher regions of magnification, resulting in only around 30 to 40 per cent of test images containing doubles. As such, we examined how the performance of the modelling methods in Chapter 4 varied by analysing 100 doubles and 100 non-doubles in the test set containing SIE lenses and HUDF sources. On average across the modelling methods, errors in Einstein radii and axis ratios for doubles were factors of 2.1 and 1.9 times larger than non-doubles (1.5 and 1.3 times larger than the combined 200-image data set), respectively. Meanwhile, orientation errors for doubles were on average a factor of 0.8 times that of non-doubles (and were almost unchanged compare to the combined data set), most likely because the set of non-doubles contained near-complete Einstein rings. However, while the modelling errors differed between doubles and non-doubles, the relative accuracies between the modelling methods remained broadly similar.

One simplification made in this work is with regard to the subtraction of lens light. In this thesis, the test images assumed an idealised light subtraction for all but the background source galaxy, in which the true parameters of the lens light profile were known. In reality, this would not necessarily be the case and the profile may also deviate somewhat from a smooth parametric profile. The point spread function (PSF) of a telescope is also never perfectly known, and so this can further increase residuals. Nevertheless, our subtraction of lens light does leave behind realistic shot noise residuals that, having been trained on, may alleviate some of the difficulty in testing the network on images with more realistic lens light subtraction.

Additionally, when dealing with real images, masks are often applied to cover residual light for all but the lensed source. This process can be automated through various means, but can often require human inspection. The performance of modelling methods may be impacted by how well light is masked or subtracted from the lens and companions in the field of view, which would require future work; for example, machine learning methods of denoising and deblending lens and source light have been developed by Madireddy et al. (2019). While PyAutoLens has been designed to accommodate masked images, it may struggle to fit the source light profile if residuals were present from incorrect masking or subtraction. Additionally, while fitting annular rings to the lensed sources greatly quickened PyAutoLens' modelling process for dealing with many thousands of images, it has been shown that such rings have the potential to bias the modelling (Nightingale et al., 2018). This may have contributed to the lower performance of PyAL (blind) in Chapter 4, however biasing from masking was not investigated in this thesis. No masks were applied to the simulated images in this work, and so the current CNN would also greatly struggle if not trained appropriately. In this way, much of the impact of masking on CNN performance would be remedied should such training also include examples where masks have been applied incorrectly. As such, it is likely that the conclusions presented here would remain approximately unchanged in the presence of masked lenses.

In addition to idealised lens light subtraction, it is worth acknowledging that the simulated data sets used in this thesis also lacked a number of features present in real astronomical images of strong lenses, limiting the current CNN's performance when applied to real images. Transient image artefacts such as hot pixels, satellite trails and cosmic rays were not included, as these were assumed to be adequately removed during image pre-processing. Additionally, bright stars and non-lensing galaxies were not included in the training or testing images, and so these would need to be taken into account during future CNN training should they not be removed during image pre-processing; the latter objects could prove challenging in this regard due to many strong lenses lying in galaxy groups. The PSF applied was identical across all images in a given data set, whereas the PSFs of telescopes can vary between observations. Such variations could lead to further distortion to the lensed source images, leading to incorrect modelling predictions. For ground-based telescopes like LSST this arises primarily due to atmospheric turbulence, and so the impact of this may be reduced somewhat through the use of adaptive optics. Despite these limitations, should they be accounted for during image pre-processing or feature in the network's training, the relative performance between the CNN and other modelling methods would likely remain the same given that these methods would also be affected. Adequate training would also allow CNN predicted uncertainties to highlight any difficulties in modelling potentially arising from such objects and artefacts. Transfer learning could also be applied in this area, involving the simulation-trained CNN being further trained on real lenses as they are discovered. However, while transfer learning would likely be effective for lens detection, such lenses would first need to be accurately modelled before they can be added to a training set for lens modelling.

Despite the limitations discussed with regard to the training set, it is worth noting that in addition to the testing of the modelling methods in Chapter 4, PyAL (blind) and the Bayesian CNN were also later tested on a real submillimetre galaxy lens, with PyAutoLens now able to predict its own uncertainties. Originally detected by the *Herschel* Space Observatory (Pilbratt et al., 2010), the chosen lens was one of the candidates produced by Nayyeri et al. (2016) from the extension to the *Herschel* Multi-tiered Extragalactic Survey (HerMES; Oliver et al., 2012), the HerMES Large Mode Survey (HeLMS; Asboth et al., 2016), with the lens known as HELMS5 (or HeLMS J234051.5-041938). Subsequent ALMA follow-up observations of this lens and others, described in Amvrosiadis et al. (2018), have allowed for uv-plane modelling of this lens using an elliptical power law mass profile with external shear by Maresca et al. (2022), with both papers providing images and further details of the lens, and the latter paper providing details of the lens light subtraction procedure. For the testing of PyAL (blind) and the Bayesian CNN, the 873 μm image was created from this ALMA data to have the same image size and resolution as the *Euclid* data set used in Chapter 4, with modelling performed identically to that chapter using an SIE mass profile. In doing so, PyAL (blind) predicted an Einstein radius, orientation, and axis ratio of 0.53±0.01 arcsec, 128±2°, and 0.88±0.01, respectively, while the Bayesian CNN predicted these as 0.52±0.01 arcsec, 118±20°, and 0.87±0.06, respectively. As can be seen, the two methods agreed very closely with one another for this real lens: all predictions matched to within their predicted uncertainties, except for the 10° difference in the orientation predictions which, while within the CNN's uncertainty, was notably beyond that of PyAL (blind). While only for a single test case, these results are a promising sign that the Bayesian CNN can produce suitable predicted parameters and uncertainties when applied to real lenses.

While the image data sets used throughout this work have contained only lensing systems, future real data sets may inadvertently contain non-lenses if the image classification process is not perfected. The expectation is that attempting to model non-lenses with PyAutoLens will produce poor evidence values which would indicate a problem with the modelling. However, the uncertainties predicted by the CNN may be of little use in this regard; while it is possible that they may increase for non-lenses due to a lack of a preferred lensing model, the CNN was not trained on such images and so their impact remains unknown without further testing. This, along with re-training the CNN on both lenses and non-lenses to ensure appropriate predicted uncertainties, is beyond the scope of this work and instead left for future investigation.

5.5 Bayesian CNN and Parametric Methods

Chapter 4 began by introducing the semilinear inversion technique PyAutoLens, chosen as the conventional parametric parameter-fitting modelling method to which the Bayesian CNN would be compared. Additionally, in Section 4.2 I presented a promising way to combine both methods through incorporating CNN-predicted parameter values and uncertainties into the priors of PyAutoLens ('PyAL + CNN' and 'PyAL + CNN (1σ)', respectively). This was done in order to create an automated method both faster and more accurate than PyAutoLens by itself when applied in an automated manner ('PyAL (blind)'). Following the work of the previous chapter, these four modelling methods were then tested on data sets of increasing complexity. The results were presented in Section 4.3, along with an analysis of their modelling speeds in Section 4.4.

5.5.1 Parametric Density Profiles

All methods reach very high accuracies when presented with images containing only SIE lenses and parametric sources, and this remains the case when the sources are replaced by HUDF galaxies. Errors are consistently below 0.014 arcsec, 7° and 0.04 for Einstein radius, orientation and axis ratio, respectively, with the CNN and PyAL (blind) achieving comparable errors on average for both of these data sets. The modelling of power law profiles (without LOS structure) results in an expected decrease in performance compared to the SIE profiles. All modelling methods experience a large scatter in their predictions of power law slope, however the CNN continues to do well at modelling these lenses despite their difficulty. Additionally, for these data sets, the combination of methods leads to significantly lower errors: PyAL + CNN (1σ) errors are 31-44 per cent lower than PyAL (blind) across the two SIE data sets, with this incorporation of uncertainties improving upon PyAL + CNN by 12-17 per cent. For the power law data set, these error reductions become 32-62 per cent and 5-39 per cent, respectively.

Compared to SIE profiles, all modelling methods struggle with power law profiles due to the inherent difficulty in accurately measuring slope values, leading to slight biasing towards overpredicting Einstein radii and underpredicting axis ratios. Results are mostly unbiased for SIE mass profiles, however for a significant number of lenses PyAL (blind) underpredicts Einstein radii and axis ratios, which generally correlate with one another, leading to a large scatter of outliers. These outliers are not present for the other modelling methods, suggesting that the CNN is more reliable in its predictions, which when incorporated into PyAutoLens help it initialise closer to the correct solution in parameter space.

5.5.2 Impact of Complex Mass Structures

The CNN was not trained on images containing LOS structure, nor to predict external convergence or shear, and hence the introduction of significant LOS structure increases modelling difficulty. Across the SIE and power law data sets, the differences between PyAL + CNN (1σ) and the CNN are greatly reduced, however both methods continue to significantly improve upon PyAL (blind), reaching 16-36 per cent and 9-35 per cent lower errors, respectively. A similar result is found when the methods are presented with complex EAGLE galaxy lenses (without LOS structure), with PyAL + CNN (1σ) and the CNN now respectively achieving 30-58 per cent and 36-55 per cent lower errors compared to PyAL (blind). With LOS structure, these improvements become almost identical: 28-48 per cent and 28-47 per cent, respectively. As such, methods involving the CNN continue to give significantly better results than PyAL (blind), despite the CNN not training on such complex lenses or LOS structure.

In general, more complex images result in larger average biases for all parameters: such values only exceed 1 per cent for Einstein radii and 5 per cent for axis ratio when presented with EAGLE galaxy lenses. The biases increased for all modelling methods, with no one method generally outperforming another. The introduction of LOS structure increased the magnitudes of biases for EAGLE galaxy lens Einstein radii, and resulted in Einstein radii being underpredicted on average for all test sets as discussed in Section 4.3, but otherwise had little effect on modelling biases.

For both SIE and power law mass profiles, the incorporation of significant LOS structure makes modelling challenging, which for this work is equivalent to that of a lens often residing within a galaxy group with many tens of galaxies along the LOS. Replacing the parametric profiles of the lenses with hydrodynamical EAGLE galaxies also serves to increase difficulty, with these complex lenses producing fewer smooth, well-defined arcs than those trained on by the CNN, which can increase the degeneracy between axis ratio and orientation when modelling (Mukherjee et al., 2018). These lenses also in general have smaller Einstein radii than the parametric lenses, corresponding to more pixelated arcs and reduced information available for modelling. Additionally, the large errors seen in axis ratios for the EAGLE galaxy lenses can be attributed to how these values often change as a function of radius. This, along with many lenses containing significant substructure in their convergence maps, especially when combined with LOS structure, prove difficult to reliably fit in order to obtain the parameter values used as 'true' values for this work.

The EAGLE convergence maps were fully utilised in the sense that their resulting deflections produced the final lensed images to which the lens modelling methods were applied. However, the fitting procedure described in Section 3.6 attempted to fit a smooth parametric profile to the convergence maps. Were lensed images to be created using these fitted parametric profiles, the resulting lensed arcs may differ to those produced using the complex mass profiles. Hence, while the fitted parametric profiles provide the best approximation of the convergence maps, their parameters may not accurately reflect those that would be obtained by modelling of the lensed source light. One alternative to this would be to generate the lensed images using deflection angles derived from the fitted parametric profile, rather than directly from the convergence map. The fitted profile parameters would therefore be accurate to the final lensed image, and hence would ensure more agreement with parameters derived from modelling of the lensed source light. However, this approach would simply produce lensed images from parametric profiles like those of the earlier test sets, with the only difference being in the distributions of parameters. As such, it would make little use of the EAGLE convergence maps and would not produce the range of possible complexities found in images of realistic lensing systems.

Another alternative approach would be to first produce the image data set containing EAGLE galaxy lenses used in this work, then fit these images rather than the convergence map. The fitting function would take distances, the source position, and the source image as its input, select parameters for the parametric profile deflection angles, and apply these deflections to the source image to output the best-fit lensed image. While this approach models the lensed source light similar to the modelling methods used in this work, it has the advantage of knowing the true distances, source position, and source appearance. However, such a fitting function would be significantly more complex and hence would likely take much longer to converge on the best-fit lens parameters.

While the CNN was not trained on images containing LOS structure or EAGLE galaxies, such images actually result in the highest improvements over PyAL (blind). This could be because the iterative method that PyAutoLens employs to navigate parameter space is more sensitive to the inclusion of these complex structures than the CNN's instantaneous approach when applied in an automated fashion. Additionally, the introduction of these structures leads to a decrease in the difference between PyAL + CNN (1σ) and the CNN. However, due to a lack of sufficient training in these areas, the CNN-predicted uncertainties also appear to be increasingly underpredicted for axis ratios and slopes. Hence, were these to be corrected through further training or otherwise, for example through the incorporation of the unbiased hierarchical inference of Wagner-Carena et al. (2021), PyAL + CNN (1σ) could produce a more pronounced increase in accuracy over the CNN for these complex lenses.

5.5.3 Modelling Speed

In addition to an increase in accuracy, the combination of methods results in faster modelling times than PyAutoLens by itself. Compared to PyAL (blind), modelling speeds are increased on average by factors of 1.19 and 1.73 using PyAL + CNN and PyAL + CNN (1σ), respectively. The priors used in the latter help to reduce the number of outlier lenses that would otherwise take much longer to model, providing a more consistent modelling time. It should be noted that for automation, PyAutoLens is only performing a single run for each image; conventionally, human inspection would be used to set these priors, which if incorrect would lead to PyAutoLens falling into local minima and require re-initialising the modelling, further slowing the process. Incorporating CNN predictions therefore takes the place of human inspection, automating the modelling process to deal with upcoming survey data.

PyAutoLens has a range of configuration options, but while these can lead to obtaining precise fits it can also be time-consuming to choose the best options. In addition, PyAutoLens requires a high-end PC to run at the speeds shown here. Meanwhile, the CNN need only run on a GPU machine for training, taking less than two hours for 100,000 training images, and while testing the CNN using a GPU machine will speed up modelling, it is not required. While the neural network alone can be rapidly trained and applied to images, its combination with PyAutoLens provides a slower yet fully automated method of obtaining more accurate results, suitable for applications to lenses that the CNN alone may struggle to model. As such, the CNN would be suitable to handle large samples of upcoming lenses, but should large uncertainties be predicted by the network, the modelling may be passed to PyAL + CNN (1σ) to refine the predictions.

5.6 Conclusions

The CNN's modelling speed makes it well-suited for large catalogues of lenses, and given the many thousands of upcoming strong lens observations such a network can very quickly predict accurate values and uncertainties for the majority of cases. When applied to smooth mass profiles, optimal performance is obtained for LSST *gri* multiband images with lens light subtracted, and while prediction accuracies for all parameters vary with both ellipticity and Einstein radius they can be improved through the stacking of images acquired over the course of future surveys. The incorporation of a fine-tuned Bayesian framework allows the network to predict its own uncertainties, and can additionally improve accuracy and reliability through repeated testing with MC dropout.

While its accuracy in this work can rival that of the combination method PyAL + CNN (1σ) for some lenses, the CNN is limited by the complications and artefacts encountered in real data, with performance heavily dependent on the size and quality of its training set. Until the CNN can be trained on vast quantities of sufficiently realistic survey data, the PyAL + CNN (1σ) combination provides a promising alternative automated modelling method. Lenses that are challenging for the CNN to model produce larger CNN uncertainties, which can then be passed to PyAutoLens (or another parametric modelling technique) to verify and refine predictions. Alternatively, any lenses that produce notably small CNN uncertainties could be passed to PyAutoLens for even further model refinement, while those producing large uncertainties (or whose predictions differ greatly between the CNN and PyAL + CNN (1σ)) are sent for human inspection.

Additionally, when automated alone, semi-linear inversion modelling methods like PyAutoLens struggle if their priors are not correctly constrained. CNN predictions can vastly simplify the search over parameter space for these methods, resulting in faster convergences and preventing them from falling into local minima. This combination of methods acts as a fully automated pipeline for application to large samples of lenses, and can achieve accurate results far quicker than conventional modelling. Rather than keeping them separate, this work highlights the importance of considering the combination of machine learning with conventional approaches in order to gain the benefits of both.

5.7 Future Work

Following the work presented in this thesis, there are a number of aspects that could see further development. The major limitation of a neural network is of course its training set, and within the strong lensing community there is currently a major need for a comprehensive realistic strong lensing simulation suite for both training and testing purposes. As such, training this network on an even larger data set containing hydrodynamical lensing galaxies and LOS structure would be a priority, with images containing other realistic features such as background galaxies, non-uniform point spread functions, and image artefacts. Multiple networks could also be used, trained on different samples of the training set, with their predictions combined in order to reduce any biases that a single network might produce. Additional lens model parameters could also be predicted by the CNN, including lens centre position, external shear and total magnification, however modelling these parameters has been explored in other works. While this work used the approximate Bayesian formalism of Levasseur et al. (2017), an updated form has since been developed by Wagner-Carena et al. (2021) that accounts for biases in the posteriors arising from differences between the training and test sets. This could therefore be incorporated into this network before it is ultimately applied to real data once upcoming surveys commence. With enough real data, any necessary modifications to the training set can be carried out and the network retrained. Such a network could also see use in probing the nature of dark matter through detecting LOS structure and substructure, either through enhanced training on these features or through the indication of their presence: they would produce large predicted uncertainties and increased mismatches between the CNN's predictions and those of the PyAutoLens methods. These detections could then be extracted from the data set for more detailed modelling in order to probe low mass scales and hence constrain dark matter's free-streaming length (see Section 1.1, and e.g. Despali et al., 2018; Gilman et al., 2019; Şengül et al., 2020).

While neural networks have seen much work in the realm of lens modelling and dark matter sub-halo measurements (e.g. Alexander et al., 2020; Lin et al., 2020), such work has generally been limited to isolated galaxy lenses. In reality, most galaxies lie in groups and little work has been done in improving the efficiency of modelling strong lenses where multiple foreground galaxies act as the lens, i.e. double lens plane systems and strong group-scale lensing. Machine learning methods like those in this thesis may provide aid in this endeavour, helping to generalise the application of neural networks to a wider range of lensing configurations. Such images could be tested on by multiple networks, each trained to predict a different number of mass profiles, with the best fit corresponding to the network predicting the lowest uncertainties. While many overlapping profiles may best fit the data, the required increase in network outputs would make training more challenging and hence fits with fewer profiles would be preferentially selected. To aid with this, an initial network could be trained to simply predict the number of profiles (and uncertainty in that number) needed to best fit the data, in turn determining which of the above networks would be applied. Alternatively, the feasibility of networks to predict convergence maps from lensed images could be explored, with application to both group-scale lenses and those with significant substructure. These maps could then be modelled more easily than lensed images, and used with time delays for potentially more accurate H_0 determination than from using fitted smooth mass profiles. However, in practice this may not be the case due to the light travel time varying so rapidly with position in regions of high magnification: the viability of predicting convergence maps in such a way remains to be investigated.

In general, neural networks require training on tens to hundreds of thousands of images. However, only a couple of thousand real lens candidates have been discovered, and of these only a few hundred are of Grade A standard (i.e. definite, clear examples of lensing systems). This therefore necessitates the use of simulations to obtain lensed images. For current neural networks, these simulations are often the bottleneck when it comes to achieving accurate results when applied to real images. Dia et al. (2019) used a generative adversarial network (GAN) to simulate galaxy images, trained on images provided by the Galaxy Zoo citizen science project, however these images are limited to low-redshift galaxies. Neural networks could provide a means of generating large data sets of high-redshift galaxy images, including mergers, quasars and galaxies containing Type Ia supernovae, for use as realistic sources in gravitational lensing simulations. To do so would likely require training a GAN using images generated by cosmological hydrodynamical simulations, such as the EAGLE simulation, along with real HUDF galaxies. Such networks, or other methods, may also have the potential to simulate dark matter profiles (in the form of convergence maps) for foreground lenses, enabling rapid generation of large training sets rather than having to solely rely on hydrodynamical simulations for realistic lenses.

While this thesis has been concerned with purely machine learning approaches, as have many other recent lens modelling methods, there is still much to be gained from the use of citizen science in this area. With regard to lens detection, Knabel et al. (2020) presented how spectroscopic, machine learning, and citizen science methods compared, and found little overlap in the three sets of identified lenses. While machine learning appeared to be the most efficient, a combination of methods would result in a more complete sample of recovered lenses, to take full advantage of the vast amounts of data generated over the next decade. Such a combination could be used to identify complex lensing systems for careful modelling, distinguishing them from those that would be easily modelled by neural networks. Similarly, we have seen in this thesis that the combination of neural networks with conventional fitting offers the highest overall accuracy, and so it is also worth exploring how machine learning and citizen science could best make use of one another for strong galaxy-scale lens modelling. For example, Küng et al. (2018) developed a form of lens modelling for use in citizen science, however this required training a select group of volunteers as opposed to being open to the general public. If the form of modelling presented in this thesis could be made accessable to a larger group, machine learning could be of potential benefit, from providing prior modelling attempts to aid volunteers, to refining their models, to using any major discrepancies between neural network and citizen science modelling as a means of identifying complex lensing systems. Alternatively, after modelling by a neural network, volunteers could be asked to examine reconstructed high-redshift sources to identify unusual shapes indicative of poor models or complex galaxy morphologies. Regardless, it is clear that the use of multiple techniques involving machine learning for both lens detection and lens modelling will be of great benefit over the coming decade.

Chapter 6

A New Era in Gravitational Lensing

In this chapter, I will expand upon the work in this thesis with regard to its potential applications, discussing the astrophysical context of the work with an emphasis on methods that employ statistical analyses of the many lensing systems made possible by future surveys like *Euclid* and Rubin LSST. First discussed in Section 1.1, a review of the scientific motivations for strong galaxy-scale lensing is continued in Section 6.1 with a focus on the above applications to specific science goals. In Section 6.2, I explore quantitatively how the results of this thesis can be applied to one of the discussed science goals by investigating how well the modelling methods can be used to constrain cosmological parameters when applied to double source plane lenses. This is followed by a discussion of the results and the conclusions of this chapter in Section 6.3.

6.1 Strong Galaxy-Galaxy Lensing Motivations

In Section 1.1.2, an overview was presented of the uses of strong galaxy-scale lenses, from detecting dark matter substructure and reconstructing high-redshift sources for aiding our understanding of galaxy evolution, to testing cosmological models through constraining parameters like the Hubble constant from gravitational time delays. This was followed in Section 1.1.3 by a description of the upcoming surveys set to discover on the order of up to 100,000 strong lenses, increasing the number of known lenses by two orders of magnitude. As such, these surveys will enable a new era in gravitational lensing whereby multiple science goals can be explored in a statistical sense rather than through the study of individual lenses. In this section, I present multiple examples of these science goals and discuss how the modelling of large samples of lenses provides a means of achieving them.

6.1.1 Dark Matter Properties from Lens Substructure

To begin, accurate lens modelling can be used to detect substructures within a lens' mass profile. In theory, CDM particles have non-relativistic velocities that allow the to form into low-mass substructures, while in comparison, warm dark matter (WDM) would suppress the number density of these substructures down to a lower mass limit dictated by its free-streaming length (e.g. Lovell et al., 2012) within the range $10^6 - 10^9 M_\odot$ (He et al., 2022a). This mass scale is therefore an important probe of the properties of dark matter, as it is currently unclear as to if the mass function of the CDM model holds at such low halo masses.

As such, much work has been done on developing the theory (Amorisco et al., 2022) and methods for detecting low-mass substructures within galaxies through gravitational lensing (Vegetti and Koopmans, 2009; Vegetti et al., 2014; Bayer et al., 2018; He et al., 2022b). While few lenses have been studied for low-mass haloes in this way (e.g. Ritondale et al., 2019; Hsueh et al., 2020), with detections below $10^{10} M_\odot$ fewer still (Vegetti et al., 2010, 2012; Hezaveh et al., 2016), approximately 50-100 lenses with measured substructure should be able to set limits on WDM mass, if not rule out CDM should no sufficient haloes be detected (Li et al., 2016b; Gilman et al., 2019; Simon et al., 2019). With the wealth of lenses expected from future surveys, automated techniques for detecting substructure have seen development over the past few years, from simulation-based inference (Brehmer et al., 2019; Wagner-Carena et al., 2022) to the use of CNNs (Alexander et al., 2020; Lin et al., 2020; Varma et al., 2020; Ostdiek et al., 2022a), in order to achieve these constraints on the properties of dark matter.

6.1.2 Galaxy Evolution from the Stellar-to-Halo Mass Relation

Strong gravitational lens modelling can often provide useful information for understanding and constraining galaxy formation and evolution, for example through the evolution of the stellar-to-halo mass relation (SHMR). In a CDM universe, galaxies form in the centres of dark matter haloes (or sub-haloes) as the infalling gas cools and condenses within their potential wells, with the well depth dictated by the halo mass. Hence, the halo mass should follow a tight correlation with properties of the galaxy within it, including the stellar mass (Girelli et al., 2020). However, the details of the mechanisms behind galaxy evolution (e.g. gravitational collapse, merging of dark-matter clumps, cooling via radiative processes) are not fully understood. The SHMR can aid in measuring the stellar initial mass function (Sonnenfeld et al., 2018) and help constrain how gas cools within galaxies, environmental effects such as ram pressure stripping, and feedback processes at play: AGN feedback limits star formation at high masses, while feedback from massive stars may eject or prevent the accumulation of gas at low masses (Wechsler and Tinker, 2018).

Large-scale hydrodynamical simulations like Illustris (Vogelsberger et al., 2014) and EAGLE (Schaye et al., 2015) can be used to model the physics governing galaxy evolution, but are limited in accuracy by poorly understood processes such as star formation and feedback (Naab and Ostriker, 2017). Many other approaches have been developed to examine how galaxy properties relate to the dark matter halo (see Wechsler and Tinker, 2018, for a review), including satellite galaxy kinematics (More et al., 2011; Li et al., 2012), weak lensing (Leauthaud et al., 2011; Velander et al., 2014), and X-ray observations of galaxy clusters (Kravtsov et al., 2018). Strong lensing provides a method of accurately measuring the total mass of a galaxy and its dark matter halo in order to compare with the stellar mass obtained from, e.g., mass-to-light ratio estimates or stellar population synthesis (Auger et al., 2010b; Küng et al., 2018; Sonnenfeld et al., 2015, 2018), and through modelling of large samples of lenses in future surveys the SHMR, and hence the physics of galaxy evolution, can be more tightly constrained.

6.1.3 Galaxy Evolution from the Lens Inner Profile Slope

An alternative approach to studying galaxy evolution is through their internal properties, such as how mass is distributed with radius, in order to answer questions about whether the internal structure and dynamics of isolated early-type galaxies evolve over time, and if their mass density profiles are accurately produced from numerical simulations. Of course, to study how these properties change with time requires modelling the mass density profiles of large samples of galaxies at a range of redshifts: strong lensing provides a suitable method for this (e.g. Bolton et al., 2012; Sonnenfeld et al., 2015; Sonnenfeld and Cautun, 2021), capable of accurate and precise modelling out to cosmological distances. As discussed in Section 1.2.5, power law mass profiles expand on SIE profiles by introducing a slope parameter, as while many strong lenses can be adequately fit with a simple SIE, having the slope as a free parameter can often lead to better fits overall.

Slope values may even evolve with redshift: while some previous studies found no trend (e.g. Barnabè et al., 2011; Cui et al., 2017), more recent work found mild negative trends such that the slope increases towards more recent times (Holanda et al., 2017; Chen et al., 2019a), suggesting dissipative processes may play a role in early-type galaxy evolution since $z = 1$ (Ruff et al., 2011).

Additionally, CDM simulations suggest dark matter haloes can be described by a Navarro–Frenk–White (NFW) profile in which the slope depends on radius (e.g. Navarro, 1996; Sonnenfeld et al., 2015). These have a 'cuspy' central slope in which the inner density scales as $\rho \propto r^{-1}$, with this slope affected by baryonic processes such as a contraction in the dark matter distribution due to the cooling and infalling of gas, or an expansion due to gas outflows caused by stellar, supernovae, or AGN feedback (Shajib et al., 2021). For example, while shallower slopes have been observed for some galaxies (e.g. Oh et al., 2011), Oldham and Auger (2018) found that most of their lenses had haloes centrally steeper than an NFW profile which suggested an early contraction of the haloes due to the initial infall of baryons. Meanwhile, Shajib et al. (2021) found that their haloes for massive ellipticals at $z \sim 0.2$ were well described by a standard NFW profile that was consistent with AGN feedback counteracting such a contraction. Accurate modelling of the inner profile slope of strong lenses, through NFW profiles or otherwise, can hence aid our understanding of these baryonic processes (e.g. Oldham and Auger, 2018; Sonnenfeld and Cautun, 2021).

6.1.4 Expansion History from Time-Delay Lenses

In addition to galaxy formation and evolution, strong lensing can also be used in the constraining of cosmological models. As mentioned in Section 1.1.2, for variable sources, time delays can exist between lensed images due to both geometric path differences of light rays on their way to an observer and the gravitational time delays from passing through different regions of the lens potential (Kochanek, 2004; Treu, 2010). Modelling of lensed quasars (e.g. Suyu et al., 2017; Bonvin et al., 2017; Chen et al., 2018; Birrer et al., 2019; Liao et al., 2019; Taubenberger et al., 2019; Birrer and Treu, 2020; Shajib et al., 2020; Wong et al., 2020; Denzel et al., 2021; Li et al., 2021) or even supernovae (Collett et al., 2019; Oguri, 2019; Foxley-Marrable et al., 2020; Bag et al., 2021; Bayer et al., 2021) can be combined with measurements of these observed time delays to measure the expansion history of the Universe through predicting a value for the Hubble constant H_0. From modelling large samples of time-delay strong lenses, the uncertainties on H_0 can be reduced to help resolve the tension between early-universe (CMB) and late-universe (e.g. lensing, supernovae, Cepheids, tip of the red-giant branch) measurements (Freedman, 2017; Mörtsell and Dhawan, 2018; Bengaly et al., 2020; Pandey et al., 2020; Vagnozzi, 2020; Riess et al., 2021). For example, the latest *Planck* measurements from the CMB provide a value of $H_0 = 67.4 \pm 0.5$ km s^{-1} Mpc^{-1} (Aghanim et al., 2020), while recent late-universe measurements are typically around 72-75: supernovae and Cepheid measurements of Riess et al. (2021) produced a value of 73.04 ± 1.04 km s^{-1} Mpc^{-1}, after quasar strong lenses of Birrer et al. (2020) gave $74.5^{+5.6}_{-6.1}$ km s^{-1} Mpc^{-1}.

To rule out systematic effects, recent work has been done to help identify and alleviate limitations in the measuring of H_0 from strong lens modelling, including the choice of mass profile (Cao et al., 2022), effects of substructure (Gilman et al., 2020) and line-of-sight structure (Fleury et al., 2021), and a degeneracy known as the mass sheet transformation (MST, Falco et al., 1985, a special case of the more general source-position transformation) which can bias H_0 while leaving imaging observables invariant (Chen et al., 2021). To determine H_0 with a 1 per cent uncertainty using a single lens would require sub-1 per cent uncertainties in measuring the time delays and in modelling the lens, while using a sample of N lenses would require these to each have uncertainties of roughly $(\sqrt{N}/2)$ per cent (Treu, 2010). Should modelling be combined with stellar kinematics to help break the MST and aid in measuring the distance to the lens, Shajib et al. (2018) concluded that a sample of roughly 40 accurately-modelled lenses at HST resolution could be enough to reach such a 1 per cent overall uncertainty. As future surveys should provide a larger sample of time-delay lenses, Park et al. (2021) applied a BNN modelling method to 200 simulated lenses at HST resolution and found that they could achieve a 9.3 per cent median precision per lens on their inferred H_0, decreasing to an unbiased 0.7 per cent (0.5 km s^{-1} Mpc^{-1}) from the combination of the full data set.

6.1.5 Expansion History from Double Source Plane Lenses

While time delays can provide accurate measures of the Hubble constant, often through assuming a ΛCDM cosmology, modelling of double source plane lenses (DSPL) offers a means of constraining other cosmological parameters independent of H_0 (Linder, 2004; Collett et al., 2012; Linder, 2016; Sharma and Linder, 2022). These systems feature a single lens that distorts the light of two separate background sources at different redshifts, resulting in two sets of lensed images and hence two measurable Einstein radii. An example of such a lens, known as the 'Jackpot' lens due to it actually containing three separate source planes, is shown in Figure 6.1. The ratio of the Einstein radii can be used to measure the matter density Ω_m when assuming a flat ΛCDM cosmology, or allows for a constraint on the dark energy equation of state parameter w when assuming a wCDM cosmology (see Section 6.2 for more details).

For example, the first and most well-studied DSPL is the aforementioned 'Jackpot' lens J0946+1006 initially discovered by Gavazzi et al. (2008). This was later used by Collett and Auger (2014) to derive $\Omega_\mathrm{m} = 0.33^{+0.33}_{-0.26}$ when assuming a flat ΛCDM cosmology, and for wCDM gave w = -1.17±0.20 based solely on this system combined with a prior from CMB measurements. Collett and Smith (2020) then reported the discovery of this system's third source, making it the only triple-source-plane lens to date, and Smith and Collett (2021) combined measurements

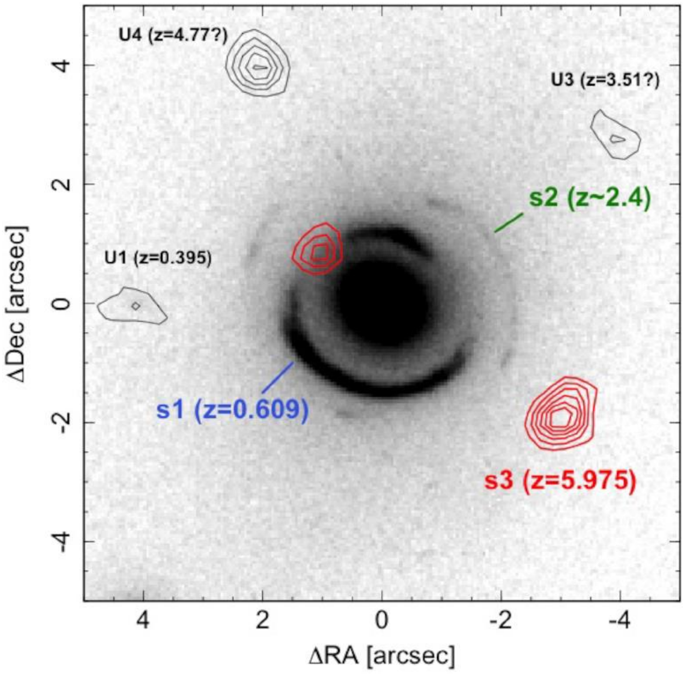

Figure 6.1: HST F814W image of the 'Jackpot' lens J0946+1006 from SLACS, indicating the discovery of a third source and unconfirmed other sources of emission from the Multi-Unit Spectroscopic Explorer (MUSE) survey (Bacon et al., 2010). (Figure credit: Collett and Smith, 2020)

with updated CMB priors and spectroscopic redshifts to obtain w = -1.04±0.20, which, along with their constraints on $\Omega_{\rm m}$, was consistent with flat ΛCDM. While a few other DSPLs have been observed (Tu et al., 2009; Tanaka et al., 2016; Schuldt et al., 2019), they lack the optimal combinations of redshifts possessed by the Jackpot lens and have not been used to constrain cosmological parameters. However, only several DSPLs may be needed to achieve competitive precision on cosmological parameters (Collett et al., 2012), with a sample of 50 expected to constrain $\Omega_{\rm m}$ and w to within 10 per cent (Gavazzi et al., 2008). With Gavazzi et al. (2008) suggesting that one in 40-80 strong lenses are DSPLs, this would imply approximately 1250-2500 DSPLs being observed in future surveys, with 300 of the highest quality from the best seeing exposures (Linder, 2016).

It is worth noting that all of these applications of strong lens modelling benefit greatly from a statistical analysis of many lens systems in order to reduce statistical uncertainty and obtain more accurate results. With future surveys like *Euclid* and

Rubin LSST, such an analysis will soon be achievable, provided sufficiently fast and accurate automated methods are employed such as the neural networks in this thesis. As these applications are performed in a statistical way, a modelling method's performance depends more on its accuracy than its precision. In Chapter 4, we have seen how the modelling methods in this work all share similar biases in general, with magnitudes increasing from small values for smooth parametric profiles (such as 0.1 per cent for Einstein radius) up to significant biases for EAGLE galaxy lenses containing substructure and LOS structure (such as 2.6 per cent and 10.1 per cent for Einstein radius and axis ratio). Hence, the level of biasing depends on the data set, although this effect should be reduced for the methods involving the CNN should the CNN be trained on such complex lenses.

6.2 Constraining Cosmological Parameters with Double Source Plane Lensing

In this section, I will investigate how the CNN's performance in Chapter 4 transfers to providing tighter constraints on cosmological parameters when applied to double source plane lenses (see Section 6.1.5), under the assumption that it models such systems with the same accuracy, precision and speed. This begins with an overview of the theory behind this technique in Section 6.2.1, followed by the analysis of the modelling methods' applicability in Section 6.2.2.

6.2.1 Theory

The expansion rate of the Universe is described by the Hubble parameter $H(z)$, itself a function of redshift z and cosmological parameters, with its normalised form given as

$$E(z) \equiv H(z)/H_0 = \sqrt{\Omega_{\rm m}(1+z)^3 + \Omega_{\rm k}(1+z)^2 + \Omega_{\rm de} f(z)} \tag{6.1}$$

for matter density $\Omega_{\rm m}$, spatial curvature density $\Omega_{\rm k}$, and dark energy density $\Omega_{\rm de}$. The function $f(z)$ depends on the chosen dark energy model: in order to explain the accelerated expansion of the Universe, a negative pressure in the form of dark energy is required. While ΛCDM models this as a cosmological constant ($\Omega_{\rm de} f(z) = \Omega_\Lambda$), other cosmologies exist such as wCDM where dark energy is modelled as a fluid or scalar field with pressure p and energy density ρ. This leads to the dark energy equation of state parameter $w = p/\rho$ that can be treated as a constant or as a function of time. A constant w produces

$$f(z) = (1+z)^{3(1+w)} \tag{6.2}$$

while for a time-varying w this becomes

$$\ln f(z) = 3 \int_0^z \frac{1+w(z')}{(1+z')} dz', \tag{6.3}$$

where the dark energy equation of state parameter is parameterised as

$$w(z) = w_0 + w_a(1 - a(z)) = w_0 + w_a \left(\frac{z}{1+z} \right) \tag{6.4}$$

for its value today w_0 and time variation w_a, with scale factor $a(z) = 1/(1+z)$ (Chevallier and Polarski, 2001; Linder, 2003; Collett et al., 2012; Tripathi et al., 2016). While $w < -1/3$ is required for accelerated expansion, a value of around -1 is typically observed (e.g. Aghanim et al., 2020) that appears to be consistent with the cosmological constant of ΛCDM. Results from Collett and Auger (2014) also greatly suggest a flat Universe ($\Omega_\Lambda = 0$, with $\Omega_{\rm m} + \Omega_{\rm de} = 1$), in which case $E(z)$ is purely a function of redshift and $\Omega_{\rm m}$.

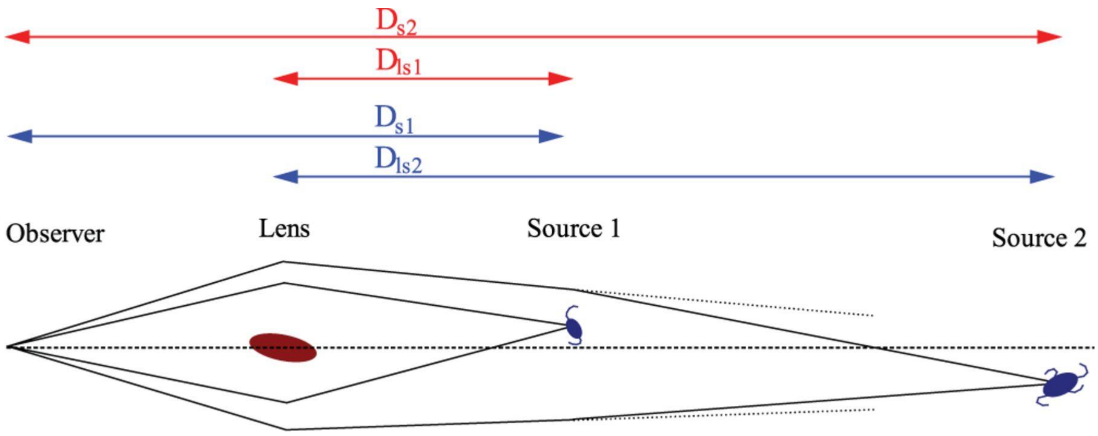

Figure 6.2: Ray diagram for a double source plane lens, in which rays from two separate sources at different redshifts are distorted by the same lens to produce two sets of lensed arcs or rings. (Figure credit: Collett et al., 2012)

From equation 1.12 in Section 1.2.3 we see that the Einstein radius of a strong gravitational lens depends on the projected mass contained within that radius and the angular diameter distances of the system. In the limit of a singular isothermal profile, this becomes Equation 1.23 which depends on the constant velocity dispersion of the lens and the ratio of the angular diameter distances between the lens and source $D_{LS}(z, z_s)$ and between the observer and the source $D_S(z_s)$, themselves functions of the lens and source redshifts, z and z_s respectively. In the limit of a flat universe, the angular diameter distance from point i to point j is defined as

$$D_{i,j} = \frac{c/H_0}{(1+z_j)} \int_{z_i}^{z_j} \frac{dz}{E(z)} \tag{6.5}$$

and is hence a function of the cosmological parameters H_0, $\Omega_{\rm m}$, and $w(z)$.

While the majority of strong galaxy-scale lenses consist of a lens and a single source, some systems contain a second source at a different redshift to the first, with the resulting double source plane lens system producing two sets of lensed arcs or rings as shown schematically in Figure 6.2. As the velocity dispersion of a singular isothermal profile is independent of radius, the ratio of these Einstein radii becomes

$$\beta(z, z_1, z_2) = \frac{\theta_{E,1}}{\theta_{E,2}} = \frac{D_{LS}(z, z_1)}{D_S(z_1)} \frac{D_S(z_2)}{D_{LS}(z, z_2)} \tag{6.6}$$

where z is the redshift of the lens, and z_1 and z_2 are the redshifts of the two sources (Linder, 2004; Collett et al., 2012; Linder, 2016; Sharma and Linder, 2022). While these angular diameter distances depend upon the value of the Hubble constant, within a flat wCDM cosmology the H_0 terms within β cancel to leave the ratio $\beta(z, z_1, z_2)$ independent of H_0 and hence only dependent on z, Ω_m and $w(z)$.

6.2.2 Application to the Modelling Methods

In Chapter 4, I presented errors and biases in the predictions of mass model parameters for four modelling methods: a Bayesian CNN, PyAL (blind), PyAL + CNN, and PyAL + CNN (1σ). For double source plane lenses, we now seek to propagate the errors and biases of the Einstein radius to those of the cosmological parameters, under the assumption that these methods model such systems with the same accuracy, precision and speed, for example through appropriate masking of each lensed source as required.

First, we examine propagating the errors to the ratio of Einstein radii, β. Beginning with the error (precision) in Einstein radius $\delta\theta_E$, simple propagation of errors yields

$$\frac{\delta\beta}{\beta} = \sqrt{\frac{\delta\theta_{E,1}}{\theta_{E,1}}^2 + \frac{\delta\theta_{E,2}}{\theta_{E,2}}^2}. \tag{6.7}$$

From Figure 2.10 in Section 2.2.1, we see that for *Euclid* images with lens light subtracted (dashed blue line), fractional errors in Einstein radii vary little in magnitude for true Einstein radii beyond 1 arcsec. As the majority of the lenses used in the test sets of Chapter 4 lie in this region, and because larger Einstein radii would generally provide the best seeing exposures and hence the highest quality measurements, for the analysis performed here I take the fractional errors in Einstein radii to be constant with respect to the Einstein radii values themselves

Table 6.1: The errors (68 per cent confidence intervals) on predicted Einstein radii, presented as percentage errors, for each modelling method and test set used in Chapter 4, computed from the distributions of differences between true and predicted parameter values across each set of 1000 test images. The first test set contained parametric source profiles while the rest contained HUDF galaxies as sources. Mass profiles used were SIE, elliptical power law (EPL), or EAGLE galaxies, with some containing line-of-sight structure (LOSS).

Test Set	CNN	PyAL (blind)	PyAL + CNN	PyAL + CNN (1σ)
SIE (parametric)	0.2%	0.2%	0.2%	0.2%
SIE	0.8%	0.8%	0.7%	0.5%
SIE + LOSS	1.2%	1.7%	1.5%	1.3%
EPL	3.1%	3.8%	3.0%	1.8%
EPL + LOSS	4.0%	5.1%	4.0%	3.4%
EAGLE	4.5%	7.9%	5.0%	4.4%
EAGLE + LOSS	5.3%	7.6%	6.3%	4.9%

($\delta\theta_E/\theta_E = constant$). Hence, the fractional error in β becomes

$$\frac{\delta\beta}{\beta} = \sqrt{\frac{\delta\theta_E}{\theta_E}^2 + \frac{\delta\theta_E}{\theta_E}^2} = \sqrt{2}\,\frac{\delta\theta_E}{\theta_E}. \tag{6.8}$$

Of course, the fractional error still depends on the modelling method and data set, with percentage errors (i.e. $100 \times \delta\theta_E/\theta_E$) for these shown in Table 6.1.

It should be noted that if one instead allows the fractional error to vary with the value of the Einstein radius, with one fractional error X times larger than the other (i.e. $(\delta\theta_{E,2}/\theta_{E,2}) = X(\delta\theta_{E,1}/\theta_{E,1})$), Equation 6.8 instead becomes $(\delta\beta/\beta) = \sqrt{X^2+1}\,(\delta\theta_{E,1}/\theta_{E,1})$. From Figure 2.4 in Section 2.1.2, the distribution of Einstein radii for *Euclid* lenses peaked around a value of 1 arcsec with a standard deviation of roughly 0.5 arcsec. Hence, it is safe to assume that one of the two sources in a DSPL system would have $\theta_E \sim 1$ arcsec. As discussed, a second source with $\theta_E \sim 1.5$ arcsec would have a similar fractional error, but if the second source instead had $\theta_E \sim 0.5$ arcsec then Figure 2.10 shows that this would increase its fractional error by a factor of approximately 1.5, resulting in $\delta\beta/\beta = \sqrt{1.5^2+1}\;\delta\theta_E/\theta_E = \sqrt{3.25}\;\delta\theta_E/\theta_E$. However, for the analysis here I will keep to Equation 6.8 for the reasons already discussed.

Having propagated precision errors to β, I now turn to propagating the biases in predicted values of Einstein radii. For predicted values $\theta_{E,p}$, true values θ_E, and fractional biases $B(\theta_E) \equiv (\theta_{E,p} - \theta_E)/\theta_E$, we have

$$\beta_{\text{measured}} = \frac{\theta_{E,1,p}}{\theta_{E,2,p}} = \frac{\theta_{E,1} + B_1(\theta_{E,1})\theta_{E,1}}{\theta_{E,2} + B_2(\theta_{E,2})\theta_{E,2}} = \frac{1 + B_1(\theta_{E,1})}{1 + B_2(\theta_{E,2})}\beta_{\text{true}} \tag{6.9}$$

where $\beta_{\text{true}} = (\theta_{E,1}/\theta_{E,2})$. As an example, if the biases were 10 per cent and 1 per cent such that $B_1(\theta_{E,1}) = 0.1$ and $B_2(\theta_{E,2}) = 0.01$, this would result in $\beta_{\text{measured}} = (1.1/1.01)\beta_{\text{true}} \simeq 1.09\beta_{\text{true}}$, hence resulting in a 9 per cent bias on the ratio. This would then need propagating through to the cosmological parameters computationally for a given set of redshifts, marginalising over each parameter in turn to determine that parameter's bias. However, for the biases presented in Chapter 4, I found negligible variation in their percentage values with increasing Einstein radius regardless of the test set or modelling method. As such, while the two Einstein radii of a DSPL system may differ in absolute bias, their fractional biases will be the same on average. In this case, following Equation 6.9 but now for the ratio of the Einstein radii where each has the same fractional bias B,

$$\beta_{\text{measured}} = \frac{1 + B_1(\theta_{E,1})}{1 + B_2(\theta_{E,2})}\beta_{\text{true}} = \frac{1 + B}{1 + B}\beta_{\text{true}} = \beta_{\text{true}} \tag{6.10}$$

and so the biases will cancel due to the use of such a ratio. Hence, biases in β only emerge if $B(\theta_E)$ changes as a function of θ_E, and as this was not the case for the results in Chapter 4, no propagation of biases was required for the work presented here.

As discussed in Section 6.1.5, approximately 1250-2500 DSPL systems are expected to be observed by future surveys, with around 300 of the highest quality. Fast, automated modelling through the use of the methods presented in this thesis would allow for a rapid determination of β for each system and hence tighter constraints on Ω_{m} and $w(z)$ than from a single system alone. While such a combi-

Table 6.2: The expected percentage precision errors for β (see Equation 6.8), for each modelling method applied to the SIE test set containing HUDF sources (without LOS structure), and for the worst cases (EAGLE + LOSS test set, except for PyAL (blind) whose worst precision was for the EAGLE test set without LOS structure). Errors are given for a single DSPL, as well as for combinations of 50 or 300 of such systems (N_{DSPL} = 1, 50, or 300, respectively).

	SIE			**Worst Case**		
Method \ N_{DSPL}	1	50	300	1	50	300
CNN	1.13%	0.16%	0.07%	7.50%	1.06%	0.43%
PyAL (blind)	1.13%	0.16%	0.07%	11.17%	1.58%	0.65%
PyAL + CNN	0.99%	0.14%	0.06%	8.91%	1.26%	0.51%
PyAL + CNN (1σ)	0.71%	0.10%	0.04%	6.93%	0.98%	0.40%

nation of measurements would not improve their accuracy, it is expected that the precision would be improved by a factor of $\sqrt{N}$ for N modelled DSPL systems.

For example, the errors for the first test set in Table 6.1 containing SIE mass profiles and parametric sources would result in a precision for β of 0.28 per cent for a single lens, regardless of the modelling method. This would then be expected to decrease to 0.04 per cent and 0.02 per cent when combining 50 or 300 DSPL systems, respectively. Such results are for the simplest test set however: those for the more realistic second SIE test set and the worst case precision errors are given in Table 6.2. As can be seen, for the SIE test set, a single lens would give β a precision of around 1 per cent, improving to less than 0.1 per cent for 300 DSPLs. In the worst cases, while the precision errors are large for a single lens, even these decrease to 0.65 per cent or less for 300 DSPLs.

We now seek to propagate these errors to those of the cosmological parameters $\Omega_{\rm m}$ and $w(z)$, assuming negligible error on redshift measurements. Due to degeneracies between the parameters, this would require propagating the errors computationally and marginalising over each parameter in turn. Fortunately, recent work by Sharma and Linder (2022) has explored the sensitivity of β to these parameters out to high redshifts, which will be used as a baseline here and is discussed below.

In Sharma and Linder (2022), they define the sensitivity of β to the parameters

$\theta = \{\Omega_\mathrm{m}, w_0, w_a\}$ through partial derivatives, as a fractional change relative to a given absolute measurement precision $\sigma(\beta)$, i.e. as $(\partial\beta/\partial\theta)/\sigma(\beta)$. The magnitude of this sensitivity is hence equal to the reciprocal of the absolute precision on a cosmological parameter $\sigma(\theta)$:

$$\sigma(\theta) = \left|\frac{\partial\beta}{\partial\theta}\right|^{-1}\sigma(\beta) = \left|\frac{(\partial\beta/\partial\theta)}{\sigma(\beta)}\right|^{-1}. \tag{6.11}$$

As we have so far used relative errors, the above can be re-written to obtain $\sigma(\theta)$ from the fractional error in β instead, $\sigma_f(\beta) \equiv \sigma(\beta)/\beta$, using $\partial \ln\beta = \partial\beta/\beta$ from propagation of errors:

$$\sigma(\theta) = \left|\frac{(\partial\ln\beta/\partial\theta)\beta}{\sigma(\beta)}\right|^{-1} = \left|\frac{(\partial\ln\beta/\partial\theta)}{\sigma_f(\beta)}\right|^{-1}. \tag{6.12}$$

This re-written version of the sensitivity of β was then determined by Sharma and Linder (2022) for each cosmological parameter as a function of redshift, shown here in Figure 6.3, with values scaled according to a measurement precision in β of 1 per cent (i.e. $\sigma_f(\beta) = 0.01$). As discussed, the reciprocal of the magnitude of the y-axis values gives the absolute precision for a cosmological parameter, and the values were determined assuming a ΛCDM cosmology with fiducial values $\Omega_\mathrm{m} = 0.3$, $w_0 = -1$, and $w_a = 0$, as well as $z_1/z = 2$ and $z_2/z_1 = 1.5$ (following Equation 6.6). These redshift ratios were chosen by Sharma and Linder (2022) motivated by the work in that paper and in Linder (2016), which argued that the ratios should correspond to the most commonly detected DSPL systems. Figure 6.3 shows that the greatest sensitivity magnitude, and hence smallest precision error, is generally found for Ω_m, followed by w_0 and finally w_a. As an example, a 1 per cent measurement precision of β produces a sensitivity to Ω_m of roughly -5.0 at $z = 1.5$, hence giving an absolute precision in Ω_m of $\sigma(\Omega_\mathrm{m}) = 1/5 = 0.2$.

Using the values from Table 6.2 as values of $\sigma_f(\beta)$, I now use Figure 6.3 to obtain the expected absolute precision errors for the cosmological parameters. For

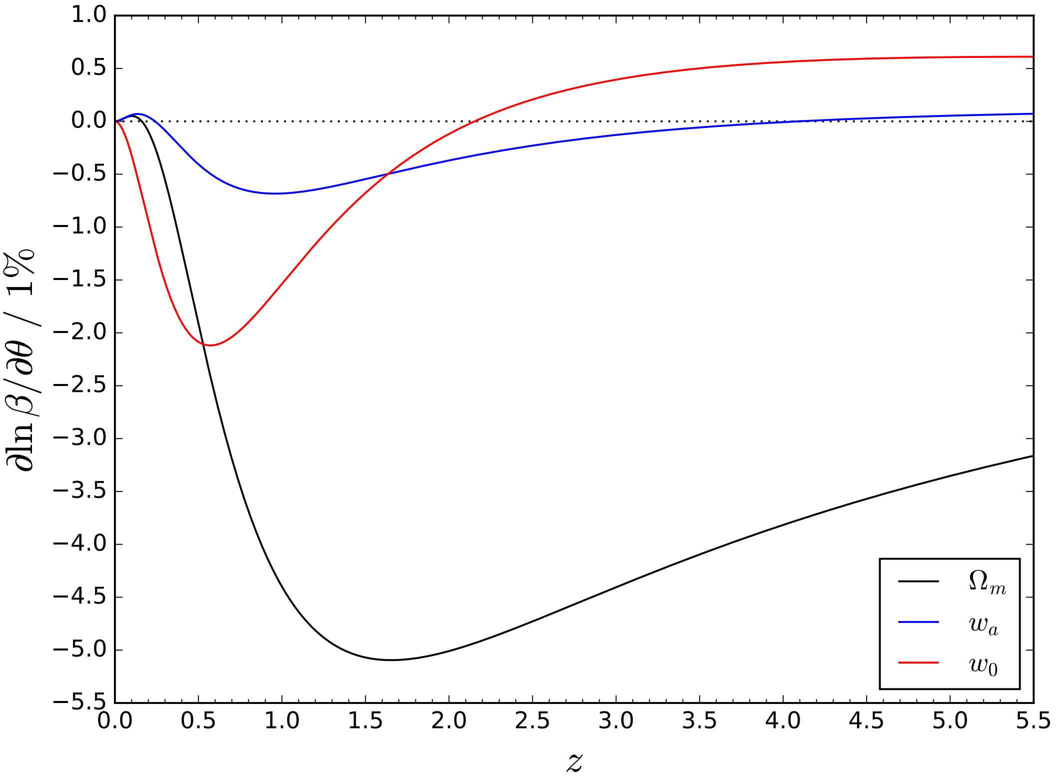

Figure 6.3: Sensitivity of the DSPL ratio β to cosmological parameters θ as functions of redshift. These parameters are Ω_{m}, w_a, and w_0 for a flat wCDM cosmology in which w is allowed to vary with redshift following Equation 6.4. Here, the sensitivity assumes a 1 per cent measurement of β. (Figure credit: Sharma and Linder, 2022)

this analysis, I use the maximum magnitudes of the sensitivity for each parameter (i.e. the troughs in the figure) and assume negligible errors on redshift measurements, indicating the optimal possible precision errors. For Ω_{m}, w_a, and w_0, these sensitivities are approximately -5.1, -2.1, and -0.7, respectively, corresponding to absolute precision errors of $\sigma(\Omega_{\mathrm{m}}) = 0.196$, $\sigma(w_0) = 0.476$, and $\sigma(w_a) = 1.429$ for a 1 per cent measurement precision in β. Scaling these by the values in Table 6.2, the resulting absolute precision errors for the cosmological parameters are given in Table 6.3.

As can be seen, given that the magnitude of each of the three parameters is typically of order unity or less, all modelling methods produce large precision errors when modelling a single DSPL system, even for the simpler SIE test set.

Table 6.3: The expected absolute precision errors for the cosmological parameters Ω_{m}, w_a, and w_0, for each modelling method applied to the SIE test set containing HUDF sources (without LOS structure), and for the worst cases (EAGLE + LOSS test set, except for PyAL (blind) whose worst precision was for the EAGLE test set without LOS structure). Errors are given for a single DSPL, as well as for combinations of 50 or 300 of such systems (N_{DSPL} = 1, 50, or 300, respectively). SIE results for Ω_{m} with N_{DSPL}=300 are given to three decimal places to better highlight their relative accuracies.

	SIE			**Worst Case**		
Method \ N_{DSPL}	1	50	300	1	50	300
Ω_{m}						
CNN	0.22	0.03	0.013	1.47	0.21	0.08
PyAL (blind)	0.22	0.03	0.013	2.19	0.31	0.13
PyAL + CNN	0.19	0.03	0.011	1.75	0.25	0.10
PyAL + CNN (1σ)	0.14	0.02	0.008	1.36	0.19	0.08
w_0						
CNN	0.54	0.08	0.03	3.57	0.50	0.20
PyAL (blind)	0.54	0.08	0.03	5.32	0.75	0.31
PyAL + CNN	0.47	0.07	0.03	4.24	0.60	0.24
PyAL + CNN (1σ)	0.34	0.05	0.02	3.30	0.47	0.19
w_a						
CNN	1.61	0.23	0.10	10.72	1.51	0.61
PyAL (blind)	1.61	0.23	0.10	15.96	2.26	0.93
PyAL + CNN	1.41	0.20	0.09	12.73	1.80	0.73
PyAL + CNN (1σ)	1.01	0.14	0.06	9.90	1.40	0.57

This is especially true for w_a, whose worst case absolute values reach almost 16. Due to the scaling performed to obtain these values, the relative performance of the modelling methods remains the same as their errors in Chapter 4, with PyAL + CNN (1σ) providing the tightest constraints on the cosmological parameters, followed by the CNN. The combination of values from 50 DSPL systems greatly reduces errors, with those of Ω_{m} and w_0 10 per cent or less of their expected values (Ω_{m} = 0.3, w_0 = -1) for the SIE test set, although for the worst cases they remain too large to be of use in constraining the parameters.

For the combined values of 300 DSPL systems, the reduction in precision errors leads to precise constraints on Ω_{m} and w_0 that rival those of the final *Planck* data release ($\Omega_{\mathrm{m}} = 0.315 \pm 0.007$ and $w_0 = -1.03 \pm 0.03$, Aghanim et al., 2020), which for the SIE test set reach values of 4 per cent or less compared to their expected values[1]. Yet, even for so many lenses, the error in w_a is still 2 or more times larger than those of w_0 due to its large degeneracy with the parameters[2]. For the worst cases, it is clear that 300 DSPL systems modelled by these methods would not be enough to reliably constrain the dark energy equation of state parameters, and even the best modelling methods would not be able to constrain precision errors of Ω_m below 0.08 (approximately 27 per cent of the expected value). However, the likelihood of these 'worst case' lenses being used for such an analysis is low, as they contain both significant LOS structures and substructures that would need to be accounted for, and to obtain β using Equation 6.6 requires the lens to have a singular isothermal profile. Hence, these systems would likely be filtered out of the data set before passing it to the modelling methods.

In terms of comparing the relative performance of the modelling methods, it can be shown that the CNN, PyAL (blind), and PyAL + CNN would respectively require 2.32, 2.33, and 1.51 times as many DSPL systems as PyAL + CNN (1σ) in order to reach the latter's precision for the parameters in the SIE test set. For example, to reach the precision of 300 DSPL systems modelled by PyAL + CNN (1σ) would require the other methods to model 695, 698, and 452 systems, respectively. For the worst cases, they would instead require 1.19, 2.43 and 1.70 times as many DSPL systems, equivalent to 357, 729, and 511 systems following the above example. From Figure 4.16 in Section 4.4, the mean average modelling times for the SIE test set are 31 minutes, 23 minutes, and 20 minutes for PyAL (blind), PyAL + CNN, and PyAL + CNN (1σ), respectively. Hence modelling of 300 DSPL systems would be expected to take on average 155 hours, 115 hours, and 100 hours, respectively, with all performed as automated processes without the need for human oversight. Applications to large numbers of lenses such as

---

[1] Following the numbers of expected DSPL systems discussed in Section 6.1.5, for 1250 combined systems these values would be roughly half of their stated values, or roughly a third for 2500 systems.

[2] This degeneracy is discussed further in Linder (2004), Linder (2016), and Sharma and Linder (2022).

this highlight the importance of seeking fast and automated modelling methods, with PyAL + CNN (1σ) greatly improving the modelling times compared to PyAL (blind), let alone PyAutoLens with human oversight, while the CNN alone would be capable of modelling all 300 DSPL systems almost instantaneously.

6.3 Discussion and Conclusions

The values presented in the previous section were propagated from the errors presented in Chapter 4, and so are also subject to the limitations discussed in Chapter 5. In addition to these, there are a number of limitations to consider regarding the use of DSPL modelling for constraining cosmological parameters. These limitations are discussed below along with suggestions for mitigating them by adjusting the observing strategy of future surveys, followed by the conclusions of this chapter.

For the analysis here, I assumed that all of the modelled DSPL systems were at a redshift corresponding to the maximum magnitude sensitivities of β to each cosmological parameter, and hence the smallest precision errors. This is of course an idealised assumption: a more realistic approach would require sampling redshifts from an expected distribution of DSPL systems and read off the sensitivities of each parameter from Figure 6.3 for each one, which is beyond the scope of this work. This would most likely increase the precision errors presented in Table 6.3, although with 1250-2500 expected DSPL systems it may be possible to alleviate this somewhat by selecting the 300 systems that would achieve the optimal sensitivities. Modelling DSPL systems for the range of lens redshifts expected by future surveys would also be of benefit, as this would aid in breaking the degeneracy between parameters (see Sharma and Linder, 2022). Additionally, while other cosmological probes such as CMB measurements are more sensitive to the matter density than the dark energy equation of state parameters, the reverse is true for the DSPL approach at low redshifts (Sharma and Linder, 2022). There is therefore a complimentarity between DSPLs and other probes for use in combining results, with Collett and Auger (2014) incorporating CMB priors in their inference of cosmological parameters using the Jackpot lens, improving upon the CMB precision by 30 per cent.

For this work, I also assumed accurate redshifts for the lens and sources, whereas initially only photometric redshifts would be rapidly available from future surveys. Early inference would hence have poorer and potentially biased constraints on the cosmological parameters: Smith and Collett (2021) found that replacing the photometric redshift of the second source in the Jackpot system with a spectroscopic redshift modified their predictions of a constant w from -1.17 ± 0.20 to -1.04 ± 0.20. However, the promising results shown in this work and others enables a suitable science justification for rapidly obtaining spectroscopic redshifts in future surveys for each DSPL system in order to increase the reliability of inferred constraints.

For constraining parameters via DSPL systems, the foreground lens is modelled as having a singular isothermal mass profile. In reality this may not always be the case, as systems can contain lens substructure or LOS structure, with the latter estimated by Collett and Auger (2014) to result in a roughly 1 per cent systematic uncertainty on β and a 25 per cent larger precision error in w. However, as mentioned, such complex systems could be filtered out of the pipeline, at least before passing them to cosmological parameter inference, perhaps using automated approaches such as another CNN: this is left for future work. Further complications arise from needing to account for any potential lensing of the further source by the nearer one, also left to future work, and overcoming the MST degeneracy (mentioned in Section 6.1.4) through incorporating dynamical information such as the lens velocity dispersion, although for a true mass sheet the MST degeneracy is largely broken given the use of the ratio β rather than individual Einstein radii

(Collett and Auger, 2014).

The results in this chapter have shown how an optimal combination of hundreds of double source plane lenses have the potential to place tight constraints on the cosmological parameters $\Omega_{\rm m}$, w_0 and w_a, that could rival those of other cosmological probes while remaining independent of H_0 (and hence avoid the so-called Hubble tension problem). This technique is made possible through the large samples of lenses expected by future surveys like Rubin LSST and *Euclid*, as well as the fast automated methods required to reliably model these systems, such as through the use of a Bayesian CNN or its combination with conventional parametric parameter fitting methods. While the methodology behind the DSPL technique has its limitations, these may be mitigated through a suitable observing strategy making use of the suggestions presented in this section:

1. As part of lens detection programs, there should be a focus on detecting DSPL systems and automatically classifying them according to their suitability for this technique. This classification would be based on the need for such lenses to contain little substructure or LOS structure to ensure their mass profiles approximate a singular isothermal profile, as well as seeking to select those at suitable lens redshifts to optimise the sensitivity, including sampling a range of redshifts to aid in breaking parameter degeneracies.
2. Rapid acquisition of spectroscopic redshifts should be ensured for these systems, which may also be used to inform the above suitability as well as aid in velocity dispersion measurements to break any MST degeneracies.
3. The sample of hundreds of systems should be modelled using fast automated techniques like the ones presented in this thesis. The modelling results of all the systems can then be combined together, along with priors from other probes (such as the latest *Planck* results), to tightly constrain cosmological parameters.

It is worth remembering that while the focus of this chapter has been on double source plane lenses, there are also a number of other applications requiring large samples of lenses, as discussed in Section 6.1. Hence, through strategies like those above, automated strong galaxy-scale gravitational lens modelling methods involving the use of neural networks promise to deeply enhance our knowledge of the Universe over the coming decade.

Bibliography

Abell, P. A., Burke, D. L., Hamuy, M., Nordby, M., Axelrod, T. S., Monet, D., Vrsnak, B., Thorman, P., Ballantyne, D., Simon, J. D., et al. (2009). LSST science book, version 2.0. Technical Report (No. SLAC-R-1031), (LSST Science Collaboration).

Ade, P. A., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., Barreiro, R., et al. (2014). Planck 2013 results. XVI. Cosmological parameters. *Astronomy & Astrophysics*, 571:A16.

Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A., Barreiro, R., Bartolo, N., Basak, S., et al. (2020). Planck 2018 results VI. Cosmological parameters. *Astronomy & Astrophysics*, 641:A6.

Alexander, S., Gleyzer, S., McDonough, E., Toomey, M. W., and Usai, E. (2020). Deep Learning the Morphology of Dark Matter Substructure. *The Astrophysical Journal*, 893(1):15.

Alexander, S., Gleyzer, S., Reddy, P., Tidball, M., and Toomey, M. W. (2021). Domain Adaptation for Simulation-Based Dark Matter Searches Using Strong Gravitational Lensing. *arXiv preprint arXiv:2112.12121.*

Allen, S., Schmidt, R., and Fabian, A. (2002). Cosmological constraints from the X-ray gas mass fraction in relaxed lensing clusters observed with Chandra. *Monthly Notices of the Royal Astronomical Society*, 334(2):L11–L15.

Amorisco, N. C., Nightingale, J., He, Q., Amvrosiadis, A., Cao, X., Cole, S., Etherington, A., Frenk, C. S., Li, R., Massey, R., et al. (2022). Halo concentration strengthens dark matter constraints in galaxy–galaxy strong lensing analyses. *Monthly Notices of the Royal Astronomical Society*, 510(2):2464–2479.

Amvrosiadis, A., Eales, S., Negrello, M., Marchetti, L., Smith, M., Bourne, N., Clements, D., De Zotti, G., Dunne, L., Dye, S., et al. (2018). ALMA observations of lensed Herschel sources: testing the dark matter halo paradigm. *Monthly Notices of the Royal Astronomical Society*, 475(4):4939–4952.

Asboth, V., Conley, A., Sayers, J., Béthermin, M., Chapman, S., Clements, D., Cooray, A., Dannerbauer, H., Farrah, D., Glenn, J., et al. (2016). HerMES: a search for high-redshift dusty galaxies in the HerMES Large Mode Survey–catalogue, number counts and early results. *Monthly Notices of the Royal Astronomical Society*, 462(2):1989–2000.

Auger, M., Treu, T., Bolton, A., Gavazzi, R., Koopmans, L., Marshall, P., Moustakas, L., and Burles, S. (2010a). The Sloan Lens ACS Survey. X. Stellar, dynamical, and total mass correlations of massive early-type galaxies. *The Astrophysical Journal*, 724(1):511.

Auger, M., Treu, T., Gavazzi, R., Bolton, A., Koopmans, L., and Marshall, P. (2010b). Dark Matter Contraction and the Stellar Content of Massive Early-type Galaxies: Disfavoring "Light" Initial Mass Functions. *The Astrophysical Journal Letters*, 721(2):L163.

Avestruz, C., Li, N., Lightman, M., Collett, T. E., and Luo, W. (2017). Automated Lensing Learner-I: An Automated Strong Lensing Identification Pipeline. *arXiv preprint arXiv:1704.02322*.

Bacon, R., McLean, I., Ramsay, S., and Takami, H. (2010). Proc. SPIE Conf. Ser. Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III. SPIE, Bellingham.

Bag, S., Kim, A. G., Linder, E. V., and Shafieloo, A. (2021). Be It Unresolved: Measuring Time Delays from Lensed Supernovae. *The Astrophysical Journal*, 910(1):65.

Bagla, J. S. (2005). Cosmological N-body simulation: Techniques, scope and status. *Current science*, pages 1088–1100.

Barnabè, M., Czoske, O., Koopmans, L. V., Treu, T., and Bolton, A. S. (2011). Two-dimensional kinematics of SLACS lenses–III. Mass structure and dynamics of early-type lens galaxies beyond $z \simeq 0.1$. *Monthly Notices of the Royal Astronomical Society*, 415(3):2215–2232.

Baron, D. and Poznanski, D. (2016). The weirdest SDSS galaxies: results from an outlier detection algorithm. *Monthly Notices of the Royal Astronomical Society*, 465(4):4530–4555.

Bassett, B. and Hlozek, R. (2010). Baryon acoustic oscillations. *Dark energy: observational and theoretical approaches*, page 246.

Baudis, L. (2018). The search for dark matter. *European Review*, 26(1):70–81.

Bayer, D., Chatterjee, S., Koopmans, L., Vegetti, S., McKean, J., Treu, T., and Fassnacht, C. (2018). Observational constraints on the sub-galactic matter-power spectrum from galaxy-galaxy strong gravitational lensing. *arXiv preprint arXiv:1803.05952*.

Bayer, J., Huber, S., Vogl, C., Suyu, S., Taubenberger, S., Sluse, D., Chan, J., and Kerzendorf, W. (2021). HOLISMOKES–V. Microlensing of type II supernovae and time-delay inference through spectroscopic phase retrieval. *arXiv preprint arXiv:2101.06229*.

Beckwith, S. V., Stiavelli, M., Koekemoer, A. M., Caldwell, J. A., Ferguson, H. C., Hook, R., Lucas, R. A., Bergeron, L. E., Corbin, M., Jogee, S., et al. (2006). The Hubble ultra deep field. *The Astronomical Journal*, 132(5):1729.

Belgacem, E., Calcagni, G., Crisostomi, M., Dalang, C., Dirian, Y., Ezquiaga, J. M., Fasiello, M., Foffa, S., Ganz, A., García-Bellido, J., et al. (2019). Testing modified gravity at cosmological distances with LISA standard sirens. *Journal of Cosmology and Astroparticle Physics*, 2019(07):024.

Bengaly, C. A., Clarkson, C., and Maartens, R. (2020). The Hubble constant tension with next-generation galaxy surveys. *Journal of Cosmology and Astroparticle Physics*, 2020(05):053.

Bertone, G., Hooper, D., and Silk, J. (2005). Particle dark matter: Evidence, candidates and constraints. *Physics reports*, 405(5-6):279–390.

Bertone, G. and Tait, T. M. (2018). A new era in the quest for dark matter. *arXiv preprint arXiv:1810.01668*.

Bignone, L. A., Pedrosa, S. E., Trayford, J. W., Tissera, P. B., and Pellizza, L. J. (2020). Non-parametric morphologies of galaxies in the EAGLE simulation. *Monthly Notices of the Royal Astronomical Society*, 491(3):3624–3642.

Birrer, S. and Amara, A. (2018). Lenstronomy: Multi-purpose gravitational lens modelling software package. *Physics of the Dark Universe*, 22:189–201.

Birrer, S., Amara, A., and Refregier, A. (2015). Gravitational lens modeling with basis sets. *The Astrophysical Journal*, 813(2):102.

Birrer, S., Shajib, A., Galan, A., Millon, M., Treu, T., Agnello, A., Auger, M., Chen, G.-F., Christensen, L., Collett, T., et al. (2020). TDCOSMO-IV. Hierarchical time-delay cosmography–joint inference of the Hubble constant and galaxy density profiles. *Astronomy & Astrophysics*, 643:A165.

Birrer, S. and Treu, T. (2020). TDCOSMO V: strategies for precise and accurate measurements of the Hubble constant with strong lensing. *arXiv preprint arXiv:2008.06157*.

Birrer, S., Treu, T., Rusu, C., Bonvin, V., Fassnacht, C., Chan, J., Agnello, A., Shajib, A., Chen, G. C., Auger, M., et al. (2019). H0LiCOW–IX. Cosmographic analysis of the doubly imaged quasar SDSS 1206+ 4332 and a new measurement of the Hubble constant. *Monthly Notices of the Royal Astronomical Society*, 484(4):4726–4753.

Blake, C., James, J. B., and Poole, G. B. (2014). Using the topology of large-scale structure in the WiggleZ Dark Energy Survey as a cosmological standard ruler. *Monthly Notices of the Royal Astronomical Society*, 437(3):2488–2506.

Blumenthal, G. R., Faber, S., Primack, J. R., and Rees, M. J. (1984). Formation of galaxies and large-scale structure with cold dark matter. *Nature*, 311(5986):517–525.

Bolton, A. S., Brownstein, J. R., Kochanek, C. S., Shu, Y., Schlegel, D. J., Eisenstein, D. J., Wake, D. A., Connolly, N., Maraston, C., Arneson, R. A., et al. (2012). The BOSS emission-line lens survey. II. Investigating mass-density profile evolution in the SLACS+ BELLS strong gravitational lens sample. *The Astrophysical Journal*, 757(1):82.

Bolton, A. S., Burles, S., Koopmans, L. V., Treu, T., Gavazzi, R., Moustakas, L. A., Wayth, R., and Schlegel, D. J. (2008a). The sloan lens ACS survey. V. The full ACS strong-lens sample. *The Astrophysical Journal*, 682(2):964.

Bolton, A. S., Burles, S., Koopmans, L. V., Treu, T., and Moustakas, L. A. (2006). The Sloan Lens ACS Survey. I. A Large Spectroscopically Selected Sample of Massive Early-Type Lens Galaxies. *The Astrophysical Journal*, 638(2):703.

Bolton, A. S., Treu, T., Koopmans, L. V., Gavazzi, R., Moustakas, L. A., Burles, S., Schlegel, D. J., and Wayth, R. (2008b). The Sloan Lens ACS Survey. VII. Elliptical galaxy scaling laws from direct observational mass measurements. *The Astrophysical Journal*, 684(1):248.

Bom, C., Makler, M., Albuquerque, M., and Brandt, C. (2017). A neural network gravitational arc finder based on the Mediatrix filamentation method. *Astronomy & Astrophysics*, 597:A135.

Bom, C., Poh, J., Nord, B., Blanco-Valentin, M., and Dias, L. (2019). Deep Learning in Wide-field Surveys: Fast Analysis of Strong Lenses in Ground-based Cosmic Experiments. *arXiv preprint arXiv:1911.06341*.

Bonvin, V., Courbin, F., Suyu, S., Marshall, P., Rusu, C., Sluse, D., Tewes, M., Wong, K., Collett, T., Fassnacht, C., et al. (2017). H0LiCOW–V. New COSMOGRAIL time delays of HE 0435- 1223: H0 to 3.8 per cent precision from strong lensing in a flat ΛCDM model. *Monthly Notices of the Royal Astronomical Society*, 465(4):4914–4930.

Boylan-Kolchin, M., Bullock, J. S., and Kaplinghat, M. (2011). Too big to fail? The puzzling darkness of massive Milky Way subhaloes. *Monthly Notices of the Royal Astronomical Society: Letters*, 415(1):L40–L44.

Boylan-Kolchin, M., Bullock, J. S., and Kaplinghat, M. (2012). The Milky Way's bright satellites as an apparent failure of ΛCDM. *Monthly Notices of the Royal Astronomical Society*, 422(2):1203–1218.

Brehmer, J., Mishra-Sharma, S., Hermans, J., Louppe, G., and Cranmer, K. (2019). Mining for Dark Matter Substructure: Inferring subhalo population properties from strong lenses with machine learning. *The Astrophysical Journal*, 886(1):49.

Brownstein, J. R., Bolton, A. S., Schlegel, D. J., Eisenstein, D. J., Kochanek, C. S., Connolly, N., Maraston, C., Pandey, P., Seitz, S., Wake, D. A., et al. (2011). The BOSS Emission-Line Lens Survey (BELLS). I. A large spectroscopically selected sample of Lens Galaxies at redshift 0.5. *The Astrophysical Journal*, 744(1):41.

Bruzual, G. and Charlot, S. (2003). Stellar population synthesis at the resolution of 2003. *Monthly Notices of the Royal Astronomical Society*, 344(4):1000–1028.

Buchner, J., Georgakakis, A., Nandra, K., Hsu, L., Rangel, C., Brightman, M., Merloni, A., Salvato, M., Donley, J., and Kocevski, D. (2014). X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue. *Astronomy & Astrophysics*, 564:A125.

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

Cabanac, R. A., Alard, C., Dantel-Fort, M., Fort, B., Gavazzi, R., Gomez, P., Kneib, J., Le Fevre, O., Mellier, Y., Pello, R., et al. (2007). The CFHTLS strong lensing legacy survey-I. Survey overview and T0002 release sample. *Astronomy & Astrophysics*, 461(3):813–821.

Canameras, R., Schuldt, S., Suyu, S., Taubenberger, S., Meinhardt, T., Leal-Taixe, L., Lemon, C., Rojas, K., and Savary, E. (2020). HOLISMOKES-II. Identifying galaxy-scale strong gravitational lenses in Pan-STARRS using convolutional neural networks. *Astronomy & Astrophysics*, 644:A163.

Cao, X., Li, R., Nightingale, J., Massey, R., Robertson, A., Frenk, C. S., Amvrosiadis, A., Amorisco, N. C., He, Q., Etherington, A., et al. (2022). Systematic Errors Induced by the Elliptical Power-law model in Galaxy–Galaxy Strong Lens Modeling. *Research in Astronomy and Astrophysics*, 22(2):025014.

Cappellari, M., Romanowsky, A. J., Brodie, J. P., Forbes, D. A., Strader, J., Foster, C., Kartha, S. S., Pastorello, N., Pota, V., Spitler, L. R., et al. (2015). Small scatter and nearly isothermal mass profiles to four half-light radii from two-dimensional stellar dynamics of early-type galaxies. *The Astrophysical Journal Letters*, 804(1):L21.

Chabrier, G. (2003). Galactic stellar and substellar initial mass function. *Publications of the Astronomical Society of the Pacific*, 115(809):763.

Charnock, T., Perreault-Levasseur, L., and Lanusse, F. (2022). Bayesian neural networks. In *Artificial Intelligence for High Energy Physics*, pages 663–713. World Scientific.

Chen, G. C., Chan, J. H., Bonvin, V., Fassnacht, C. D., Rojas, K., Millon, M., Courbin, F., Suyu, S. H., Wong, K. C., Sluse, D., et al. (2018). Constraining the microlensing effect on time delays with a new time-delay prediction model in H 0 measurements. *Monthly Notices of the Royal Astronomical Society*, 481(1):1115–1125.

Chen, G. C., Fassnacht, C. D., Suyu, S. H., Yıldırım, A., Komatsu, E., and Bernal, J. L. (2021). TDCOSMO-VI. Distance measurements in time-delay cosmography under the mass-sheet transformation. *Astronomy & Astrophysics*, 652:A7.

Chen, Y., Li, R., Shu, Y., and Cao, X. (2019a). Assessing the effect of lens mass model in cosmological application with updated galaxy-scale strong gravitational lensing sample. *Monthly Notices of the Royal Astronomical Society*, 488(3):3745–3758.

Chen, Z., Luo, W., Cai, Y.-F., and Saridakis, E. N. (2019b). New test on general relativity using galaxy–galaxy lensing with astronomical surveys. *arXiv preprint arXiv:1907.12225*.

Cheng, T.-Y., Li, N., Conselice, C. J., Aragón-Salamanca, A., Dye, S., and Metcalf, R. B. (2020). Identifying strong lenses with unsupervised machine learning using convolutional autoencoder. *Monthly Notices of the Royal Astronomical Society*, 494(3):3750–3765.

Chevallier, M. and Polarski, D. (2001). Accelerating universes with scaling dark matter. *International Journal of Modern Physics D*, 10(02):213–223.

Chianese, M. (2019). Differentiable probabilistic programming for strong gravitational lensing. *arXiv preprint arXiv:1910.06617*.

Chianese, M., Coogan, A., Hofma, P., Otten, S., and Weniger, C. (2020). Differentiable strong lensing: uniting gravity and neural nets through differentiable probabilistic programming. *Monthly Notices of the Royal Astronomical Society*, 496(1):381–393.

Chollett, F. (2017). *Deep Learning with Python*. Manning Publications.

Colaço, L., Holanda, R., and Silva, R. (2021). Probing variation of the fine-structure constant in runaway dilaton models using Strong Gravitational Lensing and Type Ia Supernovae. *The European Physical Journal C*, 81(9):1–9.

Collett, T., Montanari, F., and Räsänen, S. (2019). Model-Independent Determination of H 0 and Ω K 0 from Strong Lensing and Type Ia Supernovae. *Physical review letters*, 123(23):231101.

Collett, T. E. (2015). The population of galaxy–galaxy strong lenses in forthcoming optical imaging surveys. *The Astrophysical Journal*, 811(1):20.

Collett, T. E. and Auger, M. W. (2014). Cosmological constraints from the double source plane lens SDSSJ0946+ 1006. *Monthly Notices of the Royal Astronomical Society*, 443(2):969–976.

Collett, T. E., Auger, M. W., Belokurov, V., Marshall, P. J., and Hall, A. C. (2012). Constraining the dark energy equation of state with double-source plane strong lenses. *Monthly Notices of the Royal Astronomical Society*, 424(4):2864–2875.

Collett, T. E., Oldham, L. J., Smith, R. J., Auger, M. W., Westfall, K. B., Bacon, D., Nichol, R. C., Masters, K. L., Koyama, K., and van den Bosch, R. (2018). A precise extragalactic test of General Relativity. *Science*, 360(6395):1342–1346.

Collett, T. E. and Smith, R. J. (2020). A triple rollover: a third multiply imaged source at $z \approx 6$ behind the Jackpot gravitational lens. *Monthly Notices of the Royal Astronomical Society*, 497(2):1654–1660.

Connolly, A. J., Angeli, G. Z., Chandrasekharan, S., Claver, C. F., Cook, K., Ivezic, Z., Jones, R. L., Krughoff, K. S., Peng, E.-H., Peterson, J., Petry, C., Rasmussen, A. P., Ridgway, S. T., Saha, A., Sembroski, G., vanderPlas, J., and Yoachim, P. (2014). An end-to-end simulation framework for the Large Synoptic Survey Telescope. In Angeli, G. Z. and Dierickx, P., editors, *Modeling, Systems Engineering, and Project Management for Astronomy VI*, volume 9150 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 14.

Conselice, C. J. (2014). The evolution of galaxy structure over cosmic time. *Annual Review of Astronomy and Astrophysics*, 52:291–337.

Correa, C. A., Schaye, J., Clauwens, B., Bower, R. G., Crain, R. A., Schaller, M., Theuns, T., and Thob, A. C. (2017). The relation between galaxy morphology and colour in the EAGLE simulation. *Monthly Notices of the Royal Astronomical Society: Letters*, 472(1):L45–L49.

Crain, R. A., Schaye, J., Bower, R. G., Furlong, M., Schaller, M., Theuns, T., Dalla Vecchia, C., Frenk, C. S., McCarthy, I. G., Helly, J. C., et al. (2015). The EAGLE simulations of galaxy formation: calibration of subgrid physics and model variations. *Monthly Notices of the Royal Astronomical Society*, 450(2):1937–1961.

Cropper, M., Pottinger, S., Azzollini, R., Szafraniec, M., Awan, S., Mellier, Y., Berthé, M., Martignac, J., Cara, C., Di Giorgio, A.-M., et al. (2018). VIS: the visible imager for Euclid. In *Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave*, volume 10698, page 1069828. International Society for Optics and Photonics.

Cui, J.-L., Li, H.-L., and Zhang, X. (2017). No evidence for the evolution of mass density power-law index γ from strong gravitational lensing observation. *SCIENCE CHINA Physics, Mechanics & Astronomy*, 60(8):1–7.

Dark Energy Survey Collaboration et al. (2005). The dark energy survey. *International Journal of Modern Physics A*, 20(14):3121–3123.

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.

Davies, L., Lagos, C. d. P., Katsianis, A., Robotham, A., Cortese, L., Driver, S., Bremer, M., Brown, M., Brough, S., Cluver, M., et al. (2018). Galaxy And Mass Assembly (GAMA): The sSFR–M* relation part I–σsSFR–M* as a function of sample, SFR indicator, and environment. *Monthly Notices of the Royal Astronomical Society*, 483(2):1881–1900.

Del Popolo, A. and Le Delliou, M. (2017). Small scale problems of the ΛCDM model: a short review. *Galaxies*, 5(1):17.

Denker, J. S. and LeCun, Y. (1990). Transforming neural-net output levels to probability distributions. In *Proceedings of the 3rd International Conference on Neural Information Processing Systems*, pages 853–859.

Denzel, P., Coles, J. P., Saha, P., and Williams, L. L. (2021). The Hubble constant from eight time-delay galaxy lenses. *Monthly Notices of the Royal Astronomical Society*, 501(1):784–801.

DES Collaboration, Abbott, T. M. C., Aguena, M., Alarcon, A., Allam, S., Alves, O., Amon, A., Andrade-Oliveira, F., Annis, J., Avila, S., et al. (2021). Dark

Energy Survey Year 3 Results: Cosmological Constraints from Galaxy Clustering and Weak Lensing. *arXiv e-prints*, page arXiv:2105.13549.

Despali, G., Vegetti, S., White, S. D., Giocoli, C., and van den Bosch, F. C. (2018). Modelling the line-of-sight contribution in substructure lensing. *Monthly Notices of the Royal Astronomical Society*, 475(4):5424–5442.

Di Valentino, E., Mena, O., Pan, S., Visinelli, L., Yang, W., Melchiorri, A., Mota, D. F., Riess, A. G., and Silk, J. (2021). In the Realm of the Hubble tension − a Review of Solutions. *arXiv preprint arXiv:2103.01183*.

Dia, M., Savary, E., Melchior, M., and Courbin, F. (2019). Galaxy Image Simulation Using Progressive GANs. *arXiv preprint arXiv:1909.12160*.

Dozat, T. (2015). Incorporating Nesterov Momentum into Adam. Technical report, In Proc. ICLR Workshop.

Dutton, A. A. and Treu, T. (2014). The bulge–halo conspiracy in massive elliptical galaxies: implications for the stellar initial mass function and halo response to baryonic processes. *Monthly Notices of the Royal Astronomical Society*, 438(4):3594–3602.

Dye, S., Furlanetto, C., Dunne, L., Eales, S. A., Negrello, M., Nayyeri, H., van der Werf, P., Serjeant, S., Farrah, D., Michałowski, M., et al. (2018). Modelling high-resolution ALMA observations of strongly lensed highly star-forming galaxies detected by Herschel. *Monthly Notices of the Royal Astronomical Society*, 476(4):4383–4394.

Dye, S., Furlanetto, C., Swinbank, A., Vlahakis, C., Nightingale, J., Dunne, L., Eales, S. A., Smail, I., Oteo, I., Hunter, T., et al. (2015). Revealing the complex nature of the strong gravitationally lensed system H-ATLAS J090311. 6+ 003906 using ALMA. *Monthly Notices of the Royal Astronomical Society*, 452(3):2258–2268.

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.

Eales, S., Fullard, A., Allen, M., Smith, M. W. L., Baldry, I., Bourne, N., Clark, C. J. R., Driver, S., Dunne, L., Dye, S., et al. (2015). H-ATLAS/GAMA: quantifying the morphological evolution of the galaxy population using cosmic calorimetry. *Monthly Notices of the Royal Astronomical Society*, 452(4):3489–3507.

Etherington, A., Nightingale, J. W., Massey, R., Cao, X., Robertson, A., Amorisco, N. C., Amvrosiadis, A., Cole, S., Frenk, C. S., He, Q., et al. (2022). Automated galaxy-galaxy strong lens modelling: no lens left behind. *arXiv preprint arXiv:2202.09201*.

Falco, E., Gorenstein, M., and Shapiro, I. (1985). On model-dependent bounds on H(0) from gravitational images Application of Q0957+ 561A, B. *The Astrophysical Journal*, 289:L1–L4.

Feroz, F., Hobson, M., and Bridges, M. (2009). MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. *Monthly Notices of the Royal Astronomical Society*, 398(4):1601–1614.

Fleury, P., Larena, J., and Uzan, J.-P. (2021). Line-of-sight effects in strong gravitational lensing. *Journal of Cosmology and Astroparticle Physics*, 2021(08):024.

Flores, R. A. and Primack, J. R. (1994). Observational and theoretical constraints on singular dark matter halos. *arXiv preprint astro-ph/9402004*.

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

Foxley-Marrable, M., Collett, T. E., Frohmaier, C., Goldstein, D. A., Kasen, D., Swann, E., and Bacon, D. (2020). Observing the earliest moments of supernovae using strong gravitational lenses. *Monthly Notices of the Royal Astronomical Society*, 495(4):4622–4637.

Freedman, W. L. (2017). Cosmology at at crossroads: Tension with the Hubble Constant. *arXiv preprint arXiv:1706.02739*.

Freedman, W. L. (2021). Measurements of the Hubble Constant: Tensions in Perspective. *The Astrophysical Journal*, 919(1):16.

Frenk, C. S. and White, S. D. (2012). Dark matter and cosmic structure. *Annalen der Physik*, 524(9-10):507–534.

Fukushima, K. (1980). Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202.

Gal, Y. and Ghahramani, Z. (2016a). Bayesian convolutional neural networks with Bernoulli approximate variational inference. In *4th International Conference on Learning Representations (ICLR) workshop track*.

Gal, Y. and Ghahramani, Z. (2016b). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on Machine Learning (ICML-16)*, pages 1050–1059.

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.

Gavazzi, R., Treu, T., Koopmans, L. V., Bolton, A. S., Moustakas, L. A., Burles, S., and Marshall, P. J. (2008). The Sloan lens ACS survey. VI. Discovery and analysis of a double Einstein ring. *The Astrophysical Journal*, 677(2):1046.

Gavazzi, R., Treu, T., Rhodes, J. D., Koopmans, L. V., Bolton, A. S., Burles, S., Massey, R. J., and Moustakas, L. A. (2007). The Sloan Lens ACS survey. IV. The mass density profile of early-type galaxies out to 100 effective radii. *The Astrophysical Journal*, 667(1):176.

Geach, J. E., Ivison, R. J., Dye, S., and Oteo, I. (2018). A Magnified View of Circumnuclear Star Formation and Feedback around an Active Galactic Nucleus at z= 2.6. *The Astrophysical Journal Letters*, 866(1):L12.

Gentile, F., Tortora, C., Covone, G., Koopmans, L. V., Spiniello, C., Fan, Z., Li, R., Liu, D., Napolitano, N. R., Vaccari, M., et al. (2022). Lenses In VoicE (LIVE): Searching for strong gravitational lenses in the VOICE@ VST survey using Convolutional Neural Networks. *Monthly Notices of the Royal Astronomical Society*, 510(1):500–514.

Giani, L. and Frion, E. (2020). Testing the equivalence principle with strong lensing time delay variations. *Journal of Cosmology and Astroparticle Physics*, 2020(09):008.

Gilman, D., Birrer, S., and Treu, T. (2020). TDCOSMO-III. Dark matter substructure meets dark energy. The effects of (sub) halos on strong-lensing measurements of H0. *Astronomy & Astrophysics*, 642:A194.

Gilman, D., Birrer, S., Treu, T., Nierenberg, A., and Benson, A. (2019). Probing dark matter structure down to 107 solar masses: flux ratio statistics in gravitational lenses with line-of-sight haloes. *Monthly Notices of the Royal Astronomical Society*, 487(4):5721–5738.

Girelli, G., Pozzetti, L., Bolzonella, M., Giocoli, C., Marulli, F., and Baldi, M. (2020). The stellar-to-halo mass relation over the past 12 Gyr-I. Standard ΛCDM model. *Astronomy & Astrophysics*, 634:A135.

Goldstein, D. A., Nugent, P. E., Kasen, D. N., and Collett, T. E. (2018). Precise time delays from strongly gravitationally lensed type Ia supernovae with chromatically microlensed images. *The Astrophysical Journal*, 855(1):22.

Gott III, J. R., Weinberg, D. H., and Melott, A. L. (1987). A quantitative approach to the topology of large-scale structure. *The Astrophysical Journal*, 319:1–8.

Graves, A. (2011). Practical variational inference for neural networks. *Advances in neural information processing systems*, 24.

Haggar, R., Pearce, F. R., Gray, M. E., Knebe, A., and Yepes, G. (2021). The Three Hundred Project: Substructure in hydrodynamical and dark matter simulations of galaxy groups around clusters. *Monthly Notices of the Royal Astronomical Society*, 502(1):1191–1204.

Harvey, D., Kneib, J.-P., and Jauzac, M. (2016). Systematic or signal? How dark matter misalignments can bias strong lensing models of galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 458(1):660–665.

Hausen, R. and Robertson, B. E. (2020). Morpheus: a deep learning framework for the pixel-level analysis of astronomical image data. *The Astrophysical Journal Supplement Series*, 248(1):20.

He, Q., Nightingale, J., Massey, R., Robertson, A., Amvrosiadis, A., Cole, S., Frenk, C. S., Li, R., Amorisco, N. C., Metcalf, R. B., et al. (2022a). Testing strong lensing subhalo detection with a cosmological simulation. *arXiv preprint arXiv:2202.10191*.

He, Q., Robertson, A., Nightingale, J., Cole, S., Frenk, C. S., Massey, R., Amvrosiadis, A., Li, R., Cao, X., and Etherington, A. (2022b). A forward-modelling method to infer the dark matter particle mass from strong gravitational lenses. *Monthly Notices of the Royal Astronomical Society*, 511(2):3046–3062.

He, Z., Er, X., Long, Q., Liu, D., Liu, X., Li, Z., Liu, Y., Deng, W., and Fan, Z. (2020). Deep learning for strong lensing search: tests of the convolutional neural networks and new candidates from KiDS DR3. *Monthly Notices of the Royal Astronomical Society*, 497(1):556–571.

Hebb, D. O. (1949). The organization of behavior; a neuropsycholocigal theory. *A Wiley Book in Clinical Psychology*, 62:78.

Heitmann, K., Finkel, H., Pope, A., Morozov, V., Frontiere, N., Habib, S., Rangel, E., Uram, T., Korytov, D., Child, H., et al. (2019). The Outer Rim Simulation: A Path to Many-Core Supercomputers. *The Astrophysical Journal Supplement Series*, 245(1):16.

Hewitt, J., Turner, E., Schneider, D., Burke, B., Langston, G., and Lawrence, C. (1988). Unusual radio source MG1131+ 0456: a possible Einstein ring. *Nature*, 333(6173):537–540.

Heymans, C., Tröster, T., Asgari, M., Blake, C., Hildebrandt, H., Joachimi, B., Kuijken, K., Lin, C.-A., Sánchez, A. G., van den Busch, J. L., et al. (2021). KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints. *Astronomy & Astrophysics*, 646:A140.

Hezaveh, Y. D., Dalal, N., Marrone, D. P., Mao, Y.-Y., Morningstar, W., Wen, D., Blandford, R. D., Carlstrom, J. E., Fassnacht, C. D., Holder, G. P., et al. (2016). Detection of lensing substructure using ALMA observations of the dusty galaxy SDP. 81. *The Astrophysical Journal*, 823(1):37.

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

Hikage, C., Oguri, M., Hamana, T., More, S., Mandelbaum, R., Takada, M., Köhlinger, F., Miyatake, H., Nishizawa, A. J., Aihara, H., et al. (2019). Cosmology from cosmic shear power spectra with Subaru Hyper Suprime-Cam first-year data. *Publications of the Astronomical Society of Japan*, 71(2):43.

Hoffman, M. D., Blei, D. M., Wang, C., and Paisley, J. (2013). Stochastic variational inference. *Journal of Machine Learning Research*, 14(5).

Holanda, R., Pereira, S., and Jain, D. (2017). Constraints on a possible evolution of mass density power-law index in strong gravitational lensing from cosmological data. *Monthly Notices of the Royal Astronomical Society*, 471(3):3079–3086.

Hong, S. E., Park, S., Jee, M. J., Bak, D., and Cha, S. (2021). Weak-lensing Mass Reconstruction of Galaxy Clusters with a Convolutional Neural Network. *The Astrophysical Journal*, 923(2):266.

Hsueh, J.-W., Enzi, W., Vegetti, S., Auger, M., Fassnacht, C. D., Despali, G., Koopmans, L. V., and McKean, J. P. (2020). SHARP–VII. New constraints on the dark matter free-streaming properties and substructure abundance from gravitationally lensed quasars. *Monthly Notices of the Royal Astronomical Society*, 492(2):3047–3059.

Huang, X., Storfer, C., Gu, A., Ravi, V., Pilon, A., Sheu, W., Venguswamy, R., Banka, S., Dey, A., Landriau, M., et al. (2021). Discovering new strong gravitational lenses in the DESI legacy imaging surveys. *The Astrophysical Journal*, 909(1):27.

Huterer, D. and Shafer, D. L. (2017). Dark energy two decades after: Observables, probes, consistency tests. *Reports on Progress in Physics*, 81(1):016901.

Hyde, J. B. and Bernardi, M. (2009). The luminosity and stellar mass Fundamental Plane of early-type galaxies. *Monthly Notices of the Royal Astronomical Society*, 396(2):1171–1185.

Ibarra, A., Kavanagh, B. J., and Rappelt, A. (2019). Impact of substructure on local dark matter searches. *Journal of Cosmology and Astroparticle Physics*, 2019(12):013.

Inoue, K. T., Matsushita, S., Nakanishi, K., and Minezaki, T. (2020). ALMA 50-parsec-resolution Imaging of Jet–ISM Interaction in the Lensed Quasar MG J0414+ 0534. *The Astrophysical Journal Letters*, 892(2):L18.

Ivezić, Ž., Axelrod, T., Brandt, W., Burke, D., Claver, C., Connolly, A., Cook, K., Gee, P., Gilmore, D., Jacoby, S., et al. (2008). Large Synoptic Survey Telescope: From science drivers to reference design. *Serbian Astronomical Journal*, 176:1–13.

Ivezic, Z., Jones, L., and Lupton, R. (2010). The LSST Photon Rates and SNR Calculations, v1.2. *LSST Docushare, LSE-40.*

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

Jacobs, C., Glazebrook, K., Collett, T., More, A., and McCarthy, C. (2017). Finding strong lenses in CFHTLS using convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 471(1):167–181.

Jacobs, C., Glazebrook, K., Qin, A., and Collett, T. (2022). Exploring the interpretability of deep neural networks used for gravitational lens finding with a sensitivity probe. *Astronomy and Computing*, 38:100535.

Jedamzik, K. (2020). Primordial black hole dark matter and the LIGO/Virgo observations. *Journal of Cosmology and Astroparticle Physics*, 2020(09):022.

Jee, I., Komatsu, E., and Suyu, S. H. (2015). Measuring angular diameter distances of strong gravitational lenses. *Journal of Cosmology and Astroparticle Physics*, 2015(11):033.

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233.

Keeton, C. R. (2001). A catalog of mass models for gravitational lensing. *arXiv preprint astro-ph/0102341*.

Kelley, T., Bullock, J. S., Garrison-Kimmel, S., Boylan-Kolchin, M., Pawlowski, M. S., and Graus, A. S. (2019). Phat ELVIS: The inevitable effect of the Milky Way's disc on its dark matter subhaloes. *Monthly Notices of the Royal Astronomical Society*, 487(3):4409–4423.

Kelvin, L. S., Driver, S. P., Robotham, A. S., Graham, A. W., Phillipps, S., Agius, N. K., Alpaslan, M., Baldry, I., Bamford, S. P., Bland-Hawthorn, J., et al. (2014). Galaxy And Mass Assembly (GAMA): ugrizYJHK Sérsic luminosity functions and the cosmic spectral energy distribution by Hubble type. *Monthly Notices of the Royal Astronomical Society*, 439(2):1245–1269.

Kendall, A. and Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision? In *Advances in neural information processing systems*, pages 5574–5584.

Khullar, G., Gozman, K., Lin, J. J., Martinez, M. N., Acuña, O. S. M., Medina, E., Merz, K., Sanchez, J. A., Sisco, E. E., Stein, D. J. K., et al. (2020). COOL-LAMPS. I. An Extraordinarily Bright Lensed Galaxy at Redshift 5.04. *The Astrophysical Journal*, 906(2):107.

Kim, S. Y. and Peter, A. H. (2021). The Milky Way satellite velocity function is a sharp probe of small-scale structure problems. *arXiv preprint arXiv:2106.09050*.

Kim, S. Y., Peter, A. H. G., and Hargis, J. R. (2018). Missing Satellites Problem: Completeness Corrections to the Number of Satellite Galaxies in the Milky Way are Consistent with Cold Dark Matter Predictions. *Phys. Rev. Lett.*, 121:211302.

Klypin, A., Kravtsov, A. V., Valenzuela, O., and Prada, F. (1999). Where are the missing galactic satellites? *The Astrophysical Journal*, 522(1):82.

Knabel, S., Steele, R. L., Holwerda, B. W., Bridge, J. S., Jacques, A., Hopkins, A. M., Bamford, S. P., Brown, M. J., Brough, S., Kelvin, L., et al. (2020). Galaxy and Mass Assembly: A Comparison between Galaxy–Galaxy Lens Searches in KiDS/GAMA. *The Astronomical Journal*, 160(5):223.

Knox, L. and Millea, M. (2020). Hubble constant hunter's guide. *Physical Review D*, 101(4):043533.

Kochanek, C. (2004). The Saas Fee Lectures on strong gravitational lensing. *arXiv preprint astro-ph/0407232*.

Koopmans, L., Bolton, A., Treu, T., Czoske, O., Auger, M., Barnabe, M., Vegetti, S., Gavazzi, R., Moustakas, L., and Burles, S. (2009). The structure and dynamics of massive early-type galaxies: On homology, isothermality, and isotropy inside one effective radius. *The Astrophysical Journal Letters*, 703(1):L51.

Koopmans, L. V., Treu, T., Bolton, A. S., Burles, S., and Moustakas, L. A. (2006). The Sloan Lens ACS Survey. III. The structure and formation of early-type galaxies and their evolution since z $\approx$ 1. *The Astrophysical Journal*, 649(2):599.

Kormann, R., Schneider, P., and Bartelmann, M. (1994). Isothermal elliptical gravitational lens models. *Astronomy and Astrophysics*, 284:285–299.

Korytov, D., Hearin, A., Kovacs, E., Larsen, P., Rangel, E., Hollowed, J., Benson, A. J., Heitmann, K., Mao, Y.-Y., Bahmanyar, A., et al. (2019). CosmoDC2: A

Synthetic Sky Catalog for Dark Energy Science with LSST. *The Astrophysical Journal Supplement Series*, 245(2):26.

Kowal, C. (1968). Absolute magnitudes of supernovae. *The Astronomical Journal*, 73:1021–1024.

Kraljic, K., Arnouts, S., Pichon, C., Laigle, C., de la Torre, S., Vibert, D., Cadiou, C., Dubois, Y., Treyer, M., Schimd, C., et al. (2018). Galaxy evolution in the metric of the cosmic web. *Monthly Notices of the Royal Astronomical Society*, 474(1):547–571.

Kravtsov, A., Vikhlinin, A., and Meshcheryakov, A. (2018). Stellar mass—Halo mass relation and star formation efficiency in high-mass halos. *Astronomy Letters*, 44(1):8–34.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Krywult, J., Tasca, L., Pollo, A., Vergani, D., Bolzonella, M., Davidzon, I., Iovino, A., Gargiulo, A., Haines, C., Scodeggio, M., et al. (2017). The VIMOS Public Extragalactic Redshift Survey (VIPERS)-The coevolution of galaxy morphology and colour to z˜ 1. *Astronomy & Astrophysics*, 598:A120.

Küng, R., Saha, P., Ferreras, I., Baeten, E., Coles, J., Cornen, C., Macmillan, C., Marshall, P., More, A., Oswald, L., et al. (2018). Models of gravitational lens candidates from Space Warps CFHTLS. *Monthly Notices of the Royal Astronomical Society*, 474(3):3700–3713.

Lacey, C. and Cole, S. (1993). Merger rates in hierarchical models of galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 262(3):627–649.

Lanusse, F., Ma, Q., Li, N., Collett, T. E., Li, C.-L., Ravanbakhsh, S., Mandelbaum, R., and Póczos, B. (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.

Laureijs, R., Amiaux, J., Arduini, S., Augueres, J.-L., Brinchmann, J., Cole, R., Cropper, M., Dabin, C., Duvet, L., Ealet, A., et al. (2011). Euclid definition study report. *arXiv preprint arXiv:1110.3193*.

Leauthaud, A., Tinker, J., Bundy, K., Behroozi, P. S., Massey, R., Rhodes, J., George, M. R., Kneib, J.-P., Benson, A., Wechsler, R. H., et al. (2011). New constraints on the evolution of the stellar-to-dark matter connection: a combined analysis of galaxy–galaxy lensing, clustering, and stellar mass functions from z= 0.2 to z= 1. *The Astrophysical Journal*, 744(2):159.

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

Legin, R., Hezaveh, Y., Levasseur, L. P., and Wandelt, B. (2021). Simulation-Based Inference of Strong Gravitational Lensing Parameters. *arXiv preprint arXiv:2112.05278*.

Lemon, C. A., Auger, M. W., McMahon, R. G., and Ostrovski, F. (2018). Gravitationally lensed quasars in Gaia–II. Discovery of 24 lensed quasars. *Monthly Notices of the Royal Astronomical Society*, 479(4):5060–5074.

Levasseur, L. P., Hezaveh, Y. D., and Wechsler, R. H. (2017). Uncertainties in parameters estimated with neural networks: application to strong gravitational lensing. *The Astrophysical Journal Letters*, 850(1):L7.

Li, C., Jing, Y., Mao, S., Han, J., Peng, Q., Yang, X., Mo, H., and Van den Bosch, F. (2012). Internal Kinematics of Groups of Galaxies in the Sloan Digital Sky Survey Data Release 7. *The Astrophysical Journal*, 758(1):50.

Li, C. H. and Lee, C. (1993). Minimum cross entropy thresholding. *Pattern recognition*, 26(4):617–625.

Li, N., Becker, C., and Dye, S. (2021). The impact of line-of-sight structures on measuring H_0 with strong lensing time delays. *Monthly Notices of the Royal Astronomical Society*, 504(2):2224–2234.

Li, N., Gladders, M. D., Rangel, E. M., Florian, M. K., Bleem, L. E., Heitmann, K., Habib, S., and Fasel, P. (2016a). Pics: Simulations of strong gravitational lensing in galaxy clusters. *The Astrophysical Journal*, 828(1):54.

Li, R., Frenk, C. S., Cole, S., Gao, L., Bose, S., and Hellwing, W. A. (2016b). Constraints on the identity of the dark matter from strong gravitational lenses. *Monthly Notices of the Royal Astronomical Society*, 460(1):363–372.

Li, R., Napolitano, N., Tortora, C., Spiniello, C., Koopmans, L., Huang, Z., Roy, N., Vernardos, G., Chatterjee, S., Giblin, B., et al. (2020). New High-quality Strong Lens Candidates with Deep Learning in the Kilo-Degree Survey. *The Astrophysical Journal*, 899(1):30.

Liao, K., Shafieloo, A., Keeley, R. E., and Linder, E. V. (2019). A model-independent determination of the Hubble constant from lensed quasars and supernovae using Gaussian process regression. *The Astrophysical Journal Letters*, 886(1):L23.

Liddle, A. R. and Lyth, D. H. (2000). *Cosmological inflation and large-scale structure*. Cambridge university press.

Lin, J. Y.-Y., Yu, H., Morningstar, W., Peng, J., and Holder, G. (2020). Hunting for Dark Matter Subhalos in Strong Gravitational Lensing with Neural Networks. *arXiv preprint arXiv:2010.12960*.

Linder, E. V. (2003). Exploring the expansion history of the universe. *Physical Review Letters*, 90(9):091301.

Linder, E. V. (2004). Strong gravitational lensing and dark energy complementarity. *Physical Review D*, 70(4):043534.

Linder, E. V. (2016). Doubling strong lensing as a cosmological probe. *Physical Review D*, 94(8):083510.

Lovell, M. R., Eke, V., Frenk, C. S., Gao, L., Jenkins, A., Theuns, T., Wang, J., White, S. D., Boyarsky, A., and Ruchayskiy, O. (2012). The haloes of bright satellite galaxies in a warm dark matter universe. *Monthly Notices of the Royal Astronomical Society*, 420(3):2318–2324.

Luo, W., Yang, X., Lu, T., Shi, F., Zhang, J., Mo, H., Shu, C., Fu, L., Radovich, M., Zhang, J., et al. (2018). Galaxy–Galaxy Weak-lensing Measurements from SDSS. II. Host Halo Properties of Galaxy Groups. *The Astrophysical Journal*, 862(1):4.

Luo, W., Zhang, J., Halenka, V., Yang, X., More, S., Miller, C. J., Liu, L., and Shi, F. (2021). Emergent Gravity Fails to Explain Color-dependent Galaxy–Galaxy Lensing Signal from SDSS DR7. *The Astrophysical Journal*, 914(2):96.

MacKay, D. J. (1992). A practical Bayesian framework for backpropagation networks. *Neural computation*, 4(3):448–472.

Madireddy, S., Li, N., Ramachandra, N., Butler, J., Balaprakash, P., Habib, S., and Heitmann, K. (2019). A Modular Deep Learning Pipeline for Galaxy-Scale Strong Gravitational Lens Detection and Modeling. *arXiv preprint arXiv:1911.03867*.

Magro, D., Zarb Adami, K., DeMarco, A., Riggi, S., and Sciacca, E. (2021). A comparative study of convolutional neural networks for the detection of strong gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 505(4):6155–6165.

Maresca, J., Dye, S., Amvrosiadis, A., Bendo, G., Cooray, A., De Zotti, G., Dunne, L., Eales, S., Furlanetto, C., González-Nuevo, J., et al. (2022). Modelling high-resolution ALMA observations of strongly lensed dusty star-forming galaxies detected by Herschel. *Monthly Notices of the Royal Astronomical Society*, 512(2):2426–2438.

Maresca, J., Dye, S., and Li, N. (2021). Auto-identification of unphysical source reconstructions in strong gravitational lens modelling. *Monthly Notices of the Royal Astronomical Society*, 503(2):2229–2241.

Maturi, M., Mizera, S., and Seidel, G. (2014). Multi-colour detection of gravitational arcs. *Astronomy & Astrophysics*, 567:A111.

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

McGreer, I. D., Clément, B., Mainali, R., Stark, D. P., Gronke, M., Dijkstra, M., Fan, X., Bian, F., Frye, B., Jiang, L., et al. (2018). A bright-lensed galaxy at z= 5.4 with strong Ly α emission. *Monthly Notices of the Royal Astronomical Society*, 479(1):435–453.

Meneghetti, M. (2006). Introduction to gravitational lensing. *Lecture scripts organised in a course on web, https://www.ita.uni-heidelberg.de/∼jmerten/misc/meneghetti_lensing.pdf.*

Metcalf, R. B., Meneghetti, M., Avestruz, C., Bellagamba, F., Bom, C. R., Bertin, E., Cabanac, R., Courbin, F., Davies, A., Decenciere, E., et al. (2019). The strong gravitational lens finding challenge. *Astronomy & Astrophysics*, 625:A119.

Mollerach, S. and Roulet, E. (2002). *Gravitational lensing and microlensing*. World Scientific.

Moore, B. (1994). Evidence against dissipation-less dark matter from observations of galaxy haloes. *Nature*, 370(6491):629–631.

Moore, B., Ghigna, S., Governato, F., Lake, G., Quinn, T., Stadel, J., and Tozzi,

P. (1999a). Dark matter substructure within galactic halos. *The Astrophysical Journal Letters*, 524(1):L19.

Moore, B., Quinn, T., Governato, F., Stadel, J., and Lake, G. (1999b). Cold collapse and the core catastrophe. *Monthly Notices of the Royal Astronomical Society*, 310(4):1147–1152.

More, A., Verma, A., Marshall, P. J., More, S., Baeten, E., Wilcox, J., Macmillan, C., Cornen, C., Kapadia, A., Parrish, M., et al. (2015). Space Warps– II. New gravitational lens candidates from the CFHTLS discovered through citizen science. *Monthly Notices of the Royal Astronomical Society*, 455(2):1191—-1210.

More, S., van den Bosch, F. C., Cacciato, M., Skibba, R., Mo, H., and Yang, X. (2011). Satellite kinematics–III. Halo masses of central galaxies in SDSS. *Monthly Notices of the Royal Astronomical Society*, 410(1):210–226.

Morgan, R., Nord, B., Bechtol, K., González, S., Buckley-Geer, E., Möller, A., Park, J., Kim, A., Birrer, S., Aguena, M., et al. (2022). DeepZipper: A Novel Deep-learning Architecture for Lensed Supernovae Identification. *The Astrophysical Journal*, 927(1):109.

Morningstar, W. R., Hezaveh, Y. D., Levasseur, L. P., Blandford, R. D., Marshall, P. J., Putzky, P., and Wechsler, R. H. (2018). Analyzing interferometric observations of strong gravitational lenses with recurrent and convolutional neural networks. *arXiv preprint arXiv:1808.00011*.

Morningstar, W. R., Levasseur, L. P., Hezaveh, Y. D., Blandford, R., Marshall, P., Putzky, P., Rueter, T. D., Wechsler, R., and Welling, M. (2019). Data-driven reconstruction of gravitationally lensed galaxies using recurrent inference machines. *The Astrophysical Journal*, 883(1):14.

Mörtsell, E. and Dhawan, S. (2018). Does the Hubble constant tension call for new physics? *Journal of Cosmology and Astroparticle Physics*, 2018(09):025.

Mostoghiu, R., Knebe, A., Cui, W., Pearce, F. R., Yepes, G., Power, C., Dave, R., and Arth, A. (2019). The Three Hundred Project: The evolution of galaxy cluster density profiles. *Monthly Notices of the Royal Astronomical Society*, 483(3):3390–3403.

Mukherjee, S., Koopmans, L. V., Metcalf, R. B., Tessore, N., Tortora, C., Schaller, M., Schaye, J., Crain, R. A., Vernardos, G., Bellagamba, F., et al. (2018). SEAGLE–I. A pipeline for simulating and modelling strong lenses from cosmological hydrodynamic simulations. *Monthly Notices of the Royal Astronomical Society*, 479(3):4108–4125.

Naab, T. and Ostriker, J. P. (2017). Theoretical challenges in galaxy formation. *Annual review of astronomy and astrophysics*, 55:59–109.

Nagano, K., Fujita, T., Michimura, Y., and Obata, I. (2019). Axion dark matter search with interferometric gravitational wave detectors. *Physical review letters*, 123(11):111301.

Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814.

Narayan, R. and Bartelmann, M. (1995). Formation of Structure in the Universe. In Dekel, A. and Ostriker, J., editors, *Proceedings of the 1995 Jerusalem Winter School*. Cambridge University Press.

Narayan, R. and Bartelmann, M. (1996). Lectures on gravitational lensing. *arXiv preprint astro-ph/9606001*.

Navarro, J. F. (1996). The structure of cold dark matter halos. In *Symposium-international astronomical union*, volume 171, pages 255–258. Cambridge University Press.

Nayyeri, H., Keele, M., Cooray, A., Riechers, D., Ivison, R., Harris, A., Frayer, D., Baker, A., Chapman, S., Eales, S., et al. (2016). Candidate gravitationally lensed dusty star-forming galaxies in the Herschel wide area surveys. *The Astrophysical Journal*, 823(1):17.

Neal, R. M. (1995). *Bayesian Learning for Neural Networks*. PhD thesis, University of Toronto.

Niemi, S.-M. (2015). Euclid Visible InStrument (VIS) Python Package (VIS-PP) Documentation.

Nightingale, J., Dye, S., and Massey, R. J. (2018). AutoLens: automated modeling of a strong lens's light, mass, and source. *Monthly Notices of the Royal Astronomical Society*, 478(4):4738–4784.

Nightingale, J., Hayes, R. G., Kelly, A., Amvrosiadis, A., Etherington, A., He, Q., Li, N., Cao, X., Frawley, J., Cole, S., et al. (2021). PyAutoLens: Open-Source Strong Gravitational Lensing. *arXiv preprint arXiv:2106.01384*.

Nigoche-Netro, A., Aguerri, J., Lagos, P., Ruelas-Mayorga, A., Sánchez, L., and Machado, A. (2010). The Faber-Jackson relation for early-type galaxies: dependence on the magnitude range. *Astronomy & Astrophysics*, 516:A96.

Oguri, M. (2019). Strong gravitational lensing of explosive transients. *Reports on Progress in Physics*, 82(12):126901.

Oh, S.-H., De Blok, W., Brinks, E., Walter, F., and Kennicutt, R. C. (2011). Dark and luminous matter in THINGS dwarf galaxies. *The Astronomical Journal*, 141(6):193.

Oldham, L. J. and Auger, M. W. (2018). Dark matter contraction and stellar-mass-to-light ratio gradients in massive early-type galaxies. *Monthly Notices of the Royal Astronomical Society*, 476(1):133–150.

Oliver, S., Bock, J., Altieri, B., Amblard, A., Arumugam, V., Aussel, H., Babbedge, T., Beelen, A., Béthermin, M., Blain, A., et al. (2012). The

Herschel multi-tiered extragalactic survey: HerMES. *Monthly Notices of the Royal Astronomical Society*, 424(3):1614–1635.

Ostdiek, B., Rivero, A. D., and Dvorkin, C. (2020). Detecting Subhalos in Strong Gravitational Lens Images with Image Segmentation. *arXiv preprint arXiv:2009.06663*.

Ostdiek, B., Rivero, A. D., and Dvorkin, C. (2022a). Extracting the subhalo mass function from strong lens images with image segmentation. *The Astrophysical Journal*, 927(1):83.

Ostdiek, B., Rivero, A. D., Dvorkin, C., Berdeu, A., Langlois, M., Vachier, F., Cernicharo, J., Agúndez, M., Cabezas, C., Fuentetaja, R., et al. (2022b). Image segmentation for analyzing galaxy-galaxy strong lensing systems. *Astronomy & Astrophysics*, 657:L14.

Ostrovski, F., Lemon, C. A., Auger, M. W., McMahon, R. G., Fassnacht, C. D., Chen, G. C.-F., Connolly, A. J., Koposov, S. E., Pons, E., Reed, S. L., et al. (2017). The discovery of a five-image lensed quasar at z= 3.34 using PanSTARRS1 and Gaia. *Monthly Notices of the Royal Astronomical Society: Letters*, 473(1):L116–L120.

Paillassa, M., Bertin, E., and Bouy, H. (2020). MaxiMask and MaxiTrack: Two new tools for identifying contaminants in astronomical images using convolutional neural networks. *Astronomy & Astrophysics*, 634:A48.

Pandey, S., Raveri, M., and Jain, B. (2020). Model independent comparison of supernova and strong lensing cosmography: Implications for the Hubble constant tension. *Physical Review D*, 102(2):023505.

Park, J. W., Wagner-Carena, S., Birrer, S., Marshall, P. J., Lin, J. Y.-Y., Roodman, A., Collaboration, L. D. E. S., et al. (2021). Large-Scale Gravitational Lens Modeling with Bayesian Neural Networks for Accurate and Precise Inference of the Hubble Constant. *The Astrophysical Journal*, 910(1):39.

Pearson, J., Li, N., and Dye, S. (2019). The use of convolutional neural networks for modelling large optically-selected strong galaxy-lens samples. *Monthly Notices of the Royal Astronomical Society*, 488(1):991–1004.

Pearson, J., Maresca, J., Li, N., and Dye, S. (2021). Strong lens modelling: comparing and combining Bayesian neural networks and parametric profile fitting. *Monthly Notices of the Royal Astronomical Society*, 505(3):4362–4382.

Pécontal, E., Buchert, T., Di Stefano, P., Copin, Y., and Freese, K. (2009). Review of Observational Evidence for Dark Matter in the Universe and in upcoming searches for Dark Stars. *European Astronomical Society Publications Series*, 36:113–126.

Percival, W. J., Nichol, R. C., Eisenstein, D. J., Weinberg, D. H., Fukugita, M., Pope, A. C., Schneider, D. P., Szalay, A. S., Vogeley, M. S., Zehavi, I., et al. (2007). Measuring the matter density using baryon oscillations in the SDSS. *The Astrophysical Journal*, 657(1):51.

Perivolaropoulos, L. and Skara, F. (2021). Challenges for ΛCDM: An update. *arXiv preprint arXiv:2105.05208*.

Peter, A. H. (2012). Dark matter: a brief review. *arXiv preprint arXiv:1201.3942*.

Petrillo, C., Tortora, C., Chatterjee, S., Vernardos, G., Koopmans, L., Verdoes Kleijn, G., Napolitano, N., Covone, G., Schneider, P., Grado, A., et al. (2017). Finding strong gravitational lenses in the Kilo Degree Survey with convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 472(1):1129–1150.

Pilbratt, G., Riedinger, J., Passvogel, T., Crone, G., Doyle, D., Gageur, U., Heras, A., Jewell, C., Metcalfe, L., Ott, S., et al. (2010). Herschel Space Observatory - An ESA facility for far-infrared and submillimetre astronomy. *Astronomy & Astrophysics*, 518:L1.

Poole, G. B., Angel, P. W., Mutch, S. J., Power, C., Duffy, A. R., Geil, P. M., Mesinger, A., and Wyithe, S. B. (2016). Dark-ages Reionization and Galaxy formation simulation–I. The dynamical lives of high-redshift galaxies. *Monthly Notices of the Royal Astronomical Society*, 459(3):3025–3039.

Powell, D., Vegetti, S., McKean, J. P., Spingola, C., Rizzo, F., and Stacey, H. R. (2021). A novel approach to visibility-space modelling of interferometric gravitational lens observations at high angular resolution. *Monthly Notices of the Royal Astronomical Society*, 501(1):515–530.

Racca, G. D., Laureijs, R., Stagnaro, L., Salvignol, J.-C., Alvarez, J. L., Criado, G. S., Venancio, L. G., Short, A., Strada, P., Bönke, T., et al. (2016). The Euclid mission design. In *Space Telescopes and Instrumentation 2016: Optical, Infrared, and Millimeter Wave*, volume 9904, page 99040O. International Society for Optics and Photonics.

Radeka, V., Frank, J., Geary, J., Gilmore, D., Kotov, I., O'Connor, P., Takacs, P., and Tyson, J. (2009). LSST sensor requirements and characterization of the prototype LSST CCDs. *Journal of Instrumentation*, 4(03):P03002.

Rana, A., Jain, D., Mahajan, S., Mukherjee, A., and Holanda, R. (2017). Probing the cosmic distance duality relation using time delay lenses. *Journal of Cosmology and Astroparticle Physics*, 2017(07):010.

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P., et al. (1998). Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, 116(3):1009.

Riess, A. G., Macri, L. M., Hoffmann, S. L., Scolnic, D., Casertano, S., Filippenko, A. V., Tucker, B. E., Reid, M. J., Jones, D. O., Silverman, J. M., et al. (2016). A 2.4% determination of the local value of the Hubble constant. *The Astrophysical Journal*, 826(1):56.

Riess, A. G., Yuan, W., Macri, L. M., Scolnic, D., Brout, D., Casertano, S., Jones, D. O., Murakami, Y., Breuval, L., Brink, T. G., et al. (2021). A Comprehensive Measurement of the Local Value of the Hubble Constant with 1 km/s/Mpc Uncertainty from the Hubble Space Telescope and the SH0ES Team. *arXiv preprint arXiv:2112.04510.*

Ritondale, E., Vegetti, S., Despali, G., Auger, M., Koopmans, L., and McKean, J. (2019). Low-mass halo perturbations in strong gravitational lenses at redshift z 0.5 are consistent with CDM. *Monthly Notices of the Royal Astronomical Society*, 485(2):2179–2193.

Rivero, A. D. and Dvorkin, C. (2020). Direct detection of dark matter substructure in strong lens images with convolutional neural networks. *Physical Review D*, 101(2):023515.

Rojas, K., Savary, E., Clément, B., Maus, M., Courbin, F., Lemon, C., Chan, J., Vernardos, G., Joseph, R., Cañameras, R., et al. (2021). Strong lens systems search in the Dark Energy Survey using Convolutional Neural Networks. *arXiv preprint arXiv:2109.00014.*

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386.

Roszkowski, L., Sessolo, E. M., and Trojanowski, S. (2018). WIMP dark matter candidates and searches—current status and future prospects. *Reports on Progress in Physics*, 81(6):066201.

Rubin, D., Hayden, B., Huang, X., Aldering, G., Amanullah, R., Barbary, K., Boone, K., Brodwin, M., Deustua, S. E., Dixon, S., et al. (2018). The Discovery of a Gravitationally Lensed Supernova Ia at Redshift 2.22. *The Astrophysical Journal*, 866(1):65.

Rubin, V. C. and Ford Jr, W. K. (1970). Rotation of the Andromeda nebula from a spectroscopic survey of emission regions. *The Astrophysical Journal*, 159:379.

Ruff, A. J., Gavazzi, R., Marshall, P. J., Treu, T., Auger, M. W., and Brault, F. (2011). The SL2S Galaxy-scale Lens Sample. II. Cosmic evolution of dark and luminous mass in early-type galaxies. *The Astrophysical Journal*, 727(2):96.

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

Salmon, B., Coe, D., Bradley, L., Bradač, M., Strait, V., Paterno-Mahler, R., Huang, K.-H., Oesch, P. A., Zitrin, A., Acebron, A., et al. (2018). RELICS: A Candidate z 10 Galaxy Strongly Lensed into a Spatially Resolved Arc. *The Astrophysical Journal Letters*, 864(1):L22.

Schäfer, C., Fourestey, G., and Kneib, J.-P. (2020). Lenstool-HPC: A High Performance Computing based mass modelling tool for cluster-scale gravitational lenses. *Astronomy and Computing*, 30:100360.

Schäfer, C., Geiger, M., Kuntzer, T., and Kneib, J.-P. (2018). Deep convolutional neural networks as strong gravitational lens detectors. *Astronomy & Astrophysics*, 611:A2.

Schaye, J., Crain, R. A., Bower, R. G., Furlong, M., Schaller, M., Theuns, T., Dalla Vecchia, C., Frenk, C. S., McCarthy, I., Helly, J. C., et al. (2015). The EAGLE project: simulating the evolution and assembly of galaxies and their environments. *Monthly Notices of the Royal Astronomical Society*, 446(1):521–554.

Schneider, P. (1992). Gravitational lensing statistics. In *Gravitational Lenses*, pages 196–208. Springer.

Schneider, P. (2006). *Extragalactic astronomy and cosmology: an introduction*, volume 146. Springer.

Schneider, P., Ehlers, J., and Falco, E. E. (1992). *Gravitational Lenses*. Springer.

Schneider, P., Kochanek, C., and Wambsganss, J. (2006). *Gravitational lensing: strong, weak and micro: Saas-Fee advanced course 33*, volume 33. Springer Science & Business Media.

Schrabback, T., Hoekstra, H., Van Waerbeke, L., van Uitert, E., Georgiou, C., Asgari, M., Côté, P., Cuillandre, J.-C., Erben, T., Ferrarese, L., et al. (2021). Tightening weak lensing constraints on the ellipticity of galaxy-scale dark matter haloes. *Astronomy & Astrophysics*, 646:A73.

Schrabback, T., Schirmer, M., Van Der Burg, R. F., Hoekstra, H., Buddendiek, A., Applegate, D., Bradač, M., Eifler, T., Erben, T., Gladders, M. D., et al. (2018). Precise weak lensing constraints from deep high-resolution Ks images: VLT/HAWK-I analysis of the super-massive galaxy cluster RCS2 J 232727.7-020437 at z= 0.70. *Astronomy & Astrophysics*, 610:A85.

Schuldt, S., Chirivì, G., Suyu, S., Yıldırım, A., Sonnenfeld, A., Halkola, A., and Lewis, G. (2019). Inner dark matter distribution of the Cosmic Horseshoe (J1148+ 1930) with gravitational lensing and dynamics. *Astronomy & Astrophysics*, 631:A40.

Schuldt, S., Suyu, S., Cañameras, R., Taubenberger, S., Meinhard, T., Leal-Taixé, L., and Hsieh, B. (2020). Photometric Redshift Estimation with a Convolutional Neural Network: NetZ. *arXiv preprint arXiv:2011.12312*.

Schuldt, S., Suyu, S., Meinhardt, T., Leal-Taixé, L., Cañameras, R., Taubenberger, S., and Halkola, A. (2021). HOLISMOKES-IV. Efficient mass modeling of strong lenses through deep learning. *Astronomy & Astrophysics*, 646:A126.

Schwabe, B., Gosenca, M., Behrens, C., Niemeyer, J. C., and Easther, R. (2020). Simulating mixed fuzzy and cold dark matter. *Physical Review D*, 102(8):083518.

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

Şengül, A. Ç., Tsang, A., Rivero, A. D., Dvorkin, C., Zhu, H.-M., and Seljak, U. (2020). Quantifying the line-of-sight halo contribution to the dark matter convergence power spectrum from strong gravitational lenses. *Physical Review D*, 102(6):063502.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., and LeCun, Y. (2013). Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*.

Shajib, A. J., Birrer, S., Treu, T., Agnello, A., Buckley-Geer, E., Chan, J., Christensen, L., Lemon, C., Lin, H., Millon, M., et al. (2020). STRIDES: a 3.9 per cent measurement of the Hubble constant from the strong lens system DES J0408- 5354. *Monthly Notices of the Royal Astronomical Society*, 494(4):6072–6102.

Shajib, A. J., Treu, T., and Agnello, A. (2018). Improving time-delay cosmography with spatially resolved kinematics. *Monthly Notices of the Royal Astronomical Society*, 473(1):210–226.

Shajib, A. J., Treu, T., Birrer, S., and Sonnenfeld, A. (2021). Dark matter haloes of massive elliptical galaxies at z∼0.2 are well described by the Navarro–Frenk–White profile. *Monthly Notices of the Royal Astronomical Society*, 503(2):2380–2405.

Sharda, P., Federrath, C., da Cunha, E., Swinbank, A., and Dye, S. (2018). Testing star formation laws in a starburst galaxy at redshift 3 resolved with ALMA. *Monthly Notices of the Royal Astronomical Society*, 477(4):4380–4390.

Sharma, D. and Linder, E. V. (2022). Double Source Lensing Probing High Redshift Cosmology. *arXiv preprint arXiv:2204.03020*.

Sharon, C. E., Tagore, A. S., Baker, A. J., Rivera, J., Keeton, C. R., Lutz, D., Genzel, R., Wilner, D. J., Hicks, E. K., Allam, S. S., et al. (2019). Resolved molecular gas and star formation properties of the strongly lensed z= 2.26 galaxy SDSS J0901+ 1814. *The Astrophysical Journal*, 879(1):52.

Shu, Y., Bolton, A. S., Mao, S., Kochanek, C. S., Pérez-Fournon, I., Oguri, M., Montero-Dorta, A. D., Cornachione, M. A., Marques-Chaves, R., Zheng, Z., et al. (2016). The BOSS emission-line lens survey. IV. Smooth lens models for the BELLS GALLERY sample. *The Astrophysical Journal*, 833(2):264.

Shu, Y., Brownstein, J. R., Bolton, A. S., Koopmans, L. V., Treu, T., Montero-Dorta, A. D., Auger, M. W., Czoske, O., Gavazzi, R., Marshall, P. J., et al. (2017). The Sloan Lens ACS Survey. XIII. Discovery of 40 New Galaxy-scale Strong Lenses. *The Astrophysical Journal*, 851(1):48.

Shu, Y., Marques-Chaves, R., Evans, N. W., and Pérez-Fournon, I. (2018). SDSS J0909+ 4449: A large-separation strongly lensed quasar at z 2.8 with three images. *Monthly Notices of the Royal Astronomical Society: Letters*, 481(1):L136–L140.

Simon, J. D., Birrer, S., Bechtol, K., Chakrabarti, S., Cyr-Racine, F.-Y., Dell'Antonio, I., Drlica-Wagner, A., Fassnacht, C., Geha, M., Gilman, D., et al. (2019). Testing the nature of dark matter with extremely large telescopes. *arXiv preprint arXiv:1903.04742*.

Smith, R. J. and Collett, T. E. (2021). A fully-spectroscopic triple-source-plane lens: the Jackpot completed. *Monthly Notices of the Royal Astronomical Society*, 505(2):2136–2140.

Soldner, J. (1804). On the deflection of a light ray from its rectilinear motion, by the attraction of a celestial body at which it nearly passes by. *Berliner Astronomisches Jahrbuch*, pages 161–172.

Sonnenfeld, A. and Cautun, M. (2021). Statistical strong lensing. I. Constraints on the inner structure of galaxies from samples of a thousand lenses. *Astronomy & Astrophysics*, 651:A18.

Sonnenfeld, A., Leauthaud, A., Auger, M. W., Gavazzi, R., Treu, T., More, S., and Komiyama, Y. (2018). Evidence for radial variations in the stellar mass-to-light ratio of massive galaxies from weak and strong lensing. *Monthly Notices of the Royal Astronomical Society*, 481(1):164–184.

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.

Sonnenfeld, A., Treu, T., Marshall, P. J., Suyu, S. H., Gavazzi, R., Auger, M. W., and Nipoti, C. (2015). The SL2S Galaxy-scale Lens Sample. V. Dark matter halos and stellar IMF of massive early-type galaxies out to redshift 0.8. *The Astrophysical Journal*, 800(2):94.

Sonnenfeld, A., Verma, A., More, A., Baeten, E., Macmillan, C., Wong, K. C., Chan, J. H., Jaelani, A. T., Lee, C.-H., Oguri, M., et al. (2020). Survey of Gravitationally-lensed Objects in HSC Imaging (SuGOHI). VI. Crowdsourced lens finding with Space Warps. *Astronomy & Astrophysics*, 642:A148.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958.

Subramanian, K. and Cowling, S. A. (1986). On local conditions for multiple imaging by bounded, smooth gravitational lenses. *Monthly Notices of the Royal Astronomical Society*, 219:333–346.

Sumner, T. J. (2002). Experimental searches for dark matter. *Living Reviews in Relativity*, 5(1):1–60.

Suyu, S. H., Bonvin, V., Courbin, F., Fassnacht, C. D., Rusu, C. E., Sluse, D., Treu, T., Wong, K., Auger, M. W., Ding, X., et al. (2017). H0LiCOW–I. H 0 Lenses in COSMOGRAIL's Wellspring: program overview. *Monthly Notices of the Royal Astronomical Society*, 468(3):2590–2604.

Suyu, S. H., Marshall, P., Hobson, M., and Blandford, R. (2006). A Bayesian analysis of regularized source inversions in gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 371(2):983–998.

Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A. (2017). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In *Thirty-First AAAI Conference on Artificial Intelligence*.

Talbot, M. S., Brownstein, J. R., Dawson, K. S., Kneib, J.-P., and Bautista, J. (2021). The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: a catalogue of strong galaxy–galaxy lens candidates. *Monthly Notices of the Royal Astronomical Society*, 502(3):4617–4640.

Tanaka, M., Wong, K. C., More, A., Dezuka, A., Egami, E., Oguri, M., Suyu, S. H., Sonnenfeld, A., Higuchi, R., Komiyama, Y., et al. (2016). A spectroscopically confirmed double source plane lens system in the Hyper Suprime-Cam Subaru Strategic Program. *The Astrophysical Journal Letters*, 826(2):L19.

Tao, C. (2020). Dark Matter searches: an overview. *Journal of Instrumentation*, 15(06):C06054.

Taubenberger, S., Suyu, S., Komatsu, E., Jee, I., Birrer, S., Bonvin, V., Courbin, F., Rusu, C., Shajib, A., and Wong, K. (2019). The Hubble constant determined through an inverse distance ladder including quasar time delays and Type Ia supernovae. *Astronomy & Astrophysics*, 628:L7.

Tegmark, M., Strauss, M. A., Blanton, M. R., Abazajian, K., Dodelson, S., Sandvik, H., Wang, X., Weinberg, D. H., Zehavi, I., Bahcall, N. A., et al.

(2004). Cosmological parameters from SDSS and WMAP. *Physical review D*, 69(10):103501.

Tessore, N. and Metcalf, R. B. (2015). The elliptical power law profile lens. *Astronomy & Astrophysics*, 580:A79.

The GPyOpt authors (2016). GPyOpt: A Bayesian Optimization framework in Python. `http://github.com/SheffieldML/GPyOpt`.

To, C., Krause, E., Rozo, E., Wu, H., Gruen, D., Wechsler, R., Eifler, T., Rykoff, E., Costanzi, M., Becker, M., et al. (2021). Dark Energy Survey Year 1 Results: Cosmological Constraints from Cluster Abundances, Weak Lensing, and Galaxy Correlations. *Physical review letters*, 126(14):141301.

Tohill, C., Ferreira, L., Conselice, C., Bamford, S., and Ferrari, F. (2020). Measuring the structure of high-redshift galaxies with deep learning. *arXiv preprint arXiv:2012.09081.*

Treu, T. (2010). Strong lensing by galaxies. *Annual Review of Astronomy and Astrophysics*, 48:87–125.

Treu, T., Dutton, A. A., Auger, M. W., Marshall, P. J., Bolton, A. S., Brewer, B. J., Koo, D. C., and Koopmans, L. V. (2011). The SWELLS survey–I. A large spectroscopically selected sample of edge-on late-type lens galaxies. *Monthly Notices of the Royal Astronomical Society*, 417(3):1601–1620.

Tripathi, A., Sangwan, A., and Jassal, H. (2016). Dark energy equation of state parameter and its variation at low redshifts. *arXiv preprint arXiv:1611.01899.*

Tu, H., Gavazzi, R., Limousin, M., Cabanac, R., Marshall, P., Fort, B., Treu, T., Péllo, R., Jullo, E., Kneib, J.-P., et al. (2009). The mass profile of early-type galaxies in overdense environments: the case of the double source-plane gravitational lens SL2SJ02176-0513. *Astronomy & Astrophysics*, 501(2):475–484.

Umetsu, K. (2020). Cluster–galaxy weak lensing. *The Astronomy and Astrophysics Review*, 28(1):1–106.

Vagnozzi, S. (2020). New physics in light of the H0 tension: An alternative view. *Physical Review D*, 102(2):023518.

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., and Yu, T. (2014). scikit-image: image processing in Python. *PeerJ*, 2:e453.

Varma, S., Fairbairn, M., and Figueroa, J. (2020). Dark Matter Subhalos, Strong Lensing and Machine Learning. *arXiv preprint arXiv:2005.05353*.

Vega-Ferrero, J., Yepes, G., and Gottlöber, S. (2017). On the shape of dark matter haloes from MultiDark Planck simulations. *Monthly Notices of the Royal Astronomical Society*, 467(3):3226–3238.

Vegetti, S., Koopmans, L., Auger, M., Treu, T., and Bolton, A. (2014). Inference of the cold dark matter substructure mass function at z= 0.2 using strong gravitational lenses. *Monthly Notices of the Royal Astronomical Society*, 442(3):2017–2035.

Vegetti, S., Koopmans, L., Bolton, A., Treu, T., and Gavazzi, R. (2010). Detection of a dark substructure through gravitational imaging. *Monthly Notices of the Royal Astronomical Society*, 408(4):1969–1981.

Vegetti, S. and Koopmans, L. V. (2009). Bayesian strong gravitational-lens modelling on adaptive grids: objective detection of mass substructure in galaxies. *Monthly Notices of the Royal Astronomical Society*, 392(3):945–963.

Vegetti, S., Lagattuta, D., McKean, J., Auger, M., Fassnacht, C., and Koopmans, L. (2012). Gravitational detection of a low-mass dark satellite galaxy at cosmological distance. *nature*, 481(7381):341–343.

Velander, M., van Uitert, E., Hoekstra, H., Coupon, J., Erben, T., Heymans, C., Hildebrandt, H., Kitching, T. D., Mellier, Y., Miller, L., et al. (2014). CFHTLenS: the relation between galaxy dark matter haloes and baryons from weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 437(3):2111–2136.

Vernardos, G., Tsagkatakis, G., and Pantazis, Y. (2020). Quantifying the structure of strong gravitational lens potentials with uncertainty-aware deep neural networks. *Monthly Notices of the Royal Astronomical Society*, 499(4):5641–5652.

Villa, A. E., Masulli, P., and Rivero, A. J. P. (2016). *Artificial Neural Networks and Machine Learning–ICANN 2016: 25th International Conference on Artificial Neural Networks, Barcelona, Spain, September 6-9, 2016, Proceedings*, volume 9887. Springer.

Vogelsberger, M., Genel, S., Springel, V., Torrey, P., Sijacki, D., Xu, D., Snyder, G., Nelson, D., and Hernquist, L. (2014). Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe. *Monthly Notices of the Royal Astronomical Society*, 444(2):1518–1547.

Vogelsberger, M., Marinacci, F., Torrey, P., and Puchwein, E. (2020). Cosmological simulations of galaxy formation. *Nature Reviews Physics*, 2(1):42–66.

Wagner-Carena, S., Aalbers, J., Birrer, S., Nadler, E. O., Darragh-Ford, E., Marshall, P. J., and Wechsler, R. H. (2022). From Images to Dark Matter: End-To-End Inference of Substructure From Hundreds of Strong Gravitational Lenses. *arXiv preprint arXiv:2203.00690*.

Wagner-Carena, S., Park, J. W., Birrer, S., Marshall, P. J., Roodman, A., Wechsler, R. H., Collaboration, L. D. E. S., et al. (2021). Hierarchical Inference with Bayesian Neural Networks: An Application to Strong Gravitational Lensing. *The Astrophysical Journal*, 909(2):187.

Walsh, D., Carswell, R. F., and Weymann, R. J. (1979). 0957+ 561 A, B: twin quasistellar objects or gravitational lens? *Nature*, 279(5712):381–384.

Wang, B., Qi, J.-Z., Zhang, J.-F., and Zhang, X. (2020). Cosmological Model-independent Constraints on Spatial Curvature from Strong Gravitational Lensing and SN Ia Observations. *The Astrophysical Journal*, 898(2):100.

Warren, S. and Dye, S. (2003). Semilinear gravitational lens inversion. *The Astrophysical Journal*, 590(2):673.

Wechsler, R. H. and Tinker, J. L. (2018). The connection between galaxies and their dark matter halos. *Annual Review of Astronomy and Astrophysics*, 56:435–487.

Weinberg, S. (1989). The cosmological constant problem. *Reviews of modern physics*, 61(1):1.

Wilde, J., Serjeant, S., Bromley, J. M., Dickinson, H., Koopmans, L. V., and Metcalf, R. B. (2022). Detecting gravitational lenses using machine learning: exploring interpretability and sensitivity to rare lensing configurations. *Monthly Notices of the Royal Astronomical Society*, 512(3):3464–3479.

Winther, H. A., Casas, S., Baldi, M., Koyama, K., Li, B., Lombriser, L., and Zhao, G.-B. (2019). Emulators for the nonlinear matter power spectrum beyond Λ CDM. *Physical Review D*, 100(12):123540.

Wong, K. C., Suyu, S. H., Chen, G. C., Rusu, C. E., Millon, M., Sluse, D., Bonvin, V., Fassnacht, C. D., Taubenberger, S., Auger, M. W., et al. (2020). H0LiCOW–XIII. A 2.4 per cent measurement of H 0 from lensed quasars: 5.3 σ tension between early-and late-Universe probes. *Monthly Notices of the Royal Astronomical Society*, 498(1):1420–1439.

Wu, J. F. and Peek, J. (2020). Predicting galaxy spectra from images with hybrid convolutional neural networks. *arXiv preprint arXiv:2009.12318*.

Yang, T., Birrer, S., and Hu, B. (2020). The first simultaneous measurement of Hubble constant and post-Newtonian parameter from Time-Delay Strong Lensing. *Monthly Notices of the Royal Astronomical Society: Letters*, 497(1):L56–L61.

Zwicky, F. (1937). Nebulae as gravitational lenses. *Physical Review*, 51(4):290.