Machine Learning Strong Lensing

Stefan Schuldt

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:
Prof. Dr. rer. nat. Andreas Weiler

Prüfende der Dissertation:
1. Prof. Dr. Sherry Suyu
2. Hon.-Prof. Dr. Wolfgang Hillebrandt

Die Dissertation wurde am 20.12.2021 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 04.02.2022 angenommen.

Acknowledgements

Abstract

Over the past decades, gravitational lensing has become one of the major tools in astrophysical studies. It enables a variety of different applications, including studies of dark matter and dark energy, the measurement of cosmological parameters such as the Hubble constant H_0, analysis of the supernavae progenitor systems, and studies of high redshift galaxies and quasars. In recent years, large surveys have allowed the detection of hundreds of new strong gravitational lenses and thousands promising candidates, and machine learning has played a major role in handling the huge amount of data. As a consequence, fast and autonomous modeling for ground based imaging data is getting crucial. This will become even more relevant when new large imaging surveys, such as the Rubin Observatory Legacy Survey of Space and Time (LSST), will provide billions of high-quality images of galaxies containing around 100,000 lenses (Collett, 2015).

In this thesis, we focused on the modeling of strong lens systems using neural networks (NNs). Since one needs a good and large enough data set to train the NNs, we have developed a simulation pipeline that takes real observed images of galaxies, one for the lens and one for the background source with each a redshift measurements and a lens velocity dispersion value. The code computes then the lensing effect and adds the lensed source to the original lens image. With this procedure, we obtain ~100,000 realistic mock images in the quality of Hyper Suprime-Cam (HSC) images (Bosch et al., 2018; Aihara et al., 2019), as the light distributions and line-of-sight objects are real. These mocks are then used to train our Convolutional Neural Networks (CNNs) that predict the five parameters of a Singular Isothermal Ellipsoide (SIE) profile. Furthermore, we have explored the effect of different Einstein radii distributions and demonstrated the accuracy in image position and time delay predictions based on our network output.

Built upon this, we have improved the simulation procedure further and adopted also a SIE profile together with an external shear component. Here we include also an error prediction and make use of a residual network (ResNet) (He et al., 2015). This type of network architecture has been proposed to increase the network depth while increasing the computational time only moderately, resulting in a much more powerful network. We obtain very good results on the SIE parameters, but difficulties remain in recovering the external shear, which are minor distortions on the arcs from massive objects outside of the cutout.

As this ResNet is still trained on mock data, we carried out a detailed comparison of 32 real HSC lenses modeled with our network, and with traditional, very time and resource consuming methods. For this, we have developed a code to automate several steps to reduce the user input time. This comparison, which is the first direct comparison for ground based data, demonstrates the power and trustworthiness of the network.

Beside these developments for analyzing gravitational lenses, we present in this thesis NetZ, a novel approach to estimate photometric redshifts by using directly the images instead of

extracted quantities such as color-magnitude or size-compactness information of a galaxy. The performance of the network has been demonstrated with HSC images through a detailed comparison to another well-tested redshift code (DEmP, Hsieh & Yee, 2014) that gave the best accuracy with HSC data (Tanaka et al., 2018; Nishizawa et al., 2020). We find that our main network performs especially well in the high-redshift range without drastic cuts on e.g., the brightness, which will be very important for upcoming deep imaging surveys. One of the main advantages of this technique is the possibility of data augmentation in the training process to obtain a more uniformly distributed training sample. With this network we have predicted and published more than 34 million photometric redshifts. Given the expected similarity of HSC images to those from LSST in terms of quality and filter, this method will be immediately applicable with similar performance to LSST.

Combining the different machine learning and Bayesian analysis based software for strong lens modeling in addition to the presented photometric redshift network NetZ, we made a major step towards accurate and fast analysis of HSC data and also LSST data given their expected similarities. This enables a variety of different science applications and will be particularly useful to analyze and follow-up lensed time-variable sources discovered in the next decade by LSST.

Zusammenfassung

Über die letzten Jahrzehnte hinweg sind Gravitationslinsen zu einem der wichtigsten Instrumenten in astrophysikalischen Studien geworden. Es ermöglicht eine Vielzahl unterschiedlicher Anwendungen, einschließlich Studien zur dunklen Materie und der dunklen Energie, die Messung von kosmologischen Parameter wie der Hubble Konstante H_0, der Analyse von Supernavae Vorläuferszenarien und Studien von Galaxien und Quasaren mit hoher Rotverschiebung. In den letzten Jahren haben große Beobachtungsprogramme die Identifizierung von hunderten neuer starker Gravitationslinsen und tausenden von vielversprechenden Kandidaten ermöglicht, wobei maschinelles lernen eine große Rolle in der Handhabung diese rießige Menge an Daten gespielt hat. Als eine Konsequenz, schnelle und autonome Modellierung werden für erdgebundene Bilddaten entscheidend sein. Dies wird noch relevanter wenn neue große Beobachtungsprogramme, wie zum Beispiel das Rubin Observatory Legacy Survey of Space and Time (LSST) Projekt, Milliarden an qualitativ hochwertigen Bildern von Galaxien inklusive etwa 100,000 Gravitationslinsen (Collett, 2015) zur Verfügung stellen wird.

In dieser Dissertation haben wir uns auf die Modellierung von starker Gravitationslinsensysteme unter Ausnutzung neuronaler Netzwerke (NNe) konzentriert. Nachdem man zum trainieren von NNe einen groß genugen und realistischen Datensatz benötigt, haben wir einen Simulationscode entwickelt der echt beobachtete Bilder von Galaxien, eins für die Linse und eines für die Hintergrundquelle mit jeweils gemessener Rotverschiebung und Dispersionsgeschwindigkeitswert der Linse nutzt. Das Programm berechnet dann den Linseneffekt und fügt die durch den Linseneffekt veränderte Quelle dem Originalbild der Linse hinzu. Mit dieser Prozedur haben wir $\sim 100,000$ realistische Scheinbilder mit der Qualität von Hyper Suprime-Cam (HSC) Bildern (Bosch et al., 2018; Aihara et al., 2019) erhalten, da die Lichtverteilung und Nebenobjekten echt sind. Diese erzeugten Bilder wurden dann zum trainieren unseres faltendes neuronales Netzwerk (engl. convolutional neural network, CNN) verwendet um die fünf Parameter eines singulären isothermischen elliptischen (SIE) Profiles zu bestimmen. Desweiteren haben wir den Effekt von unerschidlichen Verteilungen der Einsteinradii erprobt und die Genauigkeit der Bildpositionen und Zeitverzögerungen basierend auf den von dem Netzwerk bestimmten Werten.

Darauf aufbauend, haben wir das Simulationsprogramm weiter optimiert und auch das SIE Profile zusammen mit externer Scherung angennommen. Hier haben wir ebenfalls eine Fehlerbestimmung eingebaut und nutzen ein Residuennetzwerk (ResNet) (He et al., 2015). Diese Art von Netzwerkarchitektur wurde vogeschlagen um die Tiefe des Netzwerkes erhöhen zu können wärend die Berechnungszeit nur geringfügig steigt, resultierend in einem deutlich leistungsfähigeren Netzwerk. Wir erhalten sehr gute Ergebnisse für die Parameter des SIE Profils, aber es bleiben Schwierigkeiten in der Erhaltung der externen Scherung vorhanden, welche geringe Verzerrungen an den Linsenbögen durch sehr massereiche Objekte außerhalb der Bildausschnittes hervorruft.

Zusammenfassung

Nachdem dieses ResNet noch an erzeugten Daten trainiert wurde, haben wir eine detailierten Vergleich anhand von 32 echten HSC Linsen durchgeführt, welche wir mit unserem Netzwerk modelliert haben und auch mit der traditionellen, sehr zeitaufwändigen und ressourcenverbrauchenden Methode. Hierfür haben wir auch ein Programm entwickelt um einige der Schritte zu automatisieren und dadurch den Zeitaufwand des Benutzers zu reduzieren. Dieser Vergleich, welcher den erste direkte Vergleich für erdgebundene Bilddaten darstellt, demonstriert die leistungsfähigkeit und Vertauenswürdigkeit des Netzwerkes.

Neben diesen Entwicklungen für die Auswertung on Gravitationslinsen, präsentieren wir in dieser Dissertation NetZ, ein neuartigen Ansatz um die Rotverschiebung photometrisch zu bestimmen wozu wir direkt die Bilddaten nutzen anstatt daraus extrahierte Werte wie Informationen zur Farb-Magnituden oder Größe-Kompaktheit der Galaxien. Die Leistung des Netzwerkes wurde demonstriert anhand HSC Bildern durch einen direkten Vergleich zu einem anderen gut getesteten Programm zur Rotverschiebungsbestimmung (DEmP, Hsieh & Yee, 2014), welches beste Genauigkeit an HSC Daten zeigte (Tanaka et al., 2018; Nishizawa et al., 2020). Wir finden das unser Hauptnetzwerk insbesondere im hohen Rotverschiebungsbereich gut funktioniert ohne starke Einschränkungen z.B. an der Helligkeit, was sehr wichtig für zukünftige tiefe Bildbeobachtungsprogramme ist. Eine der großen Vorteile dieser Methode ist die Möglichkeit von Datenerweiterung in dem Trainingsprozess um ein mehr gleichverteilteren Trainingsatz zu erhalten. Mit diesem Netzwerk haben wir mehr als 34 Millionen Rotverschiebungen photometrisch bestimmt und veröffentlicht. Aufgrund der erwarteten Ähnlichkeit zwischen HSC Bildern zu denen von LSST bezüglich Qualität und Filter, ist diese Methode direkt anwendbar auf LSST mit ähnlicher Genauigkeit.

Durch kombinieren der verschiedenen auf machinellem lernen und Bayessche Analyse basierenden Computerprogramme zur Modellierung starker Gravitationslinsen und durch zusätzlich das präsentierte Netzwerk NetZ zur photometrichen Rotverschiebungsbestimmung, haben wir einen großen Schritt in Richtung genauer und schneller Analyse von HSC Daten und auch Daten von LSST aufgrund zu erwatenden Ähnlichkeit. Dies ermöglicht eine Vielzahl an unterschiedlichen wissenschaftlichen Anwendungen und wird daher insbesondere nützlich sein um von LSST entdekte zeitlich variierende Quellen in Gravitationslinsen zu analysieren und weiter zu beobachten.

Contents

1 Introduction

1.1 Overview of gravitational lensing

Gravitational lensing describes the effect that light rays get deflected when traveling through a gravitational field. For this to happen, one needs two astrophysical objects aligned in a nearly straight line with the an observer on the earth, i.e. a light emitting source and between that source and the earth another massive object as the lens. This idea was already discussed and quantified within the Newtonian theory of gravity by considering light as massive particles with a mass going to the limit of zero. In 1915, Albert Einstein developed his Theory of General Relativity (GR) (Einstein, 1917, 1918), and predicted this effect to happen as well. Fig. 1.1 shows a schematic sketch of a lensing system.
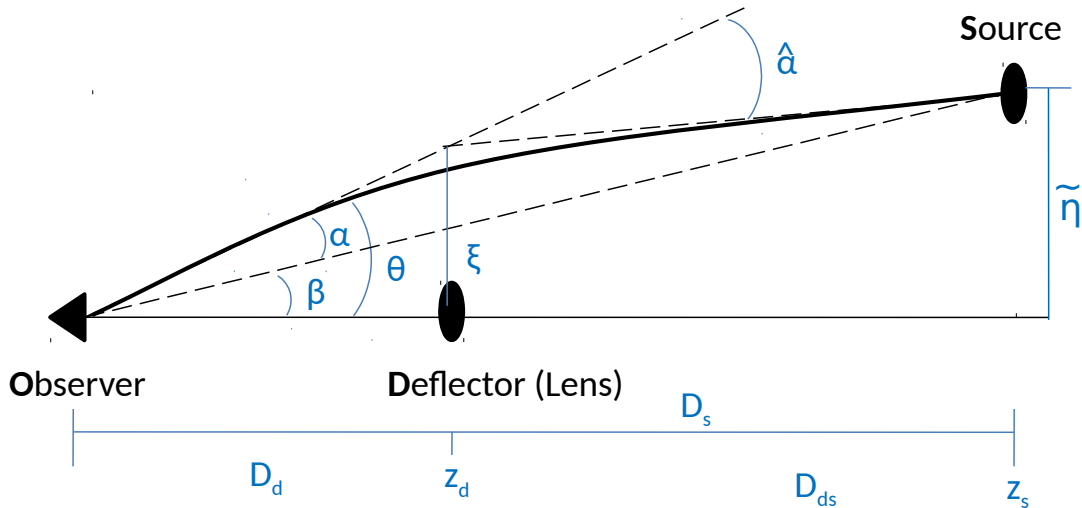
Figure 1.1: Schematic sketch of a lensing system in the thin lens approximation. Image taken from Schuldt (2018).

In this configuration, we have one deflector (lens) at an angular diameter distance D_d, corresponding to a redshift z_d, and a background source at distance D_s, corresponding to a redshift z_s, which is shifted by $\tilde{\eta}$ from a straight line through observer and lens, the so-called Line-of-sight (LOS). Einstein computed the deflection angle $\hat{\alpha}$ of a photon diverted by the gravitational field produced by a point mass M to be

$$\hat{\alpha} = \frac{4GM}{c^2 \xi} = 1.75'' \left(\frac{M}{M_\odot} \right) \left(\frac{\xi}{r_\odot} \right)^{-1} , \tag{1.1}$$

where G is the gravitational constant, c the speed of light, and ξ the impact parameter of the incoming light ray (Einstein, 1915). This raised a debate about the new theory of GR as the approximation with the Newtonian theory, which was the accepted theory of gravity at this time, predicted half of that deflection angle.

Only a few years later, the gravitational lensing effect was observed the first time and the deflection angle could be quantified experimentally. In 1919, a solar eclipse happened and allowed the measurement of the deflection of nearby stars through the mass of the sun carried out by Sir Arthur Eddington and Edwin Cottingham in Príncipe on the west coast of Africa and at the same time by Andrew Crommelin and Charles Davidson in Sobral, Brazil (Eddington, 1919). Although the deflection angle could be only quantified to be within $0.9''$ and $1.8''$ (Einstein, 1919), GR successfully overcame this test by predicting a value of $1.75''$ that matches the observed value.

After this astounding confirmation, Einstein affirmed the idea of gravitational lenses but remained very pessimistic about their observation as he thought only about star-sized lenses (Einstein, 1936). Instead, Fritz Zwicky had a much more positive view and also thought about more distant and massive objects such as galaxies. In the late 1930's, he finally proved the possibility of resolving multiple images of the same background source if the lens has a mass of a galaxy. His estimations forecast that images separation of around $10''$ are possible (Zwicky, 1937) and 1 out of 400 distant sources should be affected by lensing (Zwicky, 1937), which implies a very high probability to indeed find resolved multiple images.

Nonetheless, it took more than four decades until the first extragalactic lens was found. In 1979, Walsh et al. (1979) carried out a program to optically identify radio sources and discovered two quasars separated by $6''$ with identical redshift, color, and spectra. In the following years several more gravitational lenses were found, but mainly fortuitously.

In the late 20th century, gravitational lensing has become a prominent research field used to explore many different aspects of astrophysics and cosmology while the sample of known lenses grows slowly. During the years, three different regimes of lensing have evolved.

1.1.1 Weak gravitational lensing

In the weak lensing regime, the light of the background object gets distorted and magnified through the mass of the lens (e.g., Bartelmann & Schneider, 2001; Schneider et al., 2006), which is, however, not massive enough to produce multiple images of the source. Therefore, one cannot rely on single background objects because of their unknown intrinsic brightness and shape. Instead, weak lensing is used through a statistical inference of lensed objects in galaxy clusters or groups showing a trend in their alignment. This was observed the first time in two clusters A1689 and CL 1409+52 (Tyson et al., 1990), which were used to explore the Dark Matter (DM) distribution of the lensing cluster.

In the following years, weak lensing was detected in several other galaxy clusters or galaxy groups, and new detections are still expected in the future. Weak lensing allows to study several different aspects in astrophysics and cosmology, including the large scale structure of the Universe (e.g., Jaroszynski et al., 1990; Wambsganss et al., 1997; Schneider et al., 1998), the mass distribution of the cluster or its substructure (e.g., Lombardi & Bertin, 1998; King et al., 2001; Jee et al., 2005; Irwin & Shmakova, 2005), and the Hubble constant H_0. Moreover, the

Large Scale Structure (LSS) can produce a weak lensing effect, the so-called cosmic shear, that can be used for cosmology. It allows the measurements of the shear correlation functions (e.g., Abbott et al., 2018a; Duncan et al., 2021; Hildebrandt et al., 2017; Stafford et al., 2021) and also to estimate cosmological parameters such as the DM density parameter Ω_M or the normalization of the power spectrum σ_8 (e.g., Jain & Seljak, 1997; Hoekstra et al., 2002). Furthermore, it can be used to probe the equation of state of the Dark Energy (DE) by combining weak lensing and Cosmic Microwave Background (CMB) measurements (e.g., Baxter & Sherwin, 2021).

1.1.2 Microlensing

If the image separation of the lensed background source is so small that it cannot be resolved, we are talking about microlensing as proposed by Paczynski (1986, 1991). Even though the multiple images are not resolved, we can observe a change in the magnification resulting in a variation of the brightness over time. The first detection was reported in 1993 and used to test the DM nature (Alcock et al., 1993; Aubourg et al., 1993; Udalski et al., 1993).

In the following years, microlensing was a strong probe to test whether DM in the Milky Way is composed by compact objects, so called Massive Astrophysical Compact Halo Objects (MACHOs). For this, the brightness of a large number of stars in the Large Maggellanic Cloud (LMC) were monitored to detect fluctuations in case a MACHO would pass a star. Nowadays, micro-lensing is often used in a very similar procedure to detect planets orbiting around of stars by detecting the characteristic variation in the stars' flux (e.g., Mao & Paczynski, 1991; Tsapras et al., 2001; Bond et al., 2004; Nikolaus & Hundertmark, 2018). Moreover, microlensing can be use to study the Active Galactic Nucleus (AGN) accretion disk and AGN structure (e.g., Kochanek, 2004; Morgan et al., 2010, 2018; Chan et al., 2021).

1.1.3 Strong gravitational lensing

Stong lensing occurs when multiple images of the background source can be separately identified and distortions and magnification are prominent. To produced such a large image separation, the lens must have typically the mass of a galaxy at least and a relative good alignment with the source and the observer, i.e. a relative small offset $\tilde{\eta}$, resulting in typically tangentially stretched images which are called arcs. In an idealized scenario with $\tilde{\eta} = 0$ and a point mass, this can form a full *Einstein ring*, such that we call the image separation *Einstein radius* θ_E. A very well-known system with a nearly-full Einstein radius is the Cosmic Horseshoe (J1148+1930, Belokurov et al., 2007) shown in Fig.1.2. It is with $\sim 5.4 \times 10^{12} M_\odot$ so far the most massive galaxy observed, resulting in an Einstein radius θ_E of 5.1″. In addition to that, it contains a radial arc from another galaxy at redshift $z_{s,r} = 1.961$ (Schuldt et al., 2019), lying in between the lens ($z_d = 0.444$, Belokurov et al., 2007) and source that produce the Einstein ring ($z_{s,t} = 2.381$, Quider et al., 2009).

Strong gravitational lenses are a powerful tool for probing several different aspects of the Universe, helping to answer outstanding questions. We will discus some of them in in the following.
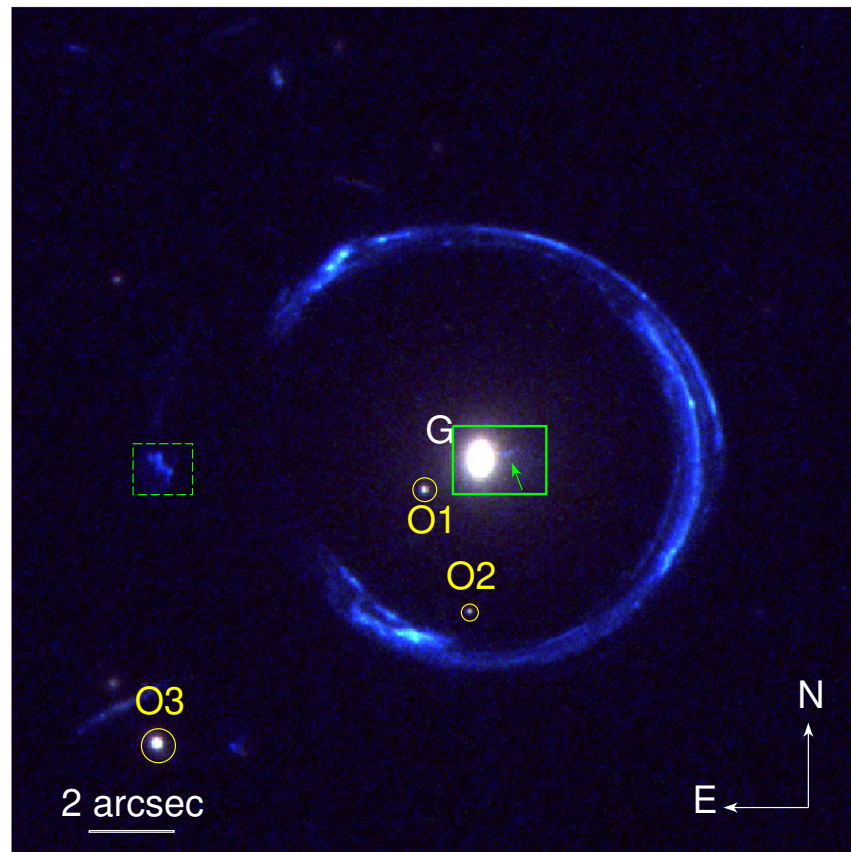
Figure 1.2: Color image of the Cosmic Horseshoe (J1148+1930) observed with Hubble Space Telescope (HST) (F475W, F606W, and F814W) as an example of a strongly lensed galaxy. The size of this image is $20'' \times 20''$, north is up and east is left. The radial arc is located on the west side of the lens galaxy *G*, marked with a green arrow. The corresponding counter image is on the left and marked with the green dashed box. Image taken from Schuldt et al. (2019).

1.1.3.1 Lens mass studies

Since the deflection of the light is independent of the nature of the mass, one can use lenses to infer the total mass of the deflector, i.e. baryonic and DM together (e.g., Belokurov et al., 2007; Treu, 2010; Treu & Ellis, 2015), independent of the nature of DM which is still unclear. So far DM is not directly observed, but lensing gives with the Bullet Cluster one of the strongest indications that DM exist (Barrena et al., 2002; Clowe et al., 2006; Tucker et al., 1998) and complement other indicators such as the Coma Cluster by considering the viral theorem (e.g., Zwicky, 1933, 1937), galactic rotation curves (e.g., López Fune, 2018), studies of the LSS (e.g., Ciarcelluti, 2005a,b; Kunz et al., 2016; Primack, 1997), or directly the CMB.

Instead of only inferring the total mass with gravitational lensing, one can also disentangle both components by assuming a mass-to-light ratio and even infer the DM distribution (e.g., Gavazzi et al., 2007; Dye et al., 2008; Suyu et al., 2012; Grillo et al., 2013; Bellagamba et al.,

2017; Collier et al., 2018; Schuldt et al., 2019). Since radial arcs are very seldom on the galaxy-scale, lensing gives typically only constraints at a distance of the Einstein radius. Therefore, the Cosmic Horseshoe with its peculiar radial arc shown in Fig. 1.2 is especially well suited for this study. In such analyses rotation curves (e.g., Hashim et al., 2014; Strigari, 2013) or velocity dispersion measurements (e.g., Mortlock & Webster, 2000; Treu & Koopmans, 2002a,b, 2004; Auger et al., 2010; van de Ven et al., 2010; Barnabè et al., 2011, 2012; Yıldırım et al., 2020) are often included, which perfectly complement the lensing data and help to break degeneracies. Through this combination, a deprojected 3D model of the mass density profile can be obtained which helps in probing cosmological models (e.g., Eales et al., 2015; Davies et al., 2018; Eales et al., 2015; Krywult et al., 2017; Birrer et al., 2020) or General Relativity (Collett et al., 2018). In addition to that, strong lensing can also be used to detect DM substructure through detailed modeling (Vegetti & Koopmans, 2009a,b; Vegetti & Vogelsberger, 2014; Vegetti et al., 2014; Hezaveh et al., 2016; Bayer et al., 2018; Brehmer et al., 2019; Ritondale et al., 2019).

1.1.3.2 High-z studies

While lensing conserves the Surface Brightness (SB), the lensed images appears sheared and magnified. Therefore, strong lensing is a perfect tool to study high-redshift sources, revealing information on the evolution of most distant galaxies so far observed (e.g., Dye et al., 2018; Lemon et al., 2018; McGreer et al., 2018; Rubin et al., 2018; Salmon et al., 2018; Shu et al., 2018). When the mass profile of the lens is well constrained, the original unlensed morphology can be reconstructed (e.g., Warren & Dye, 2003; Suyu et al., 2006; Nightingale et al., 2018; Rizzo et al., 2018; Chirivì et al., 2020; Powell et al., 2021), allowing a detailed study of their high-redshift properties. For instance, Salmon et al. (2018) report the discovery of SPT0615-JD1, a galaxy at $z \sim 10$ which is stretched into an $\sim 2.5''$ long arc by the effect of strong lensing. Since it is well resolved, they are able to extract information such as the intrinsic delensed magnitude, the stellar mass, the half-light radius suggesting a merger or accretion event, or the Star Formation Rate (SFR) indicating a typical star-forming galaxy.

1.1.3.3 Hubble constant measurements using quasar time delays

In the special case of a time-variable background object such as a quasar, a measurable gravitational time delay between two images j and k with observed positions $\boldsymbol{\theta}_j$ and $\boldsymbol{\theta}_k$ occur. This time delay is given by

$$\Delta t_{jk} = \frac{D_{\Delta t}}{c} \left[\frac{(\boldsymbol{\theta}_j - \boldsymbol{\beta})^2}{2} - \Psi(\boldsymbol{\theta}_j) - \frac{(\boldsymbol{\theta}_k + \boldsymbol{\beta})^2}{2} + \Psi(\boldsymbol{\theta}_k) \right] \qquad (1.2)$$

where $\boldsymbol{\beta}$ is the source position, Ψ the lens potential, and $D_{\Delta t}$ the so-called time delay distance. As we see from this equation, two effects are contributing to the time delay. The first one results from the difference in path length and is described by the term $(\boldsymbol{\theta} - \boldsymbol{\beta})^2/2$. The second effect is the gravitational time delay, which is proportional to the lens potential Ψ.

By monitoring the lensed quasar images, which was extensively carried out in the COSmological MOnitoring of GRAvItational Lenses (COSMOGRAIL) program (e.g., Courbin et al., 2005, 2018; Millon et al., 2020a), light curves of quasars were obtained that display the pattern

of the intrinsic brightness fluctuations with also extrinsic variations due to microlensing. As an example, the gravitationally lensed quasar system RX J1131 − 1231 is shown in Fig. 1.3 together with its light curve observed between 2003 and 2019 from the COSMOGRAIL collaboration. Such long and frequent monitoring allows to extract the light curve shift precisely, resulting in time delays between the different images with few percent uncertainty. This can be used to constrain cosmological parameters such as the Hubble constant H_0, as demonstrated extensively by the H_0 Lenses in COSMOGRAIL's Wellspring (H0LiCOW) collaboration (e.g., Rusu et al., 2020; Suyu et al., 2017; Tihhonova et al., 2020; Wong et al., 2020).

For this three ingredients are necessary: First, measured time delays Δt, second the mass along the LOS, and third a lens mass model to obtain the lens potential Ψ. With these quantities, we can use Eq. (1.2) to compute the time-delay distance defined as

$$D_{\Delta t} = (1 + z_\mathrm{d}) \frac{D_\mathrm{d} D_\mathrm{s}}{D_\mathrm{ds}} \,, \tag{1.3}$$

and with the three angular diameter distances inverse proportional to the Hubble constant H_0.

H0LiCOW's current best constraints on H_0 comes from six lensed quasars, resulting in a value of $H_0 = 73.3^{+1.7}_{-1.8} \mathrm{km\,s^{-1} Mpc^{-1}}$ assuming a flat ΛCold Dark Matter (CDM) cosmology (Wong et al., 2020) and physically motivated mass models. It is thus comparable and in agreement with results of the local Cepheid distance ladder, which currently measures $H_0 = (73.2 \pm 1.3) \mathrm{km\,s^{-1} Mpc^{-1}}$ as presented by the Supernovae, H_0, for the Equation of State (SH0ES) collaboration (Riess et al., 2021), as well as with measurements from other lensed quasars observed in the optical and Infrared (IR) (e.g., Millon et al., 2020b; Liao et al., 2019, 2020; Bonvin et al., 2017) and radio range (Qi et al., 2021). The procedure of obtaining H_0 from strongly lensed quasars is based purely on fundamental, well understood physics and is a one-step approach, in contrast to the distance ladder which adds to the uncertainty in the distance calibration and thus in H_0 as well. On the other side, strong lensing depends more strongly on the adopted cosmology compared to other late Universe probes through the higher redshifts of the lenses and quasars. Without the assumption of a flat ΛCDM cosmology the errors on H_0 can increase substantially. Nonetheless, many further investigations are done to improve the constrains through gravitational lensing and to make them less dependent on cosmological assumptions (e.g., Taubenberger et al., 2019; Wong et al., 2020; Birrer et al., 2019).

While there are nowadays several different approaches to measure the Hubble constant H_0, there were over years mainly two attempts carried out with high precision. Beside the distance ladder, also the CMB observations can be used to determine H_0 carried out extensively by the Planck Collaboration et al. (2020). While both methods where improved and the underlying measurements got more precise, their uncertainties got much smaller. This resulted finally in a discrepancy of the two measurements, as the Planck Collaboration et al. (2020) obtain a value of $H_0 = (67.4 \pm 0.5) \mathrm{km\,s^{-1} Mpc^{-1}}$, much lower than the through the distance ladder.

This debate demands new and independent approaches, where gravitational lensing is just one. We show an overview of the current measurments of H_0 from various different probes in Fig. 1.4. For better comparison, we show in that plot the newest value from the Planck Collaboration et al. (2020) as light blue bar, from gravitational lensing the current value from H0LiCOW as orange
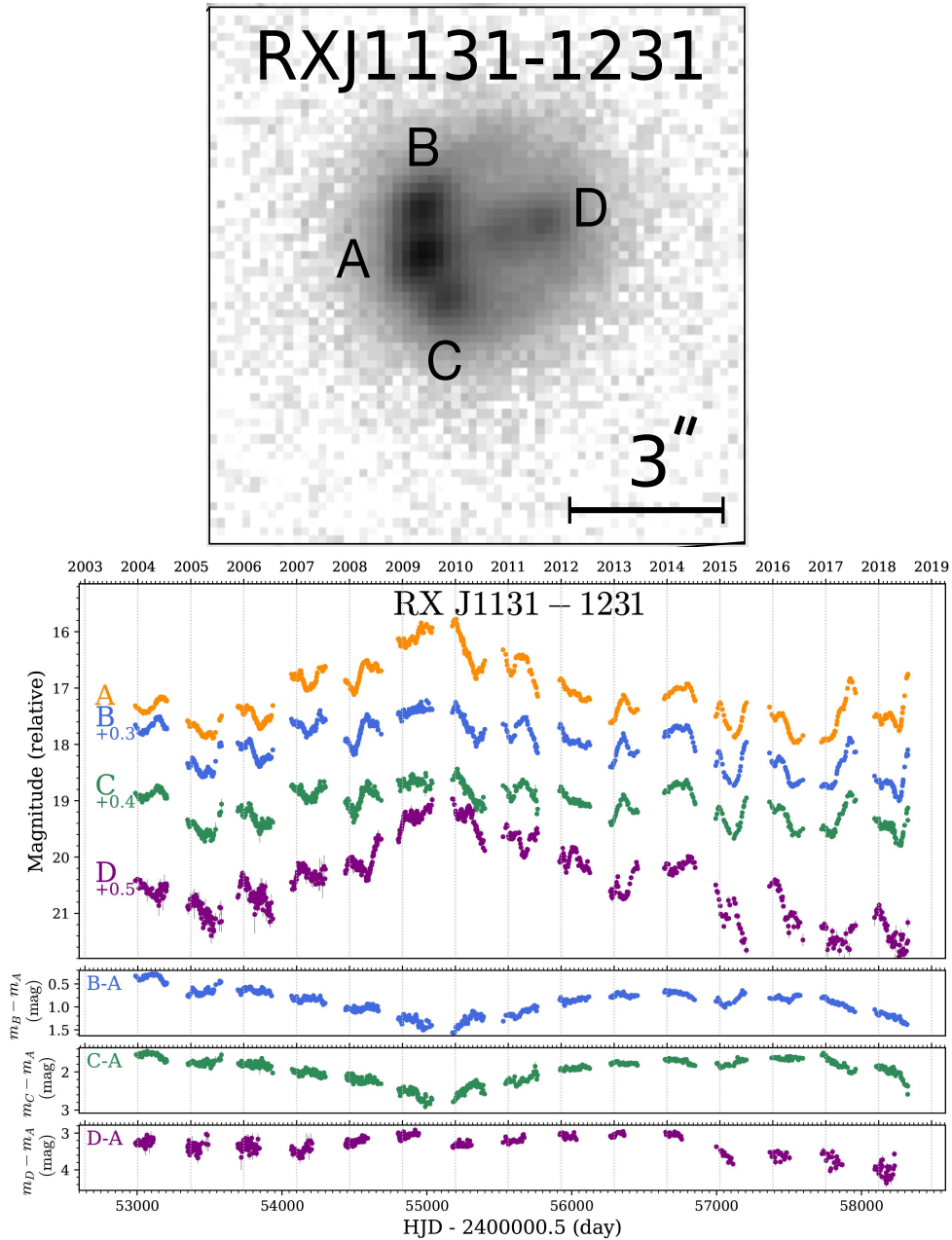
Figure 1.3: Observed *r* band image (top) and light curves for the quadruply lensed quasar system RX J1131 − 1231 (Millon et al., 2020a). The image was observed with the Leonhard Euler 1.2m telescope at La Silla Observatory and shows the four quasar images labeled with A, B, C, and D. The bottom panel shows the corresponding light curves from COSMOGRAIL together with the difference with respect to reference image A. Credit: Millon et al., 2020a A&A, 639, A101 reproduced with permission ©ESO.

bar (Wong et al., 2020), and overlaid as hatched bar the current estimates from the SH0ES team (Riess et al., 2021).

Based on the analyzed data set, the different approaches can be grouped into "early Universe" or "late Universe" measurements. The early Universe represent H_0 values obtained from the CMB power spectrum observed with the Planck and Wilkinson Microwave Anisotropy Probe (WMAP) satellites, or with the South Pole Telescope (SPT) and the Atacama Cosmology Telescope (ACT) (e.g., Aiola et al., 2020; Balkenhol et al., 2021; Dutcher et al., 2021; Planck Collaboration et al., 2020). Moreover, measurements corresponding to that group are based on Baryonic Acoustic Oscillation (BAO) and Big Bang Nucleosynthesis (BBN) and including data from Baryon Oscillation Spectroscopic Survey (BOSS) or extended BOSS (eBOSS) (e.g., Colas et al., 2020; Philcox et al., 2020; Ivanov et al., 2020; Alam et al., 2021). Since these are independent data sets, they can also be combined to obtain another measurement (e.g., Pogosian et al., 2020; Zhang & Huang, 2019). All measurements based on these data sets result in a value of $H_0 \sim 67 \mathrm{km\,s^{-1}Mpc^{-1}}$ and are in relatively good agreement to each other.

In the group of late type probes, there are also, beside the traditional distance ladder approach using Cepheids and Supernovas (SN) Type Ia (e.g., Riess et al., 2021; Breuval et al., 2020; Riess et al., 2019; Camarena & Marra, 2020), measurements extending specifically to the IR wavelength range (e.g., Dhawan et al., 2018; Freedman et al., 2012) carried out for instance by the Carnegie-Chicago Hubble Program (CCHP) (Freedman et al., 2012). Since these measurements are to some extend model dependent, Follin & Knox (2018) carried out a measurement also with an relatively model independent approach, resulting in a value of H_0 of $(73.3 \pm 1.7) \mathrm{km\,s^{-1}Mpc^{-1}}$ and therefore perfectly in agreement with the other distance ladder measurements. The uncertainty, however, is larger but still not in agreement with early type measurements. A different approach but with the same kind of data are performed by Feeney et al. (2018) with an Bayesian hierarchical approach to not rely on Gaussian distributions and by Cardona et al. (2017) with a Bayesian hyper parameter method used to avoids subjective rejection criteria of outliers and offers a possibility to test data sets for unknown systematics. As we can see directly from Fig. 1.4, both estimates are in good agreement with the other distance ladder approaches.

While these data sets include always the Cepheids as an anchor, there is also the possibility to use the Tip of the Red Giant Branch (TRGB) instead (e.g. Soltis et al., 2021; Freedman et al., 2019, 2020; Reid et al., 2019; Yuan et al., 2019). These are not fully independent measurements because of their shared SNe data set but still different measurements. Interestingly, these estimates with $H_0 \sim 70 \mathrm{km\,s^{-1}Mpc^{-1}}$ tend to lower values, i.e., in the direction of measurements relying on the CMB, BBN and BAO but still more than at least 1σ away.

Given the discrepancy between these distance ladder measurements and the early type, independent approaches and data sets became crucial for a clarification. A similar approach as the distance ladder with Cepheids or the TRGB is presented by Huang et al. (2020), who perform a dedicated search for Mira variables, which are highly evolved low-mass stars characterized by very red colors and have relatively long periods of few 100 days, using HST in NGC 1559, the host galaxy of the SN 2005df. The distance of that SN is then calibrated with the Miras, which is typically done with the Cepheids. In analogy to that, Pesce et al. (2020) consider microwave amplification by stimulated emission of radiation (maser) objects, which is an object with naturally occurring stimulated microwave emission lines. Specifically, they use megamaser, which have a very large isotropic luminosity in contrast to masers and thus can be detected even if
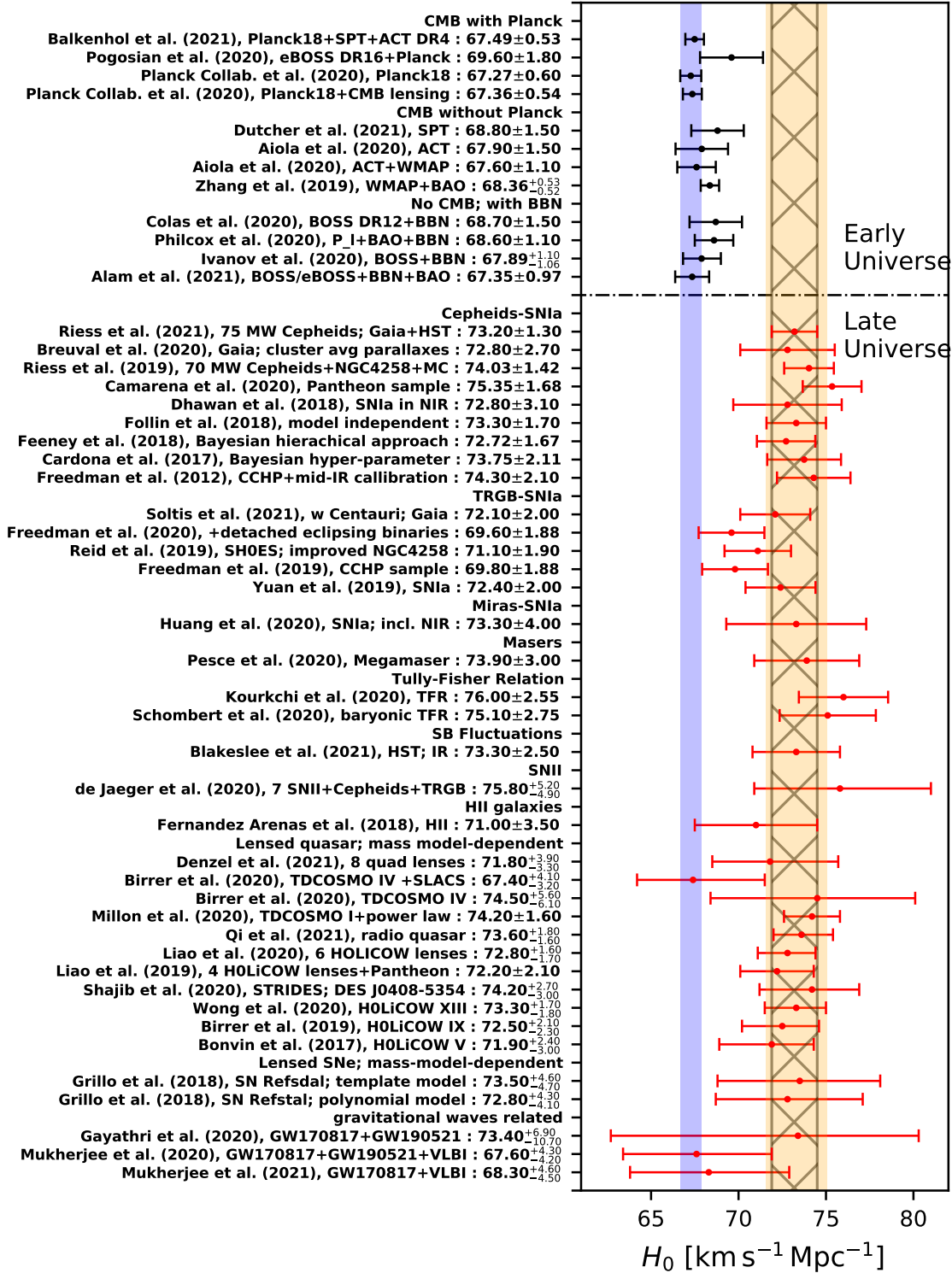
Figure 1.4: Recent measurements of the Hubble constant H_0 with different approaches and data sets. For comparison, values using the CMB (Planck Collaboration et al., 2020), gravitationally lensed quasars (H0LiCOW, Wong et al., 2020), and distance ladder with Cepheids (SH0ES program, Riess et al., 2021) are, respectively, shown as light blue, orange, and hashed bar. Adapted from Di Valentino et al. (2021).

outside of the Milky Way (MW). These maser are then again used to calibrate the distances of their host galaxies. Both measurements are in notable disagreement with the CMB measurement although the error bars are especially for the Miras larger, which might come from their relative long periods making the distance measurement more difficult.

In addition to that, Blakeslee et al. (2021) uses the SB fluctuations of 63 bright and mainly early type galaxies to reach distances in the Hubble flow. These are again combined with both the Cepheids and the TRGB to obtain finally a value of H_0, which is with $73.3 \pm 2.5 \mathrm{km\,s^{-1}Mpc^{-1}}$ in very good agreement with the current value of the SH0ES team.

Another type of approach is based on the Tully-Fisher relation, which allows to determine the distance D of spiral galaxies through their apparent brightness and their asymptotic rotation velocity measured using long-slit spectroscopy. Both Kourkchi et al. (2020) and Schombert et al. (2020) making use of this relation to measure the Hubble constant, and obtain a slightly higher value than SH0ES. Therefore, they are even further away from the CMB measurements.

While all these distance ladder approaches are using SNe Type Ia whose absolute luminosities are then calibrated differently, de Jaeger et al. (2020) present a measurement of H_0 using Cepheids, the TRGB, and 7 SNe Type II.

A further independent measurement is presented by Fernández Arenas et al. (2018) using H II regions, which are also standardizable candles. They obtained with this approach a value of $71.0 \pm 3.5 \mathrm{km\,s^{-1}Mpc^{-1}}$, which is one of the few measurements that is in relatively good agreement with both, the value from CMB and SH0ES due to the large uncertainty.

The best agreement of the later Universe probes to the CMB value is obtained using Gravitational Waves (GWs). While Gayathri et al. (2020) combines data of two GWs resulting in very large uncertainties, Mukherjee et al. (2020, 2021) combines those data additionally with the inclination angle inferred with the Very Large Baseline Interferometry (VLBI) to reduce the error bars. Nonetheless, the uncertainty is still relatively large, ranging from $H_0 \sim 63$ to $\sim 72 \mathrm{km\,s^{-1}Mpc^{-1}}$. Here are definitely more and better data necessary to reduce further the error bars. Mukherjee et al. (2020) estimates an amount of 200 GWs with current precision that are needed to bring the uncertainty below 1.3%, which is a quite ambitious goal.

In addition to this large collection of different H_0 measurements belonging to the late Universe, gravitational lensing is becoming a competitive method in addition to the distance ladder approach through continuously increasing and improving their data set (e.g., Bonvin et al., 2017; Denzel et al., 2021; Suyu et al., 2017; Birrer et al., 2019, 2020; Millon et al., 2020b; Liao et al., 2019, 2020; Shajib et al., 2020; Wong et al., 2020). While these are based on lensed quasars, also lensed SNe are possible as we will discuss in Sect. 1.1.3.4 (Refsdal, 1964). This is currently highly limited by the detected number of lensed SNe, but first results from Grillo et al. (2018) are in good agreement with the results of the lensed quasars.

In general, as we see directly from Fig. 1.4, most methods are consistent within their group, which means measurements using early type probes are in agreement with the CMB, while most methods using data from the late Universe are in agreement with the most recent value from the SH0ES team. This raises the question whether new physics is needed to resolve the discrepancy, or if just the uncertainties are underestimated or assumptions need to be re-considered.

1.1.3.4 Studies with lensed supernovae

Because of this debate, also significant efforts are currently spent on extending this procedure to lensed SNe. If a SN explodes in the background source galaxy that is gravitationally lensed into multiple images, also the light rays of the SN will be observed multiple times. In analogy to the time delay for quasars, the SN will appear at different times leading again to a measurable time delay. Sjur Refsdal proposed in 1964 that this time delay of strongly lensed SN could be also used to measure the expansion rate of the Universe (Refsdal, 1964). In comparison to lensed quasars, the well-understood light curves of SNe (e.g., Collett et al., 2019; Oguri, 2019; Foxley-Marrable et al., 2020; Bag et al., 2021; Bayer et al., 2021), and the relative minor distortions of microlensing on the light curves (e.g., Suyu et al., 2020; Huber et al., 2019) may help in tightening the constraints on H_0. On the other hand, recent studies showed that microlensing magnification influences significantly the intrinsic brightness of the SN and thus complicates the standardization (e.g., Huber et al., 2019, 2021a,b; Yahalomi et al., 2017).

Beside time-delay measurements to constrain the Hubble constant, observing lensed SNe can help to shed light on their early phase, since SNe beyond the local Universe are typically detected only at or after their peak luminosity. In case of a lensed SN, one can then use the first appearing image, together with a good mass model of the lens, to predict the time delay and location of the next image. Then a planned follow-up observation allows one to observe the reappearing SN image already in an earlier phase than usually. Given their typically short time delays of days to weeks, one has to react fast and especially to model the lens with the host galaxy in the given time. Therefore, lensed SNe Type Ia are especially promising systems for answering also outstanding questions about their progenitor system(s) that are debated through the past decades (e.g., Suyu et al., 2020). One scenario is the single degenerate case with a White Dwarf (WD) that is accreting mass from a nearby star until it reaches the Chandrasekhar mass limit (Whelan & Iben, 1973; Nomoto, 1982), such that the WD collapses and explodes as a SN. In another scenario the WD explodes already before reaching the Chandrasekhar mass, the so-called sub-Chandrasekhar detonations (Sim et al., 2010). A third option is the so-called double-degenerate scenario where two WDs are orbiting around each other, finally merging and thus exceeding the Chandrasekhar mass limit (Pakmor et al., 2010). While these are typical scenarios of SNe Type Ia, also early-phase observations for core-collapse SNe are important to study also their progenitor properties and test further current stellar evolution models. SNe Type II spectroscopy of the early phase are important to give constrains on the mass-loss history just before explosion.

To date, only three lensed SN are detected as the chance of a SN in a lensed galaxy is very slim. After Refsdal proposed the idea of lensed SN, it took more than 50 years for the first discovery. On November 10, 2014, a core-collapse SN forming an Einstein cross around an elliptical galaxy at redshift $z_d = 0.54$ was detected in HST images. It is now called SN Refsdal to honor his proposed technique (Kelly et al., 2015). The SN host galaxy is a spiral galaxy and lies at a redshift of $z_s = 1.49$. Since this lens system is part of the cluster MACSJ1149.5+222.3, the background host galaxy is additionally lensed through that gravitational potential into multiple images. Therefore, it was expected that also the SN will appear another time in those images, and many teams started directly to model that system to predict when and where the images would appear (e.g., Oguri, 2015; Sharon & Johnson, 2015; Diego et al., 2016; Treu et al., 2016). Finally, more than one year later, on December 11, 2015, a new image of the SN appeared in the

observations taken with HST (Kelly et al., 2016). This system was successfully used for various studies including constraining H_0 (Grillo et al., 2018, 2020).

The second observed strongly lensed SN is iPTF16geu, which is so far the first confirmed SN that is strongly lensed by a single, isolated galaxy at redshift $z_\mathrm{d} = 0.216$ (Goobar et al., 2017). Fig. 1.5 shows SN iPTF16geu observed with HST and Keck, a telescope located in Hawaii, where we can clearly see the four images of the SN type Ia, marked with white circles. Additionally, we see the host galaxy of the SN at a redshift $z_\mathrm{s} = 0.409$, which will remain after the SN faded away, leaving just a galaxy-galaxy lens. The very symmetric arrangement of the SN images implies a well aligned systems, with nearly no offset $\tilde{\eta}$ between the source and the LOS, resulting in similar path length for all four images and thus in a very small time-delay difference, making this system not ideal for measuring H_0. The alignment is supported by the large magnification, while the relative magnifications of the four SN images provide evidence for dust and /or sub-structures in the lensing galaxy.
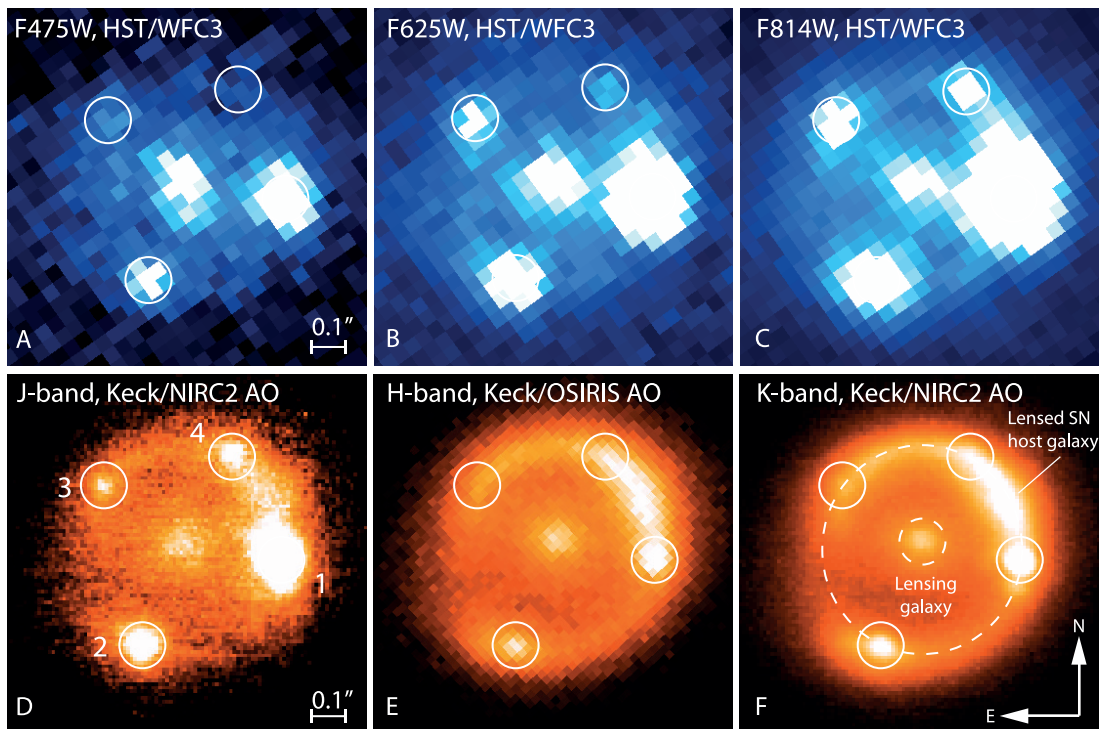
Figure 1.5: Image of the strongly lensed SN iPTF16geu observed with HST and Keck. The four images of the SN type Ia are clearly visible. Especially in the K band we can see the lensed host galaxy. From Goobar et al., 2017, Science, 356, 291. Reprinted with permission from AAAS.

The third system contains the SN AT2016jka, which is possibly a SN type Ia and got now named SN Requiem (Rodney et al., 2021). It appeared in an evolved galaxy at $z = 1.95$ and is lensed by a massive galaxy cluster. Given the huge mass of the cluster, which is mainly concentrated at the Brightest Cluster Galaxy (BCG), and the image separation, the fourth image

is predicted to appear in the year 2037 ± 2. Further model investigations will help to reduce the uncertainty, and if the SN appearance can be measured in more than two decades, this would lead to an extraordinarily precise time delay.

This list of lensed SNe shows how seldom and peculiar such events are. To investigate further in the search and study of lensed SNe, we recently initiated the Highly Optimized Lensing Investigations of Supernovae, Microlensing Objects, and Kinematics of Ellipticals and Spirals (HOLISMOKES) program (Suyu et al., 2020). In general, we follow a simple strategy to find a lensed SN: First, we detect as many as possible static strongly lensed galaxy systems and thus increase the sample of known lens candidates (Cañameras et al., 2020, 2021; Cañameras et al., in prep.; Rojas et al., 2021; Savary et al., 2021; Shu et al., in prep.). Therefore, we introduce in the following Sect. 1.2 the different lens search strategies developed and carried out by various teams. We then cross-match on a daily basis with Lasair (Smith et al., 2019), a transient alert broker, the whole list of known lens candidates with all alerts from the Zwicky Transient Facility (ZTF), an imaging survey with focus on transient detection (Bellm et al., 2019). All systems that have an alert within 5″are inspected by the cross-matcher and decided whether it is a promising lens candidate and likely a SN. If this is the case, which is unfortunately very seldom, we follow-up that system with spectroscopy and confirm the lensing nature of that candidate. In addition, we develop fast and automated modeling machineries (see Chapters 3, 4, and 5), which is the core part of this thesis, investigate into micolensing effects and algorithms to measure quickly the time delays both for SNe Type Ia (Huber et al., 2019, 2021a,b; Suyu et al., 2020) as well as SNe Type II (Bayer et al., 2021).

1.2 Searches for strong galaxy-scale lenses

Since we are very keen on increasing the sample of known galaxy-scale strong lenses to achieve our science goals and can apply the modeling pipelines presented in this thesis. Because strong lensing events are very rare with about one in 10^5 for seeing-limited ground-based data (e.g., Jacobs et al., 2019), dedicated and automated lens search programs are crucial for us to identify those systems. We give a short overview and highlight several different techniques of galaxy-galaxy scale lens searches that helped to increase the sample from only several dozen system to nowadays several hundred lenses and several thousands promising lens candidates. Especially with the current and upcoming wide-field imaging surveys, precise, fast, and automated methods are necessary that can identify lens candidates within this huge amount of galaxies. To increase the chance of being a lens, many searches focus on Luminous Red Galaxys (LRGs), which have a higher lensing cross-section (Turner et al., 1984) and smooth light profiles that help separate the foreground lens and background source emissions.

1.2.1 Pattern based searches

One of the first dedicated lens search project is the Strong Lensing Legacy Survey (SL2S), a project to search for strong lenses in the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS) sample (Cabanac et al., 2007). Cabanac et al. (2007) developed in total 3 complementary algorithms to automatically check whether lensing features are detected in the images.

The first method is an ᴀʀᴄ-detector based on a pattern recognition procedure. The second technique, called RɪɴɢFɪɴᴅᴇʀ, looks for rings in the images to detect unresolved systems, detecting blue residuals in otherwise a smooth red light distribution. The third proposed method is a multiplet detector, aimed at detecting peculiar and rare multiple arclet systems that were missed by the ᴀʀᴄ-detector. However, this method is finally not used in the paper. The RɪɴɢFɪɴᴅᴇʀ method was also successfully used in further programs (Gavazzi et al., 2012, 2014). In the same search project SL2S, More et al. (2012) present 12 group-scale lenses and 42 additional lens candidates found with the ᴀʀᴄ-detector.

1.2.2 Spectroscopic searches

Another method was proposed by Bolton et al. (2006), who looked as part of the Sloan Lens ACS (SLACS) project at spectroscopically selected galaxies from the Sloan Digital Sky Survey (SDSS) and checked if there are signs of high redshift emission lines of luminous galaxies, that could indicate a lensed object. They found in total more than 100 secure lenses or maybe lenses where further confirmation was neccessary (Bolton et al., 2006, 2008; Gavazzi et al., 2008). This made the SLACS survey to the most productive strong-lens survey at that time.

The same method was used by Brownstein et al. (2012) considering the spectra from the BOSS. They found in the BOSS Emission-Line Lens Survey (BELLS) 25 definite lenses and 11 probable lenses, which were confirmed through HST follow-up observations. Since this technique requires spectroscopic data, its applicability is highly limited to a very small fraction of observed galaxies through wide-field imaging surveys but nontheless still in use (e.g., Shu et al., 2016a; Talbot et al., 2018, 2021).

1.2.3 Search techniques using modeling

Sonnenfeld et al. (2018a) demonstrate the possibility of finding lenses though automated modeling as part of the Survey of Gravitationally-lensed Objects in HSC Imaging (SuGOHI) project. They present three complementary algorithms that were applied to Hyper Suprime-Cam (HSC) images of 43,000 LRGs detected by eBOSS. The first algorithm is called YᴀᴛᴛᴀLᴇɴs, which looks for arc-like features in the image stamp by fitting a model to it and depending on the result, it get classified as lens candidate or as non lens.

The second code is called Cʜɪᴛᴀʜ (Chan et al., 2020), which was originally developed for the search of lensed quasars (see Sect. 1.2.6). Cʜɪᴛᴀʜ requires images in four different bands $grzy$, of which it automatically pick one bluer (g/r) and one redder (z/y) bands, depending on which band has a sharper Point Spread Function (PSF). By matching these two PSFs, Cʜɪᴛᴀʜ is able to disentangle lens and arcs according to the color information to identify the lens center and the image positions. With this information, Cʜɪᴛᴀʜ models the lens image configuration with a Singular Isothermal Ellipsoid (SIE) profile. Based on the resulting χ^2, it is classified as a lens for low χ^2 values or rejected as non-lens for high χ^2 values.

The third method is following Bolton et al. (2006) as described in Sect. 1.2.2 by using emission lines in spectroscopic data. In total, Sonnenfeld et al. (2018a) find 15 definite lenses, 36 highly probable lenses and 282 possible lenses.

1.2.4 Lens searches as citizen science

With the larger imaging surveys and therefore increasing number of images, the modeling and fitting techniques get very time and resource consuming unless pre-selection and cuts on the catalog level is done. To classify a larger sample, Sonnenfeld et al. (2020) collaborated as part of the SuGOHI project with the *Space Warps* team (Marshall et al., 2016), who are conducting a citizen science project to find strong lenses. For this, Sonnenfeld et al. (2020) pre-selected $\sim 300,000$ images and included several mock lenses as well as known non-lenses that were accessible from the web and all people in the world were encouraged to help classifying them. The included mock lenses and known non-lenses were used to demonstrate the procedure and to check the performance. In parallel, they used YATTALENS, which is introduced in Sect. 1.2.3, to find lenses as well. It turned out that YATTALENS found roughly half of the number that were found by the crowd-sourcing method.

1.2.5 Machine learning based searches

Thanks to the recent wide-field imaging surveys and especially also very soon from upcoming surveys like Rubin Observatory Legacy Survey of Space and Time (LSST) and Euclid, billions of galaxy images are available in relative good quality. Within this huge amount of images, around 100,000 strongly lensed galaxies are expected (Collett, 2015). To select these specific images, one needs a fully automated and precise method, that is also fast enough to classify millions to billions of images in an acceptable amount of time. Given these requirements and the great success of Convolutional Neural Networks (CNNs) in image classification, they are perfect techniques for this task. However, since there are not enough lenses known in a specific survey, they get typically mocked up. Although the fraction in nature for a lens is quite small, the training set contains typically around 50% lenses and 50% non-lenses to give the network enough examples of lenses such that it can learn the important features of the arcs.

One of the first developed CNNs for lens identification was a so-called ensemble network performed by Jacobs et al. (2017). They applied their first network to all of the 171 deg^2 CFHTLS wide field image data (Cabanac et al., 2007), and identified 18,861 candidates including 63 known and 139 other potential lens candidates. A second search was focused on 1.4 million early-type galaxies selected on the catalog level and identified 2,465 candidates including 117 previously known lens candidates, 29 confirmed lenses/high-quality lens candidates and 266 novel probable or potential lenses.

In the same year, Petrillo et al. (2017) developed a CNN for lens identification on the Kilo Degree Survey (KiDS) sample (de Jong et al., 2015, 2017), which were applied to around 22,000 LRGs selected through a color-magnitude cut. With this network, they find 761 lens candidates, while the network is able to recover two of the three known strong lens systems.

Both teams pursued in lens recognition though CNNs; Jacobs et al. (2019) considered now data from the Dark Energy Survey (DES) and present finally a catalog of 511 lens candidates, while Petrillo et al. (2019a,b) focused still on LRGs of newer data releases of the KiDS sample.

Following this success in detection of strongly lensed galaxies, many other teams started to develop CNNs for various existing imaging surveys. For instance, Knabel et al. (2020) report a detailed comparison of spectroscopic, Machine Learning (ML), and citizen science methods for

the identification of galaxy-galaxy lens candidates. For the spectroscopic based technique data from the Galaxy And Mass Assembly (GAMA) survey were used, while for the citizen science and ML method images from the KiDS sample are considered. All three methods complement each other very well, resulting in a relative complete sample for this sky area. On the other hand, it demonstrate that a single method like a network is not able to capture nearly all lenses.

Cañameras et al. (2020) searched for lenses in the Panoramic Survey Telescope And Rapid Response System (Pan-STARRS) survey (Chambers et al., 2019) of the entire Northern sky focusing on typical high-redshift galaxies strongly lensed by massive LRGs. Due to the lower image quality of Pan-STARRS compared to previous used surveys such as HSC, KiDS, or DES, simple cuts on the photometry to select only promising candidates within the three billion sources would exclude significant fractions of interesting systems with strongly lensed arcs blended with the lens and altering its photometry. Therefore, a multi-step approach is performed to obtain a very pure sample of lens candidates. First, around 84% of the three billion sources detected in the Pan-STARRS 3π survey image stacks get excluded through conservative cuts like on their color and circular apertures, leaving still a sample of 23.1 million objects that includes 96% of the mock training systems according to the aperture magnitudes. These sample was then classified though a Neural Network (NN) on the catalog level to further reduce the number of candidates. The remaining sample contained around 1 million candidates and the corresponding *gri* images were passed through a CNN that predicted for 12,382 lens candidates a score above 0.9, containing 330 high-quality newly-discovered lenses while recovering 23 published systems selected through visual inspection.

While so far all presented networks were simple CNNs, Huang et al. (2020, 2021) made use of so called residuals neural networks presented by Lanusse et al. (2018). They apply their two networks to images of the Dark Energy Camera Legacy Survey (DECaLS) sample, resulting after visual inspection in, respectively, 335 (Huang et al., 2020) and 1210 (Huang et al., 2021) newly discovered strong lensing systems.

Rojas et al. (2021) presented again a CNN for detecting strongly lensed galaxies, focusing on LRGs of the DES sample, detecting 405 lens candidates and additionally 315 candidates that show lensing signatures but require more evidence such as higher imaging resolution or confirmation through spectroscopic data. From the 90 best lens candidates, 52 systems were selected showing only a single deflector, and modeled through an automated modeling code using traditional sampling techniques. Through this modeling, 41 systems were further justified as real lenses.

In Savary et al. (2021), a CNN committee classifier is applied to LRGs of the Canada-France Imaging Survey (CFIS) sample (Ibata et al., 2017). While for most other surveys used for lens searches each lens candidate was observed in multiple filters, typically *gri*, for the CFIS sample only *r* band data are available. This limits the network to a single-band classifier which complicates the task for the network, as color information help to distinguish between a redder lens galaxy and blue-ish arcs, coming from more distant objects. The CNN committee is finally applied to more than 2.3 million candidates, of which 9,460 obtained a score above the conservative threshold of 0.5. Through visual inspection, 133 very promising lens candidates, containing 104 completely new candidates, were found. In analogy to Rojas et al. (2021), the best 32 candidates were modeled in an autonomous way, assuming an Singular Isothermal Ellipsoid with external shear (SIE+γ_{ext}) (Kormann et al., 1994) for the lens mass and a single elliptical Sér-

sic profile for the light distribution. Finally, auto-encoders were applied to those candidates to present deblended source and light components.

There are several Deep Learning (DL) based lens search programs in the HSC data ongoing, as this survey provides, for a ground based survey, images with good quality in multiple filters and matches the expected quality of LSST quite well, making these investigations a direct preparation for LSST. Since so far several other lens search techniques were used to detect strongly lensed galaxies, this is an excellent set of data to complement the known lenses with DL techniques. Additionally, since many lenses are already known, the networks can get tested on those known systems as well as on a commonly used test set of realistic mocks shared within the different teams (More et al., in prep.).

Cañameras et al. (2021) present a residual neural network applied directly to all 62.5 million *gri* galaxy image cutouts from the HSC second Public Data Release (PDR) with an *i*-band Kron radius larger than 0.8″ to avoid strict pre-selections. This very restrictive network assign to 9651 candidates, which corresponds to 0.015% of the classified images, a score above the threshold of 0.1. Through a visual inspection, 206 newly-discovered candidates got classified as a definite or probable lens and additionally 173 known systems are recovered. The network is also able to recover 102 group- and cluster-scale lens candidates although it is not optimized for these systems. However, this demonstrates the opportunity for dedicated group- and cluster lens searches. This Residual Network (ResNet) obtains a False Positive Rate (FPR) of $\sim 0.01\%$, which is very good for the current sample size, but still high for a full image-based classification for upcoming surveys like LSST or Euclid. It turned out that many false-positives where under represented in the training sample (Cañameras et al., in prep.), such that future networks might result in even better performance. This lens search is complemented by Jaelani et al. (in prep.), carrying out also a galaxy-galaxy scale lens search in the HSC footprint. In addition, Shu et al. (in prep.) present a lens search focusing on high-redshift lens systems in the HSC sample.

While all these networks were from the supervised category, i.e. the network gets trained and tested on a specific selected data set containing lenses, typically mocked lenses, and non-lenses, there is also the option of so-called unsupervised learning which is proposed by Cheng et al. (2020) for a lens search. For this, they use a combination of a feature extractor, a convolutional autoencoder and a clustering algorithm consisting of a Bayesian Gaussian Mixture model. This network is developed using Euclid mock images and is able to recover 63% of the lensed in their sample. While this is a dedicated preparation work for the Euclid mission, previous works used real data and thus also detected new lenses that can be used for further studies. However, given that LSST is expected to have similar quality than HSC images, these developments will be extremely useful for the up coming LSST.

1.2.6 Searches of lensed quasars

Lensed quasar are especially useful to study galaxy evolution and to perform time-delay cosmography, but they are even more rare than lensed galaxies as the chance of having a bright quasar in the background source is very low. Therefore, also dedicated lens searches for strongly lensed quasar systems were carried out. The first large program targeted flat-spectrum sources in the radio wavelength range in the combination of the Cosmic Lens All-Sky Survey (CLASS) and Jodrell Bank VLA Astrometric Survey (JVAS) (Browne et al., 2003). A further large program

uses data from SDSS and is therefore called SDSS Quasar Lens Search (SQLS) (Inada et al., 2008, 2010, 2012; Oguri et al., 2006, 2008). They present an algorithm for scanning uniformly spectroscopically identified quasars by using only morphological and color selections that are generated with the standard image processing pipeline.

A further exclusive search in this field was done by Lemon et al. (2017, 2018, 2019) using data from Global Astrometric Interferometer for Astrophysics (Gaia), which is a space satellite operated by the European Space Agency (ESA). They first select candidates within the Pan-STARRS footprint that have either quasar-like WISE colors or photometric indications of quasars from SDSS, requiring either multiple detection in Gaia or a single Gaia detection near a morphological galaxy. Through modeling and follow-up observations, they confirm finally 24 (Lemon et al., 2018) and 22 (Lemon et al., 2019), respectively, lensed quasars.

Another specialized lensed quasar detection algorithm is Cʜɪᴛᴀʜ presented in Chan et al. (2020), which was also successfully used to find galaxy-galaxy lens systems (see also Sect. 1.2.3 Sonnenfeld et al., 2018a). They present 46 quasar lens candidates of which 3 are previously known. A subset of six candidates got additionally confirmed with X-Shooter spectra and got modeled uniformly with Gravitational Lens Efficient Explorer (GLEE), a widely used and well tested lens modeling software (Suyu & Halkola, 2010; Suyu et al., 2012).

A complete different procedure is presented in Chao et al. (2020a), proposing to use time variability of lens systems in the HSC transient survey to obtain a sample of lensed quasar candidates to which Cʜɪᴛᴀʜ is applied. After visual inspection seven lensed quasars, of which one was previously known, and one galaxy-galaxy lens were discovered (Chao et al., 2020b). Since also LSST will provide images of different cadence, this difference imaging method would also work for LSST.

1.3 Imaging surveys

As set out in Sect. 1.2, the galaxy-scale lenses are very seldom with one out of 10^5 (e.g., Jacobs et al., 2019) and thus dedicated lens searches are crucial to detect them. Thanks to many different ongoing and upcoming wide-field surveys, images of billions of astronomical objects are available enabling a large variety of astronomical studies, beside gravitational lensing. To identify many gravitational lenses with those algorithms introduced in Sect. 1.2, a wide-field imaging survey with relatively good image quality is preferred to cover a wide area of the sky and with good enough quality to resolve the arcs or multiple lensed images. On the other hand, to detect short-lived transients like SNe, a good cadence is crucial for the difference imaging. For detailed analysis, discovered peculiar objects are then often followed-up with space based telescopes like HST (e.g., Kinney & Maran, 1991, compare Fig. 1.2 and Fig. 1.5) that provide much better resolution ($\sim 0.04''$/pixel, depending on the filter) and therefore complement perfectly the wide-field imaging surveys.

Since these wide-field surveys are crucial for our science cases and we extensively use imaging data in this thesis, we give a short overview of current and upcoming imaging surveys. We highlight specifically HSC, a wide-field imaging surveys which we use in this thesis.

1.3.1 Current wide-field imaging surveys

In the last two decades, we experienced great changes in astronomy through large surveys including ground based wide-field imaging surveys in the optical wavelength regime. One of the first large programs to image the sky is the SDSS (York et al., 2000; Stoughton et al., 2002), which started already in the year 2000 and stopped imaging in 2009 but continued to take spectra until this day. It is using a dedicated 2.5-m wide-angle optical telescope installed at the Apache Point Observatory in New Mexico, United States. During these years, it covered around 35% of the sky and provided nearly 1 billion of images in five optical filters (g, r, i, z, and y). This was complemented by the CFHTLS (Cabanac et al., 2007), which observed at the same time (2003-2009) from Hawaii, United States of America (Boulade et al., 2003; Cuillandre et al., 2012; Hudelot et al., 2012). For this, CFHTLS used the MegaCam camera, which is able to cover a full 1×1 square degree field of view with a resolution of $0.187''$/pixel.

The same camera is now used since 2017 by the CFIS (Ibata et al., 2017), which images ~5,000 square degree of the northern sky in the r band and ~ 10,000 square degrees in the u band. This survey obtained again observing time and thus will last at least until 2022, but is most likely further extended through the newly founded collaboration between CFIS and the Pan-STARRS, another so far independent imaging survey operating since 2010. Pan-STARRS is using two wide-field 1.8 m Ritchey–Chrétien telescopes located at Haleakala in Hawaii, United States of America (Chambers et al., 2019). It reaches with the single-epoch images currently a magnitude of around 21, depending on the filter and is therefore perfectly suited for the discovery of near-earth objects.

Another imaging program dedicated for transients is the ZTF located at the Palomar Observatory in California, United States (Bellm et al., 2019). It started in 2017 and is a direct continuation of the Palomar Transient Factory (2009-2017) with an augmented camera. With its huge field of view of 47 square degrees, ZTF is able to image the entire northern sky in around three nights in one of the supported three filters g, r or i. It has an pixel size of $1''$, and is therefore perfectly suited to detect transients with rapidly changing brightness like SNe or gamma ray bursts, collision between two neutron stars, and moving objects like comets and asteroids.

There are many other surveys dedicated to image static objects. For instance the DECaLS, which started in 2014, used the Dark Energy Camera (DECam) on the Blanco 4m telescope, located at the Cerro Tololo Inter-American Observatory. Originally planned for three years, it finally were completed in March 2019. It was specifically operated to provide the optical imaging for the Dark Energy Spectroscopic Instrument (DESI) footprint, a survey focusing on spectroscopy.

These are expanded by various additional large surveys, which also cover other wavelength ranges such as the infrared, the radio, or gamma ray regime. Several surveys like the the KiDS (de Jong et al., 2013a,b) and DES (Flaugher et al., 2015; Abbott et al., 2018b; Morganson et al., 2018) combine also different regimes like optical and Near-IR (NIR). Those surveys focus typically on the covered sky area to provide billions of images with resolution of around $0.1''$ or larger, but because of the earth atmosphere normally not reaching the quality of space based telescopes, that are best suited for specific follow-up observations.

However, nowadays, also ground-based telescopes are competitive with space based telescopes like HST by using a technique called Adaptive Optics (AO). This technique is for instance

now available at the Keck telescopes (Medwadowski, 1989; Herbst et al., 1992; Wizinowich et al., 2006) located near the summit of Mauna Kea at Hawaii, United States. To achieve this, the mirror of the telescope get deformed in order to compensate the distortions introduced by the earth atmosphere to a guide star. Since the science target is typically too faint to act as the reference star, a nearby brighter star can be used as a guide star as it passed through nearly the same atmospheric turbulence (e.g., Medling et al., 2021; Nielsen et al., 2017). This however, limits the applicability of an AO system as a bright enough guide star is not always nearby. The further away the guide star is from the scientific target, the more different is the atmosphere between those two locations, which results in decreasing correction quality. This severely limits the application of this technique for astronomical observations. Another major limitation is the small field of view over which the AO correction is good. This lead to the development of multiconjugate adaptive optics where several deformable mirrors are combined to increase the field of view (e.g., Rigaut et al., 2014; Neichel et al., 2014; Johnston & Welsh, 1994).

To increase the number of systems where AO can be used, also a laser beam as an artificial guide star got installed for instance at the Keck observatory and is nowadays frequently used (e.g., Chen et al., 2016, 2021c,b; Goobar et al., 2017; Vayner et al., 2021). This technique allowed to observe the gravitational lensed SN iPTF16geu shown in the bottom panel of Fig. 1.5 (Goobar et al., 2017). The basic idea is to shoot the laser into the sky near the scientific target such that the back-scatter light, which traveled through nearly the same atmosphere, can be collected (e.g., Huang et al., 2022). Based on the received light, the mirror get deformed to compensate the atmospheric turbulence.

1.3.2 Upcoming surveys and facilities

Those surveys and many others, especially also from smaller telescopes, provide billions of images allowing a high variety of different astrophysical studies. This will reach soon another stage when the next generation of telescope start to operate and provide additional billions of images in even better quality. There will be Euclid (Laureijs et al., 2011), a new satellite mission from the ESA, which shall be launched in 2023. It will provide a huge amount of images with a resolution of $\sim 0.1''$/pixel dedicated for studies on the DE and DM content of the Universe. Moreover, it will provide insightful information on the physics of the early universe and on the initial conditions which seed the formation of the LSS. There is the James Webb Space Telescope (JWST) (Gardner et al., 2006), another satellite mission planned to be launched even before Euclid, in December 2021. It will operate around ten years and conduct high-quality data in the NIR wavelength range.

To complement these space-based telescopes, there are also several ground-based telescopes under construction. On one hand, the European Southern Observatory (ESO) is building a new Extremely Large Telescope (ELT) (Evans et al., 2015) with a primary mirror diameter of 39 meters, composed out of 798 segments. The ELT will therefore provide images with incredibly good quality, especially as it also supports AO imaging. While the ELT will provide high-quality images for specific objects, the LSST (Ivezic et al., 2008), which is also currently under construction, is expected to start operating in 2024 and provide images with an expected resolution of around $0.2''$. Although the resolution is lower than from the other new telescopes, it will photograph the entire reachable southern sky of around 18,000 square-degrees every few nights.

Moreover, it will provide color information though observations in six optical and NIR filters, whereas e.g., Euclid only supports one filter and thus give no color images. A design image of the dome with a cutaway to show the telescope is displayed in Fig. 1.6. We also show a enlarged cutout of the telescope (blue box), which will get a primary mirror with a diameter of 8.4 meters and which will be equipped with a novel three-mirror design. This allows to deliver relatively sharp images over a very wide 3.5-degree diameter field of view. The human, which is shown enlarged in the red box, demonstrates the huge size of the LSST telescope and dome.

Because of its high cadence and large area, it will be perfectly suited for the discovery of transients. It will surpass the resolution of Pan-STARRS, the current best detector for near-earth objects although its coverage is with 30,000 square degrees larger than the expected area covered by LSST. The ZTF, operating since 2018, is a similar wide-field survey dedicated for transient detection. Its field of view is more than 4.5 times that of LSST, while LSST will have a $\sim$ 30 times larger aperture. Since ZTF is reaching typically a magnitude of 21, LSST is expected with a limiting magnitude of around 24 per epoch (Huber et al., 2021a) to provide soon many more objects in even better quality then ZTF and to reach and even outperform after few years of operation HSC, the current best ground based wide-field imager, in terms of resolution. Nonetheless, both will be comparable such that the image reduction pipeline for LSST got adapted from that of HSC. Therefore, LSST which will cover the southern sky from Chile, will perfectly complement the HSC images of the northern sky.

Over its 10-year survey, LSST is expected to observe billions of images in six different filters, of which $\sim 100,000$ are expected to contain strongly lensed galaxies (Collett, 2015). This will increase the current sample of known galaxy-scale lenses by nearly two orders of magnitude, allowing a statistical analysis of galaxy properties and many other science applications. Among these lenses, forecasts predict 290 lensed SNe (Craig et al., 2021; Goldstein & Nugent, 2016; Goldstein et al., 2019; Wojtak et al., 2019) to be discovered by LSST. This will increase the current sample extremely, since, as described in Sect. 1.1.3.4, only three lensed SNe, two cluster-scale lenses and one galaxy-scale lens, are detected so far. This enables another stage of astronomical studies including precise cosmology through a sample of gravitational lensing SNe which is not possible until then. Given that LSST is starting soon its survey, it is timely to develop and test dedicated tools to handle and analyze this huge amount of data. This thesis will bring us a major step forward to achieve these ambitious goals.

1.3.3 The Hyper Suprime-Cam Subaru Strategic Program

For a newly developed software or method that is expected to be applied to any kind of specific data, for instance astronomical images, it is crucial that the software or method is tested first on a real or realistic data set with known output. Therefore, we will extensively use data collected and provided by the HSC Subaru Strategic Program (HSC-SSP, throughout the thesis just HSC) for the developments described in this thesis, .

The HSC survey (Aihara et al., 2018a,b) is a large three-layered multi-band imaging survey using the 8.2-meter Subaru telescope on the summit of Maunakea, Hawaii. The installed camera, the Hyper Suprime-Cam (Miyazaki et al., 2018; Komiyama et al., 2018; Kawanomoto et al., 2018; Furusawa et al., 2018), is a three meter tall giant digital camera. The focal plane of the camera consists of 116 Charge-Coupled Device (CCD) chips with each 2048×4096 pixels, of
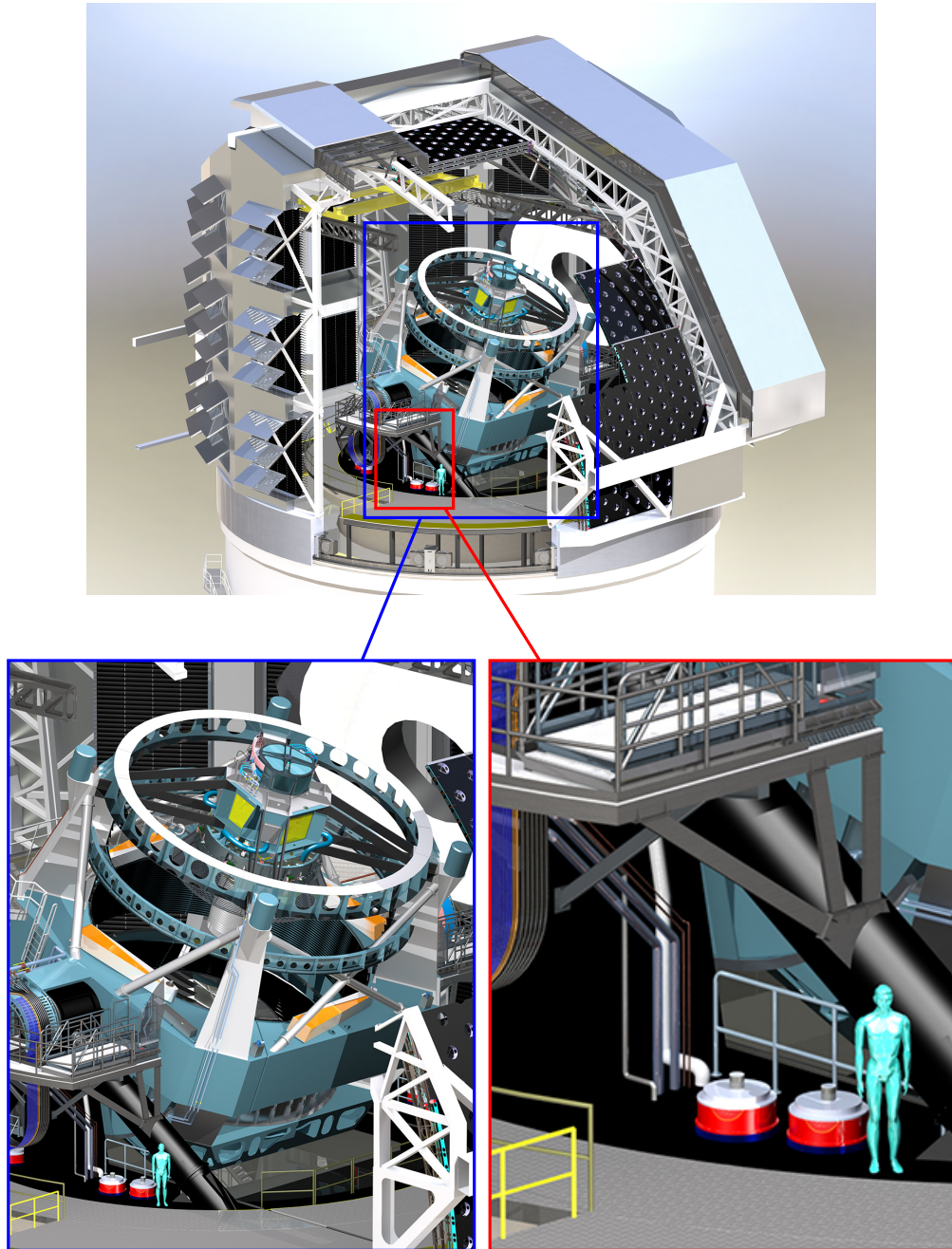
Figure 1.6: A three dimensional design image of the LSST dome with a cutaway to show the telescope facility (top panel) which is currently under construction and is expected to start operating in 2024. Below we show an enlarged image cutout of the telescope (left, blue box) and human (right, red box) that demonstrate the size of the facility. Image downloaded on Nov. 30, 2021 from `https://www.lsst.org/gallery/dome-and-telescope`. Image Credit: Rubin Obs./NSF/AURA.

which four chips are used for auto-guiding and eight chips for monitoring the focus. This results in 114 CCD chips that are used for real science observations. The camera has a field of view of 1.8 square degrees which is equivalent to 1.5 degree in diameter and able to obtain images with a resolution of 0.168″/pixel.

The survey contains three layers, Wide, Deep and UltraDeep, as indicated in Fig. 1.7. Here we show the covered area of sky in the PDR 3, where the blue regions specifies the area covered by the Wide layer, and the green regions the parts covered in the Deep and UltraDeep depth. The three red boxed indicate the regions of the final Wide survey, as the area denoted as AEGIS was only observed for calibration. Therefore, HSC imaged 1470 square degrees of the northern hemisphere with at least one filter in the Wide layer (PDR3, Aihara et al., 2021). The different shades of blue indicate the number of broad band filters in which data are available, with the darkest blue corresponding to all five filters g, r, i, z, and y. These filters are classified as broad band filters with a covered wavelength range of more than 300Å, while filters covering a range of 100Å to 300Å or smaller than 100Å are called intermediate and narrow-band filters. The choice of the filters depend on the scientific goals of the survey. While narrower passbands lead to more detailed information about the Spectral Energy Distributions (SEDs) of the observed galaxies, broad band filters increase the sensitivity of the telescope and thus allow also to detect fainter objects. The filters used by HSC for the Wide layer follow those from SDSS (Kawanomoto et al., 2018), which are commonly adopted for a good combination of different data sets. We list the exposure time, the obtained seeing, depth, saturation and covered sky area for each individual filter in Tab. 1.1, archived since 2014 over the 300 nights allocated for the HSC program.

	g	r	i	z	y
exposure (min)	10^{+2}_{-2}	10^{+2}_{-2}	20^{+3}_{-6}	20^{+3}_{-10}	20^{+3}_{-10}
seeing (arcsec)	$0.79^{+0.09}_{-0.08}$	$0.75^{+0.13}_{-0.09}$	$0.61^{+0.05}_{-0.05}$	$0.68^{+0.08}_{-0.06}$	$0.68^{+0.10}_{-0.08}$
depth (mag)	$26.5^{+0.2}_{-0.2}$	$26.5^{+0.2}_{-0.2}$	$26.2^{+0.2}_{-0.3}$	$25.2^{+0.2}_{-0.3}$	$24.4^{+0.2}_{-0.3}$
saturation (mag)	$17.4^{+0.6}_{-0.4}$	$18.1^{+0.5}_{-0.5}$	$18.3^{+0.5}_{-0.3}$	$17.5^{+0.5}_{-0.4}$	$17.0^{+0.5}_{-0.7}$
area (square degree)	1332	1298	1264	1299	1209

Table 1.1: For the Wide layer approximated exposure time, seeing, 5σ depth, and saturation magnitudes for each of the five broad band filters, averaged over the area included in the PDR2 (Aihara et al., 2019), from which the used data are. Note that, apart from the area, which is the total area covered in at least one exposure, the provided numbers are the median and quartiles of the distribution and not median and 1σ values as throughout the rest of the thesis.

In addition to these five broad band filters, HSC also covers four narrow bands (NB387, NB816, NB921, NB101). We show the bandpasses for all nine filters in Fig. 1.8, including the reflectivity of all mirrors, transmission of all optics, the atmosphere, and response of the CCD chips, assuming an airmass of 1.2. The bottom panel shows the sky emission line spectrum, from which we can see that the narrow-band filters lie in relatively dark regions of the sky spectrum.

In this thesis we use images from the Wide layer observed in g, r, i, and z in Chapters 3, 4, and 5. In Chapter 6 we additionally make use of the y band. These filters, but also the image quality and depth matches very well the expected first-year data from LSST. In addition, the image processing pipeline of HSC (Bosch et al., 2018) is a branch of the LSST pipeline. Therefore,
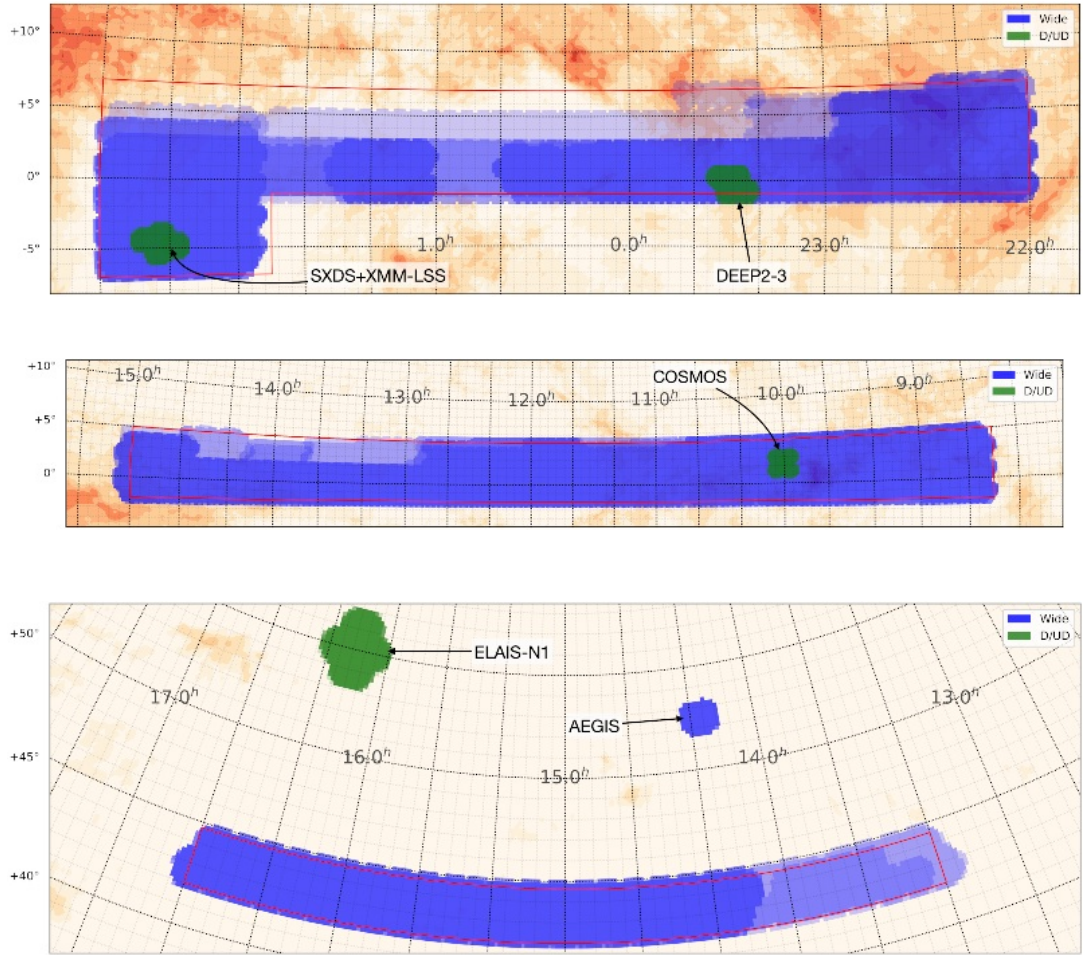
Figure 1.7: Covered area of the sky by HSC. The blue areas show the Wide layer, where the shading indicates the amount of broad band filters g, r, i, z, and y in which images are available. The green regions indicate the Deep/UltraDeep layers. In the background, the Galactic extinction map from Schlegel et al. (1998) is shown. Republished with permission of Oxford University Press-Journals, from "Second Data Release of the Hyper Suprime-Cam Subaru Strategic Program", Aihara et al., PASJ, 106, 2019; permission conveyed through Copyright Clearance Center, Inc.
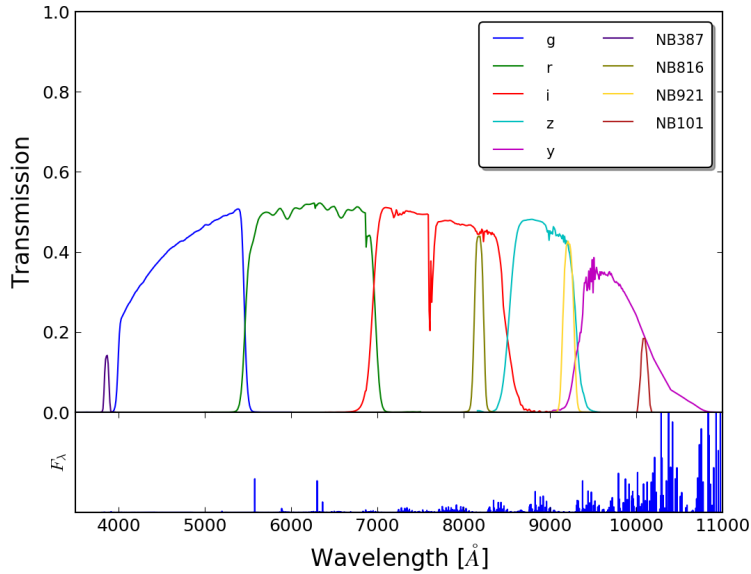
Figure 1.8: Transmission curves for the five broad band and four narrow band filters supported by HSC, including the reflectivity of all mirrors, transmission of all optics, the atmosphere, and response of the CCD chips, assuming an airmass of 1.2. The bottom panel shows the sky emission line spectrum. From this we can see that the narrow-band filters lie in relatively dark regions of the sky spectrum. Figure from Aihara et al. (2018a).Republished with permission of Oxford University Press-Journals, from "The Hyper Suprime-Cam SSP Survey: Overview and survey design", Aihara et al., PASJ, 70, S4, 2018a; permission conveyed through Copyright Clearance Center, Inc.

our results and the network performances obtained with HSC data are expected to hold also for LSST. Nonetheless, we propose to train again dedicated networks, by using the pipelines presented in this thesis, as soon as real LSST data are available to minimize effects due to small differences.

1.4 Redshifts

While the surveys introduced in Sect. 1.3 are providing an immense amount of images in different wavelength ranges and resolutions, the full exploitation of these images often require also the distance to those imaged objects. This holds not only for gravitational lensing applications mentioned in Sect.1.1, but also for many other astrophysical studies.

1.4.1 The need of redshifts

Until the early twentieth century, astronomers believed that the universe is static and its age infinite, such that a measured distance will stay the same over time. The first indication for

an expanding Universe came from Vesto Slipher, an astronomer from the Lowell Observatory, who precisely measured the radial velocities of distant galaxies including the Andromeda nebula (e.g., Slipher, 1913, 1915, 1917a,b, 1918, 1921). In the same years, Albert Einstein developed the theoretical foundation for describing an dynamical Universe with his field equations of GR (Einstein, 1908, 1915). He then introduced the cosmological constant Λ to counterbalance gravity and guarantee a static Universe, which was at that time still a requirement for the theory (Einstein, 1917). Built upon this, Friedmann (1922, 1924) demonstrated the possibility of an expanding Universe, with the consequence that all objects should move away from us. By that time, the data set of Vesto Slipher contained 41 objects of which 36 appeared to move away from us providing a first demonstration of the expansion.

The other main discovery for the breakthrough was the discovery of the relation between period and luminosity of Cepheids, which are very bright pulsating stars (Leavitt, 1908; Leavitt & Pickering, 1912). Originally discovered by Henrietta Leavitt, Edwin Hubble started to determine the distances and absolute magnitudes to several Cepheids in the Andromeda galaxy and other nearby galaxies (Hubble, 1925a,b).

However, Georges Lemaître, an Belgian astrophysicist, was the first person who combined both discoveries, the distances of Cepheids and the spectral shifts of their galaxies (Lemaître, 1927). This lead to the clear evidence that more distant objects are moving faster away from us, which implies that we are not in a static universe. To demonstrate this, Hubble measured the distance D to 24 galaxies and compare to their velocities. As we can see from Fig. 1.9, where Hubble plotted the distance against the recession velocity, the relation is linear. Therefore, Hubble defined the famous Hubble-Lemaître law

$$v = H_0 D \,. \tag{1.4}$$

Lemaître estimated the Hubble constant H_0, which we have introduced already in Sect. 1.1, to be $\sim 680 \mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$ (Lemaître, 1927) and Hubble to be $\sim 500 \mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$ (Hubble, 1929), which are both much higher than current measurements.

Further investigations also demonstrated that the ratio between velocity v and distance D is not constant, such that Eq. (1.4) got generalized with a Hubble parameter H with H_0 as the value of the present day. This implies that the universe is not only expanding, it is expanding with an accelerating speed. By introducing the scale factor a and its time-derivative $\dot{a}$, the Hubble parameter can be expressed through

$$H(t) = \frac{\dot{a}}{a} \,. \tag{1.5}$$

The scale factor can be interpreted as the ratio of the distance between two objects at some time in the past, to the distance now and therefore represents the expansion of space. This means the wavelength of the emitted photons of a distant object get shifted towards redder wavelength as they travel through an expanding space. In general, this is a relativistic effect that can be interpreted as the stretching of the electromagnetic waves (increase in wavelength) as they travel though a stretching space. Therefore, as demonstrated by Vesto Sliper (Slipher, 1913), we can measure the distance to objects by taking their spectra and comparing the wavelength λ of a specific spectral feature to the originally emitted rest-frame wavelength λ_0, resulting in the cos-
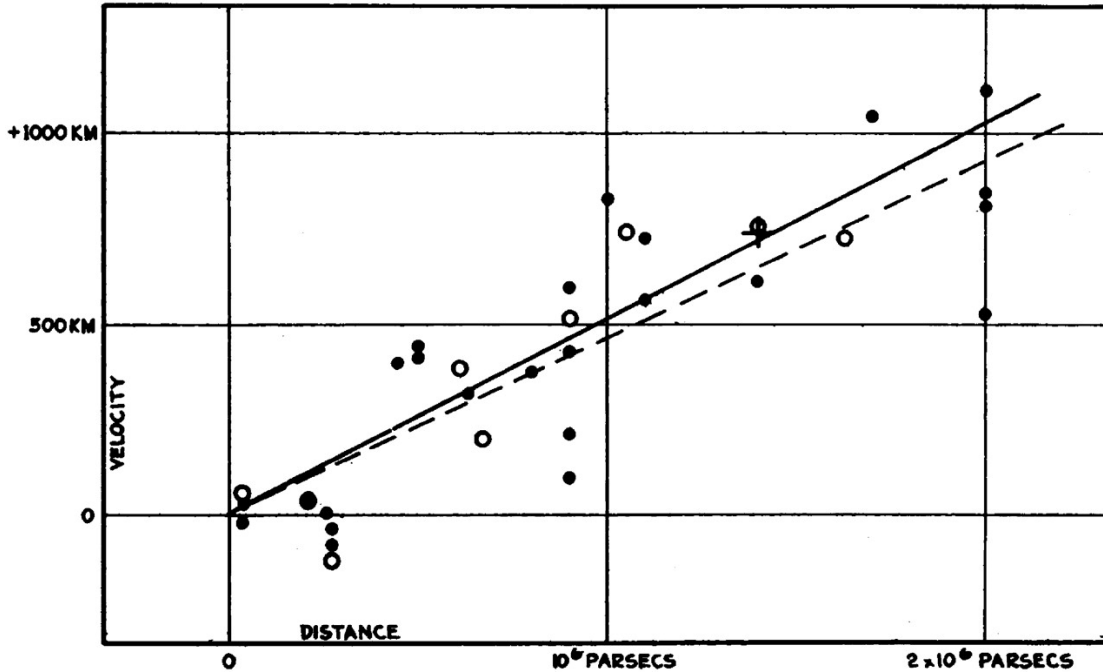
Figure 1.9: Original Hubble diagram, showing the relation between velocity and distance of extra-galactic galaxies. Filled circles and the solid line represent the solution for solar motion using the individual galaxies, while the open circles and dashed line display values of groups galaxies. The cross represent the mean velocity corresponding to the mean distance of 22 galaxies whose distances could not be estimated individually. Figure taken from Hubble (1929).

mological redshift

$$z = \frac{\lambda}{\lambda_0} - 1 = \frac{1 - a(t)}{a(t)} \,. \tag{1.6}$$

Given a cosmological model, these redshifts can be then translated to a physical distance.

1.4.2 Spectroscopic redshifts

The redshift of a specific astrophysical object such as a galaxy can be measured with a spectrograph. The obtained spectrum are predominantly formed by the approximate blackbody spectra of the individual stars the galaxy contain, producing a relatively flat spectrum. However, elements in the atmosphere of stars as well as hydrogen gas occupying the interstellar medium cause specific emission and absorption lines, as well as other prominent features such as breaks in the spectrum, which are sudden drops in the flux.

By locating these features in the spectrum, one can extract the corresponding observed wavelength λ. Using Eq. (1.6), these wavelength can be compared to the originally emitted rest-frame wavelength λ_0 to obtain the redshift. This is in general a straightforward process to obtain very

precisely the redshift of a galaxy. The main difficulty is during the identification of the emission or absorption lines, as a mis-identification can lead to a completely wrong redshift estimate.

Beside absorption and emission lines, there are in general two very prominent breaks known. One is the Balmer break, which is composed of two separate breaks that occur very closely and are therefore observed as one break. The shortest wavelength of the Balmer series, the Balmer limit, is at a wavelength of 3646Å although the mentioned Balmer break falls at around 4000Å caused by a number of metal elements in stellar atmospheres. Since the metal content of stars depend on the SFR and age of the galaxy, the Balmer break is only observed in the optical and NIR wavelength ranges for galaxies at $z \lesssim 1$.

Another prominent break is the Lyman break introduced by the hydrogen gas. It starts at the Lyman limit of 912Å, which is the ionisation wavelength of neutral hydrogen gas, and continues then towards smaller wavelength. Because of its low rest-frame wavelength, it is observed in the optical range for galaxies between $3 \lesssim z \lesssim 4$ and therefore important to measure the redshift of very distant galaxies. For closer objects, the Lyman break is observed in the Ultraviolet (UV) range and therefore can also be helpful for these objects.

Since it is crucial to have redshift measurements of scientific objects for further applications, there are in analogy to the imaging surveys introduced in Sect. 1.3 also dedicated surveys for spectroscopic observations. For instance WiggleZ (Drinkwater et al., 2010) conducted a spectroscopic survey between 2006 and 2011 targeting specifically UV bright emission line galaxies. One year later, in June 2007, the observatory Xinglong in China started the Large sky Area Multi-Object fiber Spectroscopic Telescope (LAMOST) program (Zhao et al., 2012; Luo et al., 2015) to obtain spectra. Their main scientific aspects include studies of the LSS, to find metal-poor stars in the galactic halo, and to provide information on the structure of the Milky Way. Originally planned as a 5 year survey, they obtained so far around 10 million spectra of milky way stars as well as millions spectroscopic redshifts, also called spec-z, for galaxies.

One of the largest spec-z surveys is the BOSS and the eBOSS program (Dawson et al., 2013) as part of SDSS. These surveys provide the redshift measurements of additional millions of objects with a peak in the redshift distribution at $z \sim 0.5$ for galaxies and a small sample of quasars in the range $1 \lesssim z \lesssim 3$. To expand the sample of objects with $z > 1$, the survey DEEP2 and its extension DEEP3 were carried out (Davis et al., 2003; Davis et al., 2007; Newman et al., 2013; Cooper et al., 2011; Zhou et al., 2019). They used the 10-meter class Keck telescopes to obtain around 50,000 redshifts.

There are many more surveys like the VIMOS-VLT Deep Survey (Le Fèvre et al., 2003; Le Fèvre et al., 2013) operated by ESO, DESI conducting spectra of tens of millions galaxies and quasars in the optical through scanning 14,000 square degrees of northern and southern sky (DESI Collaboration et al., 2016), zCOSMOS (Lilly et al., 2009), UDSz (Bradshaw et al., 2013; McLure et al., 2012), 3D-HST (Skelton et al., 2014; Momcheva et al., 2016), VVDS (Le Fèvre et al., 2013), VIPERS (Garilli et al., 2014), GAMA (Liske et al., 2015), PRIMUS (Coil et al., 2011; Cool et al., 2013), and many others.

Nonetheless, there are many more objects imaged of which no spec-z is available. This is mostly because spectroscopic measurements require comparably long exposure times in order to achieve sufficient signal-to-noise ratio over a wide wavelength range. Therefore, spectroscopic redshifts are much more expensive than photometric measurements making it infeasible to obtain a spec-z measurement for all objects for deep observations and for large areas of the sky. As an

example, although SDSS (York et al., 2000) has taken millions of spectra of galaxies to high precision (Ahumada et al., 2020), the same survey conducted much more detailed photometric data of galaxies in considerably less time. The objects with only photometry will grow even more when the new upcoming surveys introduced in Sect. 1.3.2 start to collect data.

1.4.3 Photometric redshifts

As a result of the huge amount of objects with only photometric observations, the idea came up whether also redshifts can be extracted from photometric data directly. The possibility was first demonstrated by Baum (1962), while Koo (1985) and Loh & Spillar (1986) were the first ones that computed photometric redshifts using data observed with CCD chips. In general, these techniques are tested on objects with both photometric data and spectroscopic data to compare the predicted photometric redshift z^{pred} to the spectroscopic redshift, acting as a reference redshift z^{ref}.

Since then, a huge amount of effort is spent on developing new and better techniques especially with the large wide-field imaging programs (see Sect. 1.3) carried out in the past two decades. The main requirement is the availability of multi-band data to provide color information to the photo-z code, which can also be obtained in different surveys or with different telescopes to cover a larger wavelength range. This helps to improve the accuracy and precision of the photo-z estimations. An example is the COSMOS2015 catalog from Laigle et al. (2016), which we also use in Chapter 6. They combined photometric observations from ~ 30 different intermediate or narrow filters spanning the UV, visible, NIR and mid-infrared wavelength range to compute very accurately photo-z values for around half a million galaxies of the 2 square degree COSMOS field.

Nonetheless, spectroscopic surveys are still necessary as this is the more accurate technique and spec-z are the reference redshifts for photo-z techniques. On the other hand, photometric redshift routines are more efficient in terms of obtaining galaxy redshifts per unit telescope time (Hildebrandt et al., 2010) and also applicable to objects that are too faint for spectroscopy (e.g., Abdalla et al., 2011; Blake et al., 2007; Connolly et al., 1995; Koo, 1985; Li et al., 2007; Oyaizu et al., 2008). In addition, for objects in the redshift range of $1.4 \lesssim z \lesssim 2.5$ the Balmer break and strong emission lines are shifted into the NIR range of the spectrum making it difficult to obtain spectroscopic redshifts in case only a spectrum in the optical range is available, where most of the wide-field surveys operate. Therefore, this redshift range is called the "redshift desert" (e.g., Steidel et al., 2004; Renzini & Daddi, 2009) where most spec-z samples lack. This means that also mixed surveys, conducting both photometric and spectroscopic data, like SDSS benefit from photo-z estimations. This makes, however, also tests of photo-z methods in that range more difficult and resulted in significant work on obtaining spectra for galaxies at all relevant regions of color-space (e.g., Masters et al., 2015).

Even with much fewer bands than used for the COSMOS2015 catalog, and also when only broad band filters are available, good photo-z estimates can be computed. Thus much more photo-z estimates are nowadays available than spec-z measurements. This is becoming even more crucial with the new upcoming surveys like Euclid and LSST, providing billions of high quality images without spectroscopic measurements. Both collaborations set very stringent criteria on the photo-z errors to accomplish their scientific goals, mostly triggered by the

weak lensing requirements. For example, Euclid requires a Root Mean Square (RMS) error $\sigma_{RMS} \leq 0.05(1 + z_{pred})$, a catastrophic outlier fraction $f_{outlier}$ below 10% at a cut of 3σ and an error in the mean redshift bin N_{bin} smaller than 0.002 (Laureijs et al., 2011). LSST requires the same performance on the σ_{RMS} and outlier fraction, while allowing a maximum bias of 0.003 (LSST Science Collaboration et al., 2009). Both surveys quote even more stringent values as their goal in addition to requesting an error estimation per system to similar accuracy.

Over the years, many different methods were developed to compute photometric redshifts. However, they all can be categorized broadly into two groups, either template fitting algorithms or empirical training algorithms using machine learning. We will briefly introduce both groups in the following.

1.4.3.1 Template fitting techniques

The original procedure of photometric redshift estimation proposed by Baum (1962) is based on observations in optical broad band filters, which allow to extract characteristics such as colors and magnitudes. These are then compared to those extracted from a set of possible SEDs, which shows the observed energy plotted against the wavelength λ.

These SED templates used for that comparison can be now either empirical or synthetic SEDs of typical galaxies. Empirical templates means they are based on real observed galaxies (e.g., Assef et al., 2010; Mannucci et al., 2001; Duncan et al., 2018a). The main problem of these empirical templates is that most observations are from local galaxies which span a relatively limited volume of the parameter space and thus limiting the templates available for fitting to observations. Therefore, often synthetic templates are used (e.g., Kriek et al., 2009; Benítez, 2000; Brammer et al., 2008), which are obtained from theoretical models of stellar population synthesis and therefore can cover the full parameter range. These parameters include the SFR, metallicity, initial mass function, interstellar reddening, flux decreases due to the Lyman alpha forest, and the limiting magnitude of each filter (e.g., Bolzonella et al., 2000). The best template is then determined through a χ^2 minimization of the difference between available template and that of the observed object. From the obtained best fit template, the corresponding redshift can be inferred.

These techniques, however, are not exempt from mis-predictions for instance due to measurement errors on the survey filter transmission curves or color-redshift degeneracies. While an advantage of template fitting techniques is that they can be applied also to fainter objects where spectral information are unachievable, template techniques are typically less reliable at high redshifts where the uncertainties in galaxy SEDs increases due to calibration on low redshift galaxies and then extrapolation. Another problem can be an incomplete template library, leading to a mis-match when fitting the observed magnitudes or colors to template SEDs. On the other hand, a library with too many templates is also disadvantageous as it possibly results in colour-redshift degeneracies (Benítez, 2000).

Therefore, template fitting algorithms are sometimes combined with Bayesian techniques. This means a set of galaxies with known spectroscopic redshifts and similar color-magnitude parameters act as a prior range for the template fitting process which lead often to improved results (e.g., Benítez, 2000; Ilbert et al., 2006; Feldmann et al., 2006)

1.4.3.2 Photo-z through machine learning

Beside the option of SED template fitting, it is also possible to use directly a sample of galaxies with known and trustworthy, e.g., spectroscopic, redshifts to calibrate an algorithm that can be quickly applied to new photometric observations. These algorithm was originally just a polynomial function between colors and redshifts (e.g., Connolly et al., 1995; Brunner et al., 1997; Wang et al., 1998) and provided surprisingly good predictions and thus enables a complete different branch of photo-z predictions.

With the increasing amount of data from both photometric and spectroscopic surveys and the great success of machine learning techniques, this procedure got extended. A lot of work was spent in recent years on these developments, resulting in a high variety of different procedures, including artificial neural networks using Fully Connected (FC) layers (e.g., Collister & Lahav, 2004; Bonnett, 2015), boosted decision trees (e.g., Gerdes et al., 2010; Henghes et al., 2021a), Gaussian process (e.g., Way et al., 2009; Bonfield et al., 2010; Gomes et al., 2018; Soo et al., 2021), kernel regression (Wang et al., 2008), k-nearest neighbors (e.g., Ball et al., 2007, 2008; Lima et al., 2008; Oyaizu et al., 2008; Henghes et al., 2021a), nonsequential neural network (de Diego et al., 2021), Quasi Newton Algorithm (e.g., Cavuoti et al., 2012; Brescia et al., 2014; Razim et al., 2021), random forest (e.g., Carliles et al., 2010; Carrasco Kind & Brunner, 2013; Li et al., 2021; Henghes et al., 2021a), spectral connectivity analysis (e.g., Freeman et al., 2009), support vector machines (e.g., Han et al., 2016; Wadadekar, 2005; Wang et al., 2008), and many others (e.g., Schmidt & Lipson, 2009; Krone-Martins et al., 2014).

Instead of using a single type of architecture, there are also nowadays codes combining different types of networks (e.g., Duncan et al., 2018b; Soo et al., 2018; Pasquet et al., 2019; Leistedt et al., 2019) which lead often to an improved performance. While these methods are all supervised, i.e., trained and tested on a specific set of objects with known redshifts, there is also the possibility of unsupervised machine learning (see also Sec. 1.5) using e.g., the so-called Self Organized Maps (SOM) (e.g., Geach, 2012; Way & Klose, 2012; Razim et al., 2021). These unsupervised networks perform similarly well as all other supervised architectures. Despite the huge number of different codes, most of the codes predict still only a point estimate without a specific uncertainty per system or a proper Probability Density Function (PDF). Another important aspect of the applicability of these networks is the redshift range. Many are only trained on the lower end, e.g., between 0 and 1 (e.g., Bonnett, 2015; Hoyle, 2016; Sadeh et al., 2016; Almosallam et al., 2016b; Pasquet-Itam & Pasquet, 2018; Pasquet et al., 2019; Eriksen et al., 2020; Campagne, 2020), while only a minority is able to predict photometric redshifts for objects at e.g., $z \sim 4$. This, however, is very important for the upcoming surveys such as Euclid and LSST providing high quality images.

The main advantages of these machine learning codes are their simple usage and applicability to different data sets as well as the possibility to include extra information such as galaxy profiles, concentration, angular sizes, or environmental properties, in addition to magnitudes or colors (e.g., Soo et al., 2018). However, these are also extracted quantities from the images, which are then fed into the network. In contrast to that, we present in Chapter 6 NetZ, a novel machine learning technique dedicated for HSC that directly uses the image cutouts and is trained on galaxies up to $z \sim 4$.

Similar methods including CNNs got now also presented for SDSS. Pasquet et al. (2019) are using a classifier to sort SDSS images into redshift bins between $0 < z < 0.4$ while Henghes et al. (2021b) are in analogy to us using a regression network but with an upper limit of z=0.3 or z=1. Henghes et al. (2021b) carry out also a detailed comparison to common techniques such as random forest algorithms. Mu et al. (2020) present a hybrid method using a combination of both a CNN and a SOM architecture resulting in much better performance than a simple k-nearest-neighbor algorithm on galaxies up to $z = 0.8$ while Campagne (2020) focused on the very low end of the redshift distribution ($z < 0.3$). Here we clearly see the limitation of the SDSS image resolution with 0.396″/pixel, which covers the area of more than 5.5 pixels of a HSC image. This demonstrates again the high quality of HSC, and the necessity of larger surveys with that or even better resolution − which is currently under construction with LSST. As a result, we also clearly see the need of new photo-z techniques that include the redshift range beyond 1 for upcoming surveys like LSST.

1.5 Machine learning

Not only in astrophysics, but also in general, we are entering an era of having huge amount of data available that need to be analyzed. And this will become even more extreme in the near future. Therefore, and also thanks to the rapid developments in computer science, ML and specifically DL obtained in the last few years huge attention, revolutionizing a well established field of data analysis. Especially the flexibility of their tasks such as pattern recognition or information extraction and their extreme low computational time when applying these fully automated and self-guided computer algorithms make them so powerful. One of the first real applications was to digitize handwritten numbers or letters (Bozinowski, 1981), and since then DL has extended to many more and much more complex tasks also thanks to newly developed types of networks.

With the great success in image pattern recognition, DL got also attention in various astrophysical studies especially in data-intensive branches. Therefore, it is no surprise, that DL was also proposed for classification of astronomical images, to e.g., sort them into different galaxy categories such as spiral galaxies, merger, dwarf galaxies and any other types of galaxies. In analogy, NNs are also used with great success to classify images of galaxies into lenses and non-lenses. Because lens classification as presented in Sec. 1.2.5 is a binary classification problem, the network predicts typically one score in the range between 0 and 1 for each image. We can then set a threshold score above which an image would get labeled as lens candidate. These scores cannot be interpreted directly as probabilities but we see a clear correlation to the purity of lenses with increasing score, such that a score of 1 is a clear lens candidate.

Beside the classification tasks on images, DL algorithms can also be used to extract a set of continuous parameter values from these images. This is the kind of networks we use in this thesis, as we predict the parameter values of our assumed mass profile of the lens (see Chapter 3 and 4). Moreover, regression networks can be used to predict photometric redshifts as introduced in Sect. 1.4. While the redshift prediction is in principle a regression task, i.e. the network should output a specific value as redshift, several teams proposed to use a classification network and predict a redshift bin for each object (e.g., Bonnett, 2015; Carrasco Kind & Brunner, 2013; Han

et al., 2016; Hoyle, 2016; Li et al., 2021; Gerdes et al., 2010; Pasquet et al., 2019; Wadadekar, 2005; Wang et al., 2008). Despite their different network types, most of them are based on catalog entries extracted beforehand from the image cutouts such as colors, magnitudes, and size-compactness of the galaxy.

1.5.1 Convolutional neural networks

In contrast to most other teams, we demonstrate in Chapter 6 the possibility to obtain an accurate photometric redshift directly from the image over a large redshift range between 0 and ~ 4. For this, we use the standard type of networks for image processing, which includes convolutions of the images to extract their features and thus is called convolutional neural network (CNN) (Fukushima & Miyake, 1982). Such NN are comprised of multiple layers, at least one input and one output layer but typically include more hidden layers in between (Lecun et al., 2015).

For CNNs, these layers include convolutional layers with the basic procedure shown in Fig. 1.10. The kernel, whose size $f \times f$ is specified in the network architecture and thus individual for each network, is applied to the input image. Following the solid arrows, it is convolved with a cutout of the image of the same size of the kernel (yellow input part in Fig. 1.10), and then get iteratively shifted by a specific amount of pixels, the so-called stride (i.e. following the dashed arrow), to cover the whole image. This produced an output reduced in size by $f - 1$ as shown on the right.
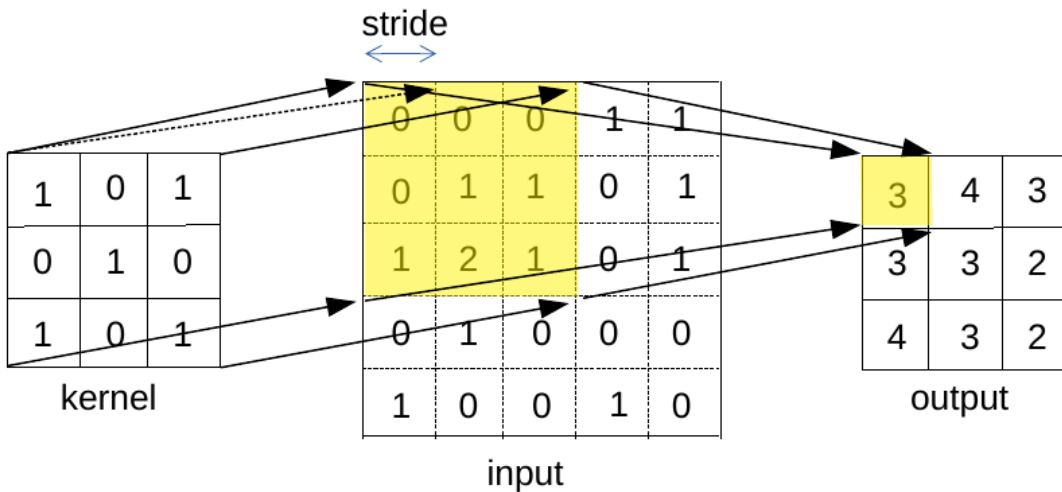
Figure 1.10: Schematic procedure of a convolutional layer.

While the CNN accepts a pre-defined amount of filters of the input images as the third dimension, which can also be just one, the convolutional layers will increase this dimension to the amount set by the network architecture through so-called feature maps. Each convolutional layer is then followed by an activation function, which give a weighting on the neurons. The most common function is the Rectified Linear Unit (ReLU) activation (Nair & Hinton, 2010).

Since the ReLU function set simply all negative values to zero, it is much faster in computation without decreasing accuracy compared to other activation functions such as a sigmoid function or the hyperbolic tangent. There are nowadays also modifications of the ReLU function proposed such as a leaky ReLU (He et al., 2015) where the negative values are multiplied by a small constant instead of zero.

Another common layer in NNs is the pooling layer, which is often applied after the activation function in CNNs. It helps to reduce the size of the maps by combining the outputs of a given space. In Fig. 1.11 we show the two options of pooling with an space of 2×2, either the average (avg-pooling) of the given space or the maximum (max-pooling) is kept.
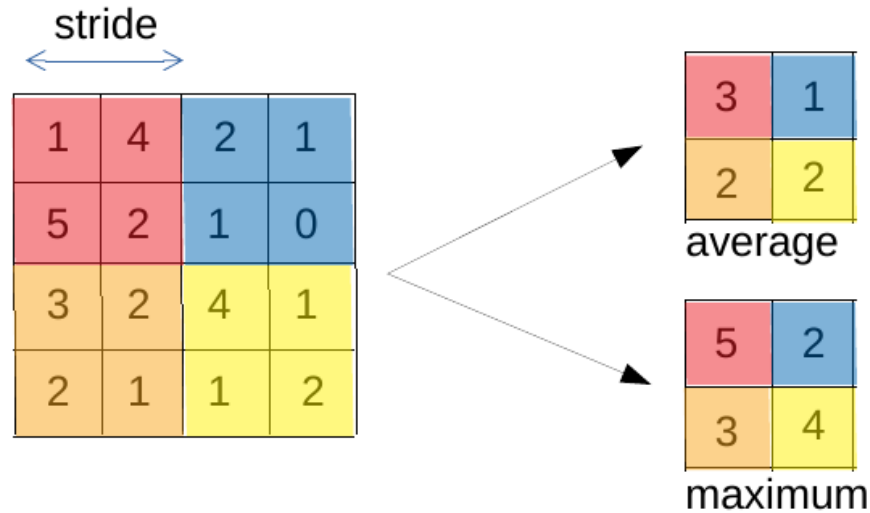
Figure 1.11: Schematic procedure of a pooling layer, either the average (avg-pooling) or the maximum (max-pooling) is kept.

The last layers of CNNs are typically FC layers, for which the remaining data cube from the previous layer gets flatten to one dimension. Unlike convolutional layers, FC layers do not share weights or a kernel, but instead contain a number of neurons that are connected to every other neuron in the following layer. These FC layers behave similarly to traditional NN using catalog entries (e.g., Bonnett, 2015; Cañameras et al., 2020; Collister & Lahav, 2004) and taking here the features extracted by the convolutional layers. The FC layers are further processing these information to reach with the final FC layer the number of parameters the network shall predict. This can be just a single neuron like for predicting the redshift of a galaxy, a few parameters like for our modeling networks, or also hundreds for instance when classifying large sets of images (e.g., He et al., 2015).

Depending on the task of the network, the output of the last FC layer is passed through a sigmoid function to map the output into the range between 0 and 1. This is for instance the case in binary classification tasks (e.g., Cañameras et al., 2020, 2021; Shu et al., in prep.) where exactly one output score is requested, but can be also used in regression networks as presented in Chapter 4, helping to better and equally optimize all parameters. For a multi-class classification

challenge, the softmax function is typically used (e.g., Fuqiang et al., 2014; Morales et al., 2021; Wang et al., 2019).

1.5.2 Residual neural networks

With the increasing amount of training data and complexity in tasks, the networks got steadily deeper in terms of layers. Nowadays, even networks with more than 1,000 layers got developed and trained (He et al., 2015). Since it is extreme difficult to train such deep networks and the computational time increases drastically, He et al. (2015) proposed a residual neural network (ResNet), a new variation of CNNs, by introducing residual blocks (pre-activated bottleneck residual units in He et al., 2015) in the network architecture. Such blocks consists of typically two or also more convolutional layers, which are additionally provided with a skip-connection (or short cut). This ensures that the convolutional layer learns residual functions with respect to the previous layer, and help avoid vanishing gradients during optimization. Thanks to these residual blocks, it is possible to increase the number of layers and performance without requesting much higher computing resources than before. Such ResNets are ideal for capturing small features and obtained the best results on the ImageNet Large Scale Visual Recognition Challenge 2015 as well as similar challenges (He et al., 2015). Therefore, we also make use of these ResNets in Chapters 4 and 5.

1.5.3 Training process for supervised learning

All the mentioned types of ML can be grouped into supervised, weakly-supervised, semi-supervised or unsupervised methods, while most of the mentioned examples, as well as our developments in this thesis, fall into the group of supervised learning. Supervised NN are optimized on a specific set of data with known output. This is called the ground truth and used to find the best network architecture and the best set of hyper-parameter values. For this, the ground truth data set is split up into training set R, validation set V, and test set T. The network is now applied during the forward propagation to the training set R where it can capture image characteristics, and the network output is compared to the expected output, the so-called labels. The differences are quantified through the loss function L, whose form depends on the task and to some extent on the developer. The weight and neurons get then updated through the back-propagation of the gradients to minimize the loss. After that, the network get applied to the validation set V to quantify how good the network perform on new data and to ensure the network is not fitting only to the training set R which is called over-fitting. Through an iterative process, the weights and neurons are optimized to solve the task in the best manner, i.e. to give the lowest validation loss. At the very end, the trained network is applied to the test set T, containing data the network has never seen before to demonstrate the performance on new data.

To minimize potential selection biases in the different splits, a common procedure is the cross-validation which we also use throughout the thesis. For this, the validation set V and the training set R get split into e.g., five sets $\{A, B, C, D, E\}$ of equal size. The network is now trained five times, while each time another split acts as validation set and the remaining ones as training set. After the training and validation of these five networks, the epoch of the minimal median validation loss get computed and a final network is trained on all five splits $\{A, B, C, D, E\}$ up

to this epoch. That network is then the final network which we apply to the test set T for a performance test and afterwards to new data without known output.

Therefore, the main requirement for supervised learning is a large enough, realistic data set with known labels. These can be either mocked up data or real data.

In case of unsupervised learning the network learns on its own the task, e.g., to group images into classes containing similar features but is not optimized on a set of examples with known labels. Examples of applications in astrophysics are redshift predictions (e.g., Geach, 2012; Way & Klose, 2012; Razim et al., 2021), classification of fast radio bursts into repeaters and non-repeaters (Chen et al., 2022), detecting extragalactic transients (Villar et al., 2021, e.g.,) or lenses (e.g., Cheng et al., 2020), and many more (e.g., Sarmiento et al., 2021). Semi-supervised learning falls in between and is an approach of ML that combines a small sample of labeled data and a larger sample without labels. A very similar branch of ML is weak-supervised learning with only noisy, limited, or imprecise labeled data. Given the success of unsupervised networks, even such data sets can lead to very precise predictions of a network. Both types of ML are also successfully used in astrophysics for different tasks (e.g., Ali-Dib et al., 2020; Balakrishnan et al., 2021; Marianer et al., 2021; Richards et al., 2012).

1.6 Thesis outline

The work presented in this thesis is ordered as follows. Since the main part of the thesis focus on gravitational lensing, we give a short summary of the lensing formalism and the typical modeling procedure in Chapter 2. In Chapter 3 we present our simulation pipeline to generate very realistic mock data as well as our modeling network where we assumed a SIE mass profile. Moreover, we show in this chapter the resulting accuracy and precision on the image position and time delays obtained with the network's model. This is followed by the modeling network assuming SIE+γ_{ext} in Chapter 4, where we also include an error estimation. A direct and detailed comparison between traditionally obtained models and those from our network is given in Chapter 5. For this comparison, we use 32 galaxy-galaxy lens systems from the SuGOHI sample. Additionally, we describe our automated modeling algorithm using conventional optimization techniques which we used to derive the comparison models. In Chapter 6 we present the photometric redshift network NetZ with our novel approach using directly the image cutouts. We finally conclude and provide an outlook in Chapter 7.

2 Lensing formalism and traditional lens modeling

In this chapter we revisit the most important aspects of the lensing formalism (Sect. 2.1), where we mainly follow Schneider et al. (2006) and the naming convention introduced in Fig. 1.1. In Sect. 2.2 we will then highlight the general procedure of lens modeling and introduce commonly adopted light and mass profiles of which we also make use in this thesis.

2.1 Lensing formalism

2.1.1 Lens equation

Normally, without lensing, the light rays from the background source would travel in a straight line (dashed in Fig. 1.1) to the observer, such that the observer would see the source under the angle $\boldsymbol{\beta}$. Therefore, $\boldsymbol{\beta}$ is called the true, unlensed source position. Because of the deflection generated by the lens, the observer sees the source at position $\boldsymbol{\theta}$ instead of $\boldsymbol{\beta}$. By assuming that $\tilde{\boldsymbol{\eta}}$ is very small compared to the distances D_d and D_s, the angles are also very small such that we can use the approximation $\sin(x) \approx x \approx \tan(x)$ and read off the geometric condition

$$\tilde{\boldsymbol{\eta}} = \frac{D_\mathrm{s}}{D_\mathrm{d}}\boldsymbol{\xi} - D_\mathrm{ds}\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) \tag{2.1}$$

from Fig. 1.1. If we introduce now angular coordinates

$$\tilde{\boldsymbol{\eta}} = D_\mathrm{s}\boldsymbol{\beta} \tag{2.2}$$
$$\boldsymbol{\xi} = D_\mathrm{d}\boldsymbol{\theta} \tag{2.3}$$

we can write Eq. (2.1) as

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \frac{D_\mathrm{ds}}{D_\mathrm{s}}\hat{\boldsymbol{\alpha}}(D_\mathrm{d}\boldsymbol{\theta}) \,. \tag{2.4}$$

By defining the scaled deflection angle as

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) \equiv \frac{D_\mathrm{ds}}{D_\mathrm{s}}\hat{\boldsymbol{\alpha}}(D_\mathrm{d}\boldsymbol{\theta}) \tag{2.5}$$

we are directly left with the famous lens equation

$$\boldsymbol{\alpha} = \boldsymbol{\theta} - \boldsymbol{\beta} \,. \tag{2.6}$$

2.1.2 Convergence and the critical mass density

As mentioned in Sect. 1.1.3, the strong lensing regime implies resolvable multiple images of the background source, which can be individual images from point like objects such as quasars, stretched arcs from extended sources like galaxies, or, if the alignment is appropriate, form a ring through merging arcs. This happens if the lens equation (2.6) has, for a specific source position $\boldsymbol{\beta}$, more than one solution. A suffcient criteria is that the convergence, sometimes also called dimensionless surface mass density,

$$\kappa = \frac{\Sigma(r)}{\Sigma_{\mathrm{crit}}} \tag{2.7}$$

is bigger or equal to 1. This can therefore be used as a criterion to distinguish between the strong and weak lensing regimes. The given expression is valid for an axisymmetric mass distribution $\Sigma = \Sigma(r)$, which can often be assumed. Here, $\Sigma(r)$ is the surface mass density and Σ_{crit} the critical mass density, a characteristic and specific quantity for each lens system, defined as

$$\Sigma_{\mathrm{crit}} = \frac{c^2}{4\pi G} \frac{D_{\mathrm{s}}}{D_{\mathrm{d}} D_{\mathrm{ds}}} \ . \tag{2.8}$$

2.1.3 Deflection angle and the lens potential

Following Schneider et al. (2006), the deflection angle for a general extended mass distribution $\rho(\boldsymbol{\xi}, z)$, with spatial coordinates (ξ_1, ξ_2, z), can be generalized to

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4G}{c^2} \int \mathrm{d}^2\xi' \int \mathrm{d}z' \rho(\xi_1', \xi_2', z') \frac{\boldsymbol{\xi} - \boldsymbol{\xi}'}{|\boldsymbol{\xi} - \boldsymbol{\xi}'|^2} \tag{2.9}$$

or

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4G}{c^2} \int \mathrm{d}^2\xi' \Sigma(\xi_1', \xi_2') \frac{\boldsymbol{\xi} - \boldsymbol{\xi}'}{|\boldsymbol{\xi} - \boldsymbol{\xi}'|^2}, \tag{2.10}$$

where

$$\Sigma(\xi_1', \xi_2') = \int \mathrm{d}z' \rho(\xi_1', \xi_2', z') \tag{2.11}$$

is the projected surface mass density of the deflector as a function of the two-dimensional position on the lens plane (ξ_1', ξ_2').

Assuming again an axisymmetric mass distribution, we can now write the scaled deflection angle with the dimensionless surface mass density as

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{1}{\pi} \int_{\mathbb{R}^2} \mathrm{d}^2\theta' \kappa(\boldsymbol{\theta}') \frac{\boldsymbol{\theta} - \boldsymbol{\theta}'}{|\boldsymbol{\theta} - \boldsymbol{\theta}'|^2} \tag{2.12}$$

or, by evaluating the spatial integral in Eq. (2.12) to 2π and taking the derivative canceling the radial integral, express the dimensionless surface mass density as a function of the deflection angle by

$$2\kappa = \frac{\partial \alpha_1}{\partial \theta_1} + \frac{\partial \alpha_2}{\partial \theta_2} = \boldsymbol{\nabla}_\theta \cdot \boldsymbol{\alpha} \ . \tag{2.13}$$

Using the mathematical relation

$$\nabla \ln(|\boldsymbol{\theta}|) = \frac{\boldsymbol{\theta}}{|\theta^2|}, \tag{2.14}$$

we can write the deflection angle α as

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \nabla \Psi(\boldsymbol{\theta}) \tag{2.15}$$

and define

$$\Psi(\boldsymbol{\theta}) \equiv \frac{1}{\pi} \int_{\mathbb{R}^2} \mathrm{d}^2 \theta' \kappa(\boldsymbol{\theta}') \ln\left(|\boldsymbol{\theta} - \boldsymbol{\theta}'|\right) \tag{2.16}$$

as the lens potential, which describes the gravitational potential generated by the lens mass distribution and the curvature of space-time created by it (Schneider, 2005). This allows us to re-write the lens equation(2.6) as

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \nabla \Psi(\boldsymbol{\theta}), \tag{2.17}$$

which directly shows the connection between the image positions on the lens plane and the first derivative of the lensing potential.

2.1.4 Surface brightness conservation, magnification, and shear

As a consequence of the Liouville theorem (e.g., Zettl, 2010) and the absence of emission and absorption of photons in gravitational lensing, the photon number and thus also the SB is conserved. However, the images of extended background sources are typically highly distorted. This can be seen for example in Fig. 1.2, where the arcs are forming nearly an Einstein ring. Also the second source, another galaxy, is lensed into a radially stretched arc (green box in Fig. 1.2), while its counter image (dashed green box in Fig. 1.2) shows a completely different structure. The magnification factor μ quantifies the distortion for a specific system.

To motivate this, suppose a source has a SB described by $I(\boldsymbol{\beta})$, then the observed SB on the image plane is

$$I(\boldsymbol{\theta}) = I^{\text{source}}(\boldsymbol{\beta}(\boldsymbol{\theta})) \,. \tag{2.18}$$

Without gravitational light deflection, the source would subtend a specific area on the sky, which we denote as $\mathrm{d}\tilde{\omega}$. Then the monochromatic flux is given by

$$\tilde{F} = I(\boldsymbol{\theta})\mathrm{d}\tilde{\omega} \,. \tag{2.19}$$

When now a massive object deflects the light, the solid angle $\mathrm{d}\omega$ will be different from $\mathrm{d}\tilde{\omega}$. Because the SB does not change, the flux of the image is given by

$$F = I(\boldsymbol{\theta})\mathrm{d}\omega \,. \tag{2.20}$$

Hence, the light deflection leads to a change of flux which we observe as distortions. This amount is given by the magnification factor μ, which is defined as the ratio of flux with and without lensing, i.e.

$$|\mu| = \frac{F}{\tilde{F}} = \frac{I(\boldsymbol{\theta})\mathrm{d}\omega}{I(\boldsymbol{\theta})\mathrm{d}\tilde{\omega}} = \frac{\mathrm{d}\omega}{\mathrm{d}\tilde{\omega}} \,. \tag{2.21}$$

39

By assuming that the source is much smaller than the angular scale on which the lens mass properties change, the lens equation can be locally linearised resulting in the Jacobian matrix

$$A = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} = \left(\delta_{jk} - \frac{\partial^2 \Psi(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right). \tag{2.22}$$

Here we have used the lens equation in form of Eq. (2.17) and the Kronecker Delta δ_{jk}. Following Schneider et al. (2006), the Jacobian matrix can further be expressed using Eq. (2.16) as

$$A = \begin{pmatrix} 1 - \kappa - \gamma_1^{\mathrm{int}} & -\gamma_2^{\mathrm{int}} \\ -\gamma_2^{\mathrm{int}} & 1 - \kappa + \gamma_1^{\mathrm{int}} \end{pmatrix} = (1 - \kappa) \begin{pmatrix} 1 - g_1 & -g_2 \\ -g_2 & 1 + g_1 \end{pmatrix}. \tag{2.23}$$

where we introduced the shear

$$\gamma^{\mathrm{int}} = \gamma_1^{\mathrm{int}} + i\gamma_2^{\mathrm{int}} \tag{2.24}$$

and the reduced shear

$$g = g_1 + ig_2 \equiv \frac{\gamma^{\mathrm{int}}}{1 - \kappa}, \tag{2.25}$$

which are both complex quantities and the key parameter in weak lensing studies. The Jacobian matrix in form of Eq. (2.23) shows also the role of the shear and convergence: the convergence is the generator of a constant isotropic distortion in the source structure, while the shear is elongating it in a preferential direction.

Finally, the determinant of the Jacobian matrix A can be related to the magnification factor

$$\mu = \frac{1}{\det(A)} = \frac{1}{(1 - \kappa)^2 - |\gamma^{\mathrm{int}}|^2}. \tag{2.26}$$

From this equation we see that for specific values of κ and γ^{int} the determinant of A vanishes, which implies that the magnification μ diverges. These locations define the so-called critical curves on the lens plane. Depending on the eigenvalues of A, they are described as tangential or radial critical curves. The mapped positions of the critical curves to the source plane are called caustic curves. Therefore, objects close to the caustic can be highly magnified, allowing us to observe them even though they would be otherwise too faint. This effect helps to study the early Universe and properties of high-redshift objects. This effect is also important on the microlensing-scale especially for transient or expanding objects like SNe, as they can cross the caustics, leading to a sudden steep increase of the magnification (e.g., Huber et al., 2019, 2021a,b).

2.1.5 Multi-plane lensing

So far we assumed the basic scenario where we have exactly one lens and one background object. This is enough for most studies, including this work, however, in some cases like the Cosmic Horseshoe shown in Fig. 1.2, more than two objects are involved. This means in detail, that the light of the most distant object at position $\boldsymbol{\theta}_N$ gets deflected by all other $N - 1$ objects towards the observer. In this case the lens equation defined in Eq. (2.6) is calculated iteratively over all

N objects and reads as

$$\boldsymbol{\theta}_j(\boldsymbol{\theta}_1) = \boldsymbol{\theta}_1 - \sum_{k=1}^{j-1} \frac{D_{kj}}{D_j} \hat{\boldsymbol{\alpha}}(\boldsymbol{\theta}_k) \,. \tag{2.27}$$

Moreover, the light of object $N - 1$ gets deflected by all remaining $N - 2$ deflectors towards the observer; and following this logic all other objects as well until the last source at position $\boldsymbol{\theta}_1$ is observed unlensed like the lens in the single-plane scenario. In other words, Eq. (2.27) gives the relation between the angular position $\boldsymbol{\theta}_j$ of a light ray in the *j*-th lens plane and the angular position in the $j = 1$ plane, which is the observed image plane. In this notation $\boldsymbol{\theta}_N$ corresponds to the source plane, and $\hat{\boldsymbol{\alpha}}(\boldsymbol{\theta}_k)$ to the deflection angle on the *k*-th plane. In analogy to Eq. (2.5), the total scaled deflection angle $\boldsymbol{\alpha}_{\text{tot}}$ is the sum over all deflection angles on all $N - 1$ lens planes

$$\boldsymbol{\alpha}_{\text{tot}} = \sum_{k=1}^{N-1} \frac{D_{kN}}{D_N} \hat{\boldsymbol{\alpha}}(\boldsymbol{\theta}_k) \,. \tag{2.28}$$

Following Eq. (2.23), the Jacobian matrix in the multi-plane scenario is given by

$$A = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\theta}} = \frac{\partial \boldsymbol{\theta}_N}{\partial \boldsymbol{\theta}_1} \,, \tag{2.29}$$

and in similar procedure the other effective quantities by replacing the deflection angle $\boldsymbol{\alpha}$ with the total deflection angle $\boldsymbol{\alpha}_{\text{tot}}$.

Since we do not perform multi-plane lens modeling in this thesis, we refrain from further discussion and only refer the reader to dedicated publications on this topic like Petters (1995a,b) and Schneider (2019).

2.2 Traditional lens modeling

After introducing the general formalism used in gravitational lensing, we further highlight the general procedure developed over the past decades to reproduce the lens observations with theoretical profiles. Such modeling is necessary for most lensing studies, including the study of lens mass properties (see Sect. 1.1.3.1), the reconstruction of the source SB for studying the background source (see Sect. 1.1.3.2), or the reconstruction of the lens potential for time-delay cosmography (see Sect. 1.1.3.3 and Sect. 1.1.3.4). The modeling is typically performed stepwise, while the exact procedure depends on the lensing regime, system size (single galaxy/-group/cluster), and on whether e.g., a quasar is lensed or only a galaxy. It also depends on the image resolution and science case which determines how detailed and complex the model shall be. This means there is not a strict procedure, especially as also each lens system is particular on its own, making detailed modeling a relatively complicated and lengthy task that requires expertise. For the modeling, there are several well established software packages available such as GLEE (Suyu & Halkola, 2010; Suyu, 2012) and its extension for Gravitational Lensing and Dynamics (GLaD) (Chirivì et al., 2020) which we extensively use in this thesis. These codes support all commonly adopted profiles, well-tested sampling algorithms to obtain the best fitting parameter values, possibilities to include measured time delays or velocity dispersions as con-

strains, as well as several other possibilities. In this section we summarize the general modeling procedure, introduce typically adopted mass and light profiles, and finally highlight optimization techniques supported by GLEE and GLaD and used in this thesis.

2.2.1 Source/Image position model and the Singular Isothermal Ellipsoid profile

For a galaxy-galaxy strong lensing system as we use in this thesis, the first step is to identify multiple components of the background source. For ground based images this is often just one set, i.e. the brightest pixel of each arc, while for high-resolution images e.g., from HST, multiple sets can be identified by taking into account that the color and intrinsic brightness of the background source is conserved (e.g., Grillo et al., 2013, 2018; Suyu et al., 2013; Schuldt et al., 2019). These identified positions can then be used to constrain the parameters of a lens mass profile through minimizing the difference between predicted positions and identified positions, either on the source plane or on the image plane, resulting in a source position or image position model, respectively. Since for the image position model the lens equation needs to be solved in every iteration, this procedure is slightly slower and computationally more expensive but on the other hand also more precise. Typically, we first run a short iteration using the source position, but from then onward optimize with the image positions.

A commonly adopted profile for the total mass is the power-law profile, for instance the Singular Power-Law Elliptical Mass Distribution (SPEMD) (Barkana, 1998) which is a softened power-law profile with a constant density within a central core of radius r_c. Its convergence as defined in Eq. (2.7), which quantity is used by GLEE to compute the deflection map, can be written as

$$\kappa(r) = E \left(r^2 + r_c'^2 \right)^{-\gamma_{PL}} \tag{2.30}$$

where γ_{PL} is the power-law index and r the elliptical radius, which can be expressed through the Cartesian coordinates x and y and the axis ratio q by

$$r = \sqrt{x^2 + \frac{y^2}{q^2}} . \tag{2.31}$$

The elliptical core radius r_c' is related to the core radius r_c by

$$r_c' = \left(\frac{2r_c}{1 + q} \right) , \tag{2.32}$$

and the whole mass distribution is then additionally rotated by the position angle θ. In this equation, E is the amplitude and following Barkana (1998) related to the Einstein radius θ_E by

$$E = \frac{2}{1 + q} \left(1 - \gamma_{PL} \right) \frac{\theta_E{}^2}{(\theta_E{}^2 + r_c'^2)^{1-\gamma_{PL}} - r_c'^{2(1-\gamma_{PL})}} \tag{2.33}$$

The case $\gamma_{PL} = 0$ corresponds to a constant mass sheet and the isothermal case is obtained with $\gamma_{PL} = 0.5$. If the number of image positions is relatively low and for a well-constrained model

the number of free parameters needs to be reduced, we often assume the isothermal case and set the core radius to zero, or, to avoid numerical instabilities in the lens center, to a very low values such as 10^{-4}, resulting in the SIE profile. The even more simple profile is the Singular Isothermal Sphere (SIS), where additionally the axis ratio is set to 1, i.e. a spherical mass distribution. In the case of an SIE, Eq. (2.30) simplifies to

$$\kappa(r) = \frac{\theta_E}{(1+q)r} \, .$$

(2.34)

Since real lensing systems are not isolated, nearby objects influence the light deflection as well. While e.g., stars, even within the Einstein radius, are not massive enough, galaxy clusters introduce notable effects on the light deflection even if they are several tens of arcseconds to arcminutes away. This mass concentration is typically outside of the image cutout (compare Fig. 1.2), such that these perturber can be described by a simple external shear strength denoted as γ_{ext} and a position angle θ_{ext}, indicating the amount and direction of the mass concentration, instead of specific mass profiles such as the power-law. These two parameters can also be converted into a complex notation through $\tilde{\gamma}_{ext} = \gamma_1 + i\gamma_2$ with

$$\gamma_1 = \gamma_{ext} \cos(2\theta_{ext})$$
$$\gamma_2 = \gamma_{ext} \sin(2\theta_{ext}) \, .$$

(2.35)

Due to the low number of constraints in the source or image position model, this external shear is often neglected and included at a later stage. However, this is not that crucial as both procedures based on the image positions alone give typically just a preliminary model for galaxy-systems, which is refined by using directly the image cutouts. These provide the full pixel information, and multiple bands can be modeled simultaneously to increase the amount of data and additionally provide color information. Here the external shear component can be included without problems.

2.2.2 Lens light modeling with the Sérsic or the chameleon profile

To model the mass distribution of the lens using the full image cutout, the lens light distribution must be modeled first. Lens galaxies are typically LRGs given their higher lensing cross-section (Turner et al., 1984). The commonly adopted profile is the Sérsic profile (Sérsic, 1963) which has a variable slope n compared to the De Vaucouleurs law (De Vaucouleurs, 1948). Assuming an elliptical profile, the intensity I of the galaxy at position

$$r = \sqrt{x^2 + \frac{y^2}{q^2}}$$

(2.36)

is given by

$$I(r) = Ae^{-\zeta(n)\left(\left(\frac{r}{r_{eff}}\right)^{1/n} - 1\right)} \, ,$$

(2.37)

with the axis ratio q, the amplitude A corresponding to the central surface density, and the effective radius r_{eff}. The constant $\zeta(n)$ depends on the exponent and can be approximated as

$$\zeta(n) = 2n - \frac{1}{3} + \frac{4}{405n} + \frac{46}{25515n^2} + \frac{131}{1148175n^3} - \frac{2194697}{30690717750n^4} \tag{2.38}$$

such that the effective radius encloses half of the projected light (Cardone, 2004; Ciotti & Bertin, 1999; Dutton et al., 2011). Therefore, the effective radius is also called half-light radius. Depending on the resolution and complexity of the lens, multiple components can be included.

An alternative to the Sérsic profile is the so-called "chameleon" profile which mimics the Sérsic profile well but allows analytic computations of lensing quantities (e.g., Dutton et al., 2011; Maller et al., 2000; Suyu et al., 2014). It is composed of a difference of two isothermal profiles with different core radii, but the same amplitude:

$$I(x, y) = \frac{I_0}{1+q_{\text{L}}} \left(\frac{1}{\sqrt{x^2 + y^2/q_{\text{L}}^2 + 4w_{\text{c}}^2/(1+q_{\text{L}})^2}} \right.$$

$$\left. - \frac{1}{\sqrt{x^2 + y^2/q_{\text{L}}^2 + 4w_{\text{t}}^2/(1+q_{\text{L}})^2}} \right) . \tag{2.39}$$

In this equation q_{L} is the axis ratio, and w_{t} and w_{c} are originally the core radii of the two isothermal profiles and now just two parameters of the profile with $w_{\text{t}} > w_{\text{c}}$ to keep $I > 0$ at every position (x, y).

Since the light distribution is typically not well enough described with one Sérsic or chameleon profile alone, often two or three such profiles are stacked together to increase the flexibility of the model. The values of these parameters describing the lens light (subscript ll) distribution are then optimized through minimizing

$$\chi^2_{\text{ll}} = \sum_{j=1}^{N_{\text{p}}} \frac{\left(I_j^{\text{obs}} - \text{PSF} \otimes I_j^{\text{pred}} \right)}{\sigma_{\text{tot},j}} , \tag{2.40}$$

which is the difference between the observed intensity I_j^{obs} of pixel j and the predicted intensity I_j^{pred} after convolving with the PSF, divided by the total noise $\sigma_{\text{tot},j}$. This is finally summed over all pixels N_{p} of the lens, which means all pixels that are not excluded through the arc mask or lens mask. These are two masks dedicated for each lensing systems to specify where the arcs and multiple images are (arcmask) and where notable light of LOS objects, such as objects O_1, O_2, and O_3 in Fig.1.2, is (lensmask) as the optimized lens light profile is not supposed to describe that light. Here the total noise $\sigma_{\text{tot},j}$ combines through a quadrature sum the background noise $\sigma_{\text{bkgr,j}}$, which can be estimated through the standard deviation from an empty patch of the cutout, and the astrophysical Poisson noise (Hasinoff, 2012).

Since the predicted light distribution shows only the light of the lens, the arcs and other bright objects in the cutout are either masked out, i.e., excluded from the sum defined in Eq. (2.40), or the uncertainty σ_{tot} for these pixel is increased drastically such that the contribution to the χ^2_{ll} is effectively zero. After optimizing the parameters for the main lens, contaminants like objects O_1, O_2, and O_3 in Fig. 1.2 can be included through an own, additional light profile.

2.2.3 Arc light modeling and source surface brightness reconstruction

After obtaining a good lens light fit, the mass profile, normally pre-optimized from the image positions, is included in the model. Instead of the identified image positions, the regions of the multiple images in the cutout are directly used to constrain the mass parameters. For this, either the mask or error map is updated such that these pixel are now also included in the χ^2_{ll} minimization, denoted then as χ^2_{arc}. In this modeling step, also the light distribution of the background source is reconstructed. One possibility for this is to adopt a profile, for instance a Sérsic profile. This light distribution is then mapped from the source plane onto the lens plane and compared to the observed arcs. The source light parameters are then chosen such that the difference on the lens plane is minimal. This principle of the source SB reconstruction is supported by GLaD and preferred for ground-based images.

The other option is to assume no light profile at all for the source and instead use a so-called pixelated source SB reconstruction. Here the code tries to reconstruct the source light distribution on a grid of pixels, where the distribution is only constrained through the observed arcs. Additionally, in the minimization of χ^2_{arc} a regression term such as

$$|s_{k-1} - 2s_k + s_{k+1}|^2 \tag{2.41}$$

is included to minimize the difference of the source intensity s_k of pixel k and the neighboring pixels on the grid (Suyu et al., 2006). This helps to obtain a smooth, physically realistic intensity map. This procedure is supported by GLEE and is typically used for high-resolution images such as HST images which provide enough details on the arcs.

2.2.4 Composite mass modeling

Depending on the science case, one might be interested in disentangling the baryonic matter from the DM component. For this, one typically assumes the baryonic mass as the light distribution scaled by a mass-to-light ratio (M/L), which is a good approximation for the optical and NIR wavelength range (e.g., Schombert & McGaugh, 2014; McGaugh, 2016), but is known not to be exact (e.g., Bernardi et al., 2018; Sonnenfeld et al., 2018b). Combining the total mass and the baryonic mass, one can construct a composite mass model of baryons and DM. The commonly adopted profile for the DM distribution is the Navarro Frenk and White (NFW) profile which was found through large numerical DM-only simulations based only on gravity (Navarro et al., 1997). The DM density is given by

$$\rho(r) = \frac{\rho_s}{\left(\frac{r}{r_s}\right)\left(1 + \frac{r}{r_s}\right)^2} \ . \tag{2.42}$$

In this equation, r_s is the scale radius and defines the radius where the slope changes: for radii much smaller than the scale radius, the density falls as $\rho_{r<<r_s} \propto r^{-1}$ and for radii much larger than the scale radius as $\rho_{r>>r_s} \propto r^{-3}$. Typically, this profile is considered with an elliptical radius

(denoted here as ENFW) and we define

$$r' \equiv \frac{r}{r_s} = \frac{\sqrt{qx^2 + \frac{y^2}{q}}}{r_s} \tag{2.43}$$

with axis ratio q. Thus, the deflection angle (Halkola et al., 2006; Meneghetti et al., 2003)

$$\alpha^{\mathrm{NFW}}(r') = \theta_E \frac{h(r')}{r'} \tag{2.44}$$

changes in the following way

$$\alpha_1^{\mathrm{ENFW}} = \alpha_1^{\mathrm{NFW}} \frac{qx}{r} \tag{2.45}$$

$$\alpha_2^{\mathrm{ENFW}} = \alpha_2^{\mathrm{NFW}} \frac{y}{qr} , \tag{2.46}$$

with

$$h(r') = \ln\left(\frac{r'}{2}\right) + \begin{cases} \frac{2}{\sqrt{1-r'^2}} \operatorname{arcosh}\left(\frac{1}{r'}\right) & \text{for } r' < 1 \\ 1 & \text{for } r' = 1 \\ \frac{2}{\sqrt{r'^2-1}} \arccos\left(\frac{1}{r'}\right) & \text{for } r' > 1 \end{cases} . \tag{2.47}$$

However, one is not limited to this profile and can instead adopt for instance a power-law profile or a generalized NFW profile, where a variable inner DM slope γ_g is introduced and the density is given by

$$\rho(r) = \frac{\rho_s}{\left(\frac{r}{r_s}\right)^{\gamma_g} \left(1 + \frac{r}{r_s}\right)^{3-\gamma_g}} . \tag{2.48}$$

2.2.5 Degeneracies and breaking them through stellar kinematics

In the lens modeling, some parameters are typically degenerate with each other, such as the ellipticity and external shear or the Einstein radius and the power-law slope. These degeneracies can be seen from 2D probability density plots of the various parameters. Moreover, there is a Mass-Sheet Degeneracy (MSD) as shown by Falco et al. (1985), which means that the mass distributions $\kappa(\boldsymbol{\theta})$ defined in Eq. (2.30) for the power-law profile and

$$\kappa_{\lambda'}(\boldsymbol{\theta}) = \lambda'\kappa(\boldsymbol{\theta}) + (1 - \lambda') \tag{2.49}$$

in combination with the same arbitrary isotropic scaling λ' of the source plane $\boldsymbol{\beta} \to \lambda'\boldsymbol{\beta}$, which is in most cases unobservable, correspond to exactly the same dimensionless observables like image positions, image shape, and magnification ratios. This means that from the observed image positions and flux ratios alone, we cannot distinguish between any of the possible $\kappa_{\lambda'}$ distributions. This MSD can not be broken through weak lensing either, as the image shapes are unaffected. Similarly, also time-delay ratios remain invariant, but the product $H_0 \times \Delta t$ is affected (Schneider & Seitz, 1995) making it impossible to obtain H_0 from gravitational lensing without breaking the MSD.

There are several ways proposed to break this degeneracy. For instance, if the magnification, which is affected by the MSD as $\mu \to \mu/\lambda'^2$, can be estimated, it can be used to break the degeneracy by constraining the value of λ'. Another possibility is to combine gravitational lensing with another, independent lens mass estimate, where velocity dispersion measurements are perfectly valid. This combination was done the first time for the model of Q0957+561 by Grogin & Narayan (1996) used to constrain the Hubble constant H_0 and got refined by Romanowsky & Kochanek (1999) who suggested to include higher-order moments of the stellar velocity. While originally velocity dispersion measurements and lensing were treated completely independently, it is nowadays common to combine them to break the MSD especially when inferring the Hubble constant (e.g., Suyu et al., 2010, 2013). Therefore, many studies on the MSD were carried out to further understand the effect and impact on H_0 inference (e.g Birrer et al., 2016; Chen et al., 2021a; Gomer & Williams, 2020; Schneider & Sluse, 2013) also under the assumption of Integrale Field Unit (IFU) stellar kinematics obtained with the JWST (Yıldırım et al., 2021). Beside breaking the MSD by including the velocity dispersion as additional data in the lens mass model, it helps also to constrain further all parameters. Therefore, our modeling software GLEE, which does not support kinematics on its own, got extended through GLaD. In addition to including the lens stellar kinematics, GLaD supports also the option of reconstructing the unlensed source kinematics simultaneously with the source SB through modeling of lensing and IFU kinematic measurements of lens and lensed source (Chirivì et al., 2020).

2.2.6 Lensed quasar modeling

The described modeling process is slightly different for modeling lensed quasars as we have to model the host galaxy of the quasar as well as the quasar images themselves, which are typically much brighter than the host (compare Fig. 1.3). The source and image position modeling procedure is the same as described in Sect. 2.2.1, where the quasar images can be used as the lensed image positions. Also the lens light modeling procedure described in Sect. 2.2.2 remains unchanged, but when including the quasar in the extended image model, a point profile is adopted for each quasar image. This point profile has only three parameters: the position of the center (x, y) and the amplitude. The light distribution is given by the PSF representing a point-like object such as a star observed through the given telescope under given circumstances. This means, when including the quasar in the lens light model, the host galaxy will be further masked to first optimize the parameters describing the bright quasar images. This, however, leads to a slight overestimate of the quasar amplitudes to compensate the light contribution of the host at these positions. Therefore, when including the host in the model, the quasar amplitudes are adjusted to slightly lower values at the first instance. Because one does not know the amount by which the quasar amplitudes need to be decreased, one trick is to increase artificially the values in the error map around the quasar positions such that the contribution of these regions to the χ^2 is smaller than that of the other parts of the host. This essentially helps in fitting better the host galaxy without introducing difficulties through the bright quasar, and after the fit of the outer parts of the host has stabilized, the uncertainties are reset to their original values and the quasar light is allowed to vary again. Apart from this, the arc light modeling follows closely the procedure described in Sect. 2.2.3, and allows to reconstruct the SB of the host galaxy. Since the mass of

the quasar is not represented though an individual mass profile, a composite mass model can be obtained through the normal procedure described in Sect. 2.2.6.

Lensed quasar systems are typically modeled to infer the Hubble constant H_0, as described in Sect. 1.1.3.3. For this, a well constrained composite mass model is indispensable such that the velocity dispersion measurements are commonly included to break the MSD. Moreover, the time-delays Δt and LOS velocity estimates are used to constrain the model further.

2.2.7 Cluster modeling

The modeling of complete clusters is different in various aspects from the described procedure for strong galaxy-galaxy lensing. Not only does the system size change from few arcseconds to tens of arcseconds, but one is also typically in the weak lensing regime, possibly in combination with strong lensing effects. However, the main difference is the number of lenses involved. While we have one main lens in a galaxy-galaxy system, a cluster contains several dozens of galaxies that need to be taken into account. Each galaxy would have a different redshift, but given the small redshift difference within the cluster, one often groups them together. This implies the multi-plane case (see Sect. 2.1.5) unless a single redshift is adopted for the whole cluster. In addition, given the size of a cluster, there are normally LOS objects at a completely different redshift. To take these additional objects into account is crucial especially for the weak lensing observations. Therefore, we often use only the image positions as constraints, which also avoids modeling of the light distribution. Since the number of lensed image sets are high, even more advanced models that include several mass profiles is well constrained. For instance, Grillo et al. (2018, 2020) modeled the lensing system SN Refsdal introduced in Sect. 1.1.3.4 using 89 multiple image systems, i.e. 178 x and y coordinates, as well as four time delays as constraints. On the other side, they have 28 free parameters describing the total mass of the cluster, 56 x and y coordinates from the background sources, three redshift families and additionally two free parameter for the Hubble constant $H_0 \in [20, 120]\,\mathrm{km\,s^{-1}Mpc^{-1}}$ and $\Omega_M = 1 - \Omega_\Lambda \in [0, 1]$ in a ΛCDM cosmology. In sum, they have 93 more data points than free parameters, making it possible to constrain the mass distribution and cosmological parameters very well.

2.2.8 Optimization algorithms

As described in the previous sections, the lens light, lens mass and possibly also the source light or neighboring objects within the cutout are described by well-motivated but adopted profiles. During the modeling process, the profile parameters are varied to find the best set of values to represent the observation. The difference between observed data and predictions is described by the χ^2 whose exact form depends on the modeling stage. For instance, the expression for the lens light fit is given by Eq. (2.40). Therefore, the task of optimizing the model is translated into minimizing the χ^2, for which various mathematical algorithms exist and are supported by our modeling software. This parameter inference is typically a Bayesian inference relying on Bayes' Theorem

$$P(C|B) = P(B|C) \times \frac{P(C)}{P(B)} \tag{2.50}$$

where $P(C)$, called prior, is the probability of observing event C and $P(C|B)$, which is called posterior, describes the probability of observing event C given that B has happened. In this terminology, we call further $P(B)$ the evidence, which is often a constant given that the observed quantity B does not change, and $P(B|C)$ the likelihood. The likelihood is then connected to the χ^2 through

$$\mathcal{L} \propto e^{-\chi^2/2}, \tag{2.51}$$

which implies that minimizing the χ^2 corresponds to maximizing the likelihood. With this theorem, we get the updated probability of C given new information from B.

A common approach to Bayesian inference is based on Markov chain Monte Carlo (MCMC) methods (e.g., Mackay, 1992, 2003). These methods sample the parameters of the model and obtain the posterior from Bayes' Theorem given in Eq. (2.50). Explicitly, we use the Metropolis Hastings algorithm (Hastings, 1970; Robert & Casella, 2004), which contains the following steps:

- Take a point in the parameter space

- Propose a direction to go through a proposal density distribution

- Calculate the posterior probability of the new point and accept the new point
  if P(new point)/P(old point) > 1 or
  if P(new point)/P(old point) < 1 and P(new point)/P(old point) > k,
  where $P(x)$ is the probability of point x and k is a random number between 0 and 1.

After a certain number of steps, the chain will converge to the final and steady state. One important characteristic of MCMC methods is that the final state does not depend on the starting point. As a proposal density distribution, it is common to use a Gaussian distribution.

An MCMC chain does not only allow to infer better fitting parameters, it also allows us to generate a covariance matrix from a full chain. This helps especially to improve the performance for sampling a high-dimensional parameter space with several degeneracies among those parameters.

Since MCMC sampling is not parallelized and with an increasing number of parameters and data points, which are the image pixels for the extended image modeling (see Sect. 2.2.2 and following modeling steps), it becomes relatively slow. We therefore also make use of the emcee software package developed by Foreman-Mackey et al. (2013). Emcee is an affine-invariant ensemble sampler for MCMC and was proposed by Goodman & Weare (2010). The main advantage is its parallelization resulting in a huge speed up, coming along with slightly lower precision. However, emcee is perfectly suited for generating a good covariance matrix from a relative long chain for another sampling sequence.

In addition to that, we also use extensively the so called simulated annealing procedure, which is adopted from thermodynamics, explicitly from annealing in metallurgy. For this, one introduces a "temperature" T' in the relation between χ^2 and the likelihood:

$$\mathcal{L} = e^{-\chi^2/2} \;\Rightarrow\; e^{-\chi^2/(2T')}. \tag{2.52}$$

In the procedure of simulated annealing the temperature is lowered iteratively, starting with a high temperature such that the probability distribution has lower peaks and it is easier to find the global maximum of $\mathcal{L}$. In each optimization iteration, the temperature is lowered to increase the peaks in the distribution, allowing to find better the exact position of the maximum. Additionally, we start with a large step size, which is reduced in each iteration as well. This helps further to first find the global maximum, and as we come closer, to fit better to the distribution with smaller step sizes. The parameters are again optimized through a short MCMC chain in each iteration, where also a covariance matrix can be included for better performance.

We further use the dual annealing optimization[1] (Tsallis, 1988; Tsallis & Stariolo, 1996; Xiang et al., 1997; Xiang & Gong, 2000; Xiang et al., 2013; Mullen, 2014), which is a similar stochastic annealing approach and combines variations of simulated annealing (Tsallis, 1988; Tsallis & Stariolo, 1996) with a strategy for applying a local search on accepted locations (Xiang et al., 1997).

Moreover, we use another stochastic algorithm called basin hopping[2] (Wales & Doye, 1997; Wales & Scheraga, 1999; Wales, 2003; Li & Scheraga, 1987), which is a two-phase method introduced by Wales & Doye (1997). It is inspired by the natural process of energy minimization of clusters of atoms. In each iteration it performs the following steps:

- random perturbation of the parameter values

- local minimization

- acceptance or rejection the new values based on an acceptance test

The acceptance test in our case is also based on the Metropolis criterion with temperature T'. The new point is accepted if it lowers the χ^2, or otherwise with a probability of

$$e^{[\chi^2(\text{old point}) - \chi^2(\text{new point})]/T'} .$$
(2.53)

In the case of $T' = 0$, the algorithm becomes a monotonic basin hopping, which means that only steps with $\chi^2(\text{new point}) < \chi^2(\text{old point})$ are accepted.

With this collection of optimization tools, we are able to infer the best fitting parameters of a given observation.

---

[1]Python Package available here: `https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.dual_annealing.html`

[2]Python Package available here: `https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.basinhopping.html`

3 Lens mass modeling through a CNN assuming SIE-only

Summary

Gravitational lensing is a very powerful tool to answer many outstanding questions in astrophysics. For nearly all applications, a mass model of the foreground object(s) is needed. With the number of current known galaxy-galaxy lenses in addition to the expected number from up-coming surveys like LSST, current modeling approaches are insufficient due to their high resource and time consuming sampling procedure. Therefore, and because of the broad applicability of ML in image processing, Hezaveh et al. (2017) suggested and demonstrated the opportunity to use a CNN to model strongly lensed images observed with HST. Since most detection are made by ground based surveys such as Pan-STARRS, HSC or in the future LSST, we developed a CNN to model ground based images from HSC published in Schuldt et al., A&A 646, A126, 2021, which is reproduced in this chapter.

To train a NN, one needs in general a large enough training set that represents the data the network shall applied to. In the case of supervised learning, which we use here, also the output values, the so-called ground truth, is needed. Since there are not enough real lenses with the mass model parameter values available, we need to mock them up. It is crucial for a good performance on real data that the training data are as realistic as possible. Therefor we developed a simulation code that uses real observed images of galaxies and only simulate the lensing effect. With this code we created around 10^5 images each in four filters. We consider also different assumptions such as image configuration or distribution and range of the Einstein radius. With those data sets, we trained CNNs to predict five values for the five SIE parameters (lens center x and y, complex ellipticity e_x and e_y, and the Einstein radius θ_E). We find very good performance of the networks, especially by training with a uniform distribution of the Einstein radius comparable to models obtained with traditional methods on ground-based images and a SIE profile.

Since the network predicts the mass parameters very fast, the network can be useful for planning follow-up observation in case e.g., a SN goes off in the background source. Therefore, we also compared the image positions and time delays using the ground truth mass parameters or the network output and find good performance.

Author contribution

I contributed the main driving force to this project. I have developed the code to simulate strong lensing images, making use of the code GLEE (Suyu & Halkola, 2010; Suyu et al., 2012). I developed the network code and have done all corresponding training and testing of the networks. For the image prediction and time delay comparison I made use of a code contributed by Aleksi Halkola to this project. I am the main author of the paper and created all included figures.

HOLISMOKES - IV. Efficient mass modeling of strong lenses through deep learning

S. Schuldt, S. H. Suyu, T. Meinhardt, L. Leal-Taixé, R. Cañameras, S. Taubenberger, and A. Halkola

ABSTRACT

Modeling the mass distributions of strong gravitational lenses is often necessary in order to use them as astrophysical and cosmological probes. With the large number of lens systems ($\gtrsim 10^5$) expected from upcoming surveys, it is timely to explore efficient modeling approaches beyond traditional Markov chain Monte Carlo techniques that are time consuming. We train a convolutional neural network (CNN) on images of galaxy-scale lens systems to predict the five parameters of the singular isothermal ellipsoid (SIE) mass model (lens center x and y, complex ellipticity e_x and e_y, and Einstein radius θ_E. To train the network we simulate images based on real observations from the Hyper Suprime-Cam Survey for the lens galaxies and from the Hubble Ultra Deep Field as lensed galaxies. We tested different network architectures and the effect of different data sets, such as using only double or quad systems defined based on the source center and using different input distributions of θ_E. We find that the CNN performs well, and with the network trained on both doubles and quads with a uniform distribution of $\theta_E > 0.5''$ we obtain the following median values with 1σ scatter: $\Delta x = (0.00^{+0.30}_{-0.30})''$, $\Delta y = (0.00^{+0.30}_{-0.29})''$, $\Delta\theta_E = (0.07^{+0.29}_{-0.12})''$, $\Delta e_x = -0.01^{+0.08}_{-0.09}$, and $\Delta e_y = 0.00^{+0.08}_{-0.09}$. The bias in θ_E is driven by systems with small θ_E. Therefore, when we further predict the multiple lensed image positions and time-delays based on the network output, we apply the network to the sample limited to $\theta_E > 0.8''$. In this case the offset between the predicted and input lensed image positions is $(0.00^{+0.29}_{-0.29})''$ and $(0.00^{+0.32}_{-0.31})''$ for the x and y coordinates, respectively. For the fractional difference between the predicted and true time-delay, we obtain $0.04^{+0.27}_{-0.05}$. Our CNN model is able to predict the SIE parameter values in fractions of a second on a single CPU, and with the output we can predict the image positions and time-delays in an automated way, such that we are able to process efficiently the huge amount of expected galaxy-scale lens detections in the near future.

3.1 Introduction

Strong gravitational lensing has become a very powerful tool for probing various properties of the Universe. For instance, galaxy-galaxy lensing can help to constrain the total mass of the lens and, assuming a mass-to-light ratio (M/L) for the baryonic matter, also its dark matter (DM) fraction. By combining lensing with other methods like measurements of the lens' velocity dispersion (e.g., Barnabè et al., 2011, 2012; Yıldırım et al., 2020) or the galaxy rotation curves (e.g., Hashim et al., 2014; Strigari, 2013), the dark matter can be better disentangled from the baryonic component and a 3D (deprojected) model of the mass density profile can be obtained. Such profiles are very helpful for probing cosmological models (e.g., Davies et al., 2018; Eales et al., 2015; Krywult et al., 2017).

Another application of strong lensing is to probe high-redshift sources thanks to the lensing magnification (e.g., Dye et al., 2018; Lemon et al., 2018; McGreer et al., 2018; Rubin et al., 2018; Salmon et al., 2018; Shu et al., 2018). In recent years, huge efforts have been made in reconstructing the surface brightness distribution of lensed extended sources. Together with redshift and kinematic measurements, these observations contain information about the evolution of galaxies at higher reshifts. If the mass profile of the lens is well constrained, the original unlensed morphology can be reconstructed (e.g., Warren & Dye, 2003; Suyu & Blandford, 2006; Nightingale et al., 2018; Rizzo et al., 2018; Chirivì et al., 2020).

Lensed supernovae (SNe) and lensed quasars are very powerful cosmological probes. By measuring the time-delays of a lensing system with an object that is variable in brightness, one can use it to constrain, for example, the Hubble constant H_0 (e.g., Refsdal, 1964; Chen et al., 2019; Rusu et al., 2020; Wong et al., 2020; Shajib et al., 2020). This helps to assess the 4.4σ tension between the Cosmic microwave background (CMB) analysis that gives $H_0 = (67.36 \pm 0.54)$ km s^{-1}Mpc^{-1} for flat Λ cold dark matter (ΛCDM; Planck Collaboration et al., 2020) and the local distance ladder with $H_0 = (74.03 \pm 1.42)$ km s^{-1}Mpc^{-1} (SH0ES project; Riess et al., 2019). To date, time-delay lensing cosmography has been mainly based on lensed quasars as the chance of a lensed supernova (SN) is substantially lower. There are currently two lensed SNe known: one core-collapse SN behind a strong lensing cluster MACS J1149.5+222.3 (SN Refsdal; Kelly et al., 2015) and one SN type Ia behind an isolated lens galaxy (iPTF16geu; Goobar et al., 2017). Thanks to the upcoming wide field surveys in the next decades, like the Rubin Observatory Legacy Survey of Space and Time (LSST, Ivezic et al., 2008), this will change. LSST is expected to detect hundreds of lensed SNe (e.g., Goldstein et al., 2019; Wojtak et al., 2019). Therefore, it is important to be prepared for such exciting transient events in a fully automated and fast way. In particular, a fast estimation of time-delay(s) is important for optimizing the observing–monitoring strategy for time-delay measurements.

In addition to time-delay measurements, observing lensed SNe type Ia can help to answer outstanding questions about their progenitor systems (Suyu et al., 2020). The basic scenario is the single degenerate case where a white dwarf (WD) is stable until it reaches the Chandrasekhar mass limit (Whelan & Iben, 1973; Nomoto, 1982) by accreting mass from a nearby star. Today there are also alternative scenarios considered where the WD explodes before reaching the Chandrasekhar mass, the so-called sub-Chandrasekhar detonations (Sim et al., 2010). Another possibility for a SN Ia is the double-degenerated scenario where the companion is another WD (e.g., Pakmor et al., 2010) and both are merging to exceed the Chandrasekhar mass limit. It is

still unclear which of the main scenarios is correct to describe the SN Ia formation, or if both are. To shed light on this debate, one possibility is to observe the SN Ia spectroscopically at very early stages, which is normally difficult because SN detections are often close to peak luminosity, past the early phase. If this SN is lensed, we can use the position of the first appearing image, together with a mass model of the underlying lens galaxy, to predict the position and time when the next images will appear. Here it is very important to react quickly, particularly to compute the mass model of the underlying lens galaxy based on imaging, as the time-delays of galaxy-galaxy strong lensing are typically on the order of days to weeks.

Since these strong lens observations are very powerful, several large surveys including the Sloan Lens ACS (SLACS) survey (Bolton et al., 2006; Shu et al., 2017), the CFHTLS Strong Lensing Legacy Survey (SL2S; Cabanac et al., 2007; Sonnenfeld et al., 2015), the Sloan WFC Edge-on Late-type Lens Survey (SWELLS; Treu et al., 2011), the BOSS Emission-Line Lens Survey (BELLS; Brownstein et al., 2012; Shu et al., 2016b; Cornachione et al., 2018), the Dark Energy Survey (DES; Dark Energy Survey Collaboration et al., 2005; Tanoglidis et al., 2020), the Survey of Gravitationally-lensed Objects in HSC Imaging (SuGOHI; Sonnenfeld et al., 2018a; Wong et al., 2018; Chan et al., 2020; Jaelani et al., 2020a), and surveys in the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS; e.g., Lemon et al., 2018; Cañameras et al., 2020) have been conducted to find lenses. So far we have detected several thousand lenses, but mainly from the lower redshift regime. However, based on newer upcoming surveys like the LSST, which will target around $20,000$ deg^2 of the southern hemisphere in six different filters (u, g, r, i, z, y), together with the Euclid imaging survey from space operated by the European Space Agency (ESA; Laureijs et al., 2011), we expect billions of galaxy images containing on the order of one hundred thousand lenses (Collett, 2015).

To deal with this huge amount of images there are ongoing efforts to develop fast and automated algorithms to find lenses in the first place. These methods are based on different identification properties, for instance on geometrical quantification (Bom et al., 2017; Seidel & Bartelmann, 2007), spectroscopic analysis (Baron & Poznanski, 2017; Ostrovski et al., 2017), or color cuts (Gavazzi et al., 2014; Maturi et al., 2014). Moreover, convolutional neural networks (CNNs) have also been extensively used in gravitational lens detection (e.g., Jacobs et al., 2017; Petrillo et al., 2017; Schaefer et al., 2018; Lanusse et al., 2018; Metcalf et al., 2019; Cañameras et al., 2020; Huang et al., 2020) as they do not require any measurements of the lens properties. Once a CNN is trained, it can classify huge amounts of images in a very short time, and is thus very efficient. Nonetheless, CNNs have limitations (e.g., completeness or accurate grading) and the performance strongly depends on the training set design as it encodes an effective prior (in the case of supervised learning). In this regard unsupervised or active learning might be promising future avenues for finding lenses.

However, these methods are only for finding the lenses; a mass model is necessary for further studies. Mass models of gravitational lenses are often described by parameterized profiles, where the parameters are optimized, for instance via Markov chain Monte Carlo (MCMC) sampling (e.g., Jullo et al., 2007; Suyu & Halkola, 2010; Sciortino et al., 2020; Fowlie et al., 2020). These techniques are very time and resource consuming as modeling one lens can take weeks or months, and they are thus difficult to scale up for the upcoming amount of data. With the success of CNNs in image processing, Hezaveh et al. (2017) showed the use of CNNs in estimating the mass model parameters of a singular isothermal ellipsoid (SIE) profile, and investigated fur-

ther error estimations (Perreault Levasseur et al., 2017), analysis of interferometric observations (Morningstar et al., 2018), and source surface brightness reconstruction with recurrent inference machines (RIMs; Morningstar et al., 2019). While they mainly consider single-band images and subtract the lens light before processing the image with the CNN, Pearson et al. (2019) presented a CNN to model the image without lens light subtraction. However, for all deep learning approaches one needs a data set that contains the images and the corresponding parameter values for training, validation, and testing the network. As there are not that many real lensed galaxies known, both groups use mock lenses for their CNNs.

We recently initiated the Highly Optimized Lensing Investigations of Supernovae, Microlensing Objects, and Kinematics of Ellipticals and Spirals (HOLISMOKES) program (Suyu et al., 2020, hereafter HOLISMOKES I). After presenting our lens search project (Cañameras et al., 2020, hereafter HOLISMOKES II), we present in this paper a CNN for modeling strong gravitationally lensed galaxies with ground-based imaging, taking advantage of four different filters and not applying lens light subtraction beforehand. In contrast to Pearson et al. (2019), we use a mocked-up data set based on real observed galaxy cutouts since the performance of the CNN on real systems will be optimal when the mock systems used for training are as close to real lens observations as possible. Our mock lens images contain, by construction, realistic line-of-sight objects as well as realistic lens and source light distributions in the image cutouts. We use the Hyper Suprime-Cam (HSC) Subaru Strategic Program (SSP) images together with redshift and velocity dispersion measurements from the Sloan Digital Sky Survey (SDSS) for the lens galaxies, and images together with redshifts from the Hubble Ultra Deep Field (HUDF) survey for the sources (Beckwith et al., 2006; Inami et al., 2017).

The outline of the paper is as follows. We describe in Sect. 3.2 how we simulate our training data, and we give a short introduction and overview of the used network architecture in Sect. 6.3. The main networks are presented in Sect. 6.4, and we give details of further tests in Sect. 3.5. We also consider the image position and time-delay differences in Sect. 3.6 for a performance test, and compare them to other modeling techniques in Sect. 3.7. We summarize and conclude our results in Sect. 3.8. Throughout this work we assume a flat ΛCDM cosmology with a Hubble constant $H_0 = 72 \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$ (Bonvin et al., 2017) and $\Omega_M = 1 - \Omega_\Lambda = 0.32$ (Planck Collaboration et al., 2020). Unless specified otherwise, each quoted parameter estimate is the median of its 1D marginalized posterior probability density function, and the quoted uncertainties show the 16th and 84th percentiles (i.e., the bounds of a 68% credible interval).

3.2 Simulation of strongly lensed images

For training a neural network one needs, depending on the network size, from tens of thousands to millions of images, together with the expected network output, which in our case are the values of the SIE profile parameters corresponding to each image. Since there are far too few known lens systems, we need to mock up lens images. While previous studies are based on partly or fully generated light distributions (e.g., Hezaveh et al., 2017; Perreault Levasseur et al., 2017; Pearson et al., 2019), we aim to produce more realistic lens images by using real observed images of galaxies and simulating only the lensing effect with our own routine. We work with the four HSC filters, $g, r, i,$ and z (respectively matched to HST filters F435W ($\bar{\lambda} = 4343.4$Å),

F606W ($\overline{\lambda} = 6000.8$Å), F775W ($\overline{\lambda} = 7702.2$Å), and F850LP ($\overline{\lambda} = 9194.4$Å)) to give the network the color information to distinguish better between lens and source galaxies. The images of HSC for these filters are very similar to the expected image quality of LSST, such that our tests and findings will also hold for LSST. Therefore, this work is in direct preparation for and an important step in modeling the expected 100,000 lens systems that will be detected with LSST in the near future.

3.2.1 Lens galaxies from HSC

For the lenses we use HSC SSP[1] images from the second public data release (PDR2; Aihara et al., 2019) with a pixel size of 0.168″. To calculate the axis ratio q_{light} and position angle θ_{light} of the lens, we use the second brightness moments calculated for the i band since redder filters follow better the stellar mass; however, the S/N is substantially lower in the z band compared to the i band. We cross-match the HSC catalog with the SDSS[2] catalog to use only images of galaxies where we have SDSS spectroscopic redshifts and velocity dispersions. With this selection we end up with a sample containing 145,170 galaxies that is dominated by luminous red galaxies (LRGs). We show in Figure 3.1 a histogram of the lens redshifts used for the simulation (in gray). We already overplot the distribution of the mock samples discussed in Sect. 6.4.
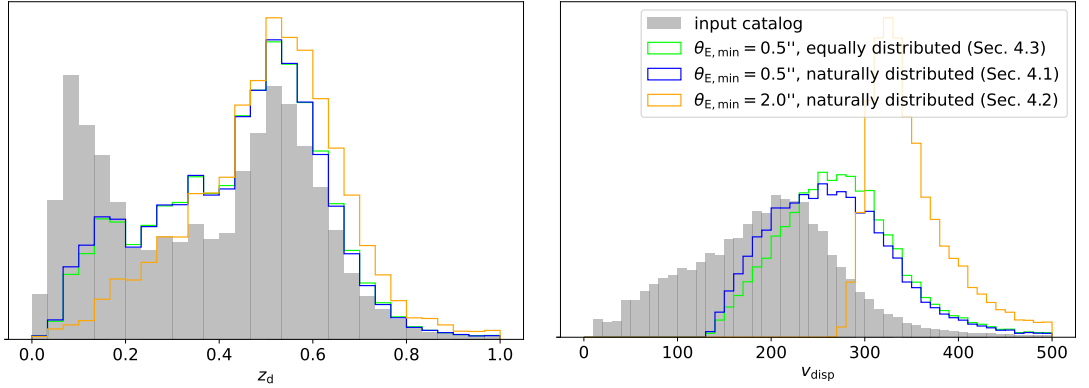

Figure 3.1: Distributions of the lens galaxy redshifts z_d (left) and velocity dispersion v_{disp} (right). Shown are the distributions of the input catalogs to the simulation code (in gray), and the distributions of the generated samples discussed in Sect. 6.4 (see inset for color-coding).

To describe the mass distribution of the lens, we adopt a SIE profile (Barkana, 1998) such that the convergence (dimensionless surface mass density) can be expressed as

$$\kappa(x, y) = \frac{\theta_E}{(1 + q)\, r} \tag{3.1}$$

---

[1] HSC SSP webpage: https://hsc-release.mtk.nao.ac.jp/doc/
[2] SDSS webpage: https://www.sdss.org/; catalog downloaded from the 14th data release page http://skyserver.sdss.org/dr14/en/tools /search/sql.aspx .

with elliptical radius

$$r = \sqrt{x^2 + \frac{y^2}{q^2}}, \tag{3.2}$$

where x and y are angular coordinates on the lens plane with respect to the lens center. In this equation θ_E denotes the Einstein radius and q the axis ratio.[3] The mass distribution is rotated by the position angle θ. The Einstein radius is obtained from the velocity dispersion v_{disp} with

$$\theta_E = 4\pi \frac{v_{disp}^2}{c^2} \frac{D_{ds}}{D_s}, \tag{3.3}$$

where c is the speed of light, and D_{ds} and D_s are respectively the angular diameter distances between the lens (deflector) and source and the observer and source. The distribution of the velocity dispersion is shown in Figure 3.1 (bottom pannel, gray histogram). We compute the deflection angles of the SIE with the lensing software GLEE (Suyu & Halkola, 2010; Suyu et al., 2012).

Based on the second brightness moments of the lens light distribution in the i band, the axis ratio q_{light} and position angle θ_{light} are obtained internally in our simulation code. Based on several studies (e.g, Sonnenfeld et al., 2018b; Loubser et al., 2020), the light traces the mass relatively well but not perfectly. Therefore, we add randomly drawn Gaussian perturbations on the light parameters, with a Gaussian width of $0.05''$ for the lens center, 0.05 for the axis ratio, and 0.17 radians (10 degrees) for the position angle, and adopt the resulting parameter values for the lens mass distribution. If the axis ratio of the mass q (i.e., with Gaussian perturbation) is above 1, we draw a second realization of the Gaussian noise. If the resulting q (from the second Gaussian perturbation) is ≤ 1, then we keep this value; otherwise, we set q to exactly 1.

While the simulation code assumes a parameterization in terms of axis ratio q and position angle θ, we parameterize for our network in terms of complex ellipticity e_c, which we define as $e_c = A\,e^{2i\theta} = e_x + i e_y$ with

$$
\begin{aligned}
e_x &= \frac{1-q^2}{1+q^2}\cos(2\theta), \\
e_y &= \frac{1-q^2}{1+q^2}\sin(2\theta).
\end{aligned} \tag{3.4}
$$

The back transformation is given by

$$
\begin{aligned}
q &= \sqrt{\frac{1-\sqrt{e_x^2+e_y^2}}{1+\sqrt{e_x^2+e_y^2}}} \\
\theta &= \begin{cases} \frac{1}{2}\arccos\left(e_y\,\frac{1+q^2}{1-q^2}\right) & \text{if } e_x > 0 \\ \frac{\pi}{2} + \left|\frac{1}{2}\arcsin\left(e_x\,\frac{1+q^2}{1-q^2}\right)\right| & \text{if } e_x < 0 \end{cases}.
\end{aligned} \tag{3.5}
$$

---

[3]The SIE mass profile introduced by Barkana (1998) allows for an additional core radius, which we set to 10^{-4}, that yields effectively a singular mass distribution without numerical issues at the lens center.

This is in agreement with previous CNN applications to lens modeling (Hezaveh et al., 2017; Pearson et al., 2019).

3.2.2 Sources from HUDF

The images for the sources are taken from HUDF[4] where the spectroscopic redshifts are also known (Beckwith et al., 2006; Inami et al., 2017). The cutouts are approximately $10'' \times 10''$ with a pixel size of $0.03''$. This survey is chosen for its high spatial resolution, and we can adopt the images without point spread function (PSF) deconvolution. Moreover, it contains high-redshift galaxies such that we can achieve a realistic lensing effect. The 1,323 relevant galaxies are extracted with Source Extractor (Bertin & Arnouts, 1996) since the lensing effect is redshift dependent and we would otherwise lens the neighboring objects as if they were all at the same redshift, which would lead to incorrect lensing features. We show a histogram of the source redshifts in Figure 3.2 (gray histogram). Since we randomly select a background source (see Sect. 3.2.3 for details), the source galaxies can be used multiple times for one mock sample, and thus the redshift distribution varies slightly between the different samples (colored histograms; see details in Sect. 6.4).
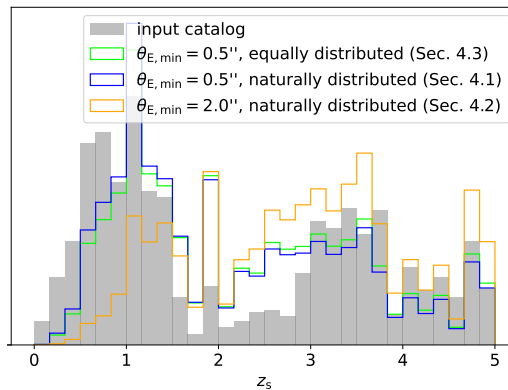

Figure 3.2: Distributions of the source redshifts z_s of the input catalog to the simulation code (gray) and of the different mock samples (colors) discussed in Sect. 6.4.

3.2.3 Mock lens systems

To train our networks we use mock images based on real observed galaxies, and only generate the lensing effect. We use HSC galaxies as lenses (see Sect. 3.2.1 for details) and HUDF galaxies as background objects (see Sect. 3.2.2) to obtain mocks that are as realistic as possible. Figure 3.3 shows a diagram of the simulation pipeline. The input has three images: the lens, the unlensed source, and the lens PSF image (top row). Together with the provided redshifts of source and lens, as well as the velocity dispersion for calculating the Einstein radius with equation (3.3),

---

[4]HUDF webpage https://www.spacetelescope.org/; downloaded on Oct. 1, 2018, from https://archive.stsci.edu/pub/hlsp/udf/acs-wfc/.

the source image can be lensed onto the lens plane (second row). For this we place a random source from our catalog randomly in a specified region behind the lens and accept this position if we obtain a strongly lensed image. Since the source images have previously been extracted, we use the brightest pixel in the *i* band to center the source. We have also implemented the option to just keep one of the two strong lens configurations, either quadruply or doubly imaged galaxies, classified based on the image multiplicity of the lensed source center. We also set a peak brightness threshold for the arcs based on the background noise of the lens. To estimate the background noise we take the corner with size 10%×10% square of the full lens image (rounded to an integer of pixel) with the lowest maximum and compute from the patch the root mean square (RMS) value used as background noise. We take the lowest maximum for each corner separately and then compute the RMS of that one because there might be line-of-sight objects in the corners that would raise the RMS values. To avoid contamination to the background estimation from the lens, we use $40'' \times 40''$ image cutouts such that each corner is $4'' \times 4''$. The peak brightness of the lensed source must then be higher than the RMS to be accepted by the simulation code.

In the next step the lensed source image with high resolution is convolved with the subsampled PSF of the lens, which is provided by HSC SSP PDR2 for each image separately. After binning the high-resolution lensed, convolved source image to the HSC pixel size and accounting for the different photometric zeropoints of the source telescope zp_{sr} and lens telescope zp_{ls}, which gives a factor of $10^{0.4(zp_{ls}-zp_{sr})}$, the lensed source image is obtained as if it had been observed through the HSC instrument (third row in Figure 3.3), i.e., on the HSC $0.168''$/pixel resolution. At this point we neglect the additional Poisson noise for the lensed arcs. Finally, the original lens and the mock lensed source images can be combined, which results in the final image (fourth row) that is cropped to a size of 64×64 pixels ($10.8'' \times 10.8''$). For better illustration, a color image based on the filters *g*, *r*, and *i* is also shown, but we generate all mock images in four bands, which we use for the network training. We show more example images based on *gri* filters in Figure 3.4. During this simulation, we set an upper limit on the Einstein radius of $5''$, which corresponds to the size of the biggest Einstein radius so far observed from galaxy-galaxy lensing (Belokurov et al., 2007).

We test the effect of different assumptions on the data set, like splitting up in quads-only or doubles-only, or different assumptions on the distribution of the Einstein radii since we found this to be crucial for the network performance. For this we generate with this pipeline new independent mock images that are based on the same lens and source images, but different combinations and alignments. The details of the different samples and the network trained on them will be discussed further in Sect. 6.4. For the set of quads-only and higher limit on the Einstein radius of $2''$ we use a modification of the conventional data augmentation in deep learning. In particular, we rotate only the lens image before adding the random lensed source image, but not the whole final image (which is done normally for data augmentation). Thus, the ground truth values are also not exactly the same values given the change in position angle and another background source with different location and redshift.
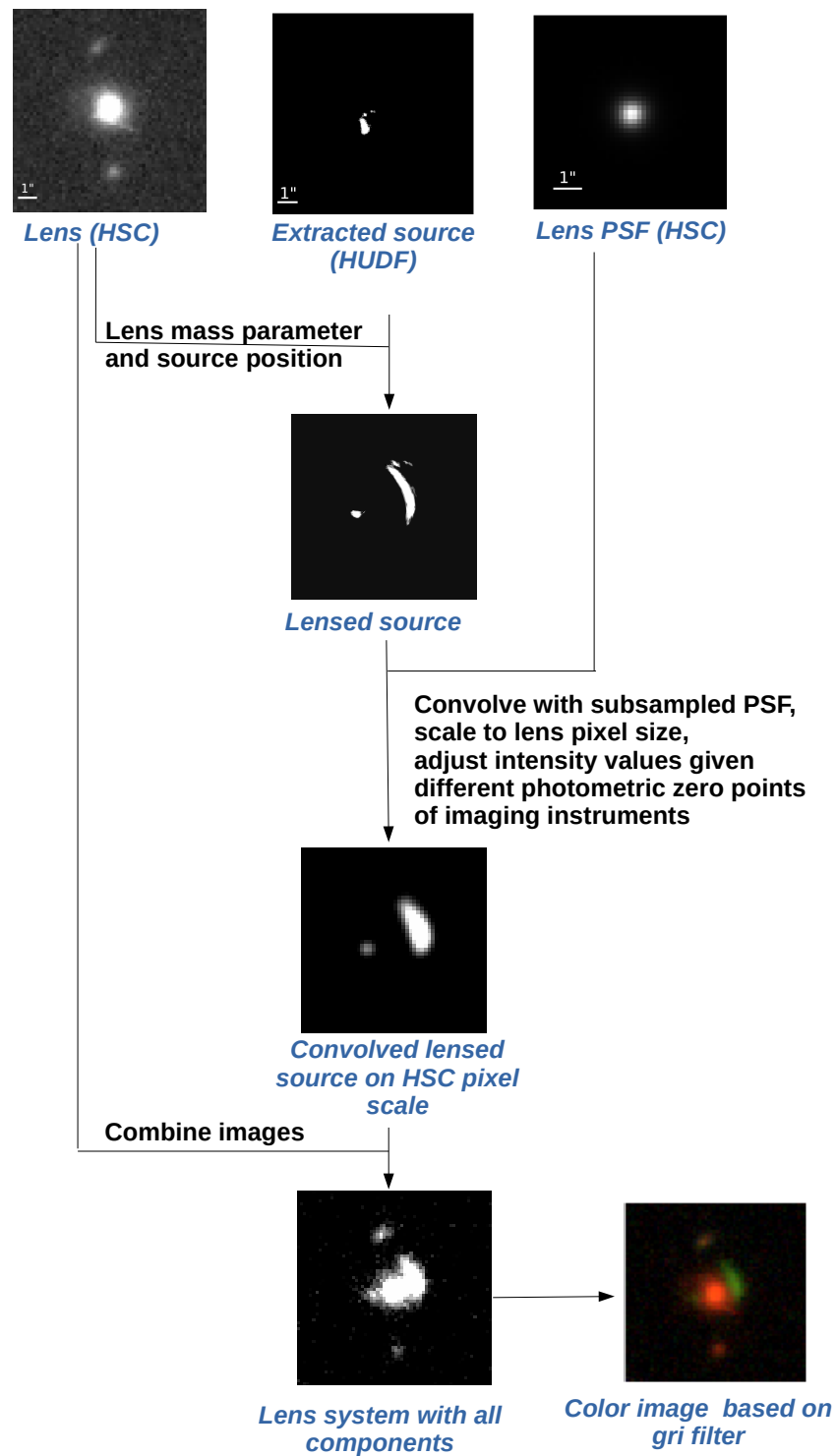
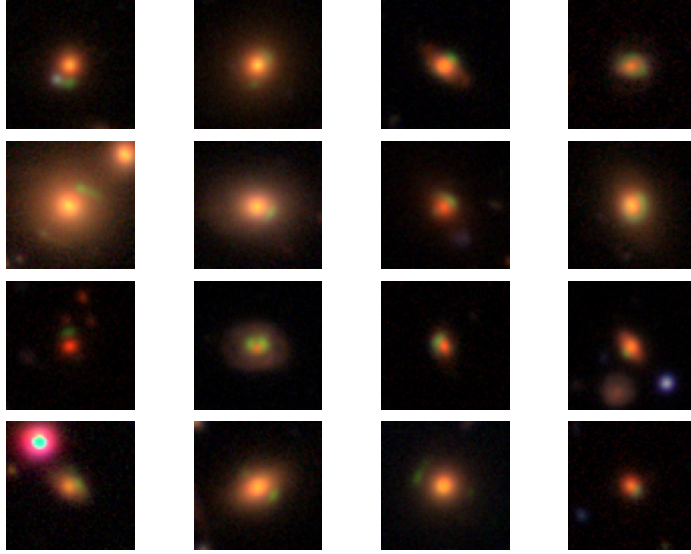**Figure 3.3:** Diagram of the simulation pipeline.

Figure 3.4: Examples of strong gravitational lens systems mocked up with our simulation code by using HUDF galaxies as sources behind HSC galaxies as lenses. Each image cutout is 10.8″ × 10.8″.

3.3 Neural networks and their architecture

Neural networks (NNs) are extremely powerful tools for a wide range of tasks, and thus in recent years broadly used and explored. Additionally, the computational time can be reduced notably compared to other methods. There are generally two types of NNs: (1) classification, where the ground truth uses different labels to distinguish between the different classes, and (2) regression, where the ground truth consists of a set of parameters with specific values. The latter is the kind we use here, which means that the network predicts a numerical value for each of the five different SIE parameters (x, y, e_x, e_y, and θ_E).

Depending on the problem the network needs to solve, there are several different types of networks. Since we are using images as data input, typically convolutional layers followed by fully connected (FC) layers are used (e.g., Hezaveh et al., 2017; Perreault Levasseur et al., 2017; Pearson et al., 2019). The detailed architecture depends on attributes such as the specific task, the size of the images, or the size of the data set. We have tested different architectures and found an overall good network performance with two convolutional layers followed by three FC layers, but no significant improvement for the other network architectures. A sketch of this is shown in Figure 6.4. The input has four different filter images for each lens system and each image a size of 64×64 pixels. The convolutional layers have a stride of 1 and a kernel size of $5 \times 5 \times C$, with $C = 4$ for the first layer and $C = 6$ for the second layer. Each convolutional layer is followed by a max-pooling layer of size $f \times f = 2 \times 2$ and stride 2. We use as activation function the rectified linear activation (ReLu) function. After the two convolutional layers, we obtain a data cube of size $13 \times 13 \times 16$, which is then passed through the FC layers after flattening to finally obtain

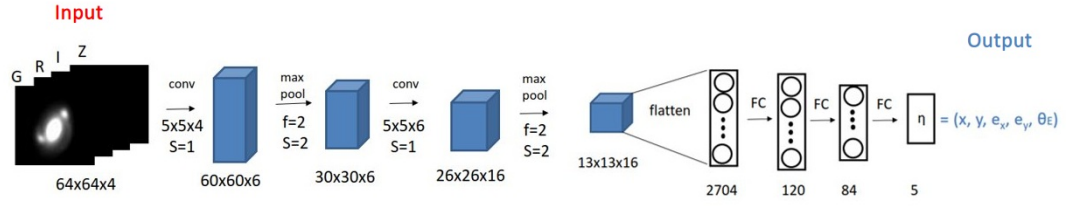the five output values. This network is coded with python 3.8.0 and uses pytorch modules (torch 1.5.1).

Figure 3.5: Overview of our main CNN architecture. The input has four different filter images for each lens system and each image a size of 64×64 pixels. The network contains two convolutional layers (conv) each followed by a max-pooling layer (max pool) with kernel size f and stride S values indicated in the figure. This is then followed, after flattening the data cube, by three fully connected (FC) layers to finally obtain the five output values of the SIE η, containing the lens center x and y, the complex ellipticity e_x and e_y, and the Einstein radius θ_E.

Independent of the exact network architecture, the network can contain hundreds of thousands of neurons or more. While initially the values of weight parameters and bias of each neuron are random, they are updated during the training. To see the network performance after the training, the data set is split into three samples: the training, the validation, and the test sets. We further divide those sets into random batches of size N. In each iteration the network predicts the output values for one batch (forward propagation), and after running over all batches from the training and validation sets, one epoch is finished. The error, which is called loss, is obtained for each batch with the loss function; we use the mean square error (MSE) defined as

$$L = \frac{1}{p \times N} \sum_{k=1}^{N} \sum_{l=1}^{p} (\eta_{k,l}^{\text{pred}} - \eta_{k,l}^{\text{tr}})^2 \times w_l, \tag{3.6}$$

where $\eta_{k,l}^{\text{tr}}$ and $\eta_{k,l}^{\text{pred}}$ respectively denotes the *lth* true and predicted parameter, in our case from $\{x, y, e_x, e_y, \theta_E\}$, of lens system k, and p denotes the number of output parameters. We incorporated in our loss function L weighting factors w_l, which are normalized such that $\sum_{l=1}^{p} w_l = p$ holds. This gives a weighting factor of 1 for all parameters if they are all weighted equally.

The loss value of that batch is then propagated to the weights and biases (back propagation) for an update based on a stochastic gradient descent algorithm to minimize the loss. This procedure is repeated in each epoch first for all batches of the training set and an average loss is obtained for the whole training set. Afterwards the steps are repeated for all batches of the validation set, while no update of the neurons is done, and an average loss for the validation set is obtained as well. The validation loss shows whether the network improved in that epoch or if a decreasing training loss is related to overfitting the neurons. A network is overfitting when it learns the training set, and not the features in the training set. After each epoch we reshuffle our whole training data to obtain a better generalization. This concludes one epoch, which is repeated iteratively to obtain a network with optimal accuracy.

This whole training corresponds to one so-called cross-validation run, where several cross-validation runs are performed by exchanging the validation set with another subset of the training set. For example, if the training set and validation set form five subsets {A, B, C, D, and E}, then we can have five independent runs of training where in each run the validation set is one of these five subsets and the training set contains the remaining four subsets. After the multiple runs, we can determine the optimal number of epochs for training by locating the epoch with the minimum average validation loss across the multiple runs. This procedure helps to minimize potential bias to certain types of lenses for a potentially unbalanced single split. The neural network trained on all five sets {A, B, C, D, E} up to that epoch is the final network, which is then applied to the test set that contains data the network has never seen before. In our case we used ~56% of the data set as the training set, ~14% as the validation set, and ~30% as the test set[5] in order to have a five-fold cross-validation for each network.

To find the best hyperparameter values for our specific problem, we test each network on its performance with several different variations of the hyperparameters. Independent of the data set, we train each cross validation run for 300 epochs, and apart from a few checks with different values, we fix the weight decay to 0.0005 and the momentum to 0.9. For the learning rate, batch size, and the initializations of the neurons, we perform a grid search, varying the learning rate $r_{\text{learn}} \in [0.1, 0.05, 0.01, 0.008, 0.005, 0.001, 0.0005, 0.0001]$ and batch size as 32 or 64 images per batch, and exploring three different network initializations. For the weighting factors of the contribution to the loss we test two options, either all parameters contribute equally (i.e., $w_l = 1 \; \forall \; l$ in Eq. (3.6)) or the contribution of the Einstein radius is a factor of 5 higher ($w_{\theta_E} = 5$). This already gives 96 different combinations of hyperparameters which we test with cross-validation and early stopping.

For a subset of the hyperparameter combinations, we test further possibilities. In particular, we explore the effect of drop-out (i.e., the omission of random neurons in every iteration) with a drop-out rate $p \in [0.1, 0.3, 0.5, 0.7, 0.9]$, but find no improvement. We further test different network architectures by adding an additional convolutional layer or fully connected layer, or by varying the number of neurons in the different layers. We also test a different weighting of the lens center parameters to the loss that is motivated by the results of our networks in Section 4.

In addition, we test the effect of five different scaling options of the input images for our data set, but assume here the learning rate $r_{\text{learn}} = 0.001$ for simplicity. First, we boost the *r* band by a factor of 10. Since the network is still able to recover the parameter values, we see that the network performance is not heavily affected by the absolute value of the images. Second, if we normalize each filter of one lens system independently of the other filters, the network fails to recover the correct parameter values. This shows us that the network is indeed able to extract the color information as the relative difference is much smaller, and thus needs the different filters. In the third and fourth option we normalize each filter with its maximum value or with the mean peak value of the different filters. Last, we also rescale the images by shifting them by the mean

---

[5]The percentiles vary slightly due to rounding effects depending on the absolute size of the simulated mocks of that sample and the assumed batch size.

value and dividing by the standard deviation[6]. Since we obtained no notable improvement with any one of these scalings, we use the images without rescaling to obtain our final networks.

3.4 Results

To train our modeling network we mock up lensing systems based on real observed galaxies with our simulation pipeline (see Sect. 3.2). Each lensing system is simulated in the four different filter *griz* of HSC to give the network color information to distinguish better between lens galaxy and lensed arcs. The network architecture assumes, as described in Sect. 6.3 in detail, images 64×64 pixels in size, which corresponds to a size of around $10'' \times 10''$.

During our network testing, we found that the distribution of Einstein radii in the training set is very important, especially as this is a key parameter of the model. Therefore, we trained a network under the assumption of different underlying data sets, for example a lower limit of the Einstein radius for the simulations or a different distribution of Einstein radii. We further tested the network performance by limiting to a specific configuration (i.e., only doubles or quads). We give an overview of the different data set assumptions in Table 3.1, as well as the best hyperparameter values that depend on the assumed data set.

We present in the following subsections our CNN modeling results for various data sets.

---

[6]The four individual images are rescaled as

$$I_{\text{scaled}} = \frac{(I - M)}{\sigma} \tag{3.7}$$

with mean

$$M = \sum_{k=0}^{f} \sum_{l,m=0,0}^{p1,p2} I_{k,l,m}/(f \times p1 \times p2) \,, \tag{3.8}$$

the number of filters f, and

$$\sigma = \sqrt{\sum_{k=0}^{f} \sum_{l,m=0,0}^{p1,p2} \frac{(I_{k,l,m} - M)^2}{(p1 \times p2 - 1)}} \,, \tag{3.9}$$

and $p1$ and $p2$ as image dimensions in pixels for the *x*- and *y*-axis, respectively. In our case we have $f = 4$ and $p1 = p2 = 64$.

Table 3.1: Overview of trained networks.

Natural distribution of Einstein radii of lenses

double	quad	$\theta_{\text{E,min}}$ ['']	$w_{\theta_{\text{E}}}$	loss	epoch	r_{learn}	N	seed	Section	Figures
✓	✓	0.5	1	0.0201	115	0.005	64	3	3.4.1	
✓	✓	0.5	5	0.0496	123	0.001	64	3	3.4.1	3.6, 3.7, 3.8, 3.13, 3.14
✓	✓	2.0	1	0.0120	85	0.01	32	3	3.4.2	
✓	✓	2.0	5	0.0209	85	0.008	32	2	3.4.2	3.7, 3.8, 3.9, 3.13, 3.14
✓		0.5	1	0.0193	242	0.008	64	1	3.5.1	
✓		0.5	5	0.0474	117	0.001	64	3	3.5.1	
✓		2.0	1	0.0118	163	0.05	64	3	3.5.1	
✓		2.0	1	0.0118	96	0.01	32	2	3.5.1	
✓		2.0	5	0.0217	62	0.008	32	3	3.5.1	
	✓	0.5	1	0.0193	151	0.008	32	2	3.5.1	
	✓	0.5	5	0.0441	69	0.001	32	2	3.5.1	
	✓	2.0	1	0.0129	267	0.01	64	2	3.5.1	
	✓	2.0	5	0.0268	285	0.005	32	1	3.5.1	

Uniform distribution of Einstein radii of lenses

double	quad	$\theta_{\text{E,min}}$ ['']	$w_{\theta_{\text{E}}}$	loss	epoch	r_{learn}	N	seed	Section	Figures
✓	✓	0.5	1	0.0223	147	0.001	32	1	3.4.3	
✓	✓	0.5	5	0.0528	112	0.0005	64	2	3.4.3	3.7, 3.8, 3.10, 3.11, 3.13, 3.14
	✓	0.5	1	0.0288	73	0.008	64	2	3.5.1	
	✓	0.5	5	0.0688	56	0.001	32	2	3.5.1	3.12

Note. The first and second columns indicate if quads and/or doubles are included in the data set. The parameter $\theta_{\text{E,min}}$ represents the lower limit on the Einstein radius in the simulation, and $w_{\theta_{\text{E}}}$ is the weighting factor of the Einstein radius in the loss function. The other parameters (lens center, ellipticity) are always weighted by a factor of 1 and the sum of all five weighting factors is normalized to the number of parameters. The fifth and sixth columns give the value of the loss of the test set and the epoch with the best validation loss. This is followed by the specific hyperparameters: learning rate r_{learn}, batch size N, and seed for the random number generator. The last two columns list the sections and the figures that present the results of the corresponding network.

3.4.1 Naturally distributed Einstein radii with lower limit 0.5″

For this network we use 65,472 mock lens images simulated following the procedure described in Sect. 3.2. Here we assume a lower limit of the Einstein radii of 0.5″ as otherwise the lensed source is totally blended with the lens and is not resolvable given the average seeing and image quality. The resulting redshift distributions are shown as the blue histograms for the lens in Figure 3.1 (top panel) and for the source in Figure 3.2. The lens redshift peaks at $z_d \sim 0.5$. Concerning the possible strong lensing configurations, the data set is dominated by doubles as expected. In addition, systems with smaller Einstein radii are more numerous than those with larger Einstein radii, as expected given the lens mass distribution, although the velocity dispersion (see Figure 3.1, bottom panel) peaks at around $v_{disp} \sim 280 \, \mathrm{km \, s^{-1}}$, and thus tends to include more massive galaxies than the input catalog (gray histogram). The distribution of the different parameters are shown in Figure 3.6, left panel; the red histogram depicts the true distribution and the blue one the predicted distribution. In the right panel we show the correlation between the true value and the predicted value for the five different parameters.

If we look at the performance on the lens center, which is measured in units of pixels with respect to the image cutout center, it seems as if the network fails totally in the first instance. We recall here how we obtain the lens mass center. In the simulation, we assume the lens *light* center to be the image center, and add a Gaussian variation (with standard deviation of 0.05″) to shift to the lens *mass* center. Thus, the ground truth (red histogram in Figure 3.6) follows a Gaussian distribution, while the predicted lens center distribution (blue) is peakier. This suggests that the network does not obtain enough information from the slight shift or distortion in the lensed arcs to correctly predict the lens mass center. We test upweighting the contribution of the lens center to the loss with a higher fraction, which results in a better performance on these two parameters, but then the performance on the other parameter deteriorates. We thus refrain from upweighting the lens center. Further difficulties on the centroid parameters are caused by all systems having the exact same lens light center (which is at the center of the image). If we assume that the lens mass perfectly follows the light distribution and the lens light center is always the same, the lens (mass) center ground truth will become a delta distribution, and the network will perform much better.Accordingly, in many automated lens modeling architectures (e.g., Pearson et al., 2019) the lens center is not even predicted. Since the difference of the center for nearly all lens systems is smaller than ±1 pixel, it does not affect the model noticeably. We nonetheless keep five parameters for generality, and suggest investigating in future work more in this direction by relaxing the strict assumption of coincidence centers of image cutout and of lens light.

Looking at the performance on the ellipticity, it turns out that most of the lens systems are approximately round (i.e., $e_x \sim e_y \sim 0$) and that the network can recover them very well. If the lens is more elliptical, the network performance starts to drop. This might be an effect of the lower number of such lens systems in the sample especially since the position angle becomes relevant, and thus the number of systems in a particular direction is again lower. We note that $e_x = \pm 0.3$ and $e_y = 0$ corresponds to an axis ratio $q = 0.73$ (i.e., quite elliptical). If the absolute value of e_x or e_y were higher, the axis ratio would be even lower, which seldom occurs in nature.

We see that the network recovers the Einstein radius better for lens systems with lower image separation than with high image separation ($\theta_E \gtrsim 2''$), which is in the first instance counterintuitive. If the lensed images are further separated, they are better resolved and less strongly
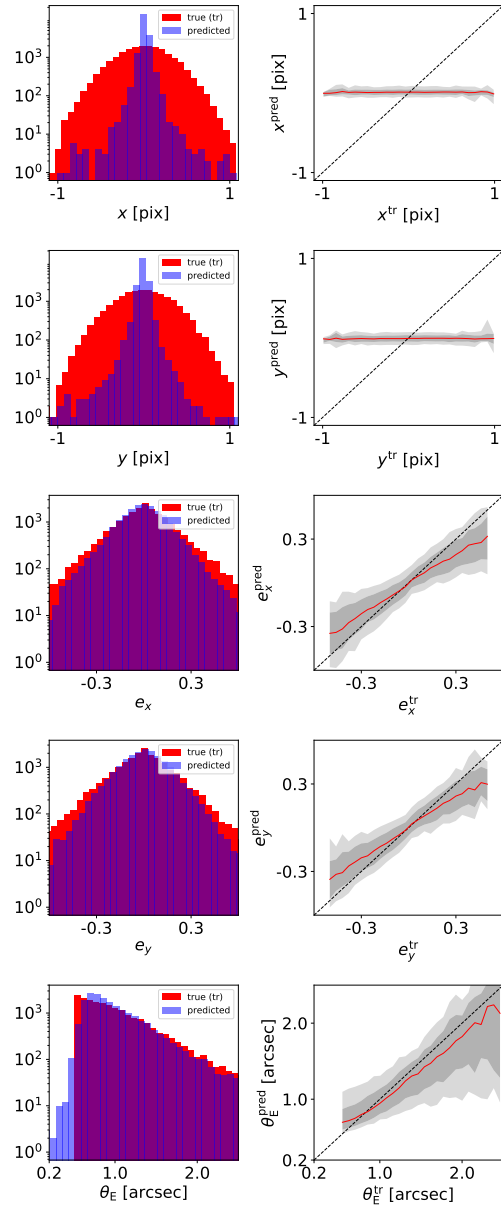
Figure 3.6: Network performance on the Einstein radius under the assumption of a lowest Einstein radius $\theta_{E,min}$ of $0.5''$and a weighting factor of $w_{\theta_E} = 5$. The left panel shows histograms of the ground truth (tr) in red and of the predicted values in blue. The right panel shows a direct comparison of the predicted against the true value, where the red line indicates the median of the distribution and the gray bands give the 1σ (16th to 84th percentile) and 2σ (2.5th to 97.5th percentile) ranges. From top to bottom are the five different model parameters, lens center x and y, complex ellipticity e_x and e_y, and Einstein radius θ_E. For all plots 30 bins over the plotting range are used.

blended with the lens, and we would expect better recovery of Einstein radii from the network. The worse network performance at larger Einstein radii can therefore only be explained by the relatively low numbers of these systems in the training data. We have more than two orders of magnitude more lens systems with $\theta_E \sim 0.5''$ than with $\theta_E \sim 2.0''$. Therefore, the network is trained to predict a small Einstein radius more often, and a larger Einstein radius less often. Since the lens systems with larger image separation are very interesting for a wide range of scientific applications, it is desirable to improve the network performance specifically on those lens systems. Therefore, we test a network with the same data set where the Einstein radius difference contributes a factor of 5 more to the loss than the other parameters. In the case of this weighted network, the prediction performance is very similar for the lens center and ellipticity, but slightly better for the Einstein radius. If we increase the contribution of the Einstein radius further, we notably worsen the performance on the other parameters.

As a further comparison of the ground truth with the predicted values of the test set, we show in Figure 3.7 the difference as normalized histograms (bottom row) and the 2D probability distributions (blue), where we find no strong correlation among the five parameters. The obtained median values with 1σ uncertainties for the different parameters are, respectively, $(0.00^{+0.31}_{-0.30})''$ for Δx, $(-0.01^{+0.29}_{-0.31})''$ for Δy, $0.00^{+0.08}_{-0.09}$ for Δe_x, $0.01^{+0.09}_{-0.08}$ for Δe_y, and $(0.02^{+0.21}_{-0.18})''$ for $\Delta\theta_E$, where Δ denotes the difference between the predicted and ground truth values. As an example, a shift of $e_x = 0.3$ to $e_x = 0.15$ with fixed $e_y = 0$ results in a shift from $q = 0.73$ to $q = 0.86$.

Finally, we show in Figure 3.8 the difference in Einstein radii as a function of the logarithm of the ratio between lensed source intensity I_s and lens intensity I_l determined in the i band, which we hereafter refer to as the brightness ratio. In the top right panel, we show the distribution of the brightness ratio. The lens intensity is defined as the sum of all the pixel values in the 64 pixels $\times$ 64 pixels cutout of the lens such that it is slightly overestimated due to light contamination from surrounding objects. The distribution peaks around -2 in logarithm to basis 10, which means that the lensed source flux is a factor 100 below that of the lens. The bottom left plot shows the median with 1σ values of the Einstein radius differences for each brightness ratio bin. Focussing on the blue curve for this section, we find a bias in the Einstein radius which is driven by the small lensing systems with $\theta_E \lesssim 0.8''$ (compare Figure 3.6). Excluding these small lensing systems, we show the corresponding plot in the lower right panel. With this limitation, we no longer find a bias, and obtain a median with 1σ values of $0.00^{+0.17}_{-0.14}''$ for the Einstein radius difference. We find a slight improvement of the performance with increasing brightness ratio for both the full sample (bottom left panel) and the sample with $\theta_E > 0.8''$ (bottom right panel).

To further improve the network performance for wide-separation lenses, we train separate networks for lens systems with Einstein radius $\theta_E > 2.0''$ in Sect. 3.4.2, and for lens systems where we artificially boost the number of lenses at the high end of θ_E in Sect. 3.4.3.

3.4.2 Naturally distributed Einstein radii with lower limit $2.0''$

Since the network presented in Sect. 3.4.1 cannot easily recover a large Einstein radius ($\theta_E \gtrsim 2''$), we test the performance of a network specialized for the high end of the distribution and set the lower limit to $\theta_{E,min} = 2''$. Because of the higher limit on the Einstein radii, the velocity dispersion (see orange histogram in Figure 3.1, bottom) is shifted towards the high end, which corresponds to more massive galaxies. We also find that the lens and source redshifts (orange
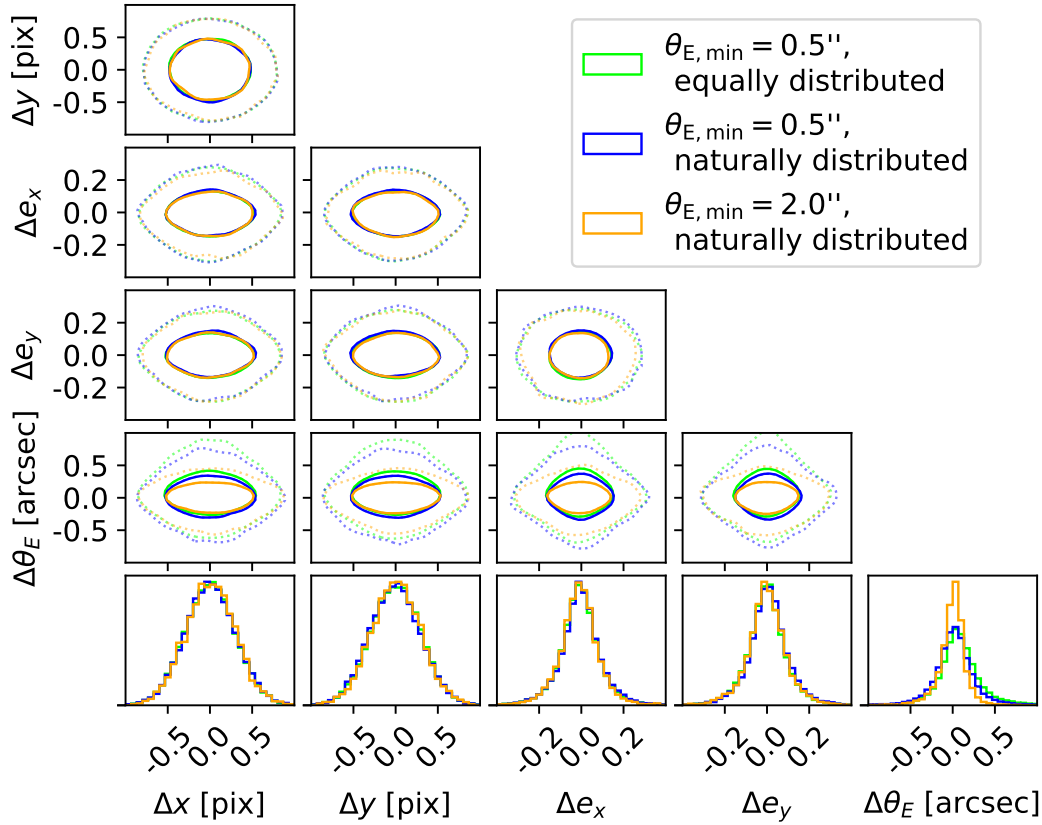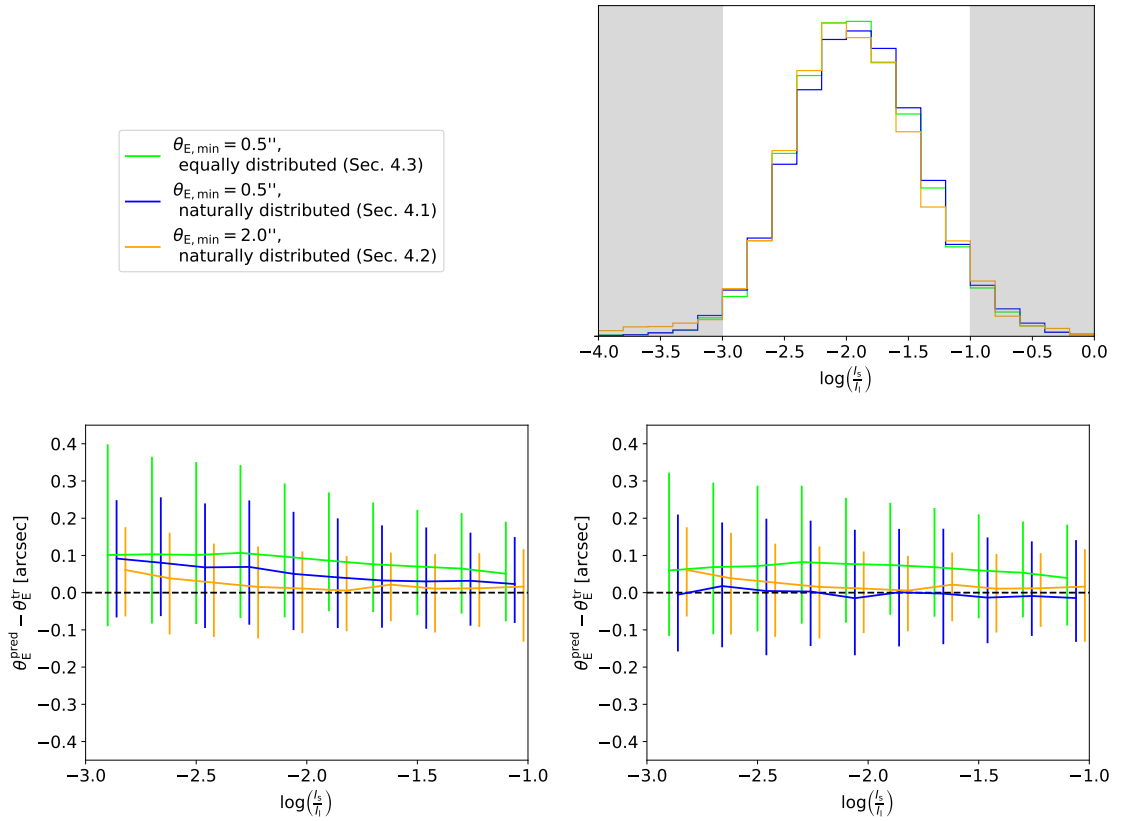
Figure 3.7: Comparison of the performance of the three networks described in Sect. 6.4. All samples include doubles and quads and a weighting factor of $w_{\theta_E} = 5$, but different Einstein radius distributions or lower limits on the Einstein radius (see legend). In the bottom row are shown the normalized histograms of the difference between predicted values and ground truth for the five parameters and above the 2D correlations distribution: 1σ contour (solid line) and 2σ contour (dotted line).

histograms in Figure 3.1 and Figure 3.2, respectively) tend to slightly higher values. Since we use the natural distribution of Einstein radii as in Sect. 3.4.1, the image-separation distribution is again bottom-heavy and the number of mock lens systems is smaller (25,623), as shown in Figure 3.9. From the blue (predicted) histogram, we see that the true distribution (red histogram) is well recovered.

In the right panel of Figure 3.9 we show the correlation of predicted and true Einstein radii. The red line, which follows quite well the diagonal dashed line, shows the median. The gray shaded regions show the 1σ and 2σ regions. We find that the network performs much better for $\theta_E \sim 2''$ than for the network trained in the full range (Sect. 3.4.1). However, this is again due to the number of lens systems that decreases towards $\theta_E \sim 4''$, and the scatter that increases dramatically for the high end of the data set.

We further show 1D and 2D probability distributions for this network in Figure 3.7 (orange) as well as the histogram of the brightness ratio, and the difference of the Einstein radii as function

Figure 3.8: Comparison of the performance of the three networks described in Sect. 6.4. All samples include doubles and quads and a weighting factor of $w_{\theta_{\mathrm{E}}} = 5$, but different Einstein radius distribution or lower limits on the Einstein radius as indicated in the legend (upper left). The upper right panel shows the histogram of the brightness ratio of lensed source and lens. The bottom panel shows for the full sample (left) and limited to $\theta_{\mathrm{E}} > 0.8''$ (right) the difference in Einstein radius as a function of the brightness ratio with the 1σ values. Shown are the Einstein radius difference in the range $-3 \leq \log\left(\frac{I_{\mathrm{s}}}{I_l}\right) \leq -1$ (white area in the histogram) where there are enough data points, and the blue and orange bars have been shifted slightly to the right for better visualization.

of the brightness ratio in Figure 3.8. While the performance for the lens center and complex ellipticity is very similar to the network presented in Sect. 3.4.1, we achieve an improvement for the Einstein radius. This is expected as the network is specifically trained for lens systems with large image separation. As we see from Figure 3.8, the larger systems do not have a higher brightness ratio on average as one might expect. As we have already seen, the network performs notably better on the Einstein radii over the whole brightness ratio range. We no longer overpredict the Einstein radius for $\log\left(\frac{I_{\mathrm{s}}}{I_l}\right) \gtrsim -2.5$, and the 1σ values are smaller as well.
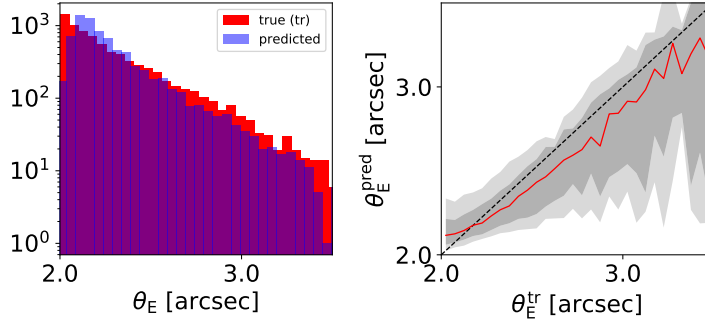
Figure 3.9: Network performance on the Einstein radius under the assumption of a smalles Einstein radius $\theta_{E,min}$ of $2.0''$. The left panel shows the histograms of the ground truth (tr) in red and of the predicted values in blue. The right panel is a 1:1 plot of predicted against true Einstein radius. The red line shows the median of the distribution and the gray bands give the 1σ and 2σ ranges. For both plots 30 bins over the plotting range are used.

3.4.3 Uniformly distributed Einstein radii with lower limit $0.5''$

Because of the extreme decrease in the number of systems towards large image separation, we test a network trained on a more uniformly distributed sample. For this, we generate more lens systems with high image separation by rotating the lens image by $n\pi/2$ with $n \in [0, 1, 2, 3]$. Here we do not reuse the same lens in the same rotation to avoid producing multiple images of lens systems that are too similar. We note that the background source and position are always different such that the lensing effect varies (see Sect. 3.2 for further details on the simulation procedure). We limit the sample to a maximum of 8,000 lens systems per $0.1''$ bin resulting in a sample of 140,812 lens systems. This results in a more uniform distribution, though the bins with the largest image separation still have fewer lens systems since it is very difficult, and a very seldom occurrence, to obtain a lensing configuration with an image separation above $\sim 2.5''$ due to the mass distribution of galaxy-scale lenses. The biggest image separation within this sample is $\sim 4.5''$, which is about 10% lower than the upper limit of $5''$ that we set for our simulations (see Sect. 3.2.3). The redshift distributions, shown as green histograms in Figure 3.1 and Figure 3.2, are similar to that of the naturally distributed sample (blue), whereas the lens velocity dispersions (Figure 3.1, bottom panel) tend to be higher (i.e., more massive galaxies), as expected.

Similar to the networks trained with natural Einstein radius distribution (see Sect. 3.4.1 and Sect. 3.4.2), in Figure 3.10 we show histograms (left column) and a 1:1 comparison (right column), but now for all five SIE parameters (from top to bottom for the lens center x and y, the complex ellipticity e_x and e_y, and the Einstein radius θ_E). For this network we obtain a median value with 1σ scatter of $(0.00^{+0.30}_{-0.30})''$ for Δx, $(0.00^{+0.30}_{-0.29})''$ for Δy, $-0.01^{+0.08}_{-0.09}$ for Δe_x, $0.00^{+0.08}_{-0.09}$ for Δe_y, and $(0.07^{+0.29}_{-0.12})''$ for the Einstein radius $\Delta \theta_E$. Comparing the performance on the Einstein radius to the network from Sect. 3.4.1 with a natural Einstein radius distribution, we see a significant improvement for the systems with larger image separation. Therefore, we can confirm that the underprediction of the Einstein radius in Sect. 3.4.1 is due to the relatively small number of large-θ_E systems in the training data. On the other hand, based on this plot the new

network seems to be slightly worse for the low-image separation systems. It tends to overpredict the Einstein radius at $\theta_E \lesssim 2.0''$ such that when we limit to $\theta_E > 0.8''$ as in Sect. 3.4.1, we only get a slight improvement in reducing the scatter and obtain $\Delta\theta_E = (0.07^{+0.25}_{-0.08})''$. Therefore, it turns out that the performance depends sensitively on the training data distribution.

We find a very similar performance on the lens center and ellipticity as for the network with the natural distribution (see Sect. 3.4.1). This is expected since the only difference is the distribution in Einstein radii. This can be further visualized with the 1D and 2D probability contours in Figure 3.7 (green) that show that overall this network performs very similarly to the network trained on the naturally distributed sample (blue). For all three networks we find minimal correlation between the different parameters.

In analogy to the previously presented networks, we show in Figure 3.8 the histogram of the brightness ratio and the Einstein radius differences as function of the brightness ratio for this network. While the distribution matches that from the sample with naturally distributed Einstein radius, we overpredict the Einstein radius more than before. This is related to the overprediction at smaller Einstein radii (see Figure 3.10), which comes from weighting higher the fraction of systems with larger image separation. We still underestimate the Einstein radius at the very high end, as already noted, but this is negligible for the overall performance compared to the amount of overestimated systems as we still have a factor of ~ 100 more of them in our sample. This is the reason why the network tends to overpredict more strongly than that trained on the naturally distributed sample (Sect. 3.4.1, and blue lines in Figure 3.7 and Figure 3.8).

Finally, we show the loss curve in Figure 3.11. The training losses (dotted lines) and validation losses (solid lines) in different colors correspond to the five different cross-validation runs. Additionally, we give the mean of the validation curves with a black solid line. This line is used to obtain the best epoch, which in this specific case is epoch 122 (vertical line). The corresponding loss is 0.0528 obtained with Eq. (3.6).

From the loss curve we see that the network does not overfit much to the training set since the validation curves do not increase much for higher epochs, but still enough to define the optimal epoch to terminate the final training. This is a sign that drop-out is not needed, which is supported by additional tests (see Sect. 6.3).

3.5 Further network tests

In addition to the networks described in Sect. 6.4, where we mainly investigated the effect of the Einstein radius distribution, in this section we discuss further tests on the training data set.

3.5.1 Data set containing double or quads only

We considered a specialized network for one of the two strong lensing options and limited our sample to either doubles or quads, where the image multiplicity is based on the centroid of the source (as the spatially extended parts of the source could have different image multiplicities depending on their positions with respect to the lensing caustics). In the case where we limited the sample to doubles only, we did our standard grid search for the different hyperparameter combinations for two samples with naturally distributed Einstein radii above 0.5'' and above
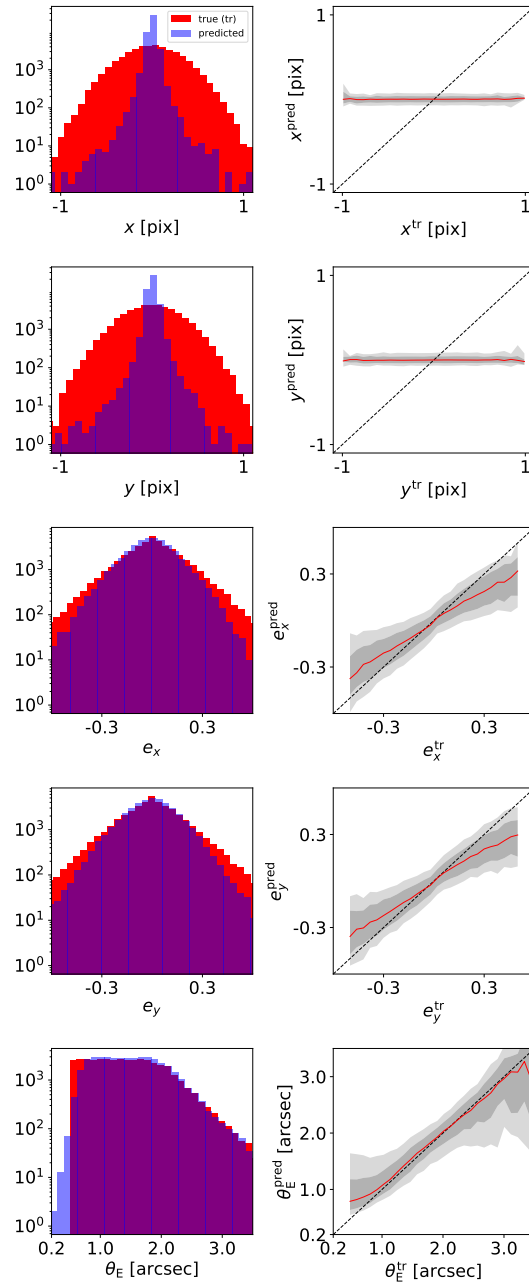
Figure 3.10: Network performance under the assumption of a lowest Einstein radius $\theta_{E,min}$ of 0.5″ but a uniform distribution up to $\sim 2''$. The left panel shows histograms of the ground truth (tr) in red and of the predicted values in blue. The right panel shows a direct comparison of the predicted against the true value. From top to bottom are the five different model parameters, lens center x and y, complex elliptcity e_x and e_y, and Einstein radius θ_E. For all plots 30 bins over the plotting range are used.
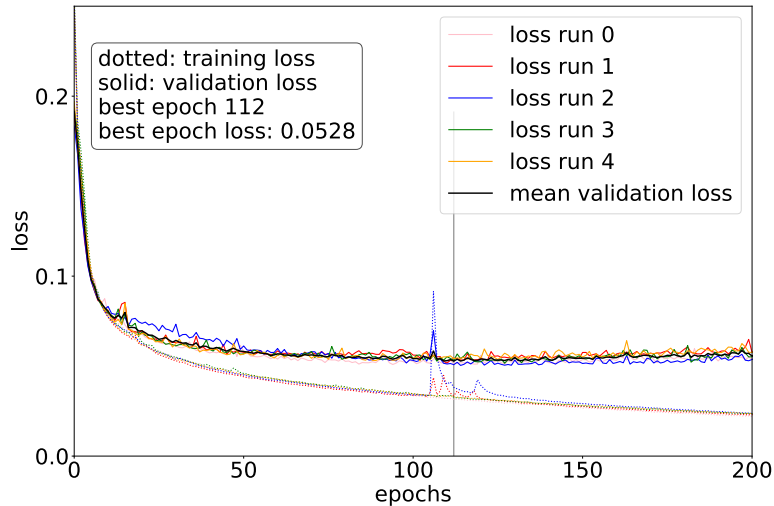
Figure 3.11: Loss curve of our best network under the assumption of equally distributed Einstein radii. The training loss is shown as dotted lines in five different colors for the five different cross-validation runs. In the same colors the validation loss is shown as solid lines together with the black curve, which is the average of the five validation curves from the cross-validation runs. From the minimum in the black curve, shown as the vertical gray line, the best epoch is found.

2.0″. With these networks we found no notable difference compared to the sample containing both doubles and quads (see Sect. 3.4.1 and Sect. 3.4.2), which was expected as the doubles dominate the sample including both doubles and quads by a factor of around 20-30 (for the different networks depending on the lower limit of the Einstein radii).

When we limited the sample to quads only, we performed our grid search again for the different hyperparameter combinations of both samples with naturally distributed Einstein radii above 0.5″ and above 2.0″ and also with equally distributed Einstein radii. Since the chance of obtaining four images is smaller than the chance of observing two images based on the necessary lensing configuration probability, the sample sizes are smaller with 42 063, 19 176, and 28 398 lensing systems. Therefore, the output has to be considered with care as this is much lower than typically used for such a network.

It turns out that these networks perform equally well on the lens center and ellipticity but better for the Einstein radius shown in Figure 3.12. By comparing this plot to Figure 3.10, we find the main improvement that the 1σ and 2σ scatters are substantially reduced and with smaller bias for systems with larger θ_{E}. An improvement on the Einstein radius is expected as the network gets the same information for the lens, but more for the lensed arcs. Even if one image is now too faint to be detected or is too blended with the lens there are three images from the quad left over to provide information on the Einstein radius.

To increase the sample we simulated a new quads-only batch with the source brightness boosted by one magnitude, which resulted in a ~ 1.5 times larger sample than before. This is still small compared to the other double or mixed samples. Now we have a brightness ratio peak at $\log\left(\frac{I_{\mathrm{s}}}{I_{\mathrm{l}}}\right) \sim -1.5$ instead of ~ -2.0 (as shown in Figure 3.8). The performance obtained

with this trained network (the loss is 0.0673 for the network with $w_{\theta_E} = 5$) is still similar to that for the quads-only network without magnitude boost (the loss is 0.0688) and no significant performance difference is observed for the individual parameters.
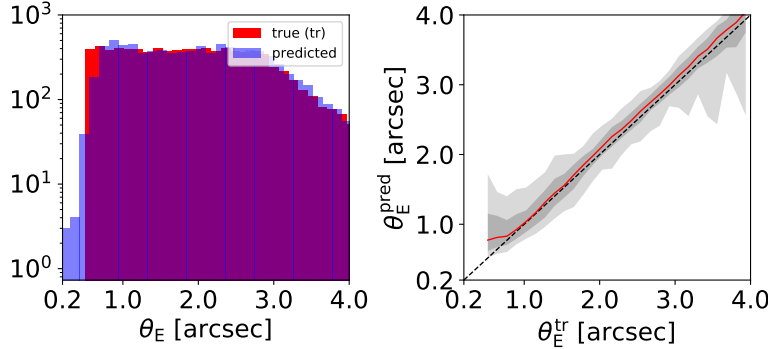

Figure 3.12: Network performance under the assumption of a lowest Einstein radius $\theta_{E,min}$ of $0.5''$ but a uniform distribution with quadruply lensed images. The left panel shows histograms of the ground truth (tr) in red and of the predicted values in blue. The right panel is a 1:1 plot of predicted against true Einstein radius. For both plots 30 bins over the plotting range are used.

3.5.2 Comparison to lens galaxy images only

As further proof of the network performance on the Einstein radius, we test how well the network is able to predict the parameters from images of only the lens galaxies (i.e., without lensed arcs). As expected, the network performs similarly well for the lens center and axis ratio, but much worse for the Einstein radius with a 1σ value of $0.41''$. This shows us that the arcs are bright enough and sufficiently deblended from the lens galaxies to be detectable by the CNN.

3.6 Prediction of lensed image position(s) and time-delay(s)

After obtaining a network for different data sets (see Table 3.1), we compared the true and predicted parameter values directly. Since the main advantage of the network is the computational speed-up compared to recent methods and the fully automated application, the network is very useful for planning follow-up observations. This needs to be done quickly in case there is, for instance, a SN or a short-lived transient occurring in the background source. We explore below how accurately we can predict the positions and time-delays of the next appearing SN images.

We used the predicted SIE parameters from the networks to predict the image positions and time-delays and compared them to those obtained with the ground-truth SIE model parameter values. This gives us a better understanding of how well the network performs and if the obtained accuracy is sufficient for such an application. For this comparison we computed the image positions of the true source center based on the true SIE parameters obtained by the simulation for the sytsems of the test set (hereafter true image positions). After removing the central

highly demagnified lensed image as this would not be observable (given its demagnification and the presence of the lens galaxy in the optical–infrared), we computed the time-delays for these systems (hereafter true time-delays Δt^{tr}) by using the known redshifts and our assumed cosmology. Based on these true image positions and time-delays, we were able to select the first-appearing image and use its true image position to predict the source position with our predicted SIE mass model. This source position was then used to predict the image positions (hereafter predicted image positions) of the next-appearing SN images based on the SIE parameter values predicted with our modeling network. The predicted image positions were then used to predict the time-delays (hereafter predicted time-delays Δt^{pred}) with the network predicted SIE parameter. We directly compared the image positions and time-delays that we obtained with the true and with the network predicted SIE parameters when we had the same number of multiple images. If the number of images did not match, which happened for 7.8% of the sample used for the network with equally balanced Einstein radii distribution containing double and quads, we omitted the candidate from this analysis as a fair comparison was not possible. Since we always remove the central image, we obtain for a double and quad, respectively, two and four images and one and three time-delays. Since the time-delays can be very different, we also compared the fractional difference between the true and predicted time-delays with respect to the true time-delays.

We chose again the three main networks from Sect. 6.4 for this comparison; they are shown in Figure 3.13. All three sets contain quads and doubles, and assume a loss weighting factor of 5 for the Einstein radius. The first set assumes a lower limit on the Einstein radius of 0.5″(blue), the second a lower limit of 2″(yellow), and the third a lower limit of again 0.5″ but with a uniform distribution on the Einstein radii instead of the natural distribution following the lensing probability (green). We plot the quantities as a function of the brightness ratio $\log\left(\frac{I_s}{I_1}\right)$ in analogy to Figure 3.7 and Figure 3.8.

In detail, Figure 3.13 contains in the upper row the median difference in the image position for the x coordinate (left) and y coordinate (right) with the 1σ value per brightness ratio bin, where only the additional image positions are taken into account as the first reference image is known, and thus they do not need to be predicted. We obtain for all three networks a median offset of nearly zero independent of the brightness ratio and whether we limit further in Einstein radii or not. The 1σ values are around 0.25″, corresponding to ~ 1.5 pixels. Explicitly, we find for the equally distributed sample applied to $\theta_{\text{E}} > 0.8″$ a median image position offset of $(0.00^{+0.29}_{-0.29})″$ and $(0.00^{+0.32}_{-0.31})″$ for the x and y coordinate, respectively. Interestingly, the 1σ values are slightly larger for quads than doubles as we would have expected that quads provide more information to constrain the SIE parameter values, and thus predict the image positions better. The reason for this is probably because quads generally have higher image magnification than doubles, and image offsets are larger with higher magnification.

The middle row of Figure 3.13 shows the legend (left) and a histogram of the difference between the predicted time-delay Δt^{pred} and the true time-delay Δt^{true}. The bottom row shows the difference in time-delay divided by the absolute value of the true time-delay per brightness ratio bin (left) and the difference of the time-delays again per brightness ratio bin (right). In terms of time-delay difference, the network trained on the natural distribution (blue) performs better than that with uniform distribution (green), but especially for the network trained for lens systems with large Einstein radius (orange) we obtain notable differences. In detail, we
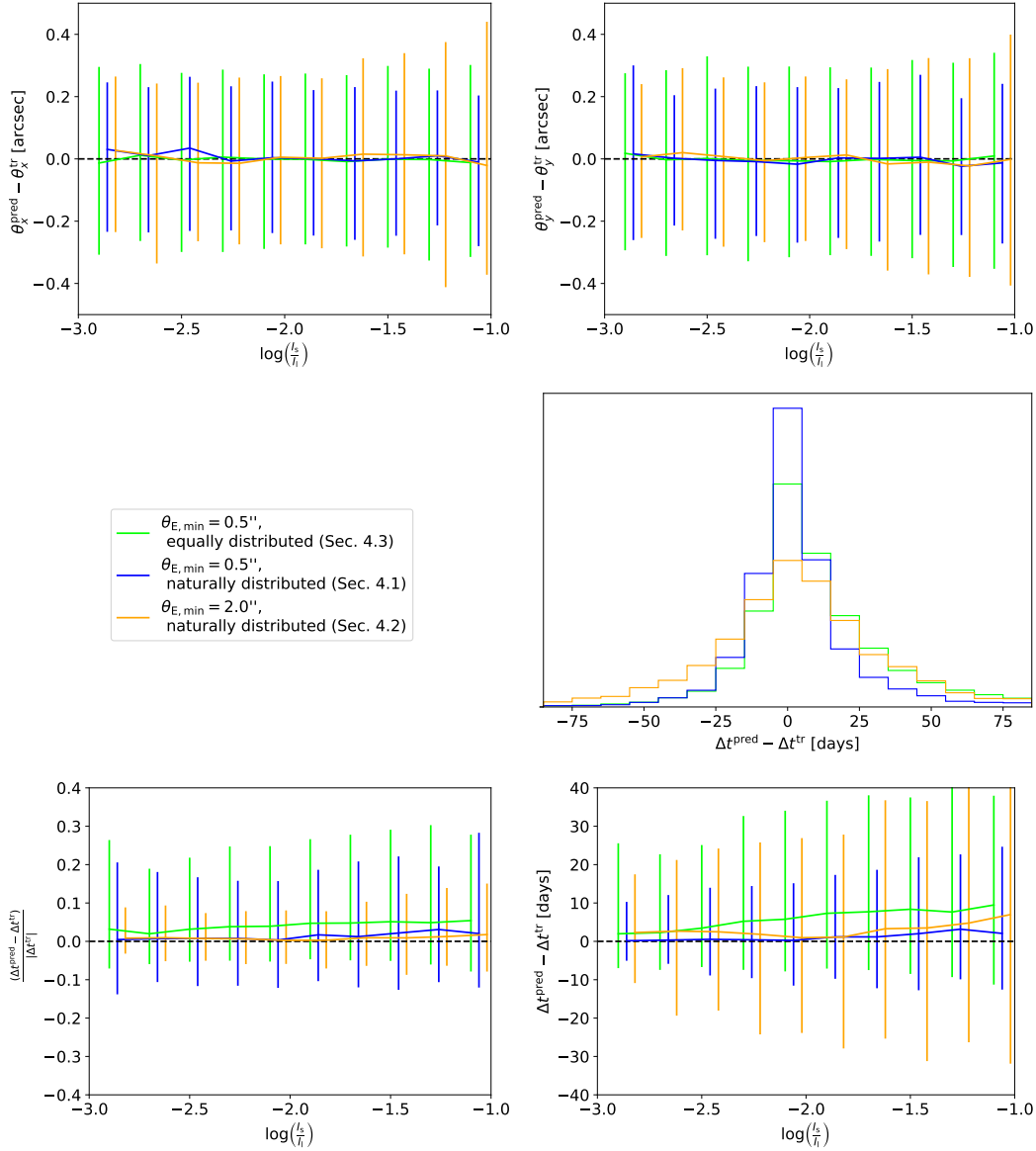
Figure 3.13: Precision of model network predictions as a function of the lens and lensed source brightness ratio in the range $-3 \leq \log\left(\frac{I_s}{I_l}\right) \leq -1$ for the three networks presented in Sect. 6.4 applied to the restricted sample with $\theta_E^{tr} > 0.8''$. The upper row shows the image position offset for the x coordinate (left) and y coordinate (right). In the middle panel is the legend (left) and a histogram of the difference in time-delay (right), while in the bottom row is shown the fraction of the time-delay difference and the true time-delay (left) and the time-delay difference (right). The curves show the median and the vertical bars the 1σ values. The blue and orange bars have been shifted slightly to the right for better visualization.

obtain a median with 1σ value for the naturally distributed sample (blue; see Sect. 3.4.1) for the time-delay difference of 2^{+18}_{-6} days and a fractional time-delay difference of $0.05^{+0.47}_{-0.09}$. Since we find a strong correlation between the offset in the Einstein radius and the time-delay offset (see Figure 3.14), we exclude again the very small Einstein radii systems ($\theta_E^{tr} < 0.8''$) and obtain for the time-delay difference 1^{+18}_{-11} days and for the fractional difference $0.01^{+0.19}_{-0.12}$. For the equally distributed sample (green; see Sect. 3.4.3) we obtain, with $\theta_E > 0.5''$ and $\theta_E > 0.8''$, respectively, for the time-delay difference 7^{+38}_{-6} and 6^{+36}_{-8} days and for the fractional time-delay difference $0.06^{+0.45}_{-0.05}$ and $0.04^{+0.27}_{-0.05}$. This restriction is easily applicable in practice since individual lensing systems are only followed up at a given time, and it is possible to check by looking at the image of the individual system whether the Einstein radius is $>0.8''$. Depending on the predicted time-delay, the model can be further improved by using traditional manual maximum likelihood modeling methods to verify the predicted time-delay.
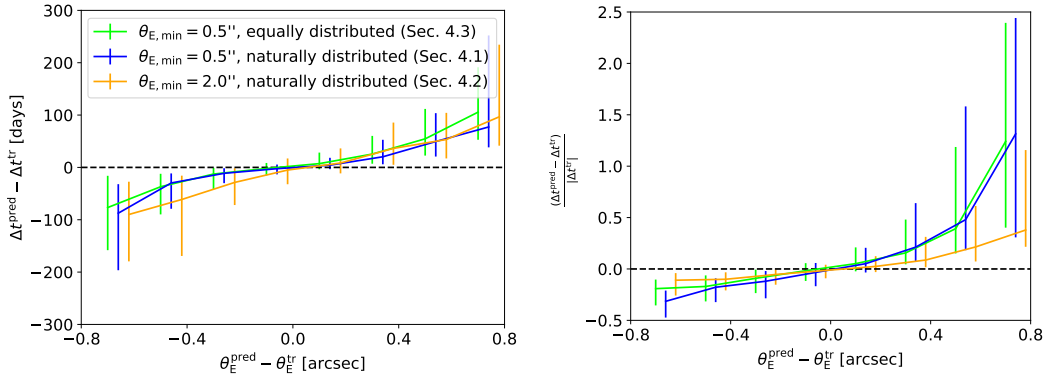
Figure 3.14: Correlation between Einstein radius offset in the range $-0.8'' \leq \theta_E^{pred} - \theta_E^{tr} \leq 0.8''$ and time-delay difference (left panel) or fractional time-delay difference (right panel) by applying the different networks to their samples after limiting to $\theta_E^{tr} > 0.8''$. The blue and orange bars have been shifted slightly to the right for better visualization.

The fractional offset in the predicted time-delays of $0.04^{+0.27}_{-0.05}$ that we achieve with our CNN for systems with $\theta_E > 0.8''$ for the uniformly distributed θ_E sample (i.e., with a symmetrized scatter of ~16%) is close to the limit that would be achievable even with detailed and time-consuming MCMC models of ground-based images. This is because the assumption of the SIE introduces additional uncertainties on the predicted time-delays in practice, even though detailed MCMC models of images would typically yield more precise and accurate estimates for the SIE parameters than our CNN. While galaxy mass profiles are close to being isothermal, the intrinsic scatter in the logarithmic radial profile slope γ' (where the 3D mass density $\rho(r_{3D}) \propto r_{3D}^{-\gamma'}$) is around ± 0.15, translating to ~ 15% scatter in the time-delays (e.g., Koopmans et al., 2006; Auger et al., 2010; Barnabè et al., 2011). In other words, if a lens galaxy has a power-law mass slope of $\gamma' = 2.1$, then our assumed SIE mass profile (with $\gamma' = 2.0$) for it would predict time-delays that are ~10% too high (e.g., Wucknitz, 2002; Suyu, 2012). While constraining the profile slope γ' with better precision than the intrinsic scatter for individual lenses is possible, this would require high-resolution imaging from space or ground-based adaptive optics (e.g., Dye & Warren, 2005; Chen et al., 2016). Given the difficulties of measuring the

power-law mass slope γ' from seeing-limited ground-based images of lens systems (although see Meng et al., 2015, for the optimistic scenario when various inputs are known perfectly such as the point spread function), we conclude that our network prediction for the delays has uncertainties comparable to that due to the unknown γ'. We expect these two sources of uncertainties to be the dominant ones in ground-based images.

We also find a decrease in the performance with increase in brightness ratio, which is in the first instance counterintuitive. If we consider the fractional offset in the left panel, we see a better performance for the sample with an Einstein radius lower limit of $\theta_{E,min} = 2''$ (orange), especially in terms of the 1σ scatter, when compared to the other two networks. This $\theta_{E,min} = 2''$ network also has minimal bias, as shown by the median line. This is understandable as the time-delays are longer for systems with a bigger Einstein radius, and therefore the fractional uncertainty is smaller. The accuracy in time-delay difference (lower right plot) is good, although the 1σ scatter is quite large, ~20 days. With this reasoning, we can also understand the worse performance of the equally distributed sample (green) compared to the naturally distributed sample (blue) as it contains a much higher fraction of systems with bigger image separation. As a higher brightness ratio ($\log(I_s/I_l)$) tends to be associated with systems with higher θ_E, the prediction of delays thus has larger scatter, as shown in the bottom right panel. Moreover, we note that we find a better performance for doubles than quads, probably because of smaller image separation and shorter time-delays of quads.

During this evaluation of the networks we should keep in mind that the main advantage of these networks is the run time: we need only a few seconds to estimate the SIE model parameters, the image positions, and the corresponding time-delays. Therefore, it is expected that we do not reach the accuracy of current modeling techniques using MCMC sampling which can take weeks. Nonetheless, the network results can serve as input to conventional modeling and can help speed up the overall modeling.

3.7 Comparison to other modeling codes

There are already several modeling codes developed, and they can be separated into two main groups. The state-of-the-art codes that rely on MCMC sampling have been widely tested and used for most of the modeling so far. The advantage of such codes are their flexibility in image cutout size or pixel size and also in terms of profiles to describe the lens light or mass distribution. With the advantage of the variety of profiles comes the disadvantage that the codes require a lot of user input which limit the applicability of such codes to a very small sample, or specific lensing systems that are modeled. Moreover, based on the MCMC sampling of the parameter space is very computationally intensive, and thus can take up to weeks per lens system, although some steps can be parallelized and run on multiple cores.

Since the number of known lens systems has grown in the past few years and will increase substantially with upcoming surveys like LSST and Euclid, the codes used to analyze individual lens systems will no longer be sufficient. Thus, the modeling process must be more automated and a speed-up will be necessary. While some newer codes (e.g., Nightingale et al., 2018; Shajib et al., 2019; Ertl et al., in prep.) are automating the modeling steps to minimize the user input, they still rely on sampling the parameter space such that the run time remains on the order of

days or weeks, and some user input per lens system. With these codes it should be possible to obtain results that are comparable with the current interactive modeling procedure.

The second new kind of modeling is based on machine learning such as that used in this work. The first network for modeling strong lens systems was presented by Hezaveh et al. (2017). While they use Hubble Space Telescope data quality, we use ground-based HSC images with similar quality to those used by Pearson et al. (2019) as most of the newly detected lens systems will be in first instance observed with ground-based facilities. Moreover, Hezaveh et al. (2017) suggest first removing the lens light and then only modeling the arcs with the network. Given the differences in the data quality, number of filters, and modeling procedure between Hezaveh et al. (2017) and our work, we cannot further compare the performance fairly.

Pearson et al. (2019) consider modeling with and without lens light subtraction, but found no notable difference; thus, we only consider modeling the lens system without an additional step to remove the lens light. Since we provide the image in four different filters, the network is able to distinguish internally between the lens galaxy and the surrounding arcs. In contrast to Pearson et al. (2019), we use the SIE profile with all five different parameters, while they only predict three parameters: axis ratio, position angle (equivalent to our complex ellipticity), and the Einstein radius. Moreover, they completely mock up their training data, assume a very conservative threshold of S/N>20 in at least one band, and do not include neighboring galaxies that can confuse the CNN; instead, we are more realistic by using real observed images as input for the simulation pipeline. We further assume an offset between lens mass distribution and lens light distribution, which complicates the CNN parameter inference further as the network has to predict the mass distribution only from the images. This way we have more realistic lens light and mass distributions and also include neighboring objects which the network has to learn to distinguish from the lensing system. Pearson et al. (2019) make use of a CNN (the same type of network that we use, and also Hezaveh et al. 2017), but they use slightly smaller input cutouts (57×57 pixels) and a different network architecture (six convolutional layers and two FC layers) than ours. They investigated mostly the effect of using a different number of filters and whether to use lens light subtraction, whereas we investigate the effect on the underlying samples and a simulation with real observed images, which means that we do not have a scenario that assumes the exact same properties. The closest scenario, from Pearson et al. (2019) the results from LSST-like *gri* images including lens light and our results based on HSC *griz* images with naturally distribution of the Einstein radii, shows that both networks are very similar in their overall performance. The reason that the performance on the Einstein radius by Pearson et al. (2019) is better and that they do not suffer from the same biases in θ_E^{pred}, even with a non-flat θ_E^{tr} distribution in their simulations, is perhaps because they use idealized simplistic simulations (high S/N, well-resolved systems, no neighbors).

There are also other recent publications related to strong lens modeling with machine learning. Bom et al. (2019) suggest a new network that predicts four parameters: the Einstein radius θ_E, the lens redshift z_d, the source redshift z_s, and the related quantity of the lens velocity dispersion v_{disp}. They adapt, as we do, a SIE profile to mock up their training data with an image quality similar to that from the Dark Energy Survey. Comparing their performance on the Einstein radius (Figure 8, left panel, assuming θ_E^{tr} on the *x*-axis and θ_E^{pred} on the *y*-axis) to our performance with the natural distribution (Figure 3.6), we find a similar trend. Both networks slightly overpredict at the very low end and underpredict at the high end. If we compare this to the network with

equal distribution, we see a clear improvement of our network on the median line for $\theta_E \sim 2-3''$. Since this code only provides the Einstein radius instead of a full SIE model, the applicability is somewhat limited.

Madireddy et al. (2019) suggest a modular network to combine lens detection and lens modeling which to date have been done with complete independent networks. In detail, they have four steps. The first is to reduce the background noise (so-called image denoising), followed by a lens light subtracting step (the deblending step), before the next network decides whether this is a lens system or not. If it detects the input image to be a lens, the module is called to predict the mass model parameter values. Each module of the network is a very deep network and both modules for detection and modeling make use of the residual neural network (ResNet) approach. They use a sample of 120,000 images, with 60,000 lenses and 60,000 non-lenses, and split this into 90% and 10% for the training and test set, respectively, without making use of the cross-validation procedure. Madireddy et al. (2019) use, as do Pearson et al. (2019), completely mocked-up images based on a SIE profile with fixed centroid to the image center such that the modeling module predicts three quantities, Einstein radius, and the two components of the complex ellipticity. Based on the different assumptions a direct comparison of the lens modeling performance is not possible. However, we see that the performance is typical for the current state of CNNs based on Pearson et al. (2019).

Comparing the network-based modeling to the traditional model using MCMC on a concrete sample is difficult as first we have to obtain the mass models for that sample with both methods. However, in general it is expected that the MCMC models are typically more precise than those obtained with neural networks because of the interactive and individual modeling procedure. In the MCMC modeling, the image residuals can be inspected to see whether the model is good and trustworthy, or if the parameters need to be refined further and different mass profile adopted (e.g., SIE plus external shear). In contrast, the fully automated procedure with CNN does not inspect the individual images and residuals in detail. However, for upcoming surveys like LSST it is impossible to model all the expected $\sim 100,000$ lens systems in the traditional MCMC way systematically given the computational time required. The only way to analyze the entire LSST lens sample will be a fully automated, fast procedure where a small fraction of outliers and (probably) slightly higher uncertainties are acceptable. Therefore, the substantial speed-up is a very important advantage of CNN modeling, as we can process one lens systems with our CNN in fractions of a second compared to weeks or months with MCMC methods. If one is interested in a specific lens system such as a lensed SN, one can consider starting with a CNN to get a good initial mass model and then refining with traditional methods to achieve a good balance between speed and accuracy.

3.8 Summary and conclusion

In this paper, we presented a convolutional neural network to model in a fully automated way and very quickly the mass distribution of galaxy-scale strong lens systems by assuming a SIE profile. The network is trained on images of lens systems generated with our newly developed code that takes real observed galaxy images as input for the source galaxy (in our case from the Hubble Ultra Deep Field), lenses the source onto the lens plane, and adds it to another real

observed galaxy image for the lens galaxy (in our case from the HSC SSP survey). We chose the HSC images as lenses and adopted their pixel sizes of $0.168''$ as this is similar to the data quality expected from LSST. With this procedure we simulated different samples to train our networks where we distinguish between the lens types (quads+doubles, doubles-only, and quads-only) and on the lower limit of the Einstein radius range. Since we find a strong dependence on the Einstein radius distribution, we also consider a uniformly distributed sample and also a weighting factor of 5 for the Einstein radius' contribution to the loss. With this we obtain eight different samples for each of the two different weighting assumptions summarized in Table 3.1.

For each sample we then perform a grid search to test different hyperparameter combinations to obtain the best network for each sample, although we find that the CNN performance depends much more critically on the assumptions of the mock training data (e.g., quads, doubles, both, or Einstein radius distribution) rather than on the fine-tuning of hyperparameters. From the different networks presented in Table 3.1, we find a good improvement for the networks trained with quads-only compared to the networks trained on both quads and doubles. If the system type is known, we therefore recommend using the corresponding network. Since the Einstein radius is a key parameter, we weighted its loss higher than for the others and, although the minimal validation loss is higher, we advocate these networks for modeling HSC-like lenses. With the network trained on both quads and doubles with the uniform distributed sample of $\theta_E > 0.5''$, we obtain for the five SIE parameters a median with 1σ value as follows: $\Delta x = (0.00^{+0.30}_{-0.30})''$, $\Delta y = (0.00^{+0.30}_{-0.29})''$, $\Delta\theta_E = (0.07^{+0.29}_{-0.12})''$, $\Delta e_x = -0.01^{+0.08}_{-0.09}$, and $\Delta e_y = 0.00^{+0.08}_{-0.09}$.

After comparing the network performance on the SIE parameter level, we tested the network performance on the lensed image and time-delay level. For this we used the first appearing image of the true mass model to predict the source position based on the predicted SIE parameter. From this source position and the network predicted SIE parameters, we then predicted the other image position(s) and time-delay(s). We find for the sample with doubles and quads a uniform distribution in Einstein radii and a weighting factor w_{θ_E} of five by applying the network to $\theta_E > 0.8''$ an average image offset of $\Delta\theta_x = (0.00^{+0.29}_{-0.29})''$ and $\Delta\theta_y = (0.00^{+0.32}_{-0.31})''$, while we achieve the fractional time-delay difference of $0.04^{+0.27}_{-0.05}$.

This is very good given that we use a simple SIE profile and need only a few seconds per lens system in comparison to state-of-the-art methods that require at least days and some user input per lens system. We anticipate that fast CNN modeling such as the one developed here will be crucial for coping with the vast amount of data from upcoming imaging surveys. For future work, we suggest investigating further into creating even more realistic training data (e.g., allowing for an external shear component in the lens mass model) and also exploring the effect of deeper or more complex network architectures. The outputs of even the network presented here can be used to prune down the sample for specific scientific studies, which can then be followed up with more detailed conventional mass modeling techniques.

4 Lens modeling with a residual neural network assuming SIE with external shear with error estimation

Note. The content of this chapter will be submitted as an individual paper to an international journal for publication. We refer to this upcoming publication with Schuldt et al. (in prep a).

4.1 Introduction

We have demonstrated in Chapter 3 the ability of modeling strongly lensed galaxy systems from HSC with a CNN. We performed extensive tests regarding the effect of different sample distributions on the Einstein radius for the ground truth, presented networks trained specifically for systems with large image separation, or only quads or doubles. Moreover, we demonstrated the accuracy and precision of image positions and time delay inferred from our network output. In this chapter we present our further investigations built upon this. Here we adopt now an external shear, describing the influence on the light deflections from mass outside the main lens potential, in addition to the SIE profile. We will refer to this combination of profiles as SIE+γ_{ext} in the remaining thesis. Adding the external shear component required an upgrade of our simulation pipeline which is introduced in Sect. 3.2.3 and used to create again our training data set. The upgrade and additionally improvements, such as including Poisson noise on the arcs, improving the lens centering and thus adding an random ±3 pixel shift, which is a common data augmentation technique to enhance the performance of CNNs Dieleman et al. (2015), are described in Sect. 4.2.

Since we include the external shear, our network predicts now seven parameters instead of five. Additionally, we include an uncertainty prediction by doubling the number of neurons in the last layer. Since this new task is much more complex than considered in Chapter 3, it requires a deeper network coping with the small distortions of the external shear, such that we rely now on ResNet (He et al., 2015). This procedure, and also the final network architecture is described in Sect. 4.3.

In Sect. 4.4 we describe then the performance of our final network and discuss the results. The extensive tests we carried our in order to find this network architecture and the best hyperparameter combination are summarized in Sect. 4.5. Moreover, we highlight several additional tests on the network to demonstrate further possibilities. We summarize and conclude our results finally in Sect. 4.6.

As already in Chapter 3, we assume a flat ΛCDM cosmology with a Hubble constant $H_0 = 72\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1}$ (Bonvin et al., 2017) and $\Omega_{\text{M}} = 1 - \Omega_{\Lambda} = 0.32$ (Planck Collaboration et al., 2020).

4.2 Simulation of strongly lensed images

To create our training, validation, and testing set for this modeling network, we follow and improve on the pipeline described in Sect. 3.2. In analogy to Chapter 3, we use as lenses HSC images (Aihara et al., 2019) of LRGs with a pixel size of 0.168″ and available spectroscopic redshifts and velocity dispersion measurements from SDSS (Abolfathi et al., 2018). In addition to our criteria in Chapter 3, we set now a lower limit of 100 km/s for the velocity dispersion, since the very low mass galaxies do not give a strong lensing configuration as we see from the right panel of Fig. 3.1. This will speed up the creation of mocks as we do not iterate over these systems anymore trying to find a suitable source. As background sources, we use again images from Hubble Ultra Deep Field (HUDF) together with their redshift measurements (Beckwith et al., 2006; Inami et al., 2017). These galaxies are extracted through the same procedure as in Sect. 3.2 using Source Extractor (Bertin & Arnouts, 1996). This background galaxy is then lensed with the software GLEE (Suyu & Halkola, 2010; Suyu, 2012) from the source plane to the lens plane and the obtained arcs are painted on top of the lens image. With this procedure we include real line-of-side objects and light distributions, making the mock images very realistic. To calculate the deflection angles, we assume a SIE profile and, in contrast to Chapter 3, we include an external shear to account for an additional mass concentration outside of the cutout. In the following we describe the changes compared to our procedure adopted in Chapter 3, and therefore show a diagram of the upgraded simulation procedure in Fig. 5.2, which is adapted from Fig. 3.3. Example images generated with our upgraded simulation pipeline are displayed in Fig. 4.2; on the left panel the final mock images and on the right panel the lensed source alone, which got added to the HSC lens image. It shows the variety of mocks and how realistic they are.

Compared to the version used in Chapter 3, we improve the lens centering and ellipticity estimation in the simulation code in the following way. The code accepts now a mask of the lens obtained for instance with the Source Extractor (Bertin & Arnouts, 1996), resulting in a more accurate extraction of the lens. To determine the lens center, we additionally apply a circular mask with radius of 20 pixels centered at the image center. This is in particular important to avoid stronger biases for the lens center through unsymmetrical extraction due to neighboring objects. From the resulting masked image, we predict the lens *light* center trough the first moments and recenter the image to the nearest pixel of the predicted lens *light* center. An interpolation for better centering is not performed to avoid manipulating the observed image. Instead, the remaining fractional pixel is taken into account through the simulation as lens *light* center, from which we determine the lens mass center. This lens mass center is drawn from a Gaussian distribution with width of ± 0.5 pixel corresponding to ±0.84″, and used later for the network training. Additionally, since we now re-centering the cutout on the lens, we shift the final mock image randomly by up to ±3 pixels in *x* and *y* direction. This ensures that the network learns to predict the lens center instead of the image center.

The axis ratio q_{light} and position angle θ_{light} of the lens *light* is determined through the second moments using the provided mask. Here is no additional circular aperture taken into account as the outer parts of the lens *light* distribution are crucial for the correct estimation of the ellipticity and axis ratio. In analogy to our mocks in Chapter 3, the axis ratio q and the position angle θ

Input

Lens (HSC)   Lens mask (i band)   Extracted source (HUDF)   Lens PSF (HSC)   Lens variance map (HSC)

calculation of lens center, q and PA in i band

Lens mass values, source position, and randomly assigned external shear

Lensed source

Convolve with subsampled PSF, scale to lens pixel size, adjust intensity values given different photometric zero points of imaging instruments, add poisson noise

Re-centering

Centered lens

Combine images, add up to +/-3 pixels shift to offset system, crop to final image size

Convolved lensed source on HSC pixel scale

Lens system with all components
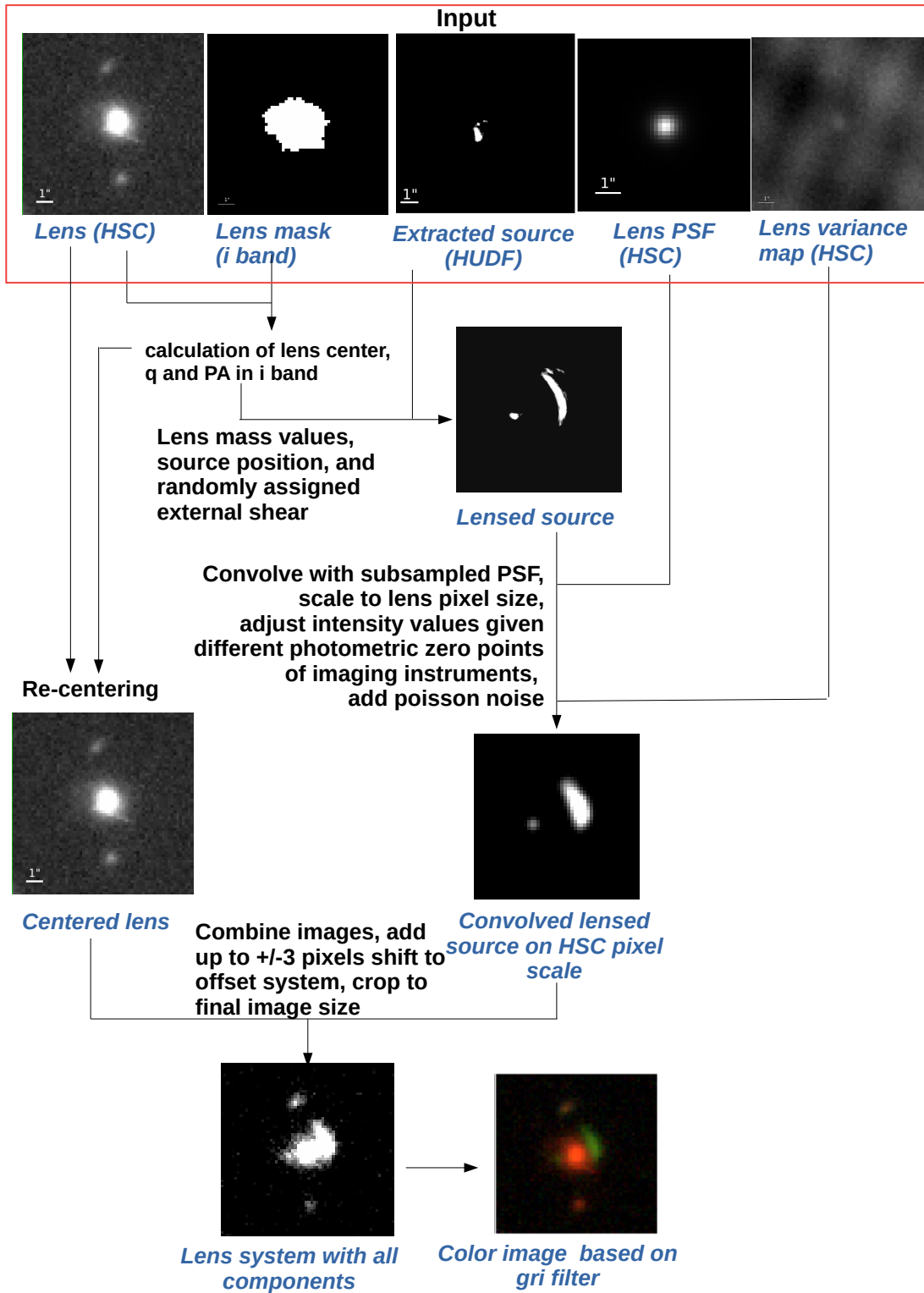
Color image based on gri filter

Figure 4.1: Flow chart of the simulation pipeline used to create our training data. Figure adapted from Fig. 3.3.
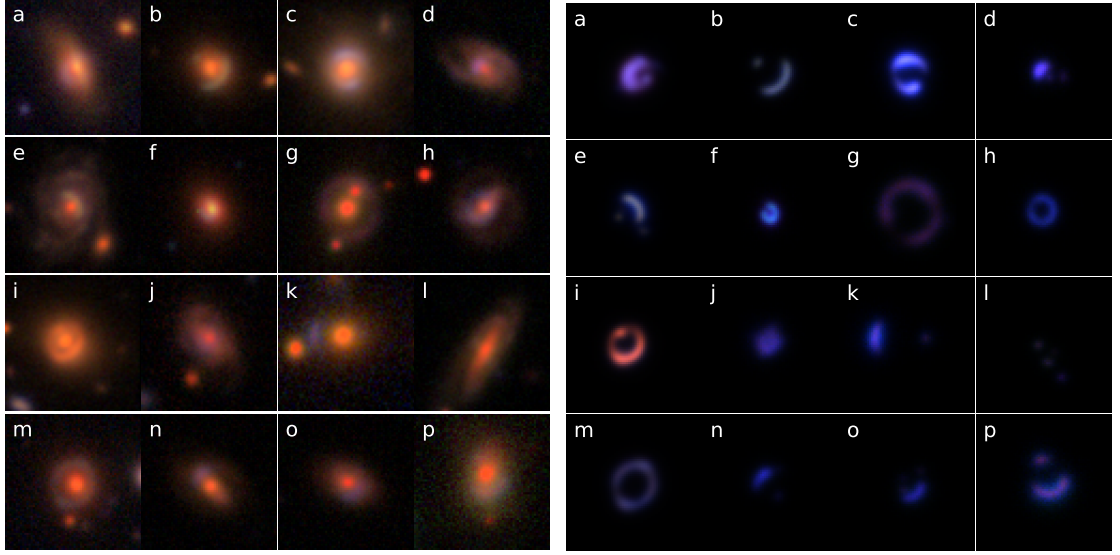
Figure 4.2: Example images of our mock lensed (left) and corresponding frames with only the lensed source (right).

of the mass distribution are drawn from Gaussian distributions centered at the lens *light* with a Gaussian width of 0.05 and 0.17 radians (equal to 10 degrees), respectively.

The Einstein radius θ_E is again obtained through Eq. (3.3) from the velocity dispersion and redshifts. For the external shear we implemented the option of either (1) randomly assigned values matching a realistic distribution obtained from galaxy-scale lenses with external shear previously estimated from detailed modeling using state-of-the-art techniques (Faure et al., 2011; Wong et al., 2011), or (2) using a flat distribution for the complex quantities γ_1 and γ_2, or (3) a flat distribution for the shear strength γ_{ext}. For the final set of mocks, we assume a flat prior in the shear strength within a range between 0 and 0.1. As described in Sect. 4.5 further, this helps the network to learn better the full parameter range of the shear strength.

Moreover, we improve our simulation code by including now Poisson noise (Hasinoff, 2012) for the arcs approximated from the variance map v_i corresponding to the individual lens image pixel i. For this we first define

$$\sigma_i = \sqrt{v_i^+} \tag{4.1}$$

with $v_i^+ = \max\{v_i, 0\}$, i.e. we reset all negative pixel values of the variance map by zero, and compute the lens background noise σ_{bkgr} defined as the minimal RMS of the four corner of the lens image. With those two quantities we then approximate the Poisson noise map of the lens as

$$\sigma_{poisson,i} = \sqrt{\sigma_i^2 - \sigma_{bkgr}^2} \tag{4.2}$$

to obtain the poisson scaling factor map

$$\alpha_{poisson,i} = \sigma_{poisson,i}^2 / I_{lens,i} \tag{4.3}$$

where I_{lens} is the lens image from HSC. Since this scaling factor α_{poisson} should be a constant, we approximate it as the median of the 10×10 pixels central region as there the lens intensity is the highest and thus the mapping is more precise than on the outer parts. The Poisson map of the arc is then simply

$$\sigma_{\text{poisson,arc},i} = \sqrt{\alpha_{\text{poisson}} \times I^+_{\text{arc},i}} \qquad (4.4)$$

where I_{arc} is the lensed source image (arcs). From this map we draw Gaussian variations representing the additional Poisson noise which are added on top of the simulated images.

In addition to these technical changes, we request, similar as described in Sect. 3.2, the lens and source pair to pass the following criteria on the brightness in at least either the *g* or *i* filter, wherever the source is brighter. This allows to accept both bluer and redder lenses, while previously a large fraction of red sources were discarded. In detail, we require the brightness peak of the lensed source (arcs), shown on the right panel in Fig. 4.2, to be at least a factor 1.5 brighter than the lens at this specific pixel, such that the main arc is detectable at this position. However, we include here no check on the extend of the arc nor the counter image(s). This implies that systems with larger Einstein radius can have fainter arcs, like e.g., image (g) and (m) in Fig. 4.2, since the lens light will be low at the lensed source brightness peak compared to systems with small image separation. Therefore, we set additionally a brightness threshold of 5 σ_{bkgr}. With this criteria we obtain mock systems with arcs above the lens background noise and thus detectable lensing features. If the arc would be too faint compared to the main lens, the network would have no chance to extract information from the arcs. This is valid, as we intend to model lenses primarily found from multiband imaging containing typically a visual inspection stage where we request to have visible indications of arcs in the color image.

Since it is hard to overcome these thresholds with sources randomly places behind the lens, we always place the source on positions with a magnification factor μ of at least 10. Additionally, if a lens source pair does not fulfill these brightness requirements, we lower iteratively the source magnitude in all bands in steps of -0.5mag by a maximum of 6 steps which is taken into account when translating the lensed source from the high resolution image to the ground-based resolution. This results still in realistic images as we can see from Fig. 4.2 with also relatively faint arcs (e.g., system (b), (g), (l), (m)).

The 16 images displayed in Fig. 4.2 are a randomly selected subset of our produced mocks. It shows the variety of included LOS objects that complicated the task for the network notably. It has to learn to distinguish between lens, arcs, and these LOS objects. Although we use only galaxies as lenses that got classified as LRGs given their higher lensing cross section, also some, probably mis-classified, edge on spirals (compare image (l) in Fig. 4.2) or face-on Early Type Galaxys (ETGs) (compare image (d) and (j) for instance) are included. Such edge on spirals or face-on ETGs are still possible lensing systems, see lens candidate HSCJ0928-0045 which we found with our residual neural network in the HSC images (Cañameras et al., 2021), but much rarer. This underrepresented systems are expected to raise difficulties for the network leading to higher uncertainties but are intended as we aim to model larger samples from current and upcoming wide-field imaging surveys that likely have these types of lens-spiral contaminants.

Image (i) in Fig. 4.2 shows untypically red arcs, matching nearly the color of the lens. This is most likely the result of Source Extractor identifying another object in the cutout close to the image center, as in the image center only a nearly indetectable object is, which although could

indeed correspond to the HUDF redshift of z_s = 5.92. The offset to the image center is not visible in the final image as each source gets automatically recentered on the brightest pixel. The lens of this system lays at a redshift z_d = 0.443 and has a velocity dispersion of 225.9km/s, resulting in an Einstein radius of 1.163″. This system shows the limitation of our simulation pipeline and introduced inaccuracies which come along with the procedure using real images. It demonstrated how careful the masking has to be done. Since these systems are very seldom in our data set, we expect their introduced difficulty for the network to be minor especially as they appear in all three splits, the training set, validation set, and test set.

Following the procedure described in this section, we create more than 90, 000 mock images. These are then used to train (50,176 mocks), validate (12,544 mocks), and test (27,392 mocks) the network, as described in Sect. 4.3.

4.3 The residual neural network architecture

As introduced in Sect. 1.5, CNNs are very powerful tools in image processing especially if an autonomous and fast method is required to cope with the huge amount of images. Since there are already thousands of known lens candidates in the HSC footprint (e.g., Jaelani et al., 2020a,b; Jaelani et al., in prep.; Sonnenfeld et al., 2019, 2020; Wong et al., 2018; Cañameras et al., 2021; Shu et al., in prep., , see also Sect. 1.2 for further details), and we expect around hundred thousand more observed in similar quality by LSST (Collett, 2015), CNNs are perfectly suited for analyzing this data.

While we used in Chapter 3 a CNN architecture based on the LeNet (Lecun et al., 1998) architecture to predict the five SIE parameters, we consider now deeper networks to predict SIE+γ_{ext} and thus make use of residual neural networks to gain notable improvement. A sketch of our network architecture is given in Fig. 4.3.

In analogy to Chapter 3, we start with a network predicting one point estimate per parameter, i.e. now seven values for the five SIE parameters plus two values for the external shear. Here we again use a Mean Square Error (MSE) loss function. We then modified very soon the network such that it predicts two values per parameter, i.e. 14 values in total. We interpret now seven values as the mean (like previously the point estimate) and the other values as the σ spread of a Gaussian function describing the error on that parameter for that specific image. Here we use a loss L given by

$$L = \sum_{k=1}^{N} \sum_{l=0}^{p} -w_l \times \log(P(\eta_{k,l}^{\text{pred}}, \eta_{k,l}^{\text{tr}}, \sigma_{k,l})) + \epsilon_l \times \log(\sigma_{k,l}^2) \qquad (4.5)$$

with a log-probability term and a regularization term to minimize the errors similar as suggested in Perreault Levasseur et al. (2017). The loss is a sum over the p different parameters η and corresponding errors σ with index l and additionally summed over each image k in the batch of size N. We introduced a weighting factor w and a regularization constant ϵ_l which can be different for each parameter l. If we can indeed interpret the second values as an 1σ error on the values, this makes the network much more powerful as it indicates how much we can trust that predictions. Perreault Levasseur et al. (2017) suggest to tune the network through the
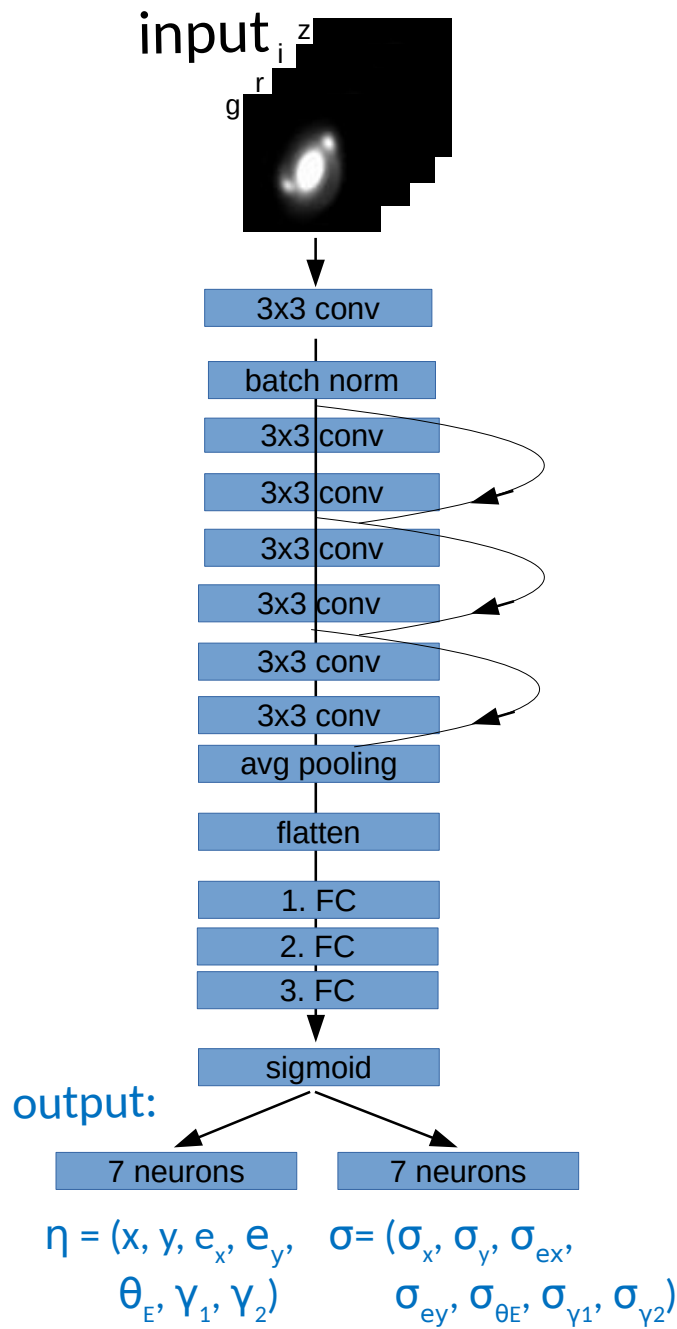
Figure 4.3: Overview of our ResNet architecture. The input are the lens images in four filters *g, r, i,* and *z*, witch are passed through a series of convolutional layers and FC layers. After a sigmoid layer to map all output values to the range [0, 1], we split them up into mean and uncertainty.

dropout rate (Srivastava et al., 2014) such that the predicted errors match the expected intervals of 68% for 1σ, 95% for 2σ and 99.7% for 3σ. Instead of using dropout for both convolutional layers and FC layers with the same rate as done in Perreault Levasseur et al. (2017), we use only dropout for the FC layers as the dropout rate typically differs significantly between convolutional layers and FC layers. Because of the low number of final FC layers (one to three layers), the effect of dropout on the percentile is minor. This makes the tuning difficult such that even if we average over the seven parameters, the right percentile were not perfectly achievable for us. It is also questionable if averaging over all seven parameters makes the predicted uncertainties interpretable as σ values as the fraction can differ significantly for one specific parameter to the expected percentile. Therefore, we investigated into network architectures with different branches of FC layers as further described in Sect. 4.5.1.

In general, we tested different setups of the ResNet architecture by varying e.g., the number of residual blocks, kernel sizes, and number of FC layers. The tests we carried out to find the best network architecture and hyper-parameter set are summarized in Sect. 4.5, while we focus here on the final network architecture and the general training procedure. As shown in Fig. 4.3, the final network architecture contains one conv layer with kernel size 3 followed by a batch normalization and three residuals blocks of each two conv layers with kernel size of 3. The standard 3×3 kernel where preferred over larger kernel sizes such as 5×5 or 7×7 as the network need to capture also very small features on the arcs to predict the external shear and our images are only 64×64 pixels large. The output of the convolutional sequence is passed after flattening through three FC layers mapping it to 14 values.

Since the parameter ranges are different for the different parameters, we include a scaling of the ranges to map them consistently to the range $[0, 1]$. This ensures an equal contribution from the different parameters to the loss, resulting in better optimization of all parameters. We assume here the following input ranges $[a, b]$: lens center $x \in [-0.6'', 0.6'']$ and $y \in [-0.6'', 0.6'']$, complex ellipticity $e_x \in [-1, 1]$ and $e_y \in [-1, 1]$, the Einstein radius $\theta_E \in [0.5'', 5'']$ as already in the simulation procedure, and the complex external shear $\gamma_1 \in [-0.1, 0.1]$ and $\gamma_2 \in [-0.1, 0.1]$. This means the ground truth get scaled as

$$\eta^{\text{scaled, tr}} = \frac{\eta^{\text{tr}} - a}{a - b}, \tag{4.6}$$

and the ourput of the network get scaled back to the original ranges through

$$\eta = (b - a) \times \eta^{\text{scaled, pred}} - a \tag{4.7}$$

and for the uncertainty

$$\sigma = (b - a) \times \sigma^{\text{scaled}}. \tag{4.8}$$

The uncertainties are not shifted by $-a$ as those are considered with respect to the predicted mean values η. This procedure implies an sigmoid layer in the network architecture, which is applied before splitting it up into 7 values for the median and 7 values for the uncertainty σ, to compute the loss with Eq. (4.5). The network parameter optimization is performed with a ReLU (Nair & Hinton, 2010) activation function and a stochastic gradient descent algorithm, adjusting the weights to minimize the loss.

To avoid unbalanced optimization of the network, we used again a 5-fold cross-validation by splitting our set of 90,112 mocks into roughly 56% training, 14%validation, and 30% testing, where rounding effects from the batch size occur, and train each run over 500 epochs. This allows us also to determine better the hyper-parameter and the final stopping epoch, defined through the minimal average validation loss. This is visualized in Fig. 4.4, where we show the loss curves of the training set (dotted) and validation set (solid) for our 5 runs (colored) and the resulting mean curve (black). The lowest mean loss is −524.14 obtained in epoch 49, such that the final run is trained over 49 epochs. It shows a relative big generalization gap (compare Fig. 3.11), which means the network performs after some amount of epochs significantly better on the training data than on the validation set. A typical procedure to reduce this effect is through dropout, which was suggested by Perreault Levasseur et al. (2017) to be used for the uncertainty. This is further discussed in Sect. 4.5.
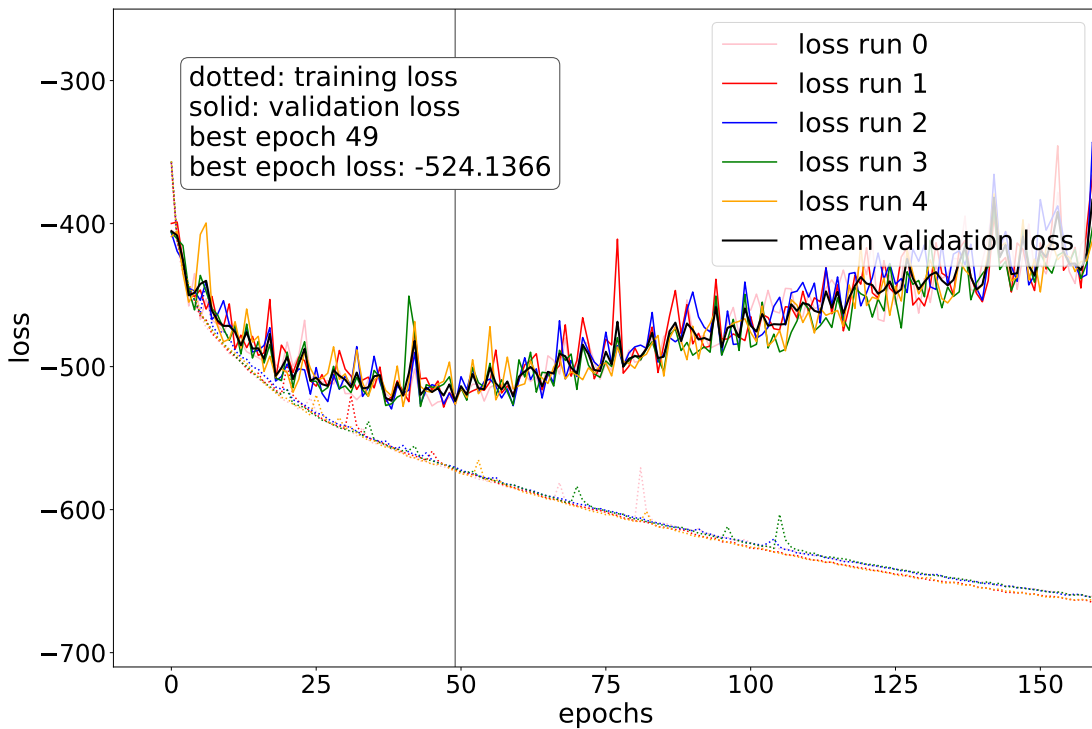

Figure 4.4: Loss curves for our 5-fold cross-validation together with its mean curve (black). The minimal validation loss is −524.14 and obtained in epoch 49.

In the presented network, we assumed a batch size of 32, a learning rate $r_{\text{learn}} = 0.00001$, a weight decay of 0.0005, a momentum of 0.9, a regularization constant ϵ of 0.5 for all different parameters and a weighting factor w of 1 for all different parameters, i.e., no up-weighting of any specific parameter such as the Einstein radius or external shear. As shown in Fig. 4.3, we include three residual blocks with a stride of 2, 1, and 1, respectively, and 24, 32, and 64 feature maps, respectively. The first layer before the residual blocks has 16 feature maps, while the input

has 4 corresponding to the 4 different filters. The FC layers connect 1024 to 32, 32 to 16 and 16 to 14 neurons, respectively. We discuss the details on the hyper-parameter search in Sect. 4.5.2.

4.4 Network results and performance

Following the work presented in Chapter 3, we present here our results obtained with a residual neural network predicting the SIE+γ_{ext} parameters and corresponding uncertainties σ. This network was trained, validated and tested on 90,112 realistic mock images created with our upgraded simulation procedure described in Sect. 4.2. The optimized network architecture is shown in Fig. 4.3, with which we obtain overall good results. We show in Fig. 4.5 a comparison between ground truth and predicted values on the test set, which are images the network has never seen before nor were they used during the cross-validation procedure. In detail, we show in the left column the performance on the lens center and the ellipticity, and in the right column the performance for the Einstein radius and the external shear, respectively, always as histogram and as median with 1σ and 2σ ranges with the true value on the x-axis and the network predictions on the y axis.

The network performs very well on the lens center, especially in contrast to our CNNs presented in Chapter 3 where we had difficulties in recovering the mass center. Here we directly see the good improvement on the lens centering resulting from the random shift of the final mock image by up to ± 3 pixels. The median with 1σ for $x^{\text{pred}} - x^{\text{tr}}$ and $y^{\text{pred}} - y^{\text{tr}}$ are, respectively, $-0.06^{+0.4}_{-0.5}$ and $0.1^{+0.6}_{-0.4}$. The complex ellipticity is also very well recovered, although the predictions tend to values closer to zero than it should. In words, the network predict the galaxy to be more round as it is, although this interpretation might be not the only reason. Another possible reason is the much higher number of training systems with values around zero such that the network guess for some systems just the most probable value. We obtain a median with 1σ values for $e_x^{\text{pred}} - e_x^{\text{tr}}$ and $e_y^{\text{pred}} - e_y^{\text{tr}}$ of $-0.009^{+0.1}_{-0.1}$ and $0.02^{+0.1}_{-0.08}$, respectively.

The Einstein radius θ_{E} is very well recovered with a mean and 1σ values of $\theta_{\text{E}}^{\text{pred}} - \theta_{\text{E}}^{\text{tr}} = 0.002^{+0.09}_{-0.09}$, although we do not up-weight the contribution to the loss as done for our CNNs presented in Chapter 3. The median line still follows closely the 1:1 line between our lower limit of $0.5''$ and around $2.5''$, dropping for systems with very large image separation. The binned 1σ and 2σ ranges indicate very precise predictions between $\sim 1''$ and $\sim 2''$, but also beyond this range the performance is good. The lower precision on the low end is most likely because of blending issues with the lens. Because of the small image separation(s), the counter image(s) are relatively more blended with the lenses such that the network cannot deblend the arcs from the lens well enough which is important for the prediction of the Einstein radius. This is confirmed by a test using the arcs alone for the network training (see Sect. 4.5.3 for details). On the other side of the range, the performance drops significantly due to their under representation in the data set. As extensively tested for our CNNs, the distribution on the Einstein radii are crucial for the performance, and a uniformly distributed sample yield the best performance considered over the full range. Therefore we artificially enforced a similar number of systems below $\sim 2''$ per bin. Since we use real measurements of the velocity dispersion and redshifts, lensing systems with an Einstein radius above $\sim 2.5''$ are very seldom even by increasing the number of iterations for testing different lens-source alignments or lens-source pairs. As a result, the number of systems
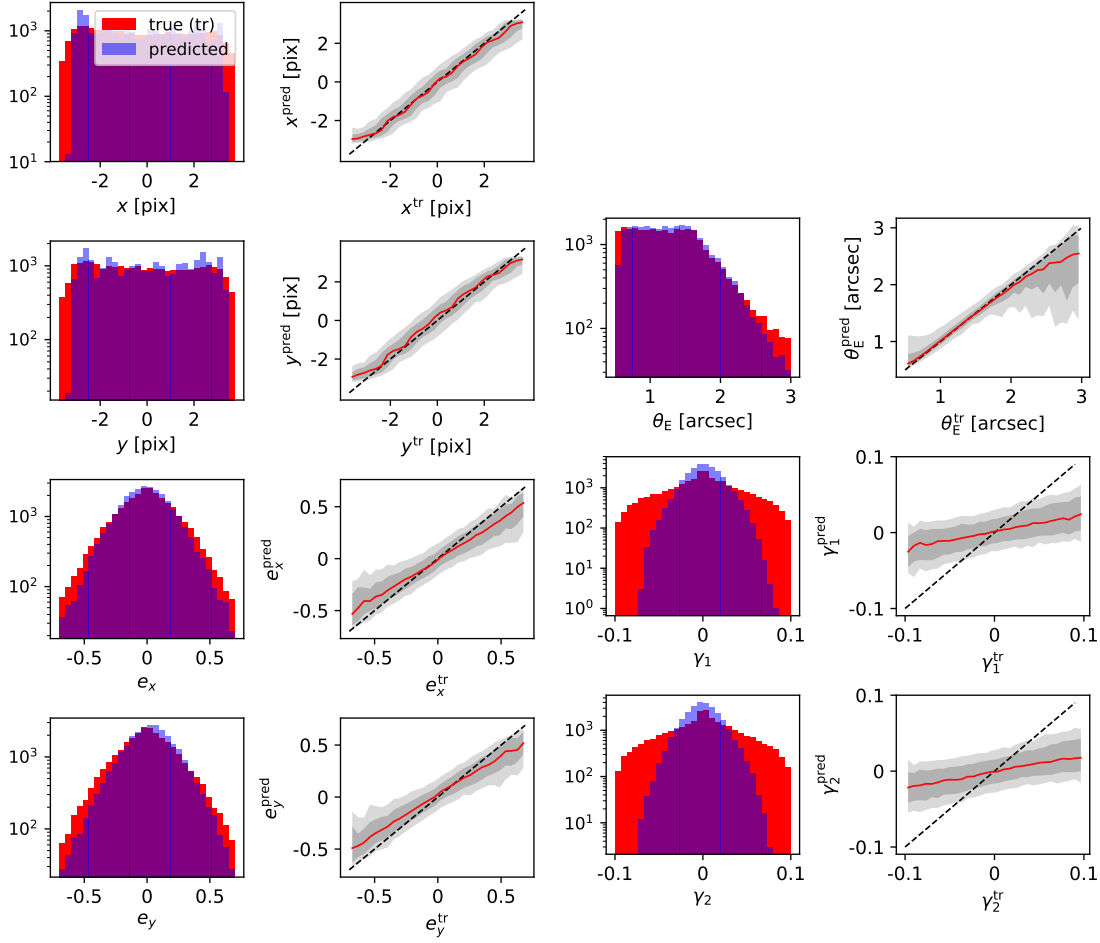
Figure 4.5: Comparison between ground truth and prediction of our final network. In the left panel of each column we show histograms of the ground truth (red) and predicted values (blue). On the right panel we plot directly the predicted value (*y*-axis) against the true value, showing the median as a red line with 1σ and 2σ ranges (gray shaded).

drop by two orders of magnitude towards $\theta_E \sim 3''$ compared to the plateau with the consequence of decreasing performance. Given the very low number of systems within each bin for $\theta_E > 3''$, we do not show them in Fig. 4.5 but keep our limit at $5''$ as largest image separation accepted by our simulation pipeline. Even if the network performs not well in this range, it has seen some of these systems and because of our introduced scaling, the network is in principle able to predict such large Einstein radii. As demonstrated with our modeling CNN in Sect. 3.4.2, we can also train a dedicated network for e.g., systems with $\theta_E \geq 2''$ giving very good performance on systems with such large image separation.

As we see from Fig. 4.5, the network is so far not able to predict the external shear well although the mean with 1σ vales are $0.001^{+0.04}_{-0.03}$ and $-0.001^{+0.03}_{-0.04}$, respectively. It tends to predict values closer to zero, resulting in a lower shear strength than it should. We tested many different

possibilities to improve on the external shear as described in Sect. 4.5, and found that blending with the lens or the low image quality compared to e.g., HST images is not the main reason for this issue. It seems that the current network cannot generalize to new systems well enough on these very minor distortions on the arcs or the degeneracy to other parameters are higher than expected, such that further investigations are necessary if one needs a good estimate of the external shear. This depends on the science case beyond the modeling. For statistical studies on the image systems such as the mass of the lenses, the external shear is expected to have only negligible influence.

We show in Fig. 4.6 the performance of the network as difference between predicted values and ground truth for the seven parameters in analogy to Fig. 4.6. In the bottom row are the normalized histograms and above the 2D correlations distribution with 1σ contour (solid line) and 2σ contour (dotted line). We find no strong correlation, also not between typically degenerated quantities such as ellipticity and external shear. However, by comparing to the performance of our CNNs in Fig. 4.6 where the plotting ranges of the SIE parameters are kept the same for a more easily comparison, we find a general better performance on the Einstein radius but with the same kind of diamond shaped 2σ contour. On the other hand, we obtain with the ResNet larger discrepancies on the lens center most likely because of our newly introduced random ± 3 pixels shift for the final mocks to ensure the network predicts the lens mass center instead of the image center. Moreover, one has to remember also our other drastic changes by introducing new parameters through the external shear, the uncertainty prediction, the change in the network architecture or introducing Poisson noise in the simulations, which makes this comparison to some extend unfair or at least difficult to infer the reasons for observed changes in performance. Since the ResNet has a much more complex task with predicting the external shear and uncertainties per parameter to solve, which it does overall good, this is definitely the more powerful network.

Beside the median value for each parameter, the network predicts also an uncertainty σ for each parameter. To interpret this as the width of a Gaussian distribution, it has to match the statistical expectations of 1σ correspond to a Convidence Intervall (CI) of 68.3%, 2σ to 95.4% CI, and 3σ to 99.7%. Given our predictions are not perfectly matching a Gaussian distribution, we scale the predicted values for the different parameters x, y, e_x, e_y, θ_E, γ_1, and γ_2 by, respectively, 1.127, 1.150, 1.185, 1.164, 1.124, 1.265, and 1.255, to match the 86.3% CI as this is the typically (e.g., in Chapter 5) interpretation. This is visualized in Fig. 4.7, where we show the coverage of the scaled uncertainty values for each parameter (gray bars) as absolute values (left) and (difference to the expectations (right). The left panel demonstrates to close match to the expected CI levels (blue dashed) especially also for the mean over all seven parameters (red dotted) and can be directly compared to the achievements from Perreault Levasseur et al. (2017, Fig. 2) and Pearson et al. (2021, Fig. 4), while the right panel highlights better the small deviations. Since we plot here the difference between achieved and expected CIs, the blue lines fall all onto exactly zero. It shows the good match of the 1σ mean and all individual parameters achieved through appropriate scaling, resulting in visible deviations for the 2σ and 3σ lines. Especially the distribution for the Einstein radius is sharper than a Gaussian distribution.

A comparison of the performance to other modeling networks are difficult given the discrepancy in assumptions. Hezaveh et al. (2017), who originally came up with this novel idea, present a network to predict e_x, e_y, and θ_E of a SIE profile for HST-like images after the lens light got
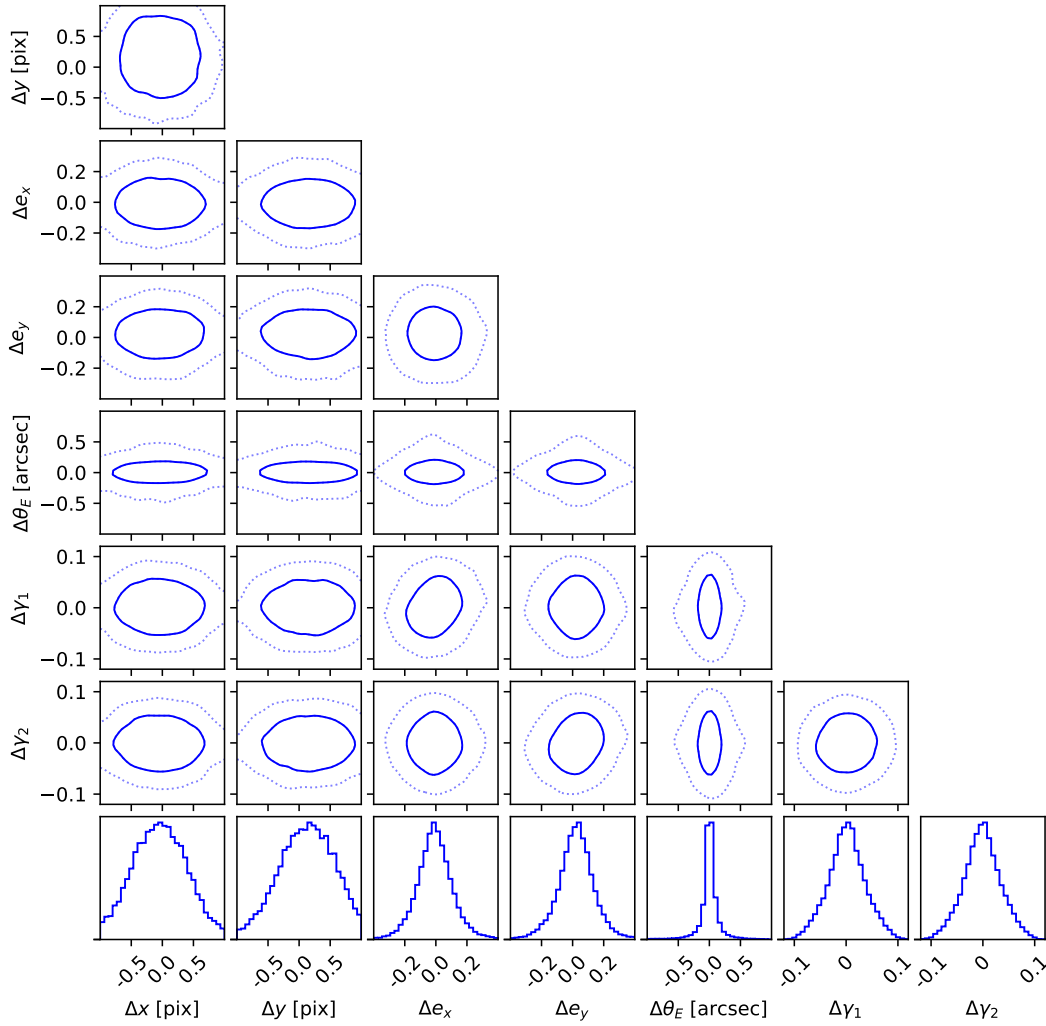
Figure 4.6: Histograms (bottom row) and 2D density plots of the difference between prediction and ground truth of our final ResNet.

subtracted to demonstrate the possibility. In Perreault Levasseur et al. (2017), they further included uncertainty predictions and an external shear component. Similar to that Pearson et al. (2021) present a network to predict e_x, e_y, and θ_E of a SIE profile for mock images with 0.1″ resolution in preparation for the Euclid space mission. Additionally, they include error estimations inspired by Perreault Levasseur et al. (2017) and explore the opportunity of a hybrid code by combining it with PyAutoLens (Nightingale et al., 2018), a fully automated non-machine learning based modeling software, for further refinement of the parameter predictions. Especially because of the difference in image resolution, number of filters, and the quality of training and
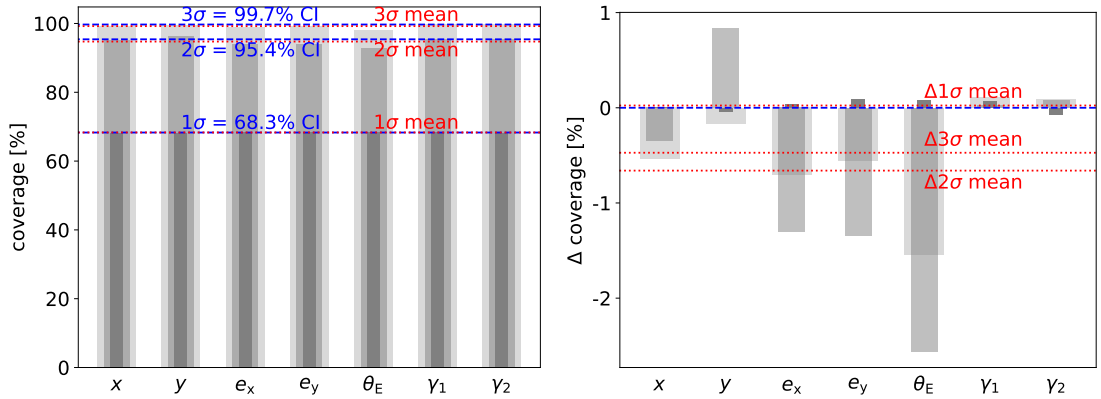
Figure 4.7: Plot of the uncertainty prediction coverage to demonstrate the interpredability of the uncertainties as 1σ values as absolute values (left panel) and relative values to their expectations (right panel).

test data makes a comparison difficult. Moreover, the different number of predicted parameters complicates the comparison given the degeneracies between the different parameters. The closest work in terms of image quality and number of filters is presented in Pearson et al. (2019), who consider CNNs to predict e_x, e_y, and θ_E of a SIE profile for Euclid, LSST r-band, and LSST gri band data. The latter is the best match to us, and as already mentioned in Sect. 3.7, comparable to the performance of our CNN from Chapter 3 and also to our upgraded ResNet network.

4.5 Network tests

To find our final network architecture as shown in Fig. 4.3 and the best set of hyper-parameters defined as the network with smallest mean validation loss (compare loss curve in Fig. 4.4), we carried out extensive tests which we summarize in this section. Throughout these tests we adopted a weight decay of 0.0005, a momentum of 0.9, and after some short tests regarding the effect on the performance, a batch size of 32 and the normalization procedure described in Sect. 4.3. As shown in Sect. 4.4, the main remaining difficulty is the prediction of the external shear which is the focus of most of the tests.

4.5.1 Network architecture

We tested several different variations of the global network architecture by varying the number of residual blocks between 2 and 6, the number of FC layers between 1 and 3, as well as the number of feature maps and strides in the convolutional layers or neurons within the different FC layers. We also tested different kernel sizes for the convolutional layers, but obtained the smallest average validation loss with the standard 3×3 kernel which is understandable given our comparable small image size.

During our tests on the network architecture, we also exchanged the pooling layer before flattening for the FC layers (compare Fig. 4.3). Here we adopted an average pooling layer, a maxima pooling layer or no pooling at all. For the two different pooling layers, we tested a kernel size of 8×8, 4×4, or 2×2. We found the best performance by using an average pooling layer with a kernel size of 8×8, as indicated in Fig. 4.3.

In addition to that, we modified the loss function by introducing the uncertainty prediction such that also σ contributes to the optimization of the weights and neurons during the training process. Here we tested different possibilities of regularization terms such as using the absolute value of σ instead of the squared term, as given in Eq. (4.5), or even to drop the regularization term completely. Moreover, we varied the regularization constant ϵ. Since changing the loss function will change the loss value for a given network, we cannot compare the obtained loss values directly. Based on a more quantitative comparison, we found no notable difference in the performance on the median predictions. Given that we use a scaling to match the expected 68.3% CI instead of tuning dropout for this, we adopted finally the regularization function proposed by Perreault Levasseur et al. (2017), which uses a squared term and a regularization constant of 0.5.

We further tested the possibility to split into multiple branches after flattening and before the FC layers (compare Fig. 4.3). Each branch consists then of one or more FC layers, such that specific FC layers predict only specific parameters. The input of the first FC layer in each branch is the full flatted data cube obtained after the pooling, i.e. the same for each branch. Here we considered three scenarios, although others are possible as well. In the first case we split into two branches, predicting each 7 values, one branch for the median values and one branch predicting the uncertainties. In the next scenario we split into 7 branches, each branch predicts the median and error for one parameter. The third considered option includes 14 branches, where each branch predicts just one value. Since we did not found an improvement through these tests, we discard this idea.

However, if not helpful to obtain better performance compared to an architecture with just a single branch of FC layers, it could be helpful when trying to tune the dropout rate as suggested by Perreault Levasseur et al. (2017) through adopting a specific dropout rate for each branch.

As completely independent test, we trained networks on pre-developed ResNet architectures, specifically on ResNet-18, ResNet-34, and ResNet-50 (He et al., 2016). We further tested both either pre-trained on the $\sim$ 14 million images from ImageNet (Deng et al., 2009; Russakovsky et al., 2015) or not pre-trained i.e. with a random initialization. Here the last FC layer, which originally outputs 1,000 values for their classification task, was replaced by a FC layer outputting only 14 neurons. Since these networks are only developed for three input channels, we used here the images only in *g*, *r* and *i* band, skipping the *y* band which has typically the lowest quality and information on the arcs. The images where then pre-processed by sub-sampling the images to the expected image size and re-scaling of the pixel values to match the expected input data from pre-training. Given this re-scaling and sub-sampling, changing drastically the last layer, and switching from a classification task to a regression task, it was no surprise that the performance were not that good. We therefore continued only with our own developed network architecture with residual blocks like in their architectures.

4.5.2 Hyper-parameter search

In addition to testing the different architecture layout, also the values for hyper-parameter need to be optimized. One of the key-quantities is the learning rate, for which we typically tested the following values: 0.1, 0.001, 0.0001, 0.00001, 0.000001 when changing any other hyper-parameter or layer. This covers a good range of plausible values. Since the changes on the weights are expected to drop over the training, we also tested the option of a "decreasing learning rate", which means we divide every 20 epoch the learning rate by a factor of 2. Because the best epoch was typically below 100, we had only a few relevant decreasing steps during training such that we adopted finally a constant learning rate again for most of our further tests.

Similar to the weighting factors introduced already for our CNN in Chapter 3 used to improve on the Einstein radius, we introduced also weighting factors in our new loss function given in Eq. (4.5). With these weighting factors we can control the contribution of the different parameters to the loss, and thus which shall be better optimized. For our CNN, we found it helpful to increase the contribution of the Einstein radius to the loss, as this is the key quantity in the SIE profile. Although it remains the key parameter in an SIE+γ_{ext} setup, we also tested the possibility of up-weighting the external shear as these two parameters are the most problematic ones. Although this helped to improve on the external shear, we discarded this option for all further tests as we gained only minor improvement on the external shear but lose notable performance on the other parameters. As already discussed in Sect. 4.4, it seems as the external shear introduces too small distortions on the arcs that can be generalized well enough with our current setup.

We further tested the effect of dropout on the FC layers even with our relatively small amount of FC layers. It helped to reduce the over-fitting, which is quite strong in our case as we see in Fig. 4.4, but resulted instead in a higher average validation loss used to select the best network. Therefore, we consider no dropout for the final network.

Although this is not a hyper-parameter that typically get tuned, we tested the effect of using different initialization of the network for a few given setups. This demonstrates how independent the final, trained network is from the original values. We found no preference of a specific considered seed and the overall performance was unaffected, but each network gave a slightly different loss, showing that it is not fully independent of the initialization. These slight changes are due to the stochastic learning process and thus commonly observed. For a few instances, the best hyper-parameter like the learning rate changed by changing the seed, indicate the importance of optimizing the hyper-parameter for a given network.

To mitigate these changes, so called ensemble learning methods can be used, were essentially the same network, i.e., with fixed architecture and hyper-parameters, get trained with different initialization and the predictions are combined afterwards. We performed such testes for a few given setups, predicted the 14 parameters with each network, and compared their average to the ground truth on the test set. In our case, this helped not to average out the outliers and decrease the scatter, which prove the similarity of the networks and their overall independence of the initialization.

4.5.3 Variations of the input data

Since we gained in our lens search projects (Cañameras et al., 2021; Cañameras et al., in prep.; Shu et al., in prep.) better performance by applying a square-root stretching, we also tested this for our modeling network although it raised the question whether this direct modification of the images is acceptable for a modeling network. This means we pass not anymore the images to the network, but rather the square-root of the images after setting all negative background pixels to zero. This helps to increase the faint arcs compared to the brighter lens in the center, but changes also the ratio between the different filters where the color information is encoded which is important for the modeling. This might be the reason that we found no improvement for the modeling network.

Another test of this kind were to subsample the images by a linear interpolation to increase the number of pixels as our images with 64×64 pixels are very small compared to images e.g., from ImageNet. Although we do not add information, we could imagine that a higher resolution helped the network to access better the small features which are important especially for the external shear. We adopted here a subsampling factor of 2, 3, or 4, but since a subsampling factor of 3 or 4 gave no improvement, we focused for further tests on a subsampling factor of two or no subsampling. A higher number of pixels allows also deeper networks when reducing the size of the data cube through larger kernel sizes or strides in the convolutional layers. Therfore, this looked promising during the development, but according to the mean validation loss the presented network does not require subsampling.

Given our difficulties with the external shear, we thought whether we can give the network some additional information that might help the network. Therefore, we tested the option to provide the Full Width at Half Maximum (FWHM) values of the PSF frames in addition to the normal set of images. These values were added to the flattened output of the convolutional layers and processed through the FC layers. Since we did not obtain any improvement with this, we considered networks accepting eight frames as we added the PSF images directly as images. Since the network architecture requires the same image size for all frames, we subsampled the PSF with a linear interpolation. Unfortunately, we also gained no improvement with this option, possibly as such networks perform well on pattern recognition with their relatively small kernel sizes rather than on analyzing very similar and completely smooth images like a PSF. Another possible reason for gaining no improvement by adding the PSF images which are likely imperfect due to the complex stacking procedure. These small errors are estimated to reach the 1% level at maximum (Aihara et al., 2018a, 2019), but for the relative small effects of the external shear on the arcs possibly relevant.

To investigate further on the external shear, we trained networks on images containing only the arcs. This was done in many other modeling networks (e.g., Hezaveh et al., 2017; Perreault Levasseur et al., 2017; Morningstar et al., 2018, 2019; Pearson et al., 2021) and seem to be helpful. Although one removes information from the lens light, one makes on the other hand the remaining information more easily accessible. We assumed here perfect lens light subtraction, which is in reality not achievable but for our test definitely good enough, which was mainly to test whether the network can now predict better the external shear which is only encoded in the arcs. As expected, the network predicts perfectly the Einstein radius with nearly no scatter over the full range, and also performs well on the lens center. This good performance on the systems

with small image separations confirms that the lower performance for our normal network is due to blending issues with the lens. In this test, we lose notable performance on the ellipticity, which is to some extend expected as this information the network can get also from the lens *light*, although it predicts the ellipticity of the mass distribution. However, also here the network cannot predict the external shear well, which suggest that the information is very hard to access and a generalization over the whole sample is currently impossible.

To answer a complete different question, we have trained network on fewer filters. Here we assumed either only the *g* band, *g* and *r* band, or *g* and *i* band. Although we performed no real optimization of the network architecture nor hyper-parameters, we found notable lose in performance with two filters. Moreover, also with just one filter the network is able to predict roughly the SIE parameters with most troubles on the ellipticity, which confirms the possibility to train CNNs or ResNets on single-band images e.g., from Euclid where the much better resolution will compensate the missing color information to some extend. As expected, the external shear is not predictable at all with just one filter and HSC image quality within our few tests in this direction.

4.5.4 Over-fitting tests

For testing whether a change in the network architectures or on the hyper-parameter is promising, we performed so-called over-fitting tests, which means that we trained the network just on a very small sample containing only 1,142 mock images. This shows if the network is able to learn all parameters by heart and, if not, exclude directly the possibility of a good enough model with this architecture. As this were only short tests, we performed no cross-validation. We show the result of one run in Fig. 4.8, from which we can see that the network predicts very good the values for the training set as expected. Given our difficulties with the external shear, it is important to remark that here the network is able to learn the external shear for a very small sample. This shows that the network is in general able to remember features of the images and can connect them to all seven parameters values although not perfectly. Especially on the shear we see some scatter, which indicates that the network not just remember the exact image and outputs the stored values, which demonstrates that the network architecture is not or not only responsible for the failure on the external shear. We further performed such over-fitting tests with networks just to predict the external shear. This helped to significantly improve on the training data. On the other hand, if we roughly double the sample size, the network has problems in recovering the seven parameters that precisely.

4.5.5 Test with fixed lens-source pairs

Since our over-fitting tests presented in Sect. 4.5.4 only shows that the network is able to predict the external shear very well on the data it was trained on, we further test whether it can predict the shear on new images if we simplify the task. For this, we considered three different stages of simplification and created a sample with 1,000 mocks each. First, we use always the same lens, but different background sources from HUDF and place them randomly behind the lens. The second scenario has always the same lens and source, but varying the position of the background source with respect to the lens plus different mass-to-light variations as in our general training
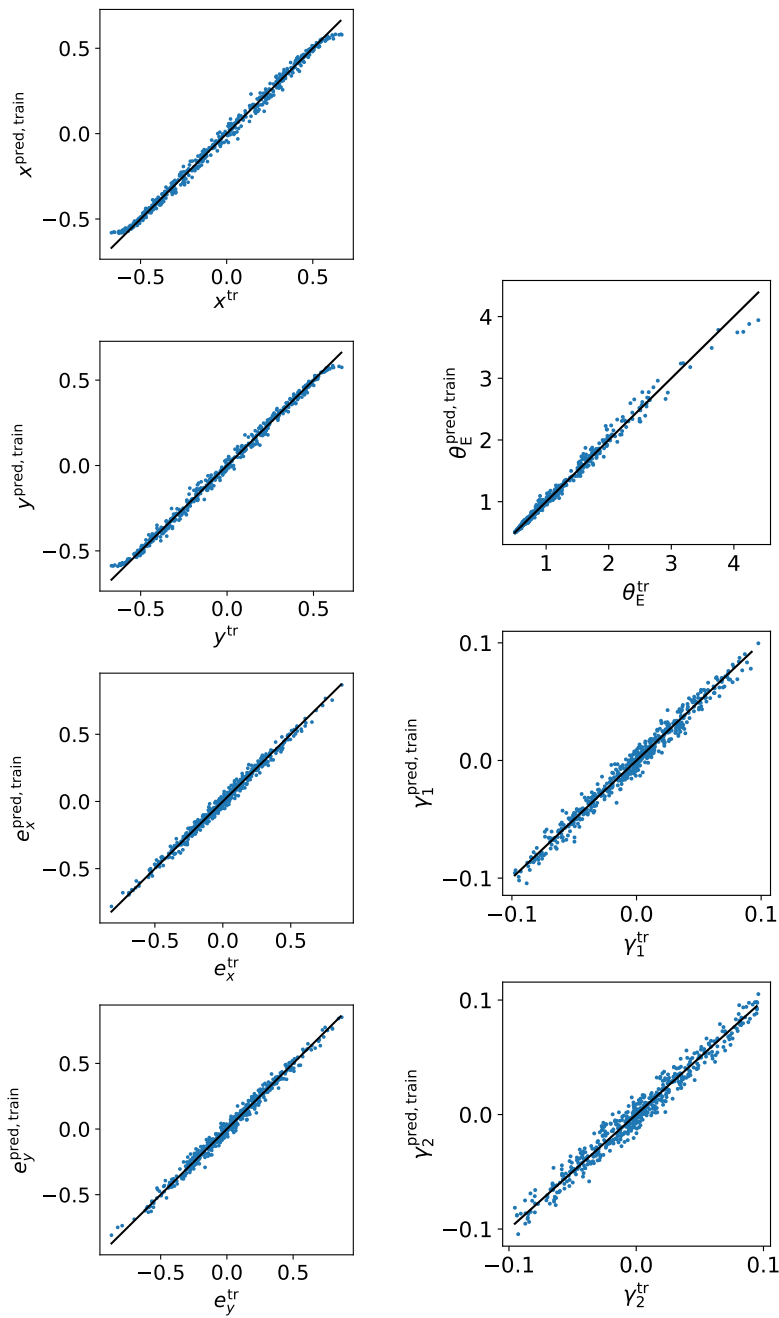
Figure 4.8: Comparison of ground truth and predictions on the training set. We train this network on a very small sample to demonstrate the ability of learning perfectly the input values for all parameters rater than obtaining a network for further applications.

data set. The third option was to keep everything fixed including the source position, and only vary the external shear. This means in the third option the arcs look always the same, with only distortions of the external shear. Since the lens does not vary and we have also restrictions on the source, resulting in very similar SIE parameter values, we excluded them in the prediction and trained the networks to only predict the external shear.

In the first and second scenario, the network is able to predict better the external shear but not perfectly, which is the case for the third option as we show in Fig. 4.9. Here the network is able to predict not only on the training set (top panel), but also on the test data equally perfectly the external shear (bottom panel). This demonstrates the ability of the network to transfer the shear extraction to completely new images and thus recover very well the external shear on the test set.
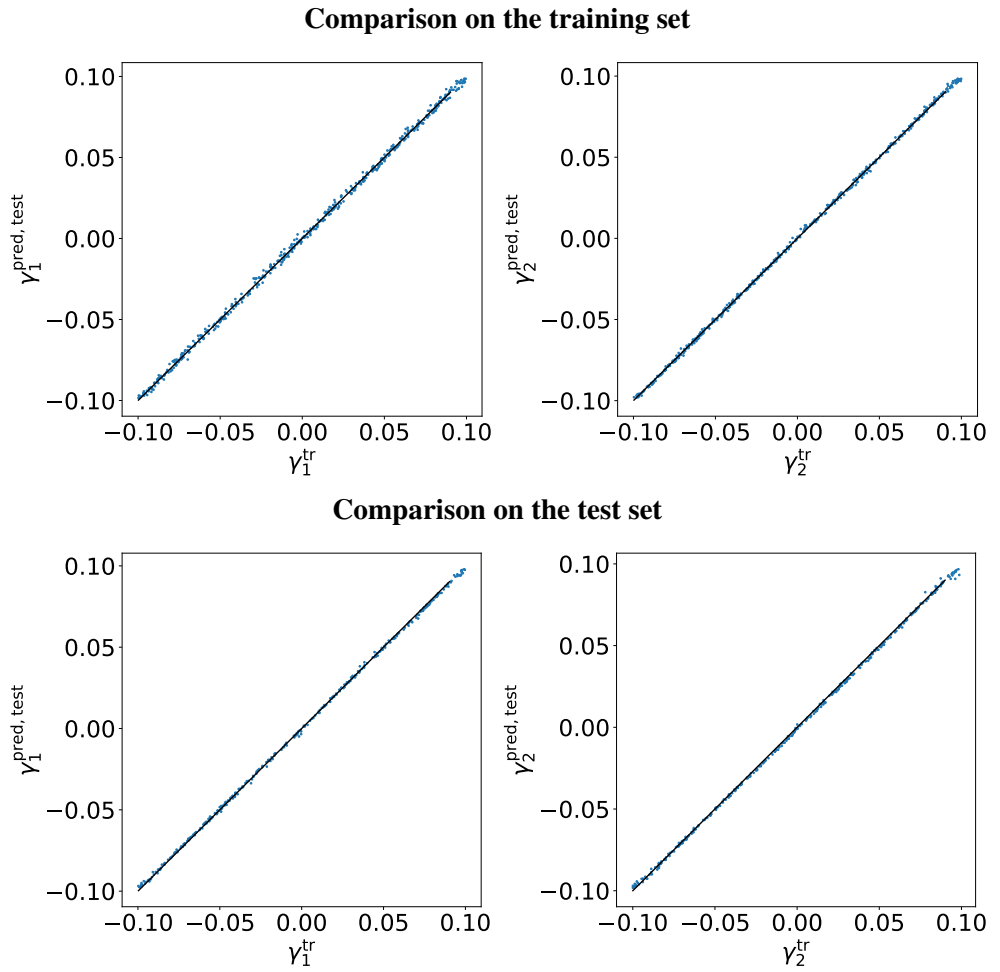
Figure 4.9: Comparison of ground truth and predictions from a network trained on 1,000 mocks with always the same lens and source pair to predict the external shear, once applied to the training set (top panel) and once to the test set (bottom panel). Under this extreme simplification the network is able to predict perfectly the external shear even on new images.

Since the lens and background source are always the same, we also exclude the possibility that the network connects other features of the image with the external shear parameters. On the other hand, we also exclude possible degeneracies between different parameters such as the ellipticity and the external shear that might explain the difficulties with the external shear. Another reason for the good performance here might be the same PSF for all systems given we are using the same lens. Normally, different lenses have slightly different PSFs, such that the arcs look differently after the convolution. This variations can introduce some difficulties for the network, as it does not know how the PSF looks for a given lens system. Especially for the external shear small effects on the arcs necessary, which get highly influenced by the different PSF shapes. However, passing the PSF frames together with the images to the network did not help (see Sect. 4.5.3).

4.6 Summary and conclusion

In Chapter 3, we have demonstrated the possibility of modeling strongly lensed ground-based galaxy-scale systems with a relatively simple CNN inspired by the LeNet architecture. Built upon these results, we showed in this chapter the possibility of modeling such systems when including additionally the external shear component to the SIE mass distribution of the lens. Moreover, we predict now also an 1σ value per parameter and lensing system. For this we make use of a residual neural network, which is a specific type of CNNs that include so called residual blocks with a skip connection. A diagram of our final network architecture is shown in Fig. 4.3 and visualize the setup of residual blocks. Because of the included error prediction, we changed the loss function from a MSE loss to a log-probability function with a regression term inspired by Perreault Levasseur et al. (2017). For an equal contribution of all parameters to the loss, we have introduced a scaling of each parameter to the range [0, 1] and therefore include the sigmoid function as last layer.

This network is trained on mock images created with our simulation pipeline introduced in Sect. 3.2, which we upgraded for this project as described in Sect. 4.2. The procedure is based on purely real observed data and only simulate the lensing effect, resulting in very realistic images as shown in Fig. 4.2. We use again as lenses images of LRGs observed with HSC, together with velocity and redshift measurements from SDSS. The background sources are again images from HUDF with provided redshifts.

Since the network shall now predict the external shear as well, these need to be also included in the simulation of the training data. Here we implemented and tested different distributions of the shear, either a flat distribution of the shear strength γ_{ext} and shear angle θ_{ext}, a shear strength distribution matching the expected distribution from lens studies, or a flat distribution in both complex quantities γ_1 and γ_2. Finally, we adopted the flat distribution in the shear strength γ_{ext}. Moreover, we include now the Poisson noise on the arcs which we approximate from the variance map provided by HSC and improve the lens center and ellipticity estimation by using a dedicated mask for each individual lens. To make sure the network predict the lens mass center, we apply a random shift of the final mock image with up to three pixels in both x and y directions.

With this procedure we created $\sim 90,000$ images each in four filters used to train our NN. Through extensive tests we found a network that is able to predict very accurately the SIE parameters compared to the ground truth values as shown in Fig. 4.5 and Fig. 4.6. Only the externals shear is very hard to predict. We carried out many tests on the external shear, such as up-weighting their loss contribution, only predicting the external shear, adding further information like the FWHM values or PSF images directly as additionally frames or subsampling. Furthermore, we are testing different network architectures by training them on a very small sample such that it completely over-fit to the training set. In this case the network is able to predict all seven parameters, i.e. also the external shear, for the training set demonstrating that the network architecture can cope with the given task. Since in that case the network can also connect other features to the values of the external shear, we considered thee further simplifications for the network. In the most extreme case, we create 1,000 mock images with always the exact same lens and source, and even force the source to the exact position being the lens, resulting in 1,000 images with slightly different arcs introduced by the external shear. In this case, as shown in Fig. 4.9, the network is able to predict perfectly the external shear also for new images in the test set. This demonstrates that the network obtain the predicted external shear from the arcs which are only different because of the variable external shear. In short, the network is able to predict the external shear from effects introduced by the external shear. Therefore, it seems as the network is in general able to extract the information from the external shear, but cannot generalize well to other systems which is probably a result of the combination of the complexity of the lensing system, image resolution, inaccurate masking of the sources affecting the mocks, for the network unknown PSF, correlations between the external shear and other parameters, or other reasons.

Independent of the difficulties on the external shear, we obtained overall very good results and a with the uncertainty prediction a much more powerful network then presented in Chapter 3, especially when taking into account the computational time for the network: it predicts in fractions of a second these parameter values, while state-of-the-art methods like GLEE & GLaD require at least days and a lot of user input to obtain these values. With this network we are able to predict the SIE+γ_{ext} values with uncertainties for all known HSC lenses or lens candidates, which are already few thousands, in an acceptable amount of time. Given the good match of HSC PDR2 images to the expected quality of LSST, our obtained performance is expected to hold also for LSST. Here we propose to generate dedicated mocks and train a separate network as soon as first data from LSST are available to omit possible lose in performance because of better but different images.

5 Direct model comparison of network and traditional method

Note. The content of this chapter will be submitted as an individual paper to an international journal for publication. We refer to this upcoming publication with Schuldt et al. (in prep b).

5.1 Introduction

We have demonstrated the possibility to modeling strong gravitational lensed galaxy-galaxy images observed with ground based telescopes through a NN. In Chapter 3, we have used a CNN inspired by the LeNet architecture (Lecun et al., 1998) to predict the five parameters from the adopted SIE profile. We then included error predictions and an external shear component as presented in Chapter 4. Both networks are trained on mock images created with our simulation pipeline introduced in Sect. 3.2.3 and Sect. 4.2, respectively. For this we use observed images of real galaxies, take their measured redshifts and the velocity dispersion of the galaxy acting as lens, and simulate only the lensing effect. This results in very realistic simulations as shown in Fig. 3.4 and Fig. 4.2, because the light distribution of the lens and source are real, but also as we include real LOS objects in the cutout. Nonetheless, they are mocks and, as discussed already, a small fraction include inaccuracies and unrealistic features. This raises the concerns that the in Chapter 3 and Chapter 4 shown performance is better than it would be on real data. Therefore, we perform in this chapter a direct comparison on a sample of 32 real galaxy-galaxy lenses found in the SuGOHI program (Sonnenfeld et al., 2018a; Wong et al., 2018; Sonnenfeld et al., 2019; Chan et al., 2020; Jaelani et al., 2020a; Sonnenfeld et al., 2020; Jaelani et al., 2021) presented in Sect. 5.2. We apply our trained ResNet to that sample, which is the first time a trained modeling network is applied to real ground-based images instead of mock images, and show our results in Sect. 5.3.

We then compare each model predicted with our network to a model that we obtained with traditional MCMC sampling methods through GLEE & GLaD (Suyu & Halkola, 2010; Suyu et al., 2012; Chirivì et al., 2020). For the traditional modeling we developed an automated code that is optimized for ground based images like from HSC or soon from LSST. For further refinement of the models we develop *gleeauto.py*, a flexible code to automate optimization steps selected by the user without assuming anything on the lens system setup. These procedures of the automation code as well as our results are described in Sect. 5.4. The comparison itself, which demonstrates how good machine learning can compete with conventional algorithms in terms of accuracy as well as time and resource cost, is given in Sect. 5.5. We finally present our conclusion in Sect. 5.6.

Similar to previous chapters, we assume throughout this chapter a flat ΛCDM cosmology with a Hubble constant $H_0 = 72 \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$ (Bonvin et al., 2017) and $\Omega_\mathrm{M} = 1 - \Omega_\Lambda = 0.32$

(Planck Collaboration et al., 2020). For the traditional modeling with GLEE & GLaD, each quoted parameter estimate is the median of its 1D marginalized posterior probability density function, and the quoted uncertainties represent the 16th and 84th percentiles (i.e., the bounds of a 68% credible interval).

5.2 Comparison data set

As comparison sample we use lenses detected within the SuGOHI program, a large and extensive lens search survey using various methods in the HSC footprint as introduced in Sect. 1.2. We select only the grade A candidates to have a very clean sample[1]. We further request galaxy-galaxy lenses as the network is trained for such systems. From the resulting sample we reject system HSCJ023307-043838 (Sonnenfeld et al., 2013; Jacobs et al., 2019; More et al., 2012; Gavazzi et al., 2014; Sonnenfeld et al., 2019; Chan et al., 2020; More et al., 2016) and HSCJ135138+002839 (Wong et al., 2018) as those look more like a cluster or group lens although listed as galaxy-galaxy system on the web-page. This results in a sample of 32 lenses which we summarize in Tab. 5.1. In that table, we also quote the redshifts for the lenses and if available for the sources. In case there is no spectroscopic redshift available, we note the photometric redshift. Since the quoted parameters are independent of the redshifts, we can model all systems although we do not have a source redshift for some of them. For our comparison we use images of the lenses in the four filters g, r, i, and z as requested by our network (compare Sect. 3.2.1 and Sect. 4.2). A mosaic of gri color images of our comparison sample is shown in Fig. 5.1.

5.3 Neural Network models

As presented in Chapter 4, we developed a NN to model HSC lens images. We apply this network to the presented sample of 32 known real lenses. It predicts within few seconds for all lenses the full set of parameter values with corresponding 1σ uncertainties. This set of parameters includes the external shear γ_1 and γ_2 which we give in Tab. 5.2 (gray shaded background) and the SIE mass parameters, the lens center x_l and y_l, the ellipticity e_x and e_y, and the Einstein radius θ_E, which we give in Tab. 5.3 (gray shaded background). The results are discussed and compared to the traditional obtained models in Sect. 5.5.

---

[1]Available on the SuGOHI web-page: http://www-utap.phys.s.u-tokyo.ac.jp/ oguri/sugohi/ .

Table 5.1: Overview of all 32 SuGOHI lenses modeled with GLEE & GLaD for a direct comparison to the network predictions of our ResNet.

Name	RA	DEC	z_d	z_s	References
HSCJ015618-010747	29.0755	−1.1298	0.542	1.167	(b), (c)
HSCJ020141-030946	30.4249	−3.1628	0.362	−	(a)
HSCJ020241-064611	30.6725	−6.7698	0.502	2.748	(a), (c)
HSCJ020955-024442	32.4809	−2.7450	0.560	−	(e), (f)
HSCJ021737-051329	34.4049	−5.2248	0.646	1.847	(c), (e), (g), (h), (k), (i)
HSCJ022346-053418	35.9423	−5.5718	0.499	1.444	(c), (e), (g), (h)
HSCJ022610-042011	36.5444	−4.3366	0.496	−	(a), (c), (h)
HSCJ023217-021703	38.0724	−2.2844	0.508	−	(a)
HSCJ023322-020530	38.3443	−2.0918	0.49	−	(e), (f)
HSCJ085046+003905	132.6942	0.6515	0.84	−	(f)
HSCJ085855-010208	134.7333	−1.0357	0.468	1.421	(a), (c)
HSCJ090429-010228	136.1239	−1.0411	0.957	3.403	(e), (j)
HSCJ094427-014742	146.1145	−1.7951	0.539	1.179	(c), (l)
HSCJ120623+001507	181.5994	0.2520	0.563	3.12	(a), (c)
HSCJ121052-011905	182.7187	−1.3181	0.7	2.295	(a), (c)
HSCJ121504+004726	183.7685	0.7906	0.642	1.297	(c), (l)
HSCJ124320-004517	190.8365	−0.7550	0.654	−	(b)
HSCJ125254+004356	193.2275	0.7323	0.649	−	(b)
HSCJ135138+002839	207.9122	0.4778	0.461	−	(b)
HSCJ141136-010215	212.9022	−1.0377	0.949	3.021	(f)
HSCJ141815+015832	214.5656	1.9756	0.556	2.139	(a), (c)
HSCJ142720+001916	216.8356	0.3211	0.551	1.266	(a), (c)
HSCJ144132-005358	220.3862	−0.8995	0.49	−	(f)
HSCJ144320-012537	220.8359	−1.4270	1.16	−	(f), (d)
HSCJ145242+425731	223.1789	42.9589	0.718	−	(b)
HSCJ150021-004936	225.0876	−0.8269	0.41	−	(f)
HSCJ150112+422113	225.3007	42.3537	0.27	−	(e)
HSCJ223733+005015	339.3897	0.8377	0.604	2.143	(a), (c)
HSCJ230335+003703	345.8965	0.6176	0.458	0.936	(b), (c), (l)
HSCJ230521-000211	346.3403	−0.0366	0.492	−	(b), (k)
HSCJ233130+003733	352.8770	0.6259	0.552	−	(b), (k)
HSCJ233146+013845	352.9434	1.6460	0.476	−	(b)

Note. From left to right we give the name used to reference each lens, right ascension (J2000), declination (J2000), lens redshift z_d, and source redshift z_s. The last column give the references: (a) Sonnenfeld et al. (2018a), (b) Wong et al. (2018), (c) Sonnenfeld et al. (2019), (d) Chan et al. (2020), (e) Jaelani et al. (2020a), (f) Sonnenfeld et al. (2020), (g) Gavazzi et al. (2014), (h) Sonnenfeld et al. (2013), (i) Cabanac et al. (2007), (j) Jaelani et al. (2020b), (k) Jacobs et al. (2019), (l) Brownstein et al. (2012)
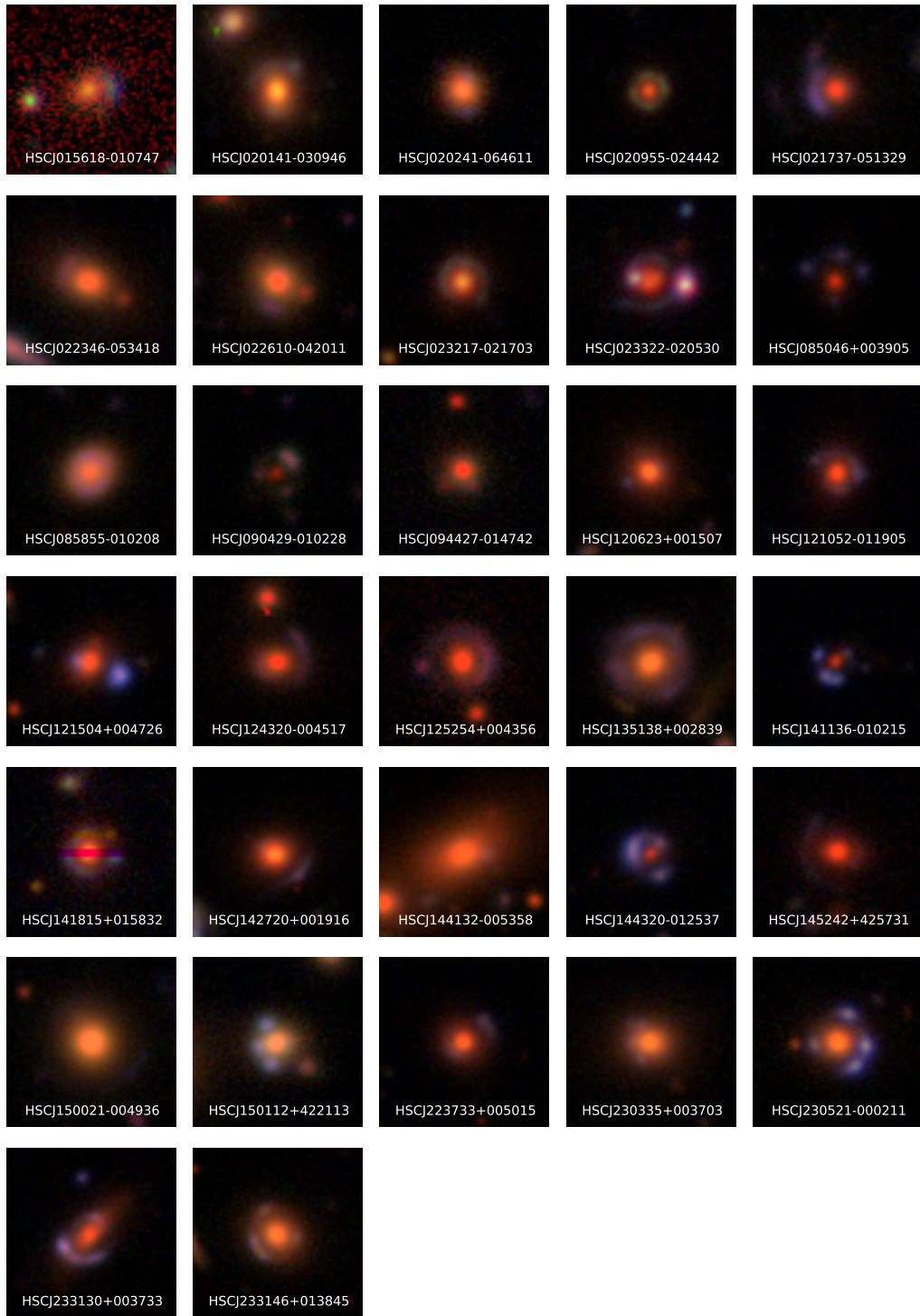
Figure 5.1: Color images based on the filters *gri* of the 32 HSC SuGOHI lenses used for direct comparison. All cutouts are $10'' \times 10''$ large and oriented such that north is up and east is left. Name of each lens given at the bottom of the image.

Table 5.2: Obtained external shear values γ_1 and γ_2 by modeling with GLEE & GLaD or with the network presented in Chapter 4 (column 2, method) for the 32 HSC SuGOHI lenses used for the direct comparison.

Name	method	γ_1	γ_2
HSCJ015618-010747	ResNet	0.01 ± 0.03	0.02 ± 0.03
	GLEE & GLaD	$0.05^{+0.01}_{-0.01}$	$-0.066^{+0.01}_{-0.008}$
HSCJ020141-030946	ResNet	-0.01 ± 0.03	0.02 ± 0.02
	GLEE & GLaD	$0.003^{+0.003}_{-0.003}$	$0.073^{+0.003}_{-0.003}$
HSCJ020241-064611	ResNet	0.03 ± 0.03	-0.01 ± 0.03
	GLEE & GLaD	$0.06^{+0.03}_{-0.03}$	$-0.10^{+0.02}_{-0.02}$
HSCJ020955-024442	ResNet	0.00 ± 0.02	0.02 ± 0.02
	GLEE & GLaD	$-0.032^{+0.005}_{-0.005}$	$-0.003^{+0.004}_{-0.004}$
HSCJ021737-051329	ResNet	0.06 ± 0.03	-0.02 ± 0.03
	GLEE & GLaD	$-0.066^{+0.008}_{-0.02}$	$-0.004^{+0.003}_{-0.003}$
HSCJ022346-053418	ResNet	0.04 ± 0.03	0.00 ± 0.03
	GLEE & GLaD	$0.005^{+0.002}_{-0.002}$	$-0.009^{+0.003}_{-0.003}$
HSCJ022610-042011	ResNet	0.00 ± 0.03	-0.02 ± 0.03
	GLEE & GLaD	$0.06^{+0.02}_{-0.02}$	$-0.018^{+0.01}_{-0.02}$
HSCJ023217-021703	ResNet	0.02 ± 0.02	0.00 ± 0.02
	GLEE & GLaD	$-0.014^{+0.008}_{-0.007}$	$-0.037^{+0.005}_{-0.005}$
HSCJ023322-020530	ResNet	-0.04 ± 0.03	0.04 ± 0.03
	GLEE & GLaD	$-0.008^{+0.001}_{-0.001}$	$0.108^{+0.001}_{-0.001}$
HSCJ085046+003905	ResNet	-0.03 ± 0.03	0.00 ± 0.03
	GLEE & GLaD	$0.058^{+0.01}_{-0.005}$	$-0.002^{+0.004}_{-0.01}$
HSCJ085855-010208	ResNet	0.04 ± 0.02	0.02 ± 0.02
	GLEE & GLaD	$-0.045^{+0.002}_{-0.002}$	$0.058^{+0.003}_{-0.003}$
HSCJ090429-010228	ResNet	0.00 ± 0.03	-0.01 ± 0.03
	GLEE & GLaD	$-0.053^{+0.002}_{-0.002}$	$-0.005^{+0.001}_{-0.001}$
HSCJ094427-014742	ResNet	-0.01 ± 0.03	-0.00 ± 0.03
	GLEE & GLaD	$0.02^{+0.01}_{-0.05}$	$0.081^{+0.005}_{-0.01}$
HSCJ120623+001507	ResNet	0.04 ± 0.03	-0.00 ± 0.03
	GLEE & GLaD	$-0.05^{+0.02}_{-0.02}$	$0.11^{+0.01}_{-0.01}$
HSCJ121052-011905	ResNet	-0.06 ± 0.02	-0.03 ± 0.02
	GLEE & GLaD	$-0.011^{+0.002}_{-0.002}$	$-0.039^{+0.005}_{-0.005}$
HSCJ121504+004726	ResNet	0.00 ± 0.04	0.00 ± 0.04
	GLEE & GLaD	$-0.055^{+0.006}_{-0.008}$	$-0.055^{+0.01}_{-0.009}$
HSCJ124320-004517	ResNet	0.00 ± 0.03	-0.01 ± 0.03

Table 5.2 Continued: Obtained external shear values γ_1 and γ_2 by modeling with GLEE & GLaD or with the network presented in Chapter 4 (column 2, method) for the 32 HSC SuGOHI lenses used for the direct comparison.

Name	method	γ_1	γ_2
	GLEE & GLaD	$-0.054^{+0.005}_{-0.005}$	$0.065^{+0.004}_{-0.003}$
HSCJ125254+004356	ResNet	-0.04 ± 0.02	-0.01 ± 0.02
	GLEE & GLaD	$0.082^{+0.004}_{-0.006}$	$0.027^{+0.007}_{-0.007}$
HSCJ135138+002839	ResNet	0.00 ± 0.02	-0.01 ± 0.02
	GLEE & GLaD	$-0.008^{+0.001}_{-0.001}$	$0.048^{+0.001}_{-0.001}$
HSCJ141136-010215	ResNet	-0.03 ± 0.03	0.02 ± 0.03
	GLEE & GLaD	$0.032^{+0.006}_{-0.007}$	$-0.006^{+0.003}_{-0.002}$
HSCJ141815+015832	ResNet	-0.06 ± 0.03	-0.01 ± 0.03
	GLEE & GLaD	$0.076^{+0.013}_{-0.014}$	$-0.050^{+0.006}_{-0.006}$
HSCJ142720+001916	ResNet	-0.03 ± 0.03	0.02 ± 0.03
	GLEE & GLaD	$-0.046^{+0.007}_{-0.007}$	$0.077^{+0.006}_{-0.008}$
HSCJ144132-005358	ResNet	-0.03 ± 0.04	-0.03 ± 0.04
	GLEE & GLaD	$0.054^{+0.02}_{-0.007}$	$-0.028^{+0.003}_{-0.003}$
HSCJ144320-012537	ResNet	0.05 ± 0.03	0.02 ± 0.03
	GLEE & GLaD	$-0.010^{+0.002}_{-0.002}$	$-0.056^{+0.001}_{-0.001}$
HSCJ145242+425731	ResNet	0.01 ± 0.03	0.01 ± 0.03
	GLEE & GLaD	$0.022^{+0.005}_{-0.005}$	$0.056^{+0.005}_{-0.004}$
HSCJ150021-004936	ResNet	-0.01 ± 0.03	-0.01 ± 0.03
	GLEE & GLaD	$0.100^{+0.001}_{-0.003}$	$-0.020^{+0.001}_{-0.002}$
HSCJ150112+422113	ResNet	0.02 ± 0.04	-0.02 ± 0.04
	GLEE & GLaD	$-0.133^{+0.001}_{-0.001}$	$0.003^{+0.002}_{-0.002}$
HSCJ223733+005015	ResNet	-0.01 ± 0.03	-0.02 ± 0.03
	GLEE & GLaD	$0.114^{+0.002}_{-0.005}$	$-0.004^{+0.002}_{-0.002}$
HSCJ230335+003703	ResNet	0.06 ± 0.03	-0.01 ± 0.03
	GLEE & GLaD	$0.010^{+0.001}_{-0.001}$	$-0.007^{+0.003}_{-0.004}$
HSCJ230521-000211	ResNet	0.05 ± 0.02	-0.02 ± 0.02
	GLEE & GLaD	$0.018^{+0.002}_{-0.002}$	$-0.046^{+0.001}_{-0.001}$
HSCJ233130+003733	ResNet	-0.04 ± 0.03	0.07 ± 0.03
	GLEE & GLaD	$0.047^{+0.002}_{-0.002}$	$0.059^{+0.001}_{-0.002}$
HSCJ233146+013845	ResNet	0.06 ± 0.02	-0.03 ± 0.02
	GLEE & GLaD	$-0.076^{+0.002}_{-0.002}$	$-0.034^{+0.002}_{-0.002}$

Table 5.3: Obtained SIE mass parameter values by modeling with GLEE & GLaD (converted to complex quantities) or with the network presented in Chapter 4 (column 2, method) for the 32 HSC SuGOHI lenses used for the direct comparison. The quoted SIE mass parameters are the lens mass center x and y, the lens mass complex ellipticity e_x and e_y, and the Einstein radius θ_E.

Name	method	$x_l['']$	$y_l['']$	e_x	e_y	$\theta_E['']$
HSCJ015618-010747	ResNet	0.1 ± 0.09	0.03 ± 0.08	0.07 ± 0.1	0.2 ± 0.1	1.6 ± 0.3
	GLEE & GLaD	$-0.076^{+0.001}_{-0.001}$	$0.070^{+0.001}_{-0.001}$	$0.562^{+0.03}_{-0.05}$	$0.417^{+0.06}_{-0.05}$	$0.99^{+0.02}_{-0.01}$
HSCJ020141-030946	ResNet	-0.12 ± 0.07	-0.03 ± 0.07	-0.11 ± 0.06	0.01 ± 0.06	1.56 ± 0.06
	GLEE & GLaD	$-0.005^{+0.001}_{-0.001}$	$-0.052^{+0.001}_{-0.001}$	$-0.211^{+0.02}_{-0.02}$	$-0.088^{+0.02}_{-0.02}$	$1.406^{+0.007}_{-0.007}$
HSCJ020241-064611	ResNet	-0.1 ± 0.1	-0.06 ± 0.09	-0.03 ± 0.06	0.00 ± 0.07	1.14 ± 0.04
	GLEE & GLaD	$-0.007^{+0.001}_{-0.001}$	$0.061^{+0.001}_{-0.001}$	$0.1^{+0.1}_{-0.1}$	$-0.158^{+0.08}_{-0.07}$	$1.26^{+0.03}_{-0.02}$
HSCJ020955-024442	ResNet	-0.14 ± 0.07	-0.17 ± 0.07	-0.05 ± 0.08	0.07 ± 0.08	1.00 ± 0.02
	GLEE & GLaD	$-0.071^{+0.001}_{-0.001}$	$0.021^{+0.002}_{-0.002}$	$-0.231^{+0.02}_{-0.02}$	$-0.084^{+0.02}_{-0.02}$	$1.043^{+0.003}_{-0.003}$
HSCJ021737-051329	ResNet	0.32 ± 0.08	0.06 ± 0.08	0.1 ± 0.1	0.1 ± 0.1	1.32 ± 0.09
	GLEE & GLaD	$-0.053^{+0.001}_{-0.001}$	$0.072^{+0.001}_{-0.001}$	$0.22^{+0.03}_{-0.02}$	$0.036^{+0.04}_{-0.009}$	$1.252^{+0.007}_{-0.02}$
HSCJ022346-053418	ResNet	-0.10 ± 0.06	-0.02 ± 0.09	0.21 ± 0.09	-0.4 ± 0.1	1.4 ± 0.1
	GLEE & GLaD	$-0.089^{+0.001}_{-0.001}$	$-0.020^{+0.001}_{-0.001}$	$0.09^{+0.007}_{-0.006}$	$-0.305^{+0.004}_{-0.004}$	$1.397^{+0.004}_{-0.003}$
HSCJ022610-042011	ResNet	-0.2 ± 0.1	-0.21 ± 0.08	-0.01 ± 0.09	-0.04 ± 0.09	1.4 ± 0.2
	GLEE & GLaD	$0.059^{+0.001}_{-0.001}$	$-0.023^{+0.001}_{-0.001}$	$-0.055^{+0.06}_{-0.06}$	$-0.065^{+0.05}_{-0.06}$	$1.16^{+0.02}_{-0.02}$
HSCJ023217-021703	ResNet	0.15 ± 0.09	-0.03 ± 0.09	0.00 ± 0.06	0.09 ± 0.07	1.30 ± 0.04
	GLEE & GLaD	$-0.058^{+0.001}_{-0.001}$	$-0.045^{+0.001}_{-0.001}$	$-0.191^{+0.03}_{-0.02}$	$0.361^{+0.02}_{-0.02}$	$1.345^{+0.009}_{-0.009}$
HSCJ023322-020530	ResNet	-0.3 ± 0.1	0.1 ± 0.1	0.2 ± 0.2	0.04 ± 0.1	1.7 ± 0.1
	GLEE & GLaD	$-0.033^{+0.002}_{-0.002}$	$0.057^{+0.001}_{-0.001}$	$-0.042^{+0.004}_{-0.004}$	$0.294^{+0.002}_{-0.003}$	$1.669^{+0.001}_{-0.001}$
HSCJ085046+003905	ResNet	0.14 ± 0.07	0.12 ± 0.08	-0.1 ± 0.1	0.0 ± 0.1	1.71 ± 0.06
	GLEE & GLaD	$-0.092^{+0.001}_{-0.001}$	$0.007^{+0.001}_{-0.001}$	$-0.336^{+0.05}_{-0.02}$	$0.087^{+0.006}_{-0.01}$	$1.750^{+0.006}_{-0.01}$
HSCJ085855-010208	ResNet	0.16 ± 0.07	-0.03 ± 0.08	0.03 ± 0.06	0.16 ± 0.07	1.00 ± 0.02

Table 5.3 Continued: Obtained SIE mass parameter values by modeling with GLEE & GLaD or with the network presented in Chapter 4.

Name	method	$x_l['']$	$y_l['']$	e_x	e_y	$\theta_E['']$
	GLEE & GLaD	$-0.007^{+0.001}_{-0.001}$	$0.015^{+0.001}_{-0.001}$	$-0.125^{+0.006}_{-0.007}$	$0.224^{+0.008}_{-0.008}$	$1.065^{+0.002}_{-0.002}$
HSCJ090429-010228	ResNet	-0.1 ± 0.1	-0.1 ± 0.2	-0.1 ± 0.1	0.2 ± 0.1	1.2 ± 0.2
	GLEE & GLaD	$-0.041^{+0.001}_{-0.001}$	$-0.107^{+0.001}_{-0.001}$	$0.205^{+0.006}_{-0.006}$	$0.033^{+0.002}_{-0.003}$	$1.232^{+0.002}_{-0.002}$
HSCJ094427-014742	ResNet	0.0 ± 0.1	-0.2 ± 0.1	-0.02 ± 0.08	0.01 ± 0.09	1.0 ± 0.2
	GLEE & GLaD	$-0.005^{+0.001}_{-0.001}$	$0.082^{+0.001}_{-0.001}$	$-0.611^{+0.02}_{-0.09}$	$-0.021^{+0.04}_{-0.08}$	$1.08^{+0.05}_{-0.01}$
HSCJ120623+001507	ResNet	0.02 ± 0.07	0.13 ± 0.08	-0.07 ± 0.09	-0.12 ± 0.09	1.28 ± 0.06
	GLEE & GLaD	$-0.060^{+0.001}_{-0.001}$	$-0.016^{+0.001}_{-0.001}$	$-0.253^{+0.08}_{-0.07}$	$0.15^{+0.04}_{-0.05}$	$1.07^{+0.01}_{-0.02}$
HSCJ121052-011905	ResNet	0.14 ± 0.08	0.07 ± 0.08	-0.19 ± 0.07	0.15 ± 0.08	1.28 ± 0.05
	GLEE & GLaD	$0.029^{+0.001}_{-0.001}$	$-0.095^{+0.001}_{-0.001}$	$-0.15^{+0.003}_{-0.003}$	$0.261^{+0.001}_{-0.001}$	$1.529^{+0.008}_{-0.008}$
HSCJ121504+004726	ResNet	0.0 ± 0.1	0.10 ± 0.09	0.2 ± 0.1	0.2 ± 0.1	0.9 ± 0.4
	GLEE & GLaD	$-0.052^{+0.001}_{-0.001}$	$0.023^{+0.001}_{-0.001}$	$-0.03^{+0.02}_{-0.03}$	$0.028^{+0.03}_{-0.03}$	$1.398^{+0.008}_{-0.007}$
HSCJ124320-004517	ResNet	0.01 ± 0.1	0.1 ± 0.1	0.16 ± 0.07	-0.03 ± 0.08	1.2 ± 0.1
	GLEE & GLaD	$0.034^{+0.001}_{-0.001}$	$0.030^{+0.001}_{-0.001}$	$-0.02^{+0.03}_{-0.02}$	$0.152^{+0.02}_{-0.02}$	$1.506^{+0.004}_{-0.005}$
HSCJ125254+004356	ResNet	-0.41 ± 0.09	0.1 ± 0.1	0.1 ± 0.1	-0.01 ± 0.09	1.88 ± 0.08
	GLEE & GLaD	$-0.040^{+0.001}_{-0.001}$	$0.066^{+0.001}_{-0.001}$	$-0.076^{+0.01}_{-0.02}$	$-0.043^{+0.03}_{-0.03}$	$1.899^{+0.007}_{-0.005}$
HSCJ135138+002839	ResNet	0.13 ± 0.09	0.0 ± 0.1	0.1 ± 0.1	0.0 ± 0.1	2.2 ± 0.1
	GLEE & GLaD	$0.068^{+0.001}_{-0.001}$	$-0.017^{+0.001}_{-0.001}$	$0.039^{+0.004}_{-0.004}$	$0.244^{+0.003}_{-0.003}$	$2.216^{+0.001}_{-0.001}$
HSCJ141136-010215	ResNet	0.2 ± 0.1	-0.02 ± 0.09	0.18 ± 0.08	0.05 ± 0.09	1.5 ± 0.3
	GLEE & GLaD	$-0.056^{+0.001}_{-0.001}$	$0.066^{+0.001}_{-0.001}$	$0.122^{+0.02}_{-0.02}$	$0.227^{+0.02}_{-0.02}$	$1.081^{+0.003}_{-0.004}$
HSCJ141815+015832	ResNet	0.2 ± 0.1	0.1 ± 0.1	-0.08 ± 0.07	0.2 ± 0.08	1.33 ± 0.07
	GLEE & GLaD	$-0.093^{+0.001}_{-0.001}$	$0.062^{+0.001}_{-0.001}$	$-0.178^{+0.05}_{-0.05}$	$-0.098^{+0.02}_{-0.02}$	$1.34^{+0.01}_{-0.01}$
HSCJ142720+001916	ResNet	0.2 ± 0.1	-0.02 ± 0.08	0.2 ± 0.08	0.05 ± 0.09	1.5 ± 0.3
	GLEE & GLaD	$-0.123^{+0.001}_{-0.001}$	$-0.09^{+0.001}_{-0.001}$	$-0.24^{+0.04}_{-0.04}$	$0.19^{+0.02}_{-0.02}$	$1.491^{+0.009}_{-0.01}$

Table 5.3 Continued: Obtained SIE mass parameter values by modeling with GLEE & GLaD or with the network presented in Chapter 4.

Name	method	$x_l['']$	$y_l['']$	e_x	e_y	$\theta_E['']$
HSCJ144132-005358	ResNet	0.05 ± 0.09	-0.10 ± 0.09	0.2 ± 0.1	0.3 ± 0.17	2 ± 1
	GLEE & GLaD	$0.001^{+0.001}_{-0.001}$	$-0.041^{+0.001}_{-0.001}$	$0.32^{+0.01}_{-0.01}$	$0.334^{+0.008}_{-0.008}$	$1.078^{+0.02}_{-0.006}$
HSCJ144320-012537	ResNet	-0.1 ± 0.1	-0.1 ± 0.1	0.1 ± 0.1	0.2 ± 0.1	1.1 ± 0.2
	GLEE & GLaD	$0.034^{+0.001}_{-0.001}$	$-0.027^{+0.001}_{-0.001}$	$0.154^{+0.007}_{-0.007}$	$0.355^{+0.004}_{-0.004}$	$1.206^{+0.001}_{-0.001}$
HSCJ145242+425731	ResNet	-0.06 ± 0.07	-0.01 ± 0.09	-0.1 ± 0.1	0.0 ± 0.1	1.9 ± 0.1
	GLEE & GLaD	$0.060^{+0.001}_{-0.001}$	$0.080^{+0.001}_{-0.001}$	$0.277^{+0.004}_{-0.005}$	$0.274^{+0.01}_{-0.01}$	$1.99^{+0.01}_{-0.01}$
HSCJ150021-004936	ResNet	0.33 ± 0.08	0.12 ± 0.08	0.01 ± 0.08	-0.01 ± 0.08	3 ± 1
	GLEE & GLaD	$0.089^{+0.001}_{-0.001}$	$0.093^{+0.001}_{-0.001}$	$0.124^{+0.007}_{-0.008}$	$-0.296^{+0.004}_{-0.009}$	$3.063^{+0.01}_{-0.008}$
HSCJ150112+422113	ResNet	-0.34 ± 0.08	0.13 ± 0.09	0.1 ± 0.2	-0.1 ± 0.2	1.0 ± 0.2
	GLEE & GLaD	$-0.066^{+0.001}_{-0.001}$	$0.008^{+0.001}_{-0.001}$	$0.489^{+0.009}_{-0.008}$	$0.273^{+0.008}_{-0.008}$	$1.117^{+0.002}_{-0.001}$
HSCJ223733+005015	ResNet	-0.0 ± 0.1	-0.1 ± 0.1	-0.01 ± 0.08	-0.08 ± 0.08	0.9 ± 0.1
	GLEE & GLaD	$0.015^{+0.001}_{-0.001}$	$0.085^{+0.001}_{-0.001}$	$0.189^{+0.002}_{-0.003}$	$0.4^{+0.003}_{-0.003}$	$1.58^{+0.006}_{-0.004}$
HSCJ230335+003703	ResNet	0.05 ± 0.06	0.00 ± 0.07	0.12 ± 0.09	0.04 ± 0.08	1.0 ± 0.1
	GLEE & GLaD	$-0.017^{+0.001}_{-0.001}$	$0.065^{+0.001}_{-0.001}$	$0.007^{+0.002}_{-0.002}$	$-0.002^{+0.001}_{-0.002}$	$1.020^{+0.003}_{-0.003}$
HSCJ230521-000211	ResNet	0.2 ± 0.2	0.1 ± 0.2	0.23 ± 0.07	0.14 ± 0.07	1.74 ± 0.05
	GLEE & GLaD	$0.076^{+0.001}_{-0.001}$	$0.080^{+0.001}_{-0.001}$	$0.44^{+0.007}_{-0.007}$	$-0.076^{+0.004}_{-0.003}$	$1.777^{+0.003}_{-0.003}$
HSCJ233130+003733	ResNet	0.09 ± 0.09	0.16 ± 0.09	-0.08 ± 0.2	0.6 ± 0.1	1.47 ± 0.05
	GLEE & GLaD	$-0.052^{+0.001}_{-0.001}$	$-0.027^{+0.001}_{-0.001}$	$-0.123^{+0.005}_{-0.005}$	$0.336^{+0.003}_{-0.003}$	$1.502^{+0.001}_{-0.001}$
HSCJ233146+013845	ResNet	-0.04 ± 0.09	0.12 ± 0.08	0.06 ± 0.06	-0.02 ± 0.06	1.49 ± 0.03
	GLEE & GLaD	$0.004^{+0.001}_{-0.001}$	$-0.030^{+0.001}_{-0.001}$	$0.067^{+0.008}_{-0.008}$	$-0.105^{+0.005}_{-0.004}$	$1.481^{+0.001}_{-0.001}$

5.4 Traditionally obtained models

We model the 32 lens systems described in Sect. 5.2 using GLEE (Suyu & Halkola, 2010; Suyu et al., 2012) and its extension GLaD (Chirivì et al., 2020), which are both well-tested modeling codes providing among others the profiles and optimization algorithms introduced in Sect. 2.2. As explained in that section as well, this traditional modeling procedure is very time and resource consuming. Especially it requires a lot of input from an user with specific modeling expertise, e.g., to create the required input files, including a configuration file specifying the adopted profiles with the initial starting values, and the optimization details such as the chain length, stepsize, and range for the different sampling methods like MCMC and simulated annealing, the prior ranges on each parameter, and several other details. Each optimization run will lead to an updated configuration file with the newest best set of parameter values. After a possible update with e.g., the prior ranges or the selection of varying parameters which get typically iteratively optimized, a new optimization run is started. This will be repeated until the sampled parameter values stabilize and represent the observation to an acceptable level. Thus, this procedure is a complete iterative process and thus the user input time is relatively high.

Therefore, we develop a code to automate the modeling procedure to minimize the user input time where we adapt partly the code and procedure presented in Ertl et al. (in prep.). The implemented procedure and decision criteria were extensively tested on the presented sample, such that the code is able to model broadly typical galaxy-galaxy lenses from ground based surveys, where most of the lenses are detected. The final procedure and criteria are presented in Sect. 5.4.1.

Since each lensing system is very special and thus require specific treatment in the modeling, the presented automated code will not obtain for every lensing system a fit that represents the observation very well. This is expected because of the huge variety of galaxy light distribution, orientations, LOS objects and other things, but provides at least an initial model for further refinement of the parameter values. For this, we introduce in Sect. 5.4.2 *gleeauto.py*, a code that automated individual optimization steps specified by the user.

The obtained best models are presented in Sect. 5.4.3 where we also discuss details of the code limitations. We finally conclude our traditional modeling in Sect. 5.4.4.

5.4.1 Automated modeling code for galaxy-galaxy lenses

Our developed automated modeling code is divided into four individual parts, where the modeling steps are again internally subdivided but do not require any further user input. The first step is simply preparing the input files which are used for the modeling. In the second step the lens light distribution is modeled following Sect. 2.2.2, where we adopt Sérsic profiles. We assume here the same structural parameters across different filters. In the third step we automated Sect. 2.2.1, which contains a short optimization of the SIE mass parameters based on the source or image positions identified beforehand from the user. In the fourth part the arclight modeling and source light reconstruction is performed (Sect. 2.2.3), where we assume one Sérsic profile to describe the source light distribution. A parameterized profile for the background source, which is supported only by GLaD, is here preferred over a pixelized source SB reconstruction, implemented in GLEE, as we use ground based images that only resolve the main structure of

the arcs. This is the only reason for making use of GLaD, although we do not include dynamical modeling, which is the key part of GLaD. Because of the relative long run-time of the last part, this part can be re-started from several different steps indicated with *m* if the code get aborted or to refine the model iteratively.

The input data with assumed nomenclature, the individual optimization steps and implemented criteria of the modeling routine are presented as a flow diagram in Fig. 5.2 and explained below in detail.

1.) **Preparation of input files**

– Creation of lens and arc masks, shown in Fig. 5.2 as the top-left and top-middle insert of the corresponding box. These masks specify the region to be modeled and can be different for each filter.

– Creation of a region file with the ds9 software[2] (Joye & Mandel, 2003) as shown in the top-right insert of the corresponding box in Fig. 5.2. Here one specifies the cutouts size, the lens center and the lens ellipticity, the image positions, and if needed a region to subtract the image background σ_{bkgr}.

– Renaming of all files according to the assumptions of the modeling code displayed on the bottom of each insert. Both, the alphanumerical ID specifying the lensing system, which is in our example 42, and a name to distinguish between the filters, which is *R* in our example, are chosen by the user.

2.) **Lens light modeling** with GLEE

– Read in the provided files, crop the lens image and error map, subtract the background if specified and save the new image and error map cutout to disk.

– The code creates now automatically the initial GLEE configuration file for the first filter. The starting values for the lens center and ellipticity are determined from the region file provided. We adopt at this stage one Sérsic profile with the following parameters and prior ranges: lens light center coordinates $x_l \in [x_l - 2'', x_l + 2'']$ and $y_l \in [y_l - 2'', y_l + 2'']$, axis ratio $q_{ll} \in [0.3, 1]$, position angle $\theta_{ll} \in [-\pi, +\pi]$, the amplitude $A_{ll} \in [0, 100]$, the effective radius $r_{eff,ll} \in [0.01'', 10'']$, and the Sérsic index $n_{ll} \in [1, 5]$.

– Since the Sérsic amplitude A_{ll} is not known a priori, the code evaluate automatically each order of magnitude between 10^{-5} and 10^5 and select the correct order of magnitude defined through the minimal χ^2. The code updates then the upper limit of the prior range to 100 times the estimated amplitude. If the new parameter range limit is lower than 10, we set it to 10. We refer to this procedure as amplitude test in the remaining chapter.

– If the reduced χ^2, χ^2_{red}, is above 2, optimize the model by running first

– a simulated annealing optimization and

---

[2]https://sites.google.com/cfa.harvard.edu/saoimageds9
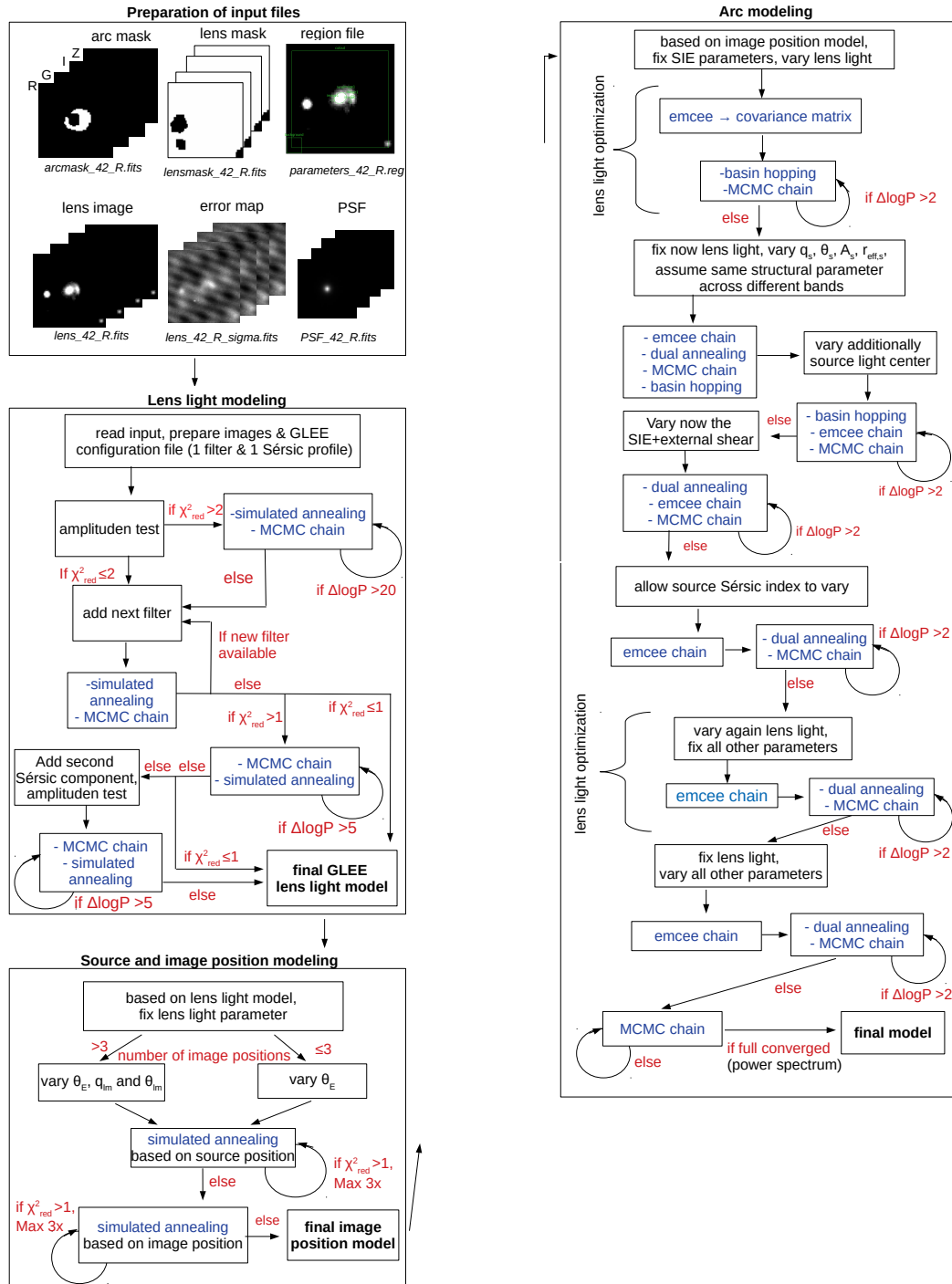
Figure 5.2: Diagram of our automated procedure for galaxy-galaxy lens modeling with GLEE & GLaD.

– then an MCMC chain to use the best model of the chain as new model parameters and also to include a covariance matrix in the next optimization sequence.

Redo both optimization steps until the MCMC chain passes the criterion of $\Delta \log P \leq$ 20 where P is the likelihood probability of the corresponding MCMC chain.

– Add now iteratively the other filters in the order specified by the user. Assume the same structural parameters across the different filters, which means only the amplitude is added as a new parameter.

  – For each new filter run first an amplitude test as described above, and

  – then a simulated annealing optimization,

  – followed by an MCMC chain to continue with the best set of parameter values of that chain.

– After adding all filters and if $\chi^2_{\mathrm{red}} > 1$, optimize all filter simultaneously further by alternating between

  – an MCMC run and

  – a simulated annealing run

  until $\Delta \log P \leq 5$.

– If still $\chi^2_{\mathrm{red}} > 1$, which is normally the case, the code adds a second but concentric Sérsic profile for each filter, assuming again the same structural parameters across the different filters.

– Determine again the order of magnitude of each new amplitude and set the prior range as specified above.

– Optimize the model further by alternating between

  – an MCMC run and

  – a simulated annealing optimization

  until $\Delta \log P \leq 5$.

– Accept this now as final lens light model obtained with GLEE.

3.) **Source and image position model** with GLEE:

– This optimization step of the SIE profile parameters is based on the multiple image positions identified by the user during the preparation stage. The mass parameters get now optimized to reproduce these image positions.

– As starting values for the SIE central coordinates, use the obtained lens light center as lens mass center and keep it fixed for now to reduce the number of free parameters. Assume the axis ratio and position angle from the lens light fit as well, which only vary if four or more image positions are specified to not underconstrain the model. The Einstein radius is estimated from the identified image positions and always allowed to vary. Assume no shear for now to minimize the number of free parameters.

- optimize the model based on the source position with simulated annealing. Perform at least one optimization but maximal three, stopping earlier if $\chi2_{\mathrm{red}} \leq 1$.

- optimize the model based on the image positions with simulated annealing. Perform at least one optimization but maximal three, stopping earlier if $\chi2_{\mathrm{red}} \leq 1$.

4.) **Arc light modeling** with GLEE & GLaD:

- Transfer the best fit values to a GLaD configuration file. Assume the source profile located at the predicted weighted source position of the specified image positions with the current mass model.

- (m=1) Perform first again a quick lens light-only optimization, to reduce the minimal differences in the model raised though differences between GLEE & GLaDin subsampling the PSF and the different usage of the masks[3].

    - For this, run first an emcee chain to obtain a new covariance matrix.

  Alternate then between

    - a basin-hopping iteration and

    - an MCMC chain, to obtain a new covariance matrix but also save the new best set of parameter values from the chain.

  until $\Delta \log P \leq 2$ is achieved in the MCMC chain.

- (m=2) Fix now all lens light parameters to the best values obtained in the previous modeling sequence. Instead, allow now the source light axis ratio $q_s \in [0.5, 1]$, the position angle $\theta_s \in [-\pi, +\pi]$, the amplitude $A_s \in [0, 50]$, and the effective radius $r_{\mathrm{eff,s}} \in [0.01, 10]$ to vary in the specified prior ranges, but assume again the same structural parameters across the different bands. Include from now on in the optimization also the regions specified in the arcmask which was previously excluded to fit only to the light from the lens.

    - Run an emcee chain to update the covariance matrix.

    - Then perform a dual annealing optimization, followed by

    - an MCMC chain and

    - a basin-hopping optimization.

- (m=3) Allow now the coordinates of the source light center $x_s \in [x_s - 1'', x_s + 1'']$ and $y_s \in [y_s - 1'', y_s + 1'']$ to vary and increase the prior range of the source axis ratio to $q_s \in [0.01, 1]$. Optimize now the model until reaching $\Delta \log P \leq 2$ by a repeated sequence of

    - a basin-hopping optimization,

    - an emcee chain to update the covariance matrix (max 10 times in total) and

---

[3]GLEE exclude directly all pixels that are specified in the mask when summing up the χ^2 (compare Eq. (2.40)), while we incorporate in GLaD the masked regions implicitly in the error map through significantly boosting of the uncertainty values such that these contripute effectively not to the χ^2.

- – an MCMC chain to save always the new best set of parameter values and update the covariance matrix.

- – (m=4) After optimizing the lens light and source light, allow now in addition to the source light the lens mass axis ratio $q_{lm} \in [0.3, 1]$, the position angle $\theta_{,lm} \in [-\pi, +\pi]$, and the Einstein radius $\theta_E \in [0.5'', 10'']$ to vary. Moreover, include from now on also an external shear component with $\gamma_{ext} \in [0, 0.2]$ and $\theta_{ext} \in [-\pi, +\pi]$. Optimize until $\Delta \log P \leq 2$ through a repeated sequence of

    - – a dual annealing iteration,

    - – an emcee chain to update the covariance matrix (max 15 times in total), and

    - – an MCMC chain to save the new best set of parameters and update the covariance matrix.

- – (m=5) Vary now additionally the source Sérsic index $n_s \in [0.5, 6]$, which was previously fixed to 3.

    - – Run an emcee chain to obtain a covariance matrix for the new set of varying parameters.

    Optimize then all varying parameters through a repeated sequence consisting of

    - – a dual annealing iteration, and

    - – an MCMC chain to save the best and update the covariance matrix

    until $\Delta \log P \leq 2$ is reached.

- – (m=6) After now all parameters got optimized, refine once more the lens light parameters which were fixed during the last optimization steps. For this, allow now the lens light to vary again, but fix instead all other parameters, i.e the lens mass, external shear, and the source light components.

    - – Run first one emcee chain to obtain a covariance matrix.

    Optimize then until $\Delta \log P \leq 2$ through a repeated sequence of

    - – a dual annealing iteration and

    - – an MCMC chain to save the best values and update the covariance matrix.

- – (m=7) Fix again all lens light parameters and vary again the source light, lens mass and external shear by using the same prior ranges as before, but update the parameter range for the source light center to again $x_s \in [x_s - 1'', x_s + 1'']$ and $y_s \in [y_s - 1'', y_s + 1'']$.

    - – Run first one emcee chain to obtain a new covariance matrix.

    Optimize until $\Delta \log P \leq 2$ through a repeated sequence of

    - – a dual annealing iteration and

    - – an MCMC chain to save the best parameter set and update the covariance matrix.

– (m=8) Double the length of the MCMC chains to $400,000$ and run chains until a chain is fully converged based on the power spectrum. Take always the best set of parameter values of the chain and update the covariance matrix. In case the tenth MCMC chain of this optimization sequence did not converge, the number of sampling steps get increased to $600,000$.

This procedure was developed through extensive testes on all our 32 SuGOHI lenses and is therefore optimized for ground based observations with a parametrized source light distribution. The code is able to predict within few hours the lens light model, within around a minute the source and image position model, and the extended image modeling with source SB reconstruction within around a day. It runs on a single core and launches automatically 60-core parallized jobs for the emcee optimizations. This allows to model uniformly a larger sample of galaxy-galaxy lensed without much user input to provide a basic model of the observations.

5.4.2 Flexible modeling code *gleeauto.py*

Using the automated procedure described in Sect. 5.4.1, we modeled all 32 SuGOHI lenses uniformly. Because every lens system is individual and different from the others, the automated procedure is not working for all of them perfectly. Since the main focus of this work is the comparison between network prediction and conventional methods rather than developing and testing this automation code in detail, we improved several models afterwards by hand. However, the automation code gives at least a very good starting point for further individual optimization with minimal user input time.

For the individual modeling, we developed *gleeauto.py*, a flexible GLEE based code to automated several optimization steps when modeling with GLEE & GLaD. This means one provides as usual a configuration file to the code that specifies the data, the number of profiles, the varying parameters, starting values, the adopted cosmology, and similar other required information. One can then specify a list of available optimization iterations that the code shall sequentially perform without further input from the user. This helps to reduce the user input time while giving the flexibility to assume any setup (e.g., number of filter or profiles, kinematic data, multiplane lensing etc.). Also, saving the best new set of parameters from an MCMC or emcee chain as well as computing the covariance matrix and updating in the configuration file, which is normally done always manually by hand, can be included in the list of tasks.

Since the code does not include any decision criteria as the code presented in Sect. 5.4.1, *gleeauto.py* can be used for any lens system configuration and does not rely on the assumption of galaxy-galaxy lensing. This means, it can be used to model ground based images like from HSC or soon from LSST, but also high-resolution images from space. Moreover, it is independent of the lensing regime, which means it is helpful for modeling any lensing system.

In addition to the sampling opportunities, we added to *gleeauto.py* several other frequently needed tools such as visualization of the obtained fits (compare Fig. A.1-Fig. A.32) with GLEE and GLaD[4], running the amplitude test introduced in Sect. 5.4.1, an update from all linked

---

[4]This plotting tool is adapted from a code provided by Dr. Giulia Chirivì with the GLaD extension (Chirivì et al., 2020).

parameters within the configuration file, and generating masks such as the required arcmask and lensmask[5].

5.4.3 Results and discussion of MCMC modeling

We model each of the presented lens sample with our automated pipeline introduced in Sect. 5.4.1, and depending on the χ^2, MCMC chain convergence, and residuals, improve them further manually where we make extensive use of *gleeauto.py*. Since we are mainly interested in the comparison to the network predictions instead of demonstrating the power of our automated code, we only report the results of the final models.

The obtained median values with 1σ uncertainties computed from our final MCMC chain for the external shear and the SIE parameters are reported in Tab. 5.2 and Tab. 5.3, respectively. Since these quantities are in complex notation in analogy to the network predictions from Sect. 5.3, we also provide the values for these quantities in the typical notation in Tab. A.1 and Tab. A.2, respectively, for more straightforward interpretation, comparison to other publications, and possible follow-up work. While the median values are directly convertible through Section 2.2.1 and Section 3.2.1, this is not that straightforward for the uncertainties. Therefore, we converted first the full MCMC chain and computed from that the median and uncertainty values. We further quote in Tab. A.1 the obtained χ^2 and the χ^2_{red} values, which gives an indication how good the obtained model is. From this list, we see that the χ^2 is only for five of the 32 lenses (HSCJ020141-030946, HSCJ023322-020530, HSCJ135138+002839, HSCJ144132-005358, HSCJ150112+422113, HSCJ230521-000211) above 1.5. This means we can fit all lens systems overall well.

In contrast to the network, we have to model with the traditional method also the lens light and source light, which also influence the quoted χ^2 and χ^2_{red}. The resulting parameter values are given in the appendix A as well. The best fitted values for the Sérsic profile amplitudes for the source light are listed in Tab. A.3, the corresponding structural parameters in Tab. A.4, and the lens light in Tab. A.5. We further show all 32 final models in the appendix A as Fig. A.1 to Fig. A.32. Each plot shows from left to right the observed image, the model, and the normalized residuals. The four rows correspond to the four different filters in the order *g*, *r*, *i*, and *z*.

Through the modeling, we made several general observations which we discussed in the following.

In several models, for instance HSCJ020955-024442, HSCJ023217-021703, HSCJ124320-004517, HSCJ125254+004356, HSCJ135138+002839, HSCJ145242+425731, the model seem to be sharper and contain more details than the observed image. Nevertheless, we use for all models the provided PSF files and can still fit the observations overall well.

In general, the lens center is in all models very well constrained ($1\sigma \leq 0.001$). The offsets with respect to the image center are also relative small, only seven systems (HSCJ022610-042011, HSCJ085046+003905, HSCJ090429-010228, HSCJ141815+015832, HSCJ142720+001916, HSCJ150021-004936, HSCJ223733+005015) have a difference larger than halve a pixel (0.084″) and not larger than one pixel (0.168″). The source center is of course not that well constrained, but surprisingly well with typically $1\sigma < 0.1″$.

---

[5]This tool was written by Dr. Yiping Shu (Shu et al., 2016b).

Several sources seem to be very elongated; twelve out of 32 have an axis ratio below 0.2 and 24 out of 32 below 0.5. This could be because of the lens search strategy from SuGOHI as they search specifically for blue arcs (compare Sect. 1.2), which come typically from star-forming galaxies. Since the arcs must be bright, galaxies with higher surface area are more likely to be detected sources. That are then typically edge-on galaxies, i.e. have very low q. This is in agreement with a relative low Sérsic index; 17 out of 32 systems have $n_s < 1$.

The estimated Einstein radius is, apart from lens HSCJ015618-010747 with $\theta_E = 0.99''$, always above $1''$. Lens system HSCJ150021-004936 has the largest image separation with $3.063''$, followed by system HSCJ135138+002839 with $2.216''$. All other systems have an Einstein radius between $1''$ and $2''$.

In the following we discuss aspects of individual lens systems that were not mentioned above.

- **HSCJ023322-020530**: This lens system has one very bright source, lensed into two images. Since those two lensed images are extremely compact, it is very hard to model them with GLaD, which is optimal for modeling extended sources but not point like objects. Additionally, it seems that the PSF is not perfectly symmetric, leading to distortions in the model. Since there is another fainter source lensed into extended arcs, we included manually here a second source at the same redshift. Given the strong residuals, also a manually optimization is very difficult such that the best model has still a relative high χ^2_{red} with 1.87. We tried several different options such as including additional profiles or relaxing assumptions on the structural parameters but obtained no notable improvement justifying to increase the complexity in the model.

- **HSCJ090429-010228**: Lens system HSCJ090429-010228 is analyzed in detail in Jaelani et al. (2020b) and it appears like it has a point like source, like an AGN. However, based on Jaelani et al. (2020b), it is a compact Lyman alpha break galaxy.

- **HSCJ135138+002839**: This lens system has overall relative low residuals, resulting in a good χ^2. The somewhat higher reduced χ^2 of 1.65 is related to the relative large part of the image in the south-west (bottom-right) that is masked out due to luminous objects. This reduces the the number modeled pixels and thus the number of degrees of freedom notable. Although those masked pixel are not taken into account when computing the χ^2, the reduction in the number of degrees of freedom increases effectively the χ^2_{red} which need to be taken into account when comparing to the χ^2_{red} of other lens systems. Nonetheless, it shows that the model is not perfectly representing the observations. There are two additional areas in the image masked, one on the south-east of the lens and the other on the north-west. Given the orientation, shape and color (compare Fig.5.1, this could be from a second source behind the lens. To confirm this, either further multiplane model analysis, which is however beyond of the scope of this work, or spectroscopic observations would be needed.

- **HSCJ141815+015832**: Two images (g and r band) of this lensing system are unfortunately slightly corrupted. Nonetheless, we modeled this system masking out the specific regions. We find that the remaining lensing information is sill enough to constrain the parameter values and provide a reasonable fit, most likely as we model the different filter

simultaneously and assume to have the same structural parameters for the lens light. This is, however, not obvious as it belongs the most relevant filters and both masked areas go directly through the lens and arcs.

- **HSCJ150112+422113**: The best model of this system represent the observed structure in lens and arcs relatively well, but shows still notable residuals, both in the lens light as well as in the arcs, and thus result in a higher reduced χ^2 of 1.63. To improve the model a flexible lens and source center across the different bands might help.

- **HSCJ230521-000211**: The final model of this lens system represent the observed structure in lens and arcs, but has slight differences especially also in the positions. Therefore, a different lens center for each band might improve the fit slightly but is beyond of the uniform modeling.

5.4.4 Conclusion of MCMC modeling

For modeling our sample of 32 SuGOHI lenses, we developed an automated modeling pipeline presented in Sect.5.4.1. This automation helps to reduce the user input drastically, as the remaining parts are only to prepare the input region file and masks. These steps are hard to automate while not including the possibility of masking wrong parts or identifying wrong objects as lensed images. Unfortunately, this limit the applicability to only smaller samples of around dozens of lenses but is not scalable to hundreds or thousands of lenses. The automation reduces significantly the computational time compared to fully manual modeling, but remains at few days in total, which limit also the applicability to moderate sample sizes. Since the main code is not parallelized and require not much memory, one can, however, easily model many lensing systems at the same time and thus increase effectively the amount of modeled lenses per day.

As expected, such automated modeling codes to uniformly model larger samples provide typically a good to reasonable fit depending on the complexity of the individual lens system. It provides a basic model of the full sample, which can then be used for individual follow up modeling if needed. For this we present *gleeauto.py*, a python code that allows e.g., running several different optimization steps with GLEE & GLaD through one submission without any assumption of the lensing system setup.

With the combination of both pipelines, we are able to obtain good fits for all 32 lens systems and reconstruct the SB of the background source which is described by a single Sérsic profile. The lens light is described by two concentric Sérsic profiles, where we additionally assume the same structural parameters across the different filters. The lens mass is parametrized through a SIE profile, where we include also an external shear.

5.5 Comparison and Discussion

After modeling our comparison sample with our ResNet from Chapter 4 and with GLEE & GLaD as the traditional method, we compare now directly the obtained SIE+γ_{ext} parameter values. For this, we show them in Fig. 5.3 as histograms (left) and plotted against each other (right),

with the traditional obtained values on the *x* axis and network predictions on the *y* axis. We further show the difference between the traditionally and network based values as a histogram in Fig. 5.4.
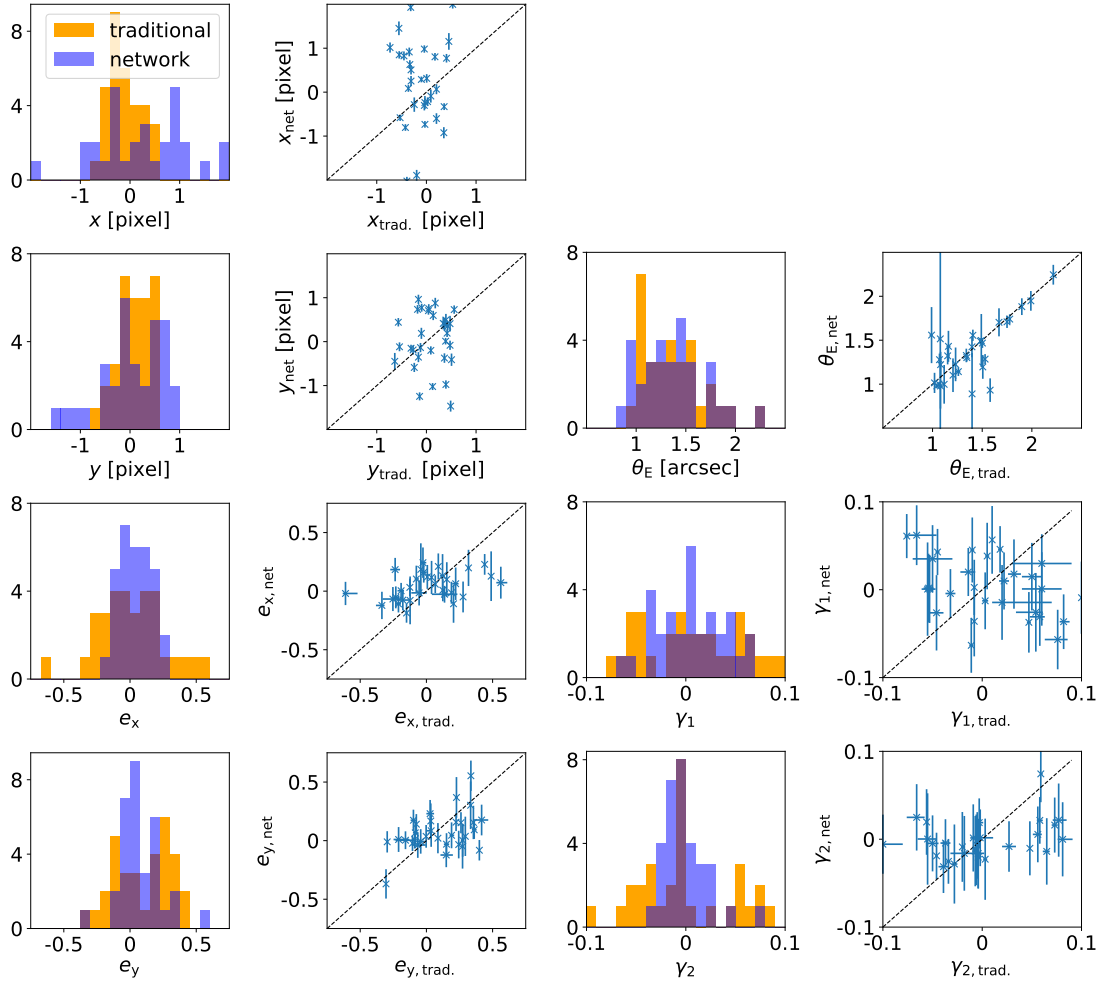
Figure 5.3: Comparison of the SIE+γ_{ext} values obtained with GLEE & GLaD as the traditional method (orange histogram) and our ResNet (blue histogram). We show for each parameter a histogram (left) and also plot them agains each other (right).

As we can see from Fig. 5.3, the Einstein radius is very well constrained through the traditional procedure because we use at the beginning the image positions as constrains. These image positions are constraining already good the Einstein radius, which is then further refined through the extended image modeling. The network has more difficulties, especially on faint images, but performs overall very well. Especially the systems with wide image separation matches perfectly the predictions from the traditional modeling. This is not obvious given the performance on the test set for the under represented systems with wide-image separations as we see in Fig. 4.5.
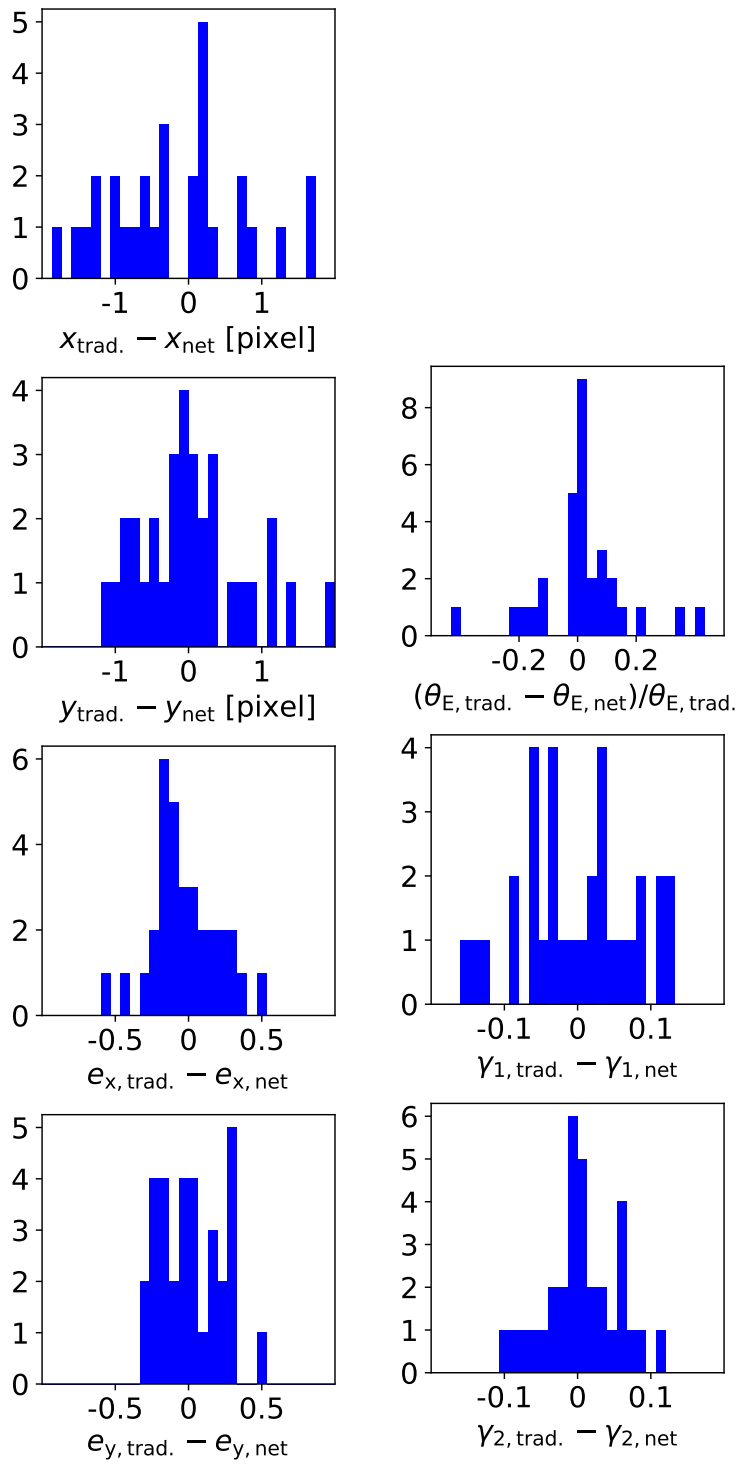
Figure 5.4: Difference of the SIE+γ_{ext} values obtained with GLEE & GLaD as the traditional method and our ResNet.

Also for the systems with Einstein radius between $\sim 1''$ and $\sim 1.5''$ the network performs good, although a few are out of the 1σ range.

It is interesting to see that the network assigns to two systems very large error bars; lens system HSCJ150021-004936 got a value of 3 ± 1 and lens HSCJ144132-005358 a value of 2 ± 1 from the ResNet. When we look at their color image in Fig 5.1, we see essentially no arcs, explaining the difficulty for the network and reason for the large uncertainties. Also, if we look at each filter alone in Fig. A.23 and Fig. A.26, the arcs are very faint. Both systems have only a measured lens redshift but no source redshift. Lens system HSCJ150021-004936 was found through the Space Warps citizen science project but missed by YATTALENS (see also Sect. 1.2.3) while system HSCJ144132-005358 where missed by both Space Warps and YATTALENS and just noticed serendipitously (Sonnenfeld et al., 2020). The non-detection from YATTALENS, a method that looks for arc-like features and classify the images based on a fit, demonstrates the difficulty to model them in an automated way. Therefore, it is understandable that the network has difficulties on these Einstein radii and predicts an extremely large uncertainty. This is actually a very good sign, as the network seems to be able to adjust the uncertainty prediction.

The network assigned also a relative large error to system HSCJ121504+004726 with 0.9 ± 0.4, while we obtained with GLEE & GLaD a very precise estimate of $1.398^{+0.008}_{-0.007}$. The reason for creating troubles for the network is probably the asymmetric alignment of the images which we can see in Fig. 5.1 and also Fig. A.16. The brightest image on the south-west side is much further away than the fainter counter image on the east side, possibly confusing the network.

The coordinates of the lens center x and y are very well constrained by both methods. The traditional modeling predicts a lens center very close to the image center, i.e. within ±1 pixel. Here we have to remember that we assume the light center to be coincident with the lens mass center, which is not the case for the network. Since the lens light has a relative large influence on the χ^2 and thus on the lens center, the predicted value will be highly influenced by the lens light. A possible offset to the true mass center can be compensated through a change in the external shear. This could be a reason that the network predicts for several systems a larger offset to the image center, especially for the x coordinate. The largest offset is $-0.34''$ for system HSCJ150112+422113 correspond to around two pixels, where the ResNet performs very well on the test data (see Fig. 4.5). This lens system has also larger residuals from the manual modeling, resulting in an χ^2_{red} of 1.63, such that also the parameter values obtained with GLEE & GLaD are not that trustworthy. The fact that we can model all lenses well with GLEE & GLaD by assuming a coincidence lens light and mass center, implies that we could also adopt this assumption when generating our network training data. Moreover, if we assume the traditional obtained value to be more accurate, a lens center offset of ±1 pixel is enough instead of an offset of ±3 pixels currently used in the simulations. This would simplify slightly the task for the network and thus increase the performance of the network, including also the other parameters. On the other hand, most lens systems got also from the traditional procedure a slight offset to the cutout center. Therefore, it is good to include a lens center and to predict five parameters for the SIE profile, instead of assuming that the lens light and mass center fall both exactly on the cutout center and thus predicting only three parameters (ellipticity and Einstein radius) as done in other modeling networks (Hezaveh et al., 2017; Perreault Levasseur et al., 2017; Pearson et al., 2019, 2021). This implies a more complex model and task for the network, increasing possibly the

inaccuracies and uncertainties on the prediction. This is worth it although the influence of the lens center, if within ~ 1 pixel, is expected to be minor.

The ellipticity instead shows better agreement as most of the 1σ bars, which are typically much larger predicted by the network, reach the 1:1 line for e_y. In general, the network predicts values closer to zero than the values obtained with the traditional modeling, which was expected from the performance on the test data (compare Fig. 4.5). This is most likely a result of having nearly two orders of magnitude more systems with ellipticity ~ 0 than ~ ±0.5. Since the ellipticity is set by real observations, a more flatter distribution, which would lead to an improvement on the currently underrepresented values, is difficult to achieve. Given this, it is good to see that the network can predict, for system HSCJ144132-005358, an ellipticity component of 0.3 ± 0.17 that matches perfectly the value $0.334^{+0.008}_{-0.008}$ obtained with GLEE & GLaD.

The external shear is as expected very difficult to estimate. This is especially valid for the ResNet, but also the traditionally obtained error bars for γ_1 are often relatively large indicating the difficulty to constrain that parameter. For γ_2 we find a good cluster of systems that got ~ 0 assigned from both methods.

In general, it is interesting and surprising that the values often matches better for one axis than the other one, although there is no real dependency. The histograms of the difference in Fig. 5.4 are much peakier, indicating better agreement between the two methods, for y and γ_2 compared to x and γ_1, respectively.

While the conventional method is considered to give better predictions in general, few of those models show residuals in their fit as we can see from Fig. A.1 - Fig. A.32, resulting in an higher $\chi^2_{\rm red}$ as noted in Sect. 5.4.3. For instance, lens systems HSCJ023322-020530 and HSCJ150112+422113 remained with a $\chi^2_{\rm red}$ of 1.87 and 1.63, respectively, for which the network predict very similar Einstein radii but quite different lens center coordinates and ellipticities as well as a different external shear, possibly to some extend because the traditional model is not that accurate. System HSCJ023322-020530 was difficult to model with GLEE & GLaD, as it shows two very bright objects, for which we had to adopt at least a second Sérsic profile to describe it with acceptable accuracy. Moreover, it shows an extended, relative faint arc which seems to be offset to the source giving the bright images, such that we modeled those arcs with a separate Sérsic component. Nonetheless, it resulted in visible residuals (see Fig. A.9) and higher $\chi^2_{\rm red}$ as noted above.

Special consideration is also required for system HSCJ141815+015832, as it has corrupted data which we try to avoid in the training of the network. As we noted already in Sect. 5.4.3, the traditional modeling worked quite well regardless of the missing data. This surprisingly holds also for the network in general. The lens center is not in agreement within 1σ, but the ellipticity, the Einstein radius, and the external shear matches very good without larger uncertainties than other systems. This demonstrates that our network is able to handle such cases it was not trained on.

5.6 Summary and conclusion

In this chapter, we compared the predictions of our residual neural network from Chapter 4 to the SIE+$\gamma_{\rm ext}$ parameter values obtained through modeling with GLEE & GLaD. For this

comparison we selected known galaxy-galaxy lenses detected in HSC, as the network is trained for these kind of systems and for this image quality. This resulted in a sample of 32 grad-A lenses, which we presented in Sect. 5.2. We then apply our trained network to that sample, and report the obtained values with 1σ uncertainties in Tab. 5.2 and Tab. 5.3.

We further model the full sample of 32 lenses with GLEE & GLaD, a software based on Basian optimizations algorithms such as simulated annealing and MCMC sampling and thus referred to as traditional, non-machine learning technique. Because of the iterative sampling, this procedure is very time and resource consuming. To minimize the user input time, we automated most of the modeling steps and developed a dedicated procedure to model galaxy-galaxy strong lensing systems from ground based surveys like those in our comparison sample. Because this code is specifically optimized for ground based observations, we adopted a parameterized source SB reconstruction rather than a pixelated reconstruction. This means we adopt one Sérsic profile to describe the light distribution of the background source.

Since each lens is unique especially also because of the environment, the presented uniform modeling sequence does not produce a perfect fit for all lenses. Because we are mainly interested in the comparison, and thus want to have very good models for each individual lens, we further refine some of the models by manually modeling. For this, we develop *gleeauto.py*, a software package that accepts an configuration file for GLEE & GLaD and a list of optimization algorithms. These specified optimization steps are then performed directly after each other without any further input of the user.

With both codes, we were able to model all 32 lens systems to an acceptable quality and in an acceptable amount of time, allowing us to compare directly the parameter estimates. We find in general very good agreement for the Einstein radius, especially for the wide-separation systems ($\theta_E \geq 1.5''$). This is not completely expected given the performance on the test set (compare Fig. 4.5) because of the under-representation of these systems in the training set. It is very interesting and good to see that the network predicts high uncertainties for the two systems HSCJ150021-004936 and HSCJ144132-005358 where the arcs are extremely faint. The predicted Einstein radii from the traditional modeling is comparably very good constrained, which comes at least partly from using our visually identified image positions as constrains to get a first estimate.

For the lens center, all values predicted through the traditional modeling procedure are within ±1 pixel with respect to the cutout center, while the network predicts for some systems larger offsets. This could partly be a result from our assumption of a coincidence between lens light and mass center for the traditional modeling, which is not the case when generating our training data. The ellipticity is relative well constrained by both techniques, while the network tend to predict values closer to zero compared to GLEE & GLaD. This is in agreement with the network performance on the test set and a result of a realistic, but non-uniform distribution in the training sample. As expected, the external shear is not well predicted by the network, resulting in the prediction of larger uncertainties. Nonetheless, the performance of the network is overall very good, especially when taking into account of the user in put time and final computational time. We were able to predict all seven SIE+γ_{ext} values for the full sample within a fraction of a minute, while the traditional modeling, even with our automated code, requires more than one day per lens in addition to possible follow-up modeling.

In total, we confirm with our comparison that the network preforms similar well on real lenses than on our test set. This demonstrates that our mocks are indeed very realistic and that we can expect similar performance on a large sample of hundreds to thousands systems, which can be modeled easily with our network. This would allow a detailed statistical analysis of lens mass properties. In contrast to that, we are able to model a sample of dozens of lenses with our automated pipeline to better accuracy and we can also confirm the quality of the fit in terms of a χ^2 which is not possible for the network output. The *gleeauto.py* code enables us to refine further the models obtained with our fully automated procedure for galaxy-galaxy lenses or also other dedicated automated modeling codes (e.g., Ertl et al., in prep.; Hezaveh et al., 2017; Nightingale et al., 2018, 2021a,b; Pearson et al., 2019, 2021; Perreault Levasseur et al., 2017). The combination of all three codes enable us to handle all different sample sizes of lenses, and thus bring us a huge step forward in handling the new detected lenses in current (e.g., Cañameras et al., 2020, 2021; Cañameras et al., in prep.; Jaelani et al., in prep.; Knabel et al., 2020; Rojas et al., 2021; Savary et al., 2021; Shu et al., in prep.) and upcoming wide-field imaging surveys such as HSC, LSST and Euclid.

6 Photometric redshift estimation with a convolutional neural network: NetZ

Summary

Redshifts of astrophysical objects such as galaxies are a key quantity required for nearly all astrophysical studies. This redshift can be determined precisely through analysis of the corresponding spectrum. Since those objects are first detected trough imaging surveys, this would require additional resources such as time of a telescope and an expert to analyze the spectrum. Given the past and current wide-field imaging surveys, spectroscopic redshifts are only available for a very small fraction. Therefore, several photometric redshift techniques were developed in the past decades. Such methods will be also crucial for forthcoming surveys such as LSST, detecting billions of galaxies that need redshifts assigned. Therefore, we developed NetZ, a machine learning based photo-z method published in Schuldt et al., A&A 651, A55, 2021, which is reproduced in this chapter of the thesis.

While most of the other photo-z methods use photometric quantities such as color-magnitude or size-compactness measurements that are extracted from the image and thus often limited to a narrow redshift range (e.g., $z \leq 1$), we use directly the images as input to the network NetZ. To demonstrate this possibility, we train a CNN on HSC images using all five available filter g, r, i, z and y to provide color information to the network. As ground truth are either spectroscopic redshifts or ~ 30 band photometric redshifts used that pass our criteria listed in Sect.6.2. In detail, we train three different networks; first, NetZ$_{\mathrm{main}}$, a network trained on the full redshift range between 0 and ~ 4 of our data set, second, NetZ$_{\mathrm{LRG}}$, a network trained on only LRGs, and third, NetZ$_{\mathrm{lowz}}$, a network trained only in the range $0 < z \leq 1$ as many other photo-z networks. We compare the precision and accuracy for all three networks in detail to our reference redshifts, and also compare NetZ$_{\mathrm{main}}$ to another photo-z method Direct Empirical Photometric (DEmP) (Hsieh & Yee, 2014), that was specifically run on our data set to obtain a fair comparison. With this comparison, we find overall very good performance of NetZ$_{\mathrm{main}}$ that is comparable to DEmP. One of the main advantages of NetZ is the opportunity of simple data augmentation by rotating/mirroring/flipping the image. By using this kind of data augmentation to increase the fraction of photo-z galaxies such that our training sample result in a more equally distributed sample, we obtain notable better performance in the high-z range, while the performance in the lower range remains nearly stable. We finally run NetZ$_{\mathrm{main}}$ on more than 34 million galaxies and publish our photometric redshifts.

Author contribution

I contributed the main driving force to this project. I have developed the network code and done all corresponding training and testing of the networks. I have created all figures included in this publication and I am also the main author of it.

Photometric redshift estimation with a convolutional neural network: NetZ

S. Schuldt, S. H. Suyu, R. Cañameras, S. Taubenberger, T. Meinhardt, L. Leal-Taixé, and B. C. Hsieh

ABSTRACT

Galaxy redshifts are a key characteristic for nearly all extragalactic studies. Since spectroscopic redshifts require additional telescope and human resources, millions of galaxies are known without spectroscopic redshifts. Therefore, it is crucial to have methods for estimating the redshift of a galaxy based on its photometric properties, the so-called photo-z. We have developed NetZ, a new method using a convolutional neural network (CNN) to predict the photo-z based on galaxy images, in contrast to previous methods that often used only the integrated photometry of galaxies without their images. We use data from the Hyper Suprime-Cam Subaru Strategic Program (HSC SSP) in five different filters as the training data. The network over the whole redshift range between 0 and 4 performs well overall and especially in the high-z range, where it fares better than other methods on the same data. We obtained a precision $|z_{\mathrm{pred}} - z_{\mathrm{ref}}|$ of $\sigma = 0.12$ (68% confidence interval) with a CNN working for all galaxy types averaged over all galaxies in the redshift range of 0 to ~4. We carried out a comparison with a network trained on point-like sources, highlighting the importance of morphological information for our redshift estimation. By limiting the scope to smaller redshift ranges or to luminous red galaxies (LRGs), we find a further notable improvement. We have published more than 34 million new photo-z values predicted with NetZ. This shows that the new method is very simple and swift in application, and, importantly, it covers a wide redshift range that is limited only by the available training data. It is broadly applicable, particularly with regard to upcoming surveys such as the Rubin Observatory Legacy Survey of Space and Time, which will provide images of billions of galaxies with similar image quality as HSC. Our HSC photo-z estimates are also beneficial to the Euclid survey, given the overlap in the footprints of the HSC and Euclid.

6.1 Introduction

Past imaging surveys have detected billions of galaxies over the sky, a number that will grow sub-
stantially with forthcoming wide-field surveys, such as the Rubin Observatory Legacy Survey of
Space and Time (LSST). In most applications for which galaxies are used, redshifts are needed,
but spectroscopic redshifts are available only for a small fraction of them. Therefore photometric
redshift techniques (hereafter photo-z, see Hildebrandt et al., 2010, and references therein) were
developed and improved over the last decades (e.g., Coupon et al., 2009; Hildebrandt et al., 2008,
2012; Dahlen et al., 2013; Bonnett et al., 2016; Tanaka et al., 2018). Typically, photometry in
multiple wavelength bands has been used to minimize the difference between spectroscopically
confirmed redshifts and the predicted photometric redshifts.

Today, there are two main families of photo-z methods, namely: template fitting and machine
learning (ML) methods. They are complementary to one another and both are capable of pre-
dicting very precise photo-z. Template fitting codes (e.g., Arnouts et al., 1999; Bolzonella et al.,
2000; Feldmann et al., 2006; Brammer et al., 2008; Duncan et al., 2018b) are mainly based on
galaxy spectral energy distribution (SED) template libraries. This method is physically moti-
vated and well studied thus far. The templates are used to match the observed colors with the
predicted ones (via the so-called nearest neighbor algorithms). Such an approach represents the
opportunity to provide photo-z estimates in regions of color-magnitude space where no refer-
ence redshifts are available. Additionally, ML provides another approach to get very precise
and fast photo-z estimates (e.g., Tagliaferri et al., 2003; Collister & Lahav, 2004; Lima et al.,
2008; Wolf, 2009; Carliles et al., 2010; Singal et al., 2011; Hoyle, 2016; Tanaka et al., 2018;
Bonnett, 2015; D'Isanto & Polsterer, 2018; Eriksen et al., 2020; Schmidt et al., 2020). The main
requirement is a training sample with known (i.e., spectroscopic or very good photo-z) reference
redshifts, which should match the expected redshift distribution. Depending on the network ar-
chitecture, ML codes generally look for specific features in the training sample and try to extract
the important information. So far, most algorithms are based on photometric parameters like
color-magnitude measurements or also size-compactness measurements and often limited to a
narrow redshift range, for example, up to $z = 1$ (e.g., Bonnett, 2015; Hoyle, 2016; Sadeh et al.,
2016; Almosallam et al., 2016b; Pasquet-Itam & Pasquet, 2018; Pasquet et al., 2019; Eriksen
et al., 2020; Campagne, 2020).

Based upon the success of CNNs in image processing, we move on to our investigation of
a network that estimates photo-z based directly on images of galaxies. This is similar to the
work done by Hoyle (2016), where images of galaxies are converted into magnitude images and
pixel color maps to feed the architecture, however, our network accepts the images directly as
observed. Moreover, while Hoyle (2016) used a classification network whereby the galaxies are
sorted into redshift bins between 0 and 1, we use a regression network. This means our network
predicts one specific number for the galaxy redshift. The work presented by D'Isanto & Polsterer
(2018) and Pasquet et al. (2019) explores both networks with CNN layers on SDSS galaxies
to obtain a probability density function (PDF). D'Isanto & Polsterer (2018) tested networks for
either quasars or galaxies as well as a combination of stars, quasars, and galaxies. For the galaxy
sample, they limited their study to $0 < z < 1$, although most galaxies are at the lower end, such
that Pasquet et al. (2019) directly limit the range up to $z = 0.4$. In comparison to those two
networks, we have many more galaxies with higher redshifts ($z \sim 1 - 3$) and thus we do not set

limits on the redshift range for the purpose of obtaining a more powerful network that is directly applicable to the expected redshift range covered by LSST. Based on the available reference redshifts, we tested the performance up to a redshift of 4. Since we provide images of different filters, our CNN is able to extract the color and magnitude parameters internally and output a photo-z value at the end. It is trained on images observed in five different filters, specifically on Hyper Suprime-Cam Subaru Strategic Program (HSC SSP, hereafter HSC; Aihara et al., 2018a) *grizy* images of galaxies with known spectroscopic or reliable ~30-band photometric redshifts.

The outline of the paper is as follows. In Sect. 6.2, we describe the training data we applied and we give a short introduction and overview of the network architecture we used in Sect. 6.3. Our main network, NetZ$_{main}$, is presented in Sect. 6.4 and we compare our results to other model techniques in Sect. 6.5. We show, in Sect. 6.6, our results based on the network NetZ$_{LRG}$, which is specialized for Luminous Red Galaxies (LRGs) and NetZ$_{lowz}$, which is specialized for the low redshift range. We summarize our results in Sect. 6.7.

6.2 Training data

We use images from PDR2 of the HSC-SSP[1] survey (Aihara et al., 2019) for the training of the CNN. The HSC is a wide-field optical camera with a field of view of 1.8 square degrees (1.5 degree in diameter) installed at the 8.2m Subaru Telescope. The data release covers over 300 square degrees of the night sky in five optical filters known as *grizy*. The exposure time is 10 minutes for the filters *g* and *r* and 20 minutes for *i*, *z*, and *y*, yielding limiting magnitudes of around 26. The pixel size is 0.168″, such that our cutouts with 64 × 64 pixels result in images of around 10″ × 10″. The median seeing in the i-band is 0.6″.

The catalog of all available galaxies from HSC PDR2 in the wide area that pass the following criteria:

- {*grizy*}_cmodel_flux_flag is *False*

- {*grizy*}_pixelflags_edge is *False*

- {*grizy*}_pixelflags_interpolatedcenter is *False*

- {*grizy*}_pixelflags_saturatedcenter is *False*

- {*grizy*}_pixelflags_crcenter is *False*

- {*grizy*}_pixelflags_bad is *False*

- {*grizy*}_sdsscentroid_flag is *False*

includes around 190 Million galaxies and is represented by a green box in Figure 6.1. The corresponding HSC images can be used as input data for the network NetZ.

As ground truth, we use the spectroscopic redshifts provided by the HSC team, which is a collection from various spectroscopic surveys (zCOSMOS DR3 (Lilly et al., 2009), UDSz (Bradshaw et al., 2013; McLure et al., 2012), 3D-HST (Skelton et al., 2014; Momcheva et al.,

---

[1]HSC webpage: https://hsc-release.mtk.nao.ac.jp/doc/
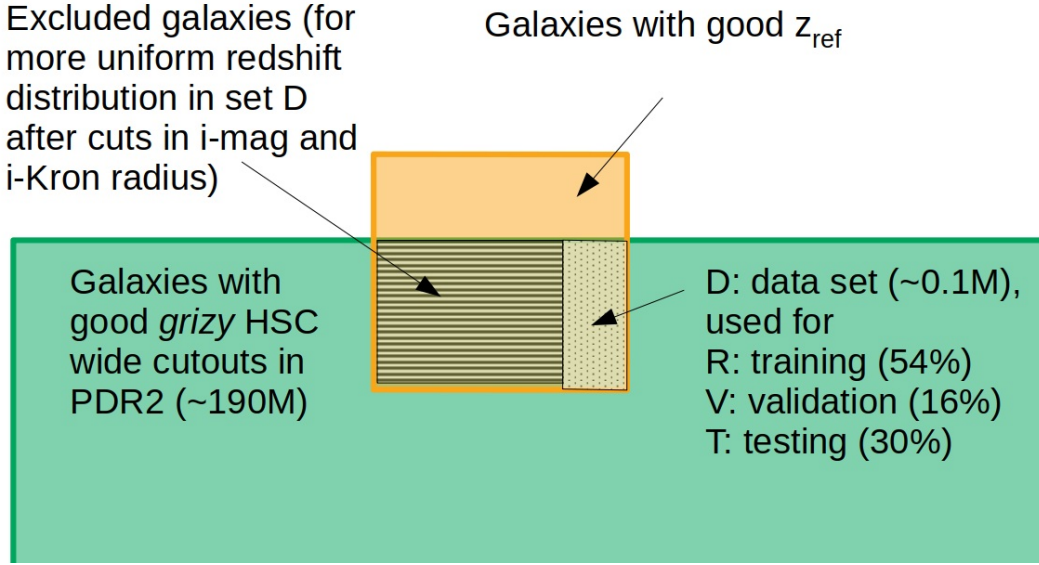
Figure 6.1: Sketch of the available data and the intersection of the data D (dotted) used for training (R), validation (V), and testing (T) of the main network NetZ$_{main}$, as presented in Sect. 6.4.

2016), VVDS (Le Fèvre et al., 2013), VIPERS PDR1 (Garilli et al., 2014), SDSS DR14 (Alam et al., 2015), GAMA DR2 (Liske et al., 2015), DEEP3 (Davis et al., 2003; Newman et al., 2013), PRIMUS DR1 (Coil et al., 2011; Cool et al., 2013)). Since we aim to obtain a network that is applicable to all morphological types, the above list does not include spectroscopic surveys that are most strongly biased towards specific galaxy types of similar morphology. Specifically, we do not consider objects from SDSS BOSS/eBOSS to train our main network NetZ$_{main}$ as those surveys explicitly target LRGs at $z<1$ and known quasars. We do consider training exclusively on LRGs in our separate network NetZ$_{LRG}$. Furthermore, we do not include the WiggleZ catalog (Drinkwater et al., 2010), which targets UV bright emission line galaxies and which would further steepen the redshift distribution of the training set at low-redshift (see below). Despite unavoidable biases due to the selection function of each survey, we expect that this collection of spectroscopically-confirmed redshifts has limited morphological pre-selection. This spec-z sample is cleaned with the following criteria:

- source type is *GALAXY* or *LRG*

- $z > 0$

- $z \neq 9.9999$[2]

- $0 < z_{err} < 1$

---

[2]This is the upper limit of the catalog and thus treated as no spec-z available, i.e. excluded

- the galaxy identification number (*ID*) is unique

- specz_flag_homogeneous is *False* (homogenized spec-*z* flag from HSC team)

This spec-*z* sample is used in combination with COSMOS2015 (Laigle et al., 2018), a photo-*z* catalog of the COSMOS field based on around 30 available filters, where we enforce the following criteria:

- flag_capak is *False*

- type = 0 (only galaxies)

- χ^2(gal) < χ^2(star) and χ^2(gal)/Nbands < 5 (fits are reasonable and better than stellar alternatives)

- ZP_2< 0 (no secondary peak)

- $\log(M_\star)$ > 7.5 (stellar mass successfully recovered)

- 0 < *z* < 9

- $\max(z_{84} - z_{50}, z_{50} - z_{16})$ < 0.05(1 + *z*) (1σ-redshift dispersion < 5%)

This selection primarily follows the criteria from the other HSC photo-*z* methods (Tanaka et al., 2018; Nishizawa et al., 2020). We then select galaxies with *i*-band magnitudes brighter than 25 mag and a Kron radius larger than 0.8″ in the *i* band. The limit on the Kron radius is chosen with the aim of obtaining a set that best represents the sample that we are applying NetZ to. These criteria ensure that we have accurate and reliable reference redshifts for our training, validation, and testing. While such criteria could lead to potential selection bias in the objects, our combination of photo-*z* and spec-*z* helps mitigate selection biases. Furthermore, we verify that the color space spanned by the objects from the cleaned catalog is similar to that of the objects in the HSC PDR2 with a Kron radius above 0.8". This allows us to apply the trained NetZ based on the reference redshifts to those HSC PDR2 galaxies. The cleaned catalog used for training, validation, and testing is shown as a yellow box in Figure 6.1, and the overlap with available good HSC images in all five filters (green box) contains 406,540 galaxies.

Based on various tests during the development stage, we found a significant improvement by masking the background and surrounding objects next to the galaxy of interest with the source extractor (Bertin & Arnouts, 1996) before feeding them into the CNN. As a boundary, we use the 3σ level of the background. Fully deblended neighboring objects in the field can be excluded by requesting the object center to be within five pixels of the image center. With this method, we keep only the central galaxy(ies) in the image cutout. At the end we convolve the extracted image with a gaussian kernel of size 3×3 pixels and a width of 1.5 pixels to smooth out the boundaries very slightly. We show color images of random galaxies from our NetZ$_{main}$ test sample in Figure 6.2 as examples. The masked background is shown in blue and has pixel values set to zeros in the image. We provide the reference redshift, which can be either a spectroscopic or photometric redshift, and our predicted redshift at the top of each image. The HSC identification number is given in the bottom of each image. From these examples, it can be seen

that the extraction procedure works well overall, but has its limitations; for instance, the first image of the second row is partly truncated because of a masked bright neighbouring object. Since this procedure is aimed at masking only obvious and well deblended companions, while beeing purposely conservative and retaining blended galaxies. Therefore, the third image in the first row in Figure 6.2 is expected.
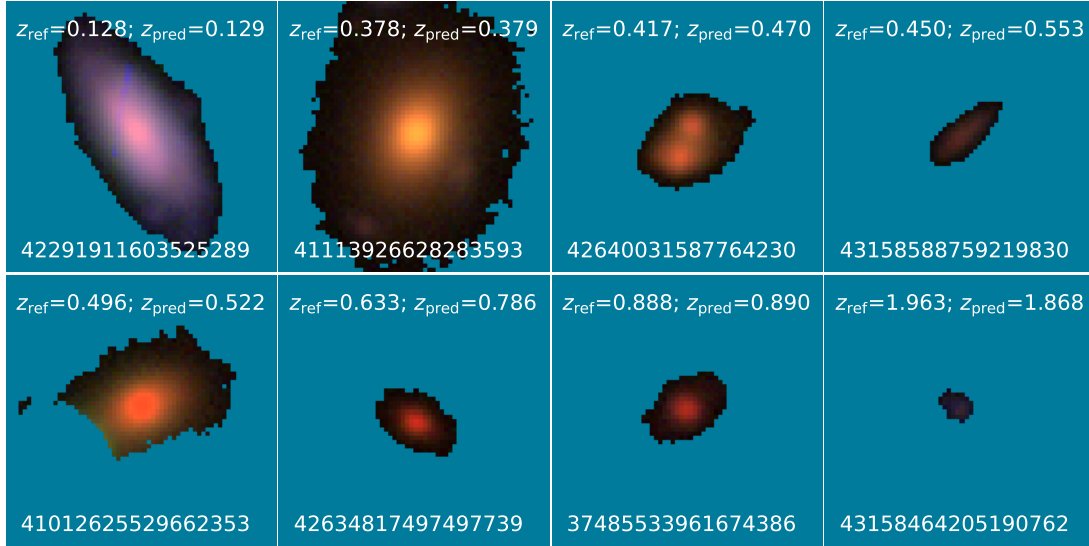
Figure 6.2: Overview of galaxies from our data set. The masked neighbouring objects and background are shown in blue and have pixel values of zeros in the image. The images are $10.75'' \times 10.75''$ (64×64 pixels) and based on the three filters g, r, and i. In each panel, the reference and predicted redshifts of the object are indicated at the top and the HSC identification number is at the bottom.

The reference redshift selection criteria described above give us a sample of galaxies with accurate reference redshifts, z_{ref}. Since the sample D is dominated by galaxies with $z_{\mathrm{ref}} < 1$, we test the effect of data augmentation. Explicitly, we include rotated images for $z_{\mathrm{ref}} > 1$, and in addition, mirrored images for $z_{\mathrm{ref}} > 2$. An alternative to data augmentation is to introduce weights for the galaxies. For example, Lima et al. (2008) proposed a relative weighting of galaxies in order to match their spectroscopic sample to observables of the photometric sample. Although we could also adapt a similar weighting scheme to balance the redshift distribution, we favor the data augmentation technique that is commonly used in neural networks.

Since the distribution of the reference redshifts in the training set is very important for the network and still dominated by the lower redshift end, we limit each redshift bin of width 0.01 to have no more than 1000 galaxies from those passing the above criteria. With this limit, we obtain a uniform distribution up to $z_{\mathrm{ref}} \sim 1.5$ This essentially limits the number of low-redshift galaxies that would otherwise be over-represented in the training set. As a result, the redshift distribution becomes more uniform and allows the CNN to learn and predict redshifts for the full redshift range rather than only the lower redshift end. The excluded galaxy sample is marked in the underlying yellow box with lines in Figure 6.1, while sample D is used for our main network

NetZ$_{main}$, shown with a red histogram in Figure 6.3. We show the distribution of the augmented sample as a black dashed histogram in Figure 6.3.
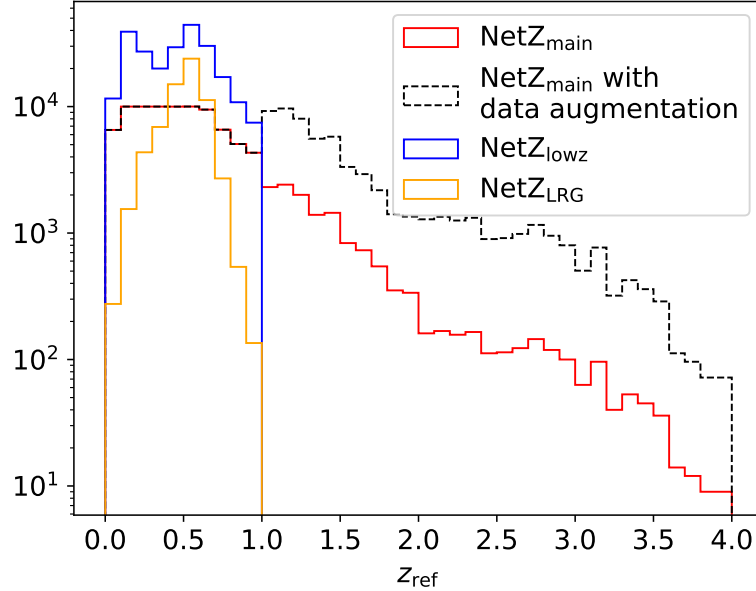
Figure 6.3: Histograms of the redshift samples used in this work. For NetZ$_{main}$, we show the original redshift distribution in red, and the data augmented distribution in dashed black (with more galaxies at $z_{ref} > 1$) that was used for our final network. The distribution used to train our two specialized networks (see Sect. 6.6 for details) is overplotted for NetZ$_{lowz}$ in blue and for NetZ$_{LRG}$ in orange.

6.3 Deep learning and the network architecture

Neural networks (NN) are very powerful tools that serve many different tasks, especially in works involving a huge amount of data. Substantial efforts have therefore been dedicated to deep learning (DL) developments in recent years. In general, for supervised learning, it is necessary to have a data set where the input and output, that is, the so-called ground truth, are known. On this data, the network is trained and can afterwards be applied to new data where the output is not known. The main advantages of NN include the variety of architectures and thus the broad range of problem they can be applied to, as well as the speed of those networks in comparison to other methods. Generally, there are two kinds of networks: classification networks distinguish between different classes of objects, whereas regression networks predict specific numerical quantities. The latter is the kind of network we are using here, namely it is the network that predicts a specific value for the redshift of a galaxy.

Depending on the task, there are different types of networks. Since our input consists of images of galaxies, a typical type is the CNN where the fully-connected (FC) layers are preceded

by a number of convolutional (conv) layers. The detailed architecture depends on various parameters such the specific task, the size of the images, and the size of the data set. We tested different architectures and found an overall good network behavior with two convolutional layers followed by three FC layers. We tested different constructions of CNN architecture by varying the number of convolutional or FC layers, strides, and kernel sizes but with no improvement. A sketch of the final architecture is shown in Figure 6.4. The input consists of five different filters for each galaxy and each image has a size of 64×64 pixels, corresponding to an image size of around $10'' \times 10''$. The convolutional layers have stride $s = 1$ and a kernel size of $5 \times 5 \times C$, where $C = 5$ in the first convolutional layer, and $C = 32$ in the second layer. We used 32 kernels and 64 kernels in the first and second convolutional layers, respectively. Each convolutional layer is followed by max pooling of size 2×2 and of stride 2. This results in a data cube of size $13 \times 13 \times 64$, which, after flattening, is passed on to the FC layers to obtain the single output value, namely, the redshift of the galaxy.
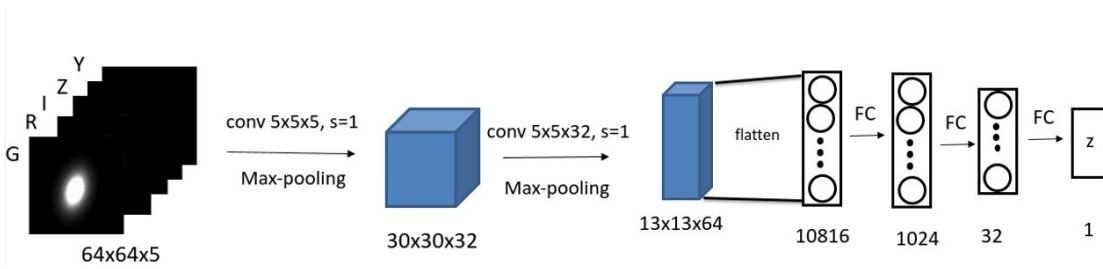
Figure 6.4: Overview of the CNN architecture. It contains two convolutional (conv) layers with max pooling and three fully connected (FC) layers. The input consists of images of size 64×64 pixels in five different filters (*grizy*). The output displays the predicted photometric redshift.

Independent of the network architecture, the network can contain hundreds of thousands (or more) neurons. Even though at the beginning, the values of the weight parameters and bias of each neuron are random, they are updated at every iteration of the training. To see the network performance after the training, we need to split the data into three sets, the training set R, the validation set V, and the test set T (see Figure 6.1). In our case, we used 56% of the data set as training set, 14% as validation set, and 30% as test set. We trained over 300 epochs and divided each epoch into a number of iterations by splitting the training, validation, and test set into batches of a size N. In each iteration, a batch is passed through the entire network to predict the redshifts z_{pred} (forward propagation). The difference between those predicted values and the ground truth is quantified by the loss function L, where we use the mean-square-error (MSE) defined as [3]

$$L = \frac{1}{N} \sum_{k=1}^{N} (z_{\text{pred},k} - z_{\text{ref},k})^2 . \tag{6.1}$$

After completing the forward propagation and computing the loss for the batch, the information is propagated to the weights and biases (back propagation) that are then modified using a stochastic gradient descent algorithm with a momentum of 0.9. This procedure is repeated for

---

[3]This definition is for only one parameter, which in our case is the redshift. For a general expression, one would also sum over the different parameters.

all batches in the training set and a total training loss for this epoch is thus obtained. Afterwards, the loss is computed within the validation set to determine the improvement of the network, which concludes the epoch.

We perform a so-called cross validation to minimize bias in the validation set, which comprises training the NN on the training set and using the validation set to validate the performance after each epoch as described above. These steps are repeated by exchanging the validation set within the training set, such that we have with our splitting five cross-validation runs. In the end, the network is trained on training and validation set together and terminated at the best epoch of all cross validation runs. The best epoch is defined as the epoch with the minimal average validation loss. This network is then applied to the test set, which contains data the network has never seen before.

6.4 Main Redshift Network NetZ_{main}

In this section, we present our main network $NetZ_{main}$ which is trained in the full redshift range $(0 < z \lesssim 4)$. We find that this CNN is overall very precise in predicting redshifts. Figure 6.5 shows a comparison of our final network predictions z_{pred} to the reference redshifts z_{ref} of the test set T. In detail, the left panel of this plot shows a histogram of the reference redshifts (red) and predicted values (blue). On the right panel, a 1:1 comparison of reference and predicted redshifts is plotted. The red line shows the median and the gray bands the 1σ and 2σ confidence levels.
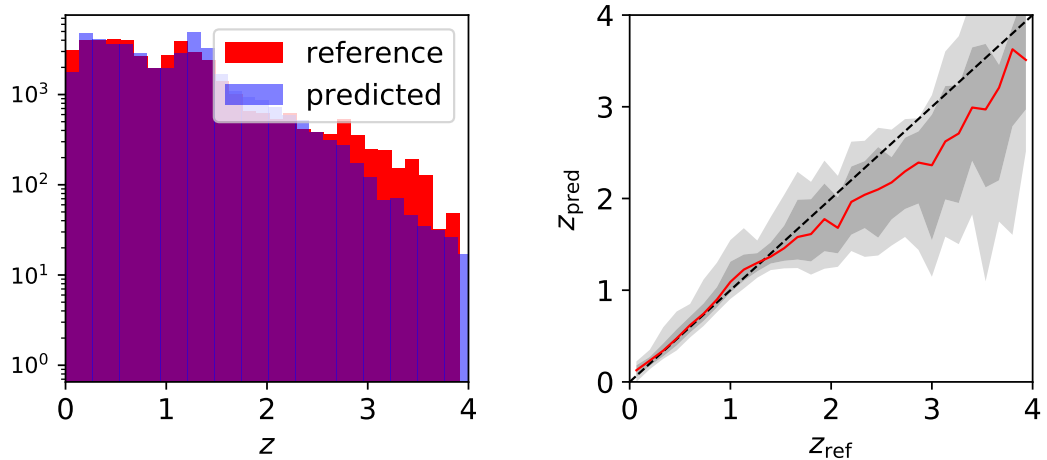
Figure 6.5: Performance of the final network on the test set T. On the left hand side, histograms of the reference and predicted redshift distributions are shown in red and blue, respectively. On the right hand side, a 1:1 comparison of reference and predicted redshifts is plotted. The red line shows the median predicted redshift per bin and the gray bands the 1σ and 2σ confidence levels. While the red line follows the black dashed reference line for low redshift very nicely, $NetZ_{main}$ tends to underpredict the high end.

While the network performance is good in the redshift range between 0 and ~2, the network starts to underestimate the higher redshifts. This is understandable as the network is trained on many more images in the lower reshift range as we can see directly from the histogram. The reason is the limited amount of available training data (reference redshifts) above $z \sim 2$. Moreover, these distant galaxies are typically faint and small in extent, which complicates the learning procss with regard to their morphological features.

As described in Sect. 6.3, we use cross-validation and train always over 300 epochs. We do not see much overfitting from the loss curve, where overfitting means that after a certain number of epochs the network learns to predict the redshifts better for the training set than for the validation set. Based on our testing of different hyper-parameters such as batch size or learning rate, the best moment for terminating the training of NetZ$_{main}$ is at epoch 135 with a loss of 0.1107 according to the loss function L. This network has a learning rate of 0.0005 and a batch size of 128. We also tested drop-out, which means to ignore during each training epoch a new random set of neurons. This can help to reduce overfitting and balance the importance of the neurons in the network. We carried out a test using a dropout rate of 0.5 between the FC layers, but it turned out that drop-out was not necessary for this network.

We tested the network performance by varying the masking, such as the deblending threshold and the kernel for the smoothing. The difference of $\lesssim 0.01$ in the predictions is small compared to the typical photo-z uncertainty (as we see in the scatter of Figure 6.5). This network stability is important in case the extraction is not done perfectly as planned and done for the training. The masking is done in the exact same way for the newly predicted photo-z values as for training and testing.

It turns out that the network predicts similar but slightly different values for the augmented images, which shows that the network does not identify the rotated or mirrored images as duplicates. The possibility to use such data augmentation and hence boost the performance at high redshifts is a major strength of NetZ.

As a further test, we replaced the image cutouts of the galaxies with point-like sources using the corresponding PSF images and scaling them to the correct magnitudes. This way, the images contain only the information available from the catalogs (as used in typical photo-z methods) but exclude any morphological information such as the galaxy shapes from the real cutouts. We tested a few different hyper-parameter combinations by varying the learning rate and also the number of convolutional layers, but with the result of worse performance in predicting redshifts, as shown in Figure 6.6. For more detail, we show on the left panel the performance of the network trained on the point-like images in analogy to the right panel of Figure 6.5; and on the right panel, we make the direct comparison between the network trained on the correct image cutouts (red) and the network trained on images of point-like sources (blue). We can directly see the smaller scatter in the predicted redshift when using the correct cutouts, especially on the high-redshift range even with our use of data augmentation in this redshift range. This test shows the importance of the morphological information for this method and that it contributes significantly to the robustness of the photo-z predictions.
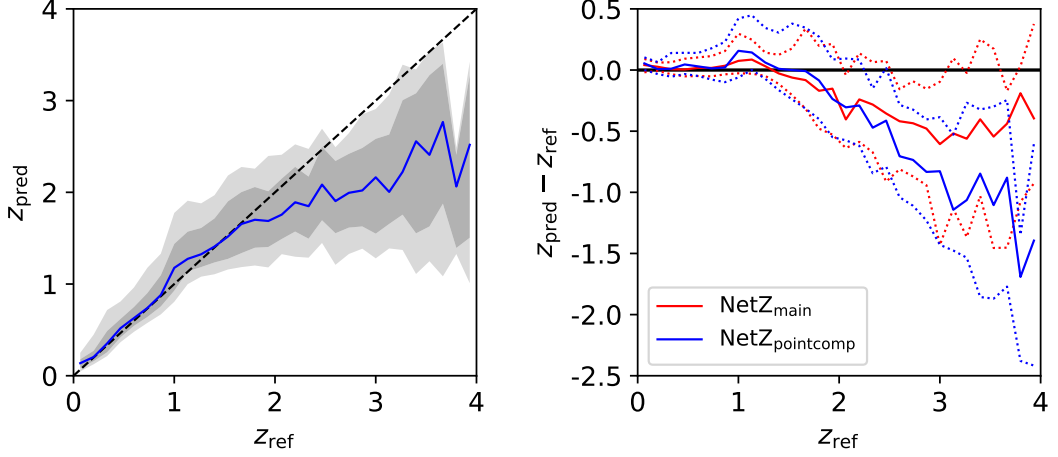
Figure 6.6: Performance of the network trained on images of point-like sources in place of galaxies (blue) with 1 and 2σ ranges on the left (gray), and as a comparison to NetZ$_{\text{main}}$ (red) on the right panel with 1σ ranges (dotted). We directly see that the original galaxy images and thus their morphological information improve the network.

6.5 Comparison of NetZ$_{\text{main}}$ to other photo-z methods

6.5.1 Detailed comparison to HSC method DEmP

Since there are already different photo-z methods developed and applied to the HSC data, we show a comparison here. It is very important to use the same data set for a fair comparison. Thus, we can only make a comparison with the DEmP (Hsieh & Yee, 2014) method, where we have a predicted photo-z value for each galaxy within our test set T and using identical training and validation sets as for NetZ$_{\text{main}}$, without data augmentation – since DEmP also relies on the reference distribution. DEmP is one of the best-performing methods from the HSC photo-z team (Tanaka et al., 2018; Nishizawa et al., 2020) and, thus, it stands as a good performance benchmark. DEmP is a hybrid photo-z method by combining polynomial fitting and a N-nearest neighbor method based on photometric values on a catalog level. Therefore, the input data are totally different from those of NetZ, which is based on the pixelated image cutouts of the galaxies.

For the comparison, we adopted three quantities from the HSC photo-z papers (Tanaka et al., 2018; Nishizawa et al., 2020), which are defined as follows for each redshift bin:

$$\textbf{Bias:} \quad \text{Median}\,(\Delta z_i) = \text{Median}\left(\tfrac{z_{\text{pred},i}-z_{\text{ref},i}}{1+z_{\text{ref},i}}\right), \tag{6.2}$$

$$\textbf{Dispersion:} \quad \sigma = 1.48 \times \text{MAD}(\Delta z_i) =$$
$$= 1.48 \times \text{Median}\,(|\Delta z_i - \text{Median}(\Delta z_i)|), \tag{6.3}$$

$$\textbf{Outlier rate:} \quad f_{\text{outlier}} = \tfrac{N(|\Delta z_i|>0.15)}{N_{\text{bin}}}. \tag{6.4}$$

Figure 6.7: NetZ$_{main}$ (black points) performance in terms of bias (top-left panel), dispersion (bottom-left panel), and outlier rate (bottom-right panel) as functions of the reference redshift z_{ref} in comparison to DEmP (red bars). Definitions of bias, dispersion and outlier rate are given in Eqs. (6.2)-(6.4). We show also with blue bars the results from a network where we do not use data augmentation to increase the number of $z$high-z galaxies. The values in dashed bars are based on limited number of galaxies. The histogram in the top-right panel shows the number of galaxies as a function of redshift in the test set T used for the comparison. NetZ performs substantially better than DEmP at $z_{ref} \gtrsim 2$, with smaller bias, lower dispersion and lower outlier rate, by up to a factor of 2.

where i denotes the ith galaxy in the redshift bin, z_{pred} the predicted photometric redshift, z_{ref} the reference redshift, N the number of galaxies satisfying the specified condition, and N_{bin} the total number of galaxies in the bin. The dispersion is defined using the median absolute deviation (MAD), as expressed above. The multiplication factor comes from statistics and is the relation factor for normally distributed data between MAD and the standard deviation (Rousseeuw & Croux, 1993).

The comparison is shown in Figure 6.7, with black triangles showing the performance of NetZ$_{main}$ and red bars showing the performance of DEmP. Since we use data augmentation (rotation and mirroring of images; see Sect. 6.2 for details) to increase the number of high-z galaxies, which is not possible for DEmP, we also trained a network without data augmentation – we show both in Figure 6.7 for comparison. For the range $z_{ref} \lesssim 1.5$, the performances of both methods are very good especially for the bias, with DEmP performing slightly better than NetZ.

If we compare the range $z_{\mathrm{ref}} \gtrsim 1.5$, the performance of both methods decreases, but NetZ with data augmentation now performs noticeably better than DEmP.

A decrease in performance in the redshift range around $z \approx 2$ is expected, as the used filter set *grizy* does not cover the prominent 4000Å break but, in contrast to the other methods presented in Tanaka et al. (2018), NetZ and partly DEmP can at least break the degeneracy between low redshifts ($z < 0.5$) and the redshift range around 3 to 4, which is not the case for Mizuki from Nishizawa et al. (2020) and several methods from Tanaka et al. (2018) that used HSC images with similar reference redshifts. A more detailed and direct comparison is difficult since the predicted redshifts from these methods are not publicly available for our whole test set; moreover, some of the galaxies in our test set could be in the training data of these methods, which would artificially improve their performance.

The very low dispersion of DEmP in the highest-redshift bin comes from DEmP underestimating consistently most of the redshifts, and hence the outlier rate is large. Although the outlier rate is high in the range $z_{\mathrm{ref}} \gtrsim 2$ in general, the performance is primarily limited by the number of existing reference redshifts in this range. While DEmP is developed and tested with a big enough training sample also for higher redshifts, we used here for DEmP the exact same data set as for NetZ for a fair comparison. Since there is no sufficient training sample at $z_{\mathrm{ref}} > 2$ for both NetZ without data augmentation and DEmP, we plot these values in dashed because they are less reliable. It is nonetheless encouraging to see the significant reduction in the bias, dispersion, and outlier rate of NetZ with data augmentation for the high-z range, up to a factor of 2 relative to DEmP, thanks to the use of the spatial information from the galaxy images in addition to photometry. Especially for upcoming surveys such as LSST, which will provide relatively deep images, it is important to have methods prepared and tested on the higher redshift range.

As a further comparison, we show in Figure 6.8 a scatter plot of z_{pred} versus z_{ref} for DEmP (Hsieh & Yee, 2014) and our neural network NetZ$_{\mathrm{main}}$ with data augmentation. From this plot we can again see the good performance for the low-z range, where we note that the number of outliers from NetZ is negligible compared to the number of galaxies in the bins, which is also evident in the outlier rate. If we assume that all catastrophic outliers for $z_{\mathrm{ref}} < 1.5$ are misplaced at high redshift, which is very conservative, then $> 77\%$ of the galaxies predicted to be at $z_{\mathrm{pred}} > 1.5$ are actually at $z_{\mathrm{ref}} > 1.5$. In the high-z range, NetZ tends to predict too low redshifts, but it does not have the cluster of catastrophic outliers at $z_{\mathrm{pred}} \sim 0.5$ and z_{ref} between 3 and 3.5 that DEmP does and this is due to the Lyman-break or Balmer-break misclassification. Even for the galaxies where the NetZ redshifts are classified as outliers, these redshifts are closer to the true redshift than for DEmP. The outlier rate for NetZ is dominated by blue-star-forming galaxies and galaxies with a small spatial extent (covering $\approx 20 - 30$ pixels) that provide little information for the CNN to extract features. We therefore note that galaxies covering a small number of pixels are more prone to be catastrophic outliers in their redshifts and should be treated with caution.

6.5.2 Photo-z with morphological information

The studies presented by, for instance, Soo et al. (2018) and Wilson et al. (2020) aim to include morphological information of galaxies to improve photo-z estimations. In particular, Wilson
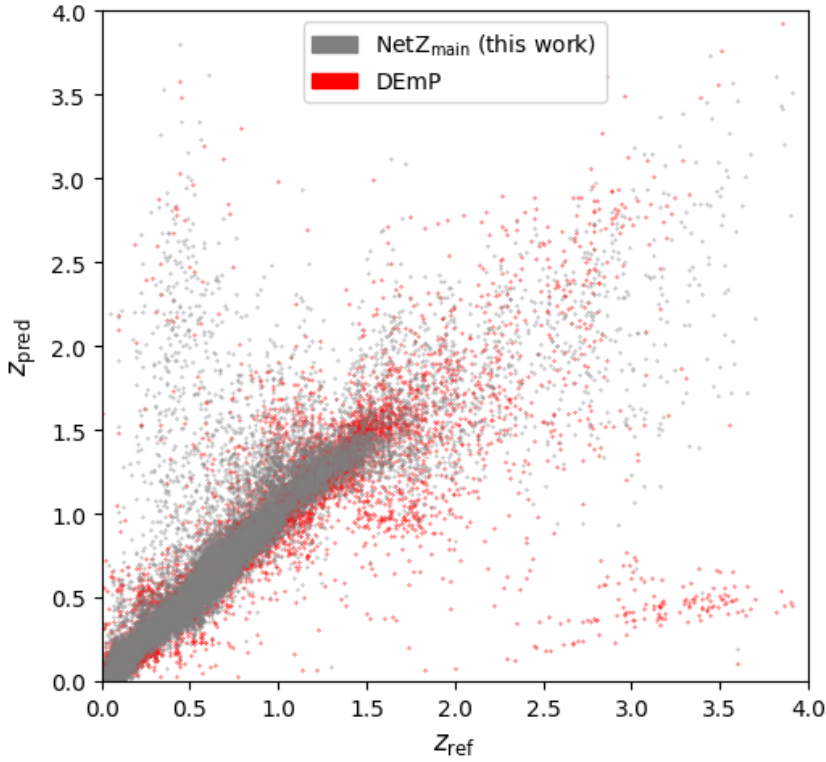
Figure 6.8: Network performance as scatter plot comparing the predicted with the reference redshifts for NetZ$_{main}$ (this work) and DEmP (Hsieh & Yee, 2014). The scatter looks overall comparable at $z_{ref} \lesssim 2$, while NetZ$_{main}$ does not contain the catastrophic outliers at $z_{ref} \sim 3$ and $z_{pred} \sim 0.5$ that DEmP has.

et al. (2020) make use of optical and near infrared observations, some of which are obtained with the Hubble Space Telescope (HST). Therefore the considered data cover a wider wavelength range and are additionally of better spatial resolution than our ground based HSC images. However, Wilson et al. (2020) consider only photometric measurements and four morphological measurements (half-light radius, concentration, asymmetry, and smoothness), rather than the pixels directly. By working directly with the image pixels in our CNN, we use the maximal amount of information and we are independent of the pipelines and uncertainties when extracting morphological measurements. Moreover, Wilson et al. (2020) limit the range to $0 < z < 2$, which makes the network not directly applicable to deep imaging surveys, especially since we are focusing on the high-redshift range. As we show in the next section, we also obtain very good results within a limited range. With these differences in the assumptions and data sets, it is difficult to directly compare the results of Wilson et al. (2020) and NetZ. Nonetheless, if we compare our outlier fraction with ~0.05 up to $z \sim 1.7$ (see Figure 6.7) to that from Wilson et al. (2020), which is called OLF, with ~0.1-0.2 up to $z = 2$ (see Tables 2 and 3 of Wilson et al. (2020)), NetZ yields a good improvement. While Soo et al. (2018) and Wilson et al. (2020) find

that morphological measurements do not provide a notable improvement in photo-z predictions when compared to using only multi-band photometric measurements, our NetZ results (Figure 6.6) show that the pixels in the image cutouts that contain morphological information are useful. This suggests that a promising avenue for future developments of photo-z methods is to combine photometric measurements (as typically used for current photo-z methods) with direct image cutouts (as used for NetZ) instead of morphological measurements.

6.5.3 Photo-z estimates for LSST

Schmidt et al. (2020) present a collection of different photo-z methods tested on LSST mock data. In particular, they compare 12 different codes, of which three methods are based on template fitting (BPZ, Benítez (2000); EAZY, Brammer et al. (2008); LePhare, Arnouts et al. (1999)), seven are based on machine learning (ANNz2, Sadeh et al. (2016); CMN,Graham et al. (2018); FlexZBoost, Izbicki et al. (2016); GPz, Almosallam et al. (2016a); METAPhoR, Cavuoti et al. (2017); SkyNet, Graff et al. (2014); TPZ, Carrasco Kind & Brunner (2013)), one is a hybrid method (Delight, Leistedt & Hogg, 2017), and one is a pathological photo-z PDF estimator method (trainZ, Schmidt et al., 2020). The last method trainZ is designed to serve as an experimental control, and not a competitive photo-z PDF method. It assigns to each galaxy set a photo-z PDF by effectively performing a k-nearest neighbor procedure. As training data, they use $< 10^7$ LSST like mock data limited to $0 < z < 2$ and an i band magnitude limit of 25.3 to match the LSST gold sample (for further details see Schmidt et al., 2020). The main advantage of these methods in Schmidt et al. (2020) compared to the current version of NetZ is the probability density function estimates, whereas NetZ does not require photo-z pre-selection and shows a good performance over a broader redshift range ($0 < z < 4$). Based on the different redshift range and data sets, a detailed and fair comparison is not possible. If we compare Figure 6.5 to Figure B1 of Schmidt et al. (2020) quantitatively, we see an overall similar performance, but most of the LSST methods have a cluster of outliers at $z_{\mathrm{ref}} \sim 0.5$ and $z_{\mathrm{ref}} \sim 1.7$ which we do not see with NetZ. The kink at $z_{\mathrm{ref}} \sim 1.7$ might be related to the drop of data points and an edge effect near the end of the assumed range since we observe a similar effect with NetZ for higher redshifts ($z_{\mathrm{ref}} \sim 3$). Comparing the machine learning methods is difficult as well. The network architectures, as with nearest-neighbour algorithms, random forests, prediction trees, or sparse Gaussian processes, which are presented in Schmidt et al. (2020), are simply too different from the image-based CNN we present with NetZ.

6.6 Limited-range and LRG-only redshift network

During our testing, we found we could obtain substantial improvement by restricting the redshift range. We explored, for instance, networks with redshift ranges limited to $0 < z < 1$ and $1 < z < 2$, but not to higher redshift intervals, due to the limitations in available reference redshifts for $z > 2$. Limiting to $0 < z < 1$ is also done in several other works (e.g., Hoyle, 2016; Pasquet et al., 2019; Campagne, 2020). To benefit from these refined networks in any practical application, we would first need to predict the correct redshift range and then these networks could be used in a specified range. We also considered combining multiple networks

and iteratively refining the photo-z predictions, that is, start with NetZ$_{\text{main}}$ to predict z_{pred} and then, based on the value of z_{pred}, we could subsequently apply a network that is trained in a narrower range around z_{pred} to refine the z_{pred} estimate. However, we find that outliers from NetZ$_{\text{main}}$ limit the gain we can achieve in refining z_{pred}. A practical possibility to use a redshift network for the lower-redshift end of the distribution would be to restrict the sample by the galaxy brightness. If we restrict our data set D to galaxies with an apparent *i*-band AB magnitude brighter than 22, the catalog includes only 1.3% objects with $z_{\text{ref}} \geq 1$ and we miss 12.9% of all galaxies from the original set D with $z_{\text{ref}} \leq 1$. For the training of NetZ$_{\text{lowz}}$ itself we limit only to a narrow redshift range but not in magnitude. The performance of NetZ$_{\text{lowz}}$ is shown in Figure 6.9, on the left a histogram of the reference (red) and predicted (blue) redshifts. The distribution of the predicted redshift follows that of the reference redshift very well. On the right side, we show a 1:1 correlation plot, with the median as a red line and in gray the 1σ and 2σ areas. If we compare the two (Figure 6.5), we can see that NetZ$_{\text{lowz}}$ performs significantly better than NetZ$_{\text{main}}$, as expected.
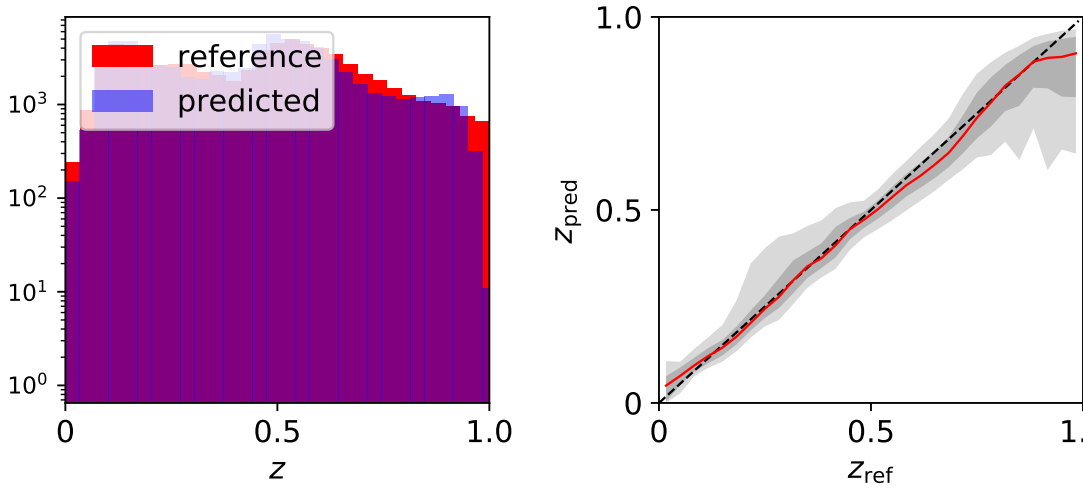
Figure 6.9: Performance of the network NetZ$_{\text{lowz}}$ trained on all types of galaxies in the range $0 < z_{\text{ref}} \leq 1$. On the left hand side, histograms of the redshift distributions are shown, in red the distribution of the reference redshifts used to train the network (ground truth) and in blue the predicted redshift distribution. On the right panel, a 1:1 comparison of reference and predicted redshifts is plotted. The red lines show the median and the gray bands show the 1σ and 2σ confidence levels.

We further show in Figure 6.10 the bias, dispersion, and outlier rate for NetZ$_{\text{lowz}}$ (red). If we compare this performance to NetZ$_{\text{main}}$ applied to the same galaxies for a fair comparison (blue), we find a good improvement in the bias and, with a factor of ~ 2 reduction, in the dispersion. Only the outlier rate is comparable. If we compare the performance of NetZ$_{\text{main}}$ on the full test set that of the network, we still see an improvement for the network NetZ$_{\text{lowz}}$ without the *i* magnitude limitation. This confirms that the improvement is related to the network range. A scatter plot of NetZ$_{\text{lowz}}$ is shown in Figure 6.11.
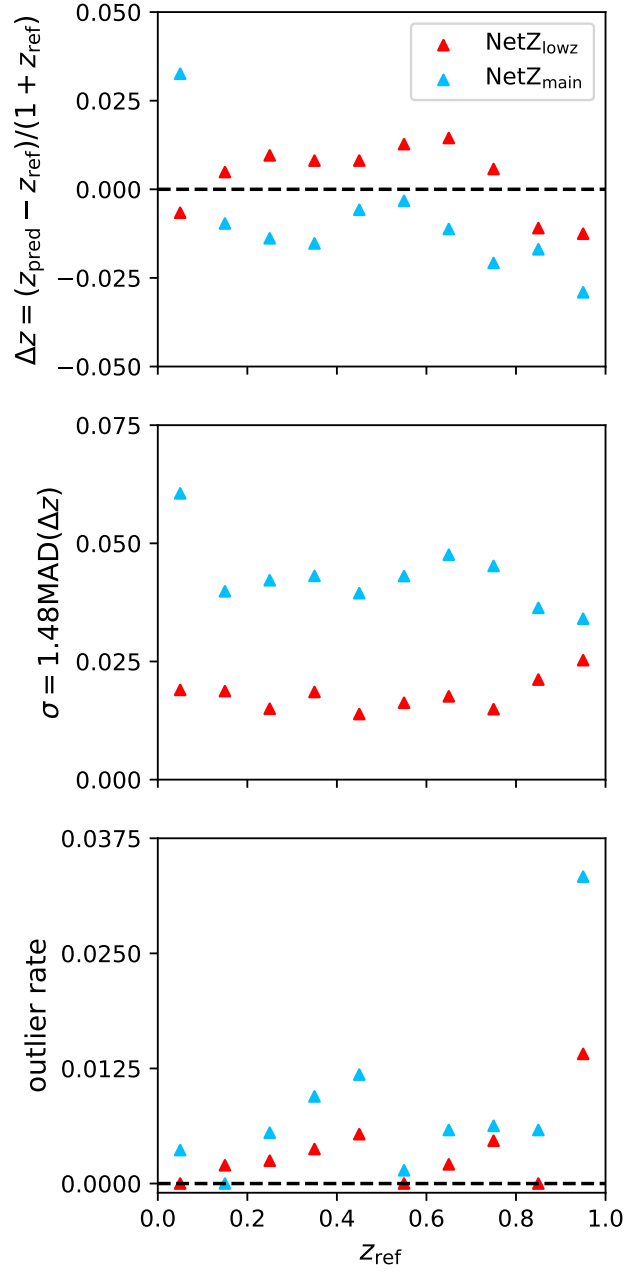
Figure 6.10: Network performance of NetZ$_{lowz}$ compared to NetZ$_{main}$ in terms of bias, dispersion, and outlier rate (see eq. (6.2)-(6.4)) as functions of the reference redshift z_{ref}. For this comparison, we use the overlap between both test sets and only galaxies with an *i*-band magnitude brighter than 22 as NetZ$_{lowz}$ would be applied only to them.
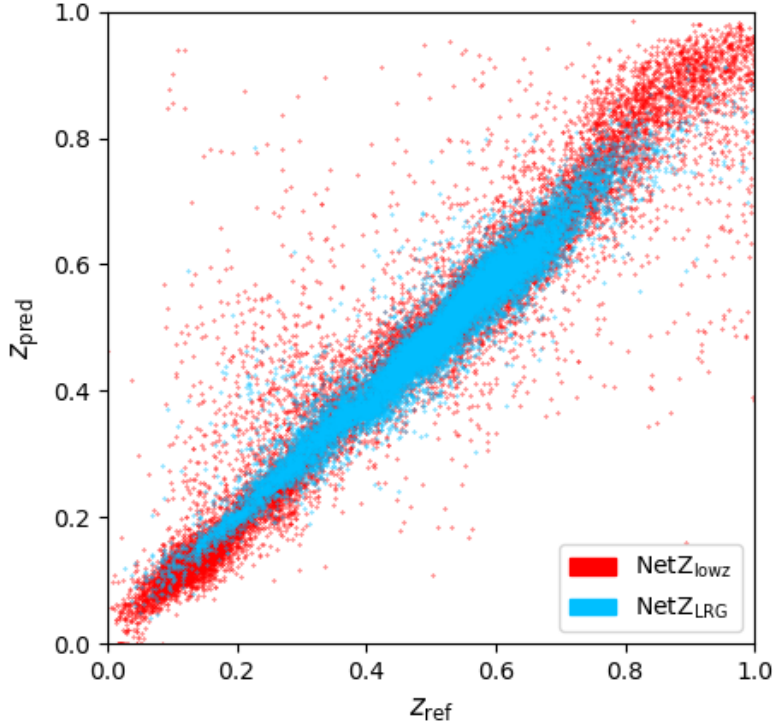
Figure 6.11: Predicted redshifts z_{pred} against the reference redshifts z_{ref} for the networks $\mathrm{NetZ_{LRG}}$ and $\mathrm{NetZ_{lowz}}$ of their test set. We see directly a lower outlier rate for $\mathrm{NetZ_{LRG}}$ than $\mathrm{NetZ_{lowz}}$.

Instead of applying networks trained for specific redshift ranges, which is difficult to do in practice, we can consider specific classes of galaxies that can be selected a priori, such as LRGs. Therefore we investigate a redshift estimation network specialized on LRGs that are useful for various studies including strong lensing and baryon acoustic oscillations. Since nearly all LRGs out of our reference sample have $z_{\mathrm{pred}} < 1$, we show here the network performance of $\mathrm{NetZ_{LRG}}$ in comparison to the network $\mathrm{NetZ_{lowz}}$ trained on all galaxy types. Figure 6.12 shows on the left a histogram and on the right the 1:1 comparison of z_{ref} and z_{pred}.

We show further the bias, dispersion, and outlier rate (defined in eq (6.2)-(6.4)) in Figure 6.13. The network $\mathrm{NetZ_{LRG}}$ performs better in most redshift bins. Finally, in Figure 6.11 we show a scatter plot of this network without magnitude limitation in comparison to $\mathrm{NetZ_{lowz}}$. From this we can see again the redshift limits of the LRG sample and also the good improvement.

Both networks $\mathrm{NetZ_{LRG}}$ and $\mathrm{NetZ_{lowz}}$ show that photo-z for subsamples of galaxies does overall better than the main network $\mathrm{NetZ_{main}}$ that is trained on all galaxies. Therefore, for specific subsamples, it would be beneficial to train a CNN specific to that sample.
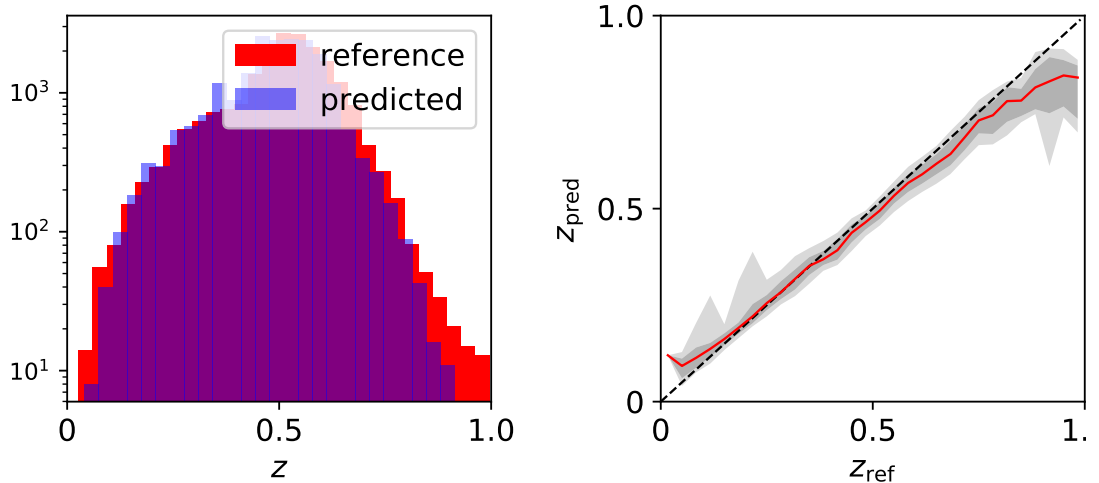
Figure 6.12: Performance of the network NetZ$_{LRG}$ trained on LRGs only. On the left-hand side, histograms of the redshift distributions are shown, in red we show the distribution of the reference redshifts used to train the network (ground truth) and in blue the predicted redshift distribution. On the right panel, a 1:1 comparison of reference and predicted redshifts is plotted. The red lines show the median and the gray bands the 1σ and 2σ confidence levels.

6.7 Summary and conclusions

With current and upcoming imaging surveys, we anticipate that billions of galaxies will be the subject of observations, while just a small fraction of them will have spectroscopically confirmed redshifts. Therefore, it is necessary to have tools to obtain good photometric redshifts, especially for the higher redshift range as the upcoming surveys will provide deeper images where previous photo-z methods have strong limitations. With the success of ML and especially CNNs in image processing, we investigated a new CNN based technique to estimate the photo-z of a galaxy. The method is very general; it accepts directly cutouts of the observed images and predicts the corresponding redshift. Therefore it is directly applicable to all HSC cutouts after applying simple cuts on the Kron radius and i band magnitude observables.

For training the network and testing the performance, we carry out a comparison with to reference redshifts from various, mostly magnitude-limited surveys. In this paper, we focus on HSC data with a pixel size of $0.168''$ and use the available five filters *grizy*, which are also part of the upcoming LSST[4]. In principle, it is also possible to include additional filters, such as the near-infrared (NIR) range from the same or a different telescope which would improve the performance even more, as shown by Gomes et al. (2018), for the low redshift range ($z \lesssim 0.6$). The only constraint from the CNN is the constant pixel size over all different filters. Since NIR images have typically larger pixel sizes, an interpolation and resampling to the same pixel resolution as the optical images would be necessary. What remains to be seen is how much NetZ could benefit from such additional filters, especially in the high-z range, and this would need to be tested. In addition, NetZ could be trained on additional Euclid images that are high-resolution

---

[4]LSST has in addition u-band observations.
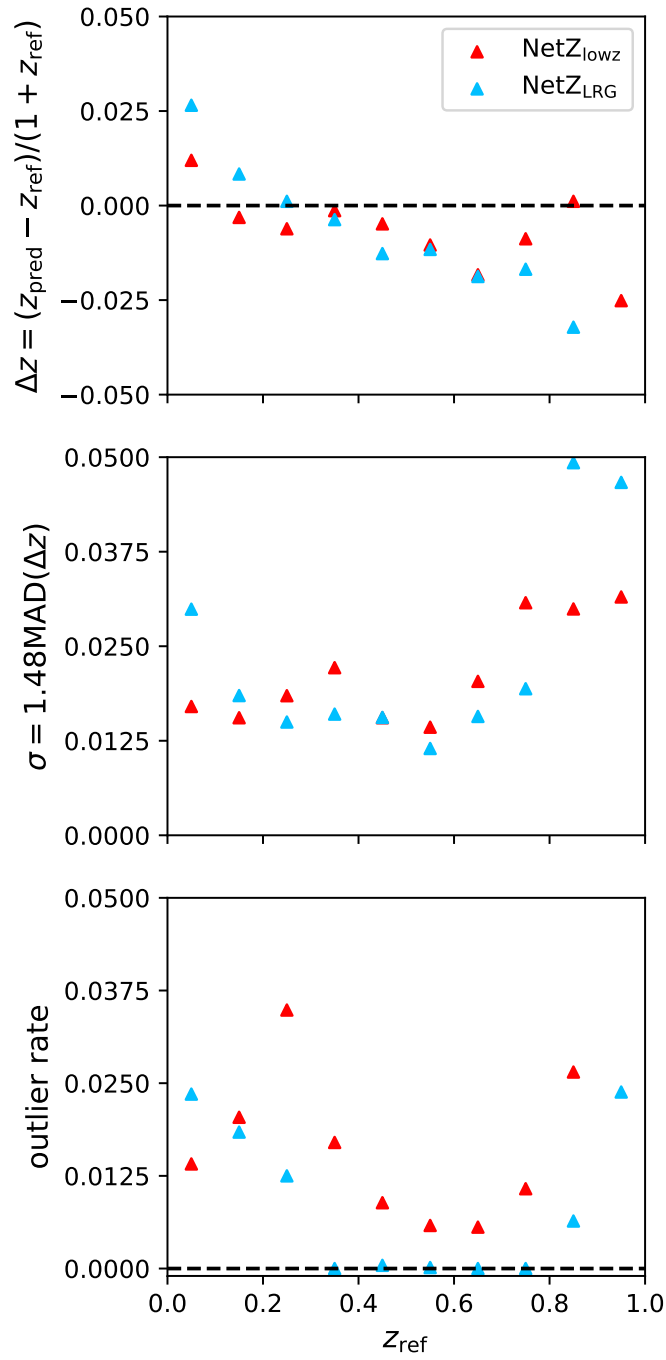
Figure 6.13: Network performance of NetZ$_{LRG}$ compared to NetZ$_{lowz}$ applied to LRGs only in terms of bias, outlier rate, and dispersion (see eq. (6.2)-(6.4)) as functions of the reference redshift, z_{ref}.

from space in the visible and infrared range. Even without combining Euclid with ground-based images, our photo-z estimates from HSC are useful for Euclid given the overlap in the footprints of HSC and Euclid.

With our trained network on HSC images, we find an overall very good performance of the network with a 1σ uncertainty of 0.12 averaged over all galaxies from the whole redshift range. Our CNN provides a point estimate for each galaxy with uncertainties adopted from the scatter in each redshift bin of the test set. Based on the amount of available data, the network performs better in the redshift range below $z = 2$. In the range above $z = 2$, we are using, as a way of gaining an advantage over state-of-the art methods like DEmP, data augmentation by rotating and mirroring the images. While the bias for DEmP and NetZ$_{\text{main}}$ as well as the dispersion for DEmP increases significantly in this range, with NetZ$_{\text{main}}$ we obtain by using data augmentation similar values as for the lower redshift range. We also obtain better outlier rates for the highest redshift bins by using data augmentation but the improvement is less pronounced. In particular, NetZ does not under-predict the redshifts of galaxies with $z_{\text{ref}} \sim 3 - 3.5$ by as much as DEmP and other methods due to the Lyman and Balmer break misclassification. The main limitations that all photo-z methods face when predicting redshifts for distant galaxies is the low number of reference redshifts. In our case, the number drops by a factor of around 1000 compared to the range where $z < 2$. Therefore, using the image cutouts gives a good advantage as we can use data augmentation by rotating and mirroring the images. The effect is impressive as one can see in Fig 6.7. Since this is not possible for other photo-z methods, several of them focus only on the lower redshift range $z < 1$ or even lower (e.g., Hoyle, 2016; Pasquet et al., 2019; Campagne, 2020). If we also limit the redshift range to $z < 1$, we find a substantial improvement in our network's performance.

In cases where we set our focus on a specific galaxy type like LRGs, we find a further improvement with regard to the network. This is understandable as the network can learn better the specific features of this galaxy type. Based on the small number of LRGs with redshift above $z = 1$, we limit the range of NetZ$_{\text{LRG}}$ to $0 < z < 1$ and compare it to a network trained on all galaxy types in the same redshift range for a fair comparison.

This paper provides a proof of concept for using a CNN for photo-z estimates. Based on the encouraging results of NetZ particularly at high redshifts, we propose further investigations along the lines of combining our CNN with a nearest-neighbor algorithm or a fully-connected network that ingests catalog-based photometric quantities (see Leal-Taixé et al., 2016). There are several methods, like DEmP and other methods (e.g. D'Isanto & Polsterer, 2018; Schmidt et al., 2020), which provide a probability distribution function for the redshifts. Further developments of our CNN approach to provide a probability distribution function of the photo-z require more complex networks such as Bayesian neural networks (e.g., Perreault Levasseur et al., 2017) or mixture density networks (D'Isanto & Polsterer, 2018; Hatfield et al., 2020; Eriksen et al., 2020). While this is beyond the scope of the current paper, such Bayesian or mixture networks are worth exploring.

In this work, we show that a very simple convolutional neural network is able to predict accurate redshifts directly from the observed galaxy images. NetZ therefore has the advantage of using maximal information from the intensity pixels in the galaxy images, rather than relying on photometric or morphological measurements that could be prone to uncertainties and biases, especially for images of blended galaxies. We ran NetZ$_{\text{main}}$ on 34,414,686 galaxies from the

HSC public data release 2 (PDR2) wide survey and provide the catalog here[5] . We flagged all negative predictions and clear catastrophic outliers ($z_{pred} > 5$), which are 15,043 and 3,314 objects, respectively, as −99. In Figure 6.14, we show a histogram of the newly available photo-z values (blue filled), whose distribution resembles the magnitude-limited sample of the cleaned COSMOS2015 (Laigle et al., 2018, orange histogram), which was scaled by a factor of 1010, to have the same sample size for the purposes of making a direct comparison. This check shows that our NetZ predictions indeed produce a realistic galaxy redshift distribution expected for a depth similar to that of LSST.
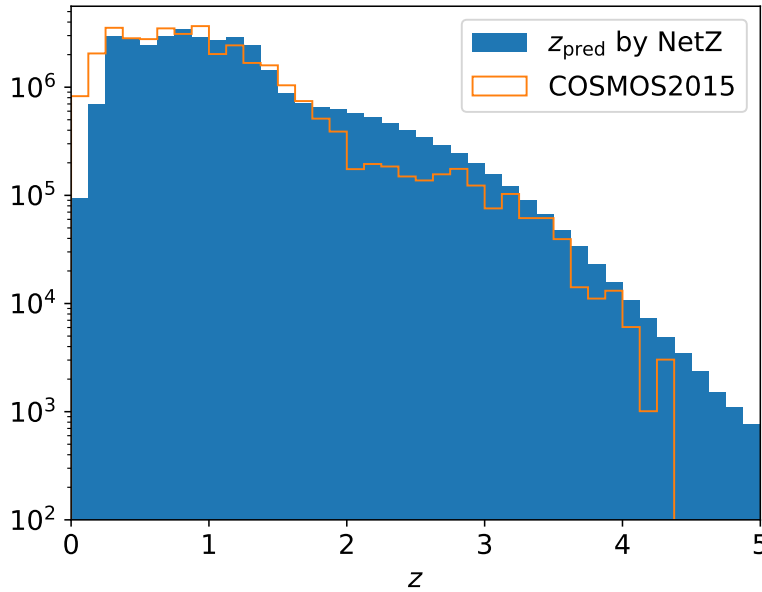
Figure 6.14: Histogram of the newly predicted photo-z values with NetZ based on images of the HSC PDR2 (blue filled) and, for comparison, the distribution of COSMOS2015 (Laigle et al., 2018) scaled by a factor of 1010 to have the same sample size (orange open). The similarity in the two distributions shows that NetZ produces a realistic galaxy redshift distribution.

As the image quality, depth and processing of HSC and LSST first-year data are expected to be similar (the image processing pipeline of HSC is a branch of the LSST pipeline), the method we have developed here will be directly applicable and beneficial to the glslinkLSSTLSST. The additional u-filter in the LSST will likely further improve photo-z predictions. When applying our method to the LSST data, we do not expect to necessarily have to test the network architecture, however, it is likely that some hyper-parameter combinations would need testing. Since training is more optimally carried out on real images, rather than on mock images as done, for instance, in Schmidt et al. (2020, and references therein), we suggest that it is optimal

---

[5]The catalog is available at `https://www.dropbox.com/sh/grjfo0gkcxsj9n2/AAD-B7D6m7_1i6GGTX0Ionwja?dl=0`

to train a new network on LSST images as soon as they are available.

7 Conclusion & Outlook

Strong gravitational lensing with its versatile nature has become a very powerful tool leading to remarkable achievements in astrophysics and cosmology. It enables the opportunity to probe many different aspects and help to answer outstanding questions in those research fields. This thesis aims to model galaxy-galaxy lensing systems in a fast and efficient way, which is highly needed to cope with the current and upcoming wide-field imaging surveys from which we expect 100,000 lenses. The presented modeling networks and pipelines allow us to cope with this amount of lenses, and thus help to pave the way for further dedicated studies of the lenses. Especially lensed time-variable objects such as SNe require fast and dedicated modeling of their host galaxies to predict when and where the next strongly lensed image will appear. This allows us to observe the SN spectra before the peak luminosity is reached. Early spectra are crucial to shed light onto their progenitor systems, which are still unclear. Moreover, strongly lensed SNe allows to measure the Hubble constant H_0 through the time delay between the different appearing images. This helps to resolve the current debate between early and late Universe measurements and clarifies whether new physics is required for this.

For nearly all studies in astrophysics and cosmology, a redshift of the analyzed object, like a SN, quasar, star or directly a full galaxy, is needed. This holds also for lensing studies, where a lens and a source redshift is required and a possibility to confirm the lensing nature. Therefore, well tested photo-z methods with good performance are indispensable for the huge amount of observed objects. This is the second aspect of the thesis, where we present an extensively tested novel approach to estimate their photo-z.

7.1 Lens mass modeling

In Chapter 3, we presented a CNN to predict the SIE parameters of strongly lensed galaxy-galaxy lens systems. Since most of the lenses are detected with ground based facilities, our network is trained specifically for images from HSC, which is a wide-field imaging survey in multi-band and with very good image quality. To train the network, it is crucial to use realistic mock data to obtain trustworthy performance on the test set, which is also simulated. For this we developed a simulation pipeline to essentially lens the image of a real observed galaxy and paint these arcs to another real observed image. In our case, we use images from HSC PDR2 as lens image, together with measured velocity dispersion and redshifts from SDSS. As background sources we used images and redshifts from HUDF by taking advantage of their high redshift and very good resolution compared to HSC. Given the HSC pixel size of $0.168''$, we set a lower limit on the Einstein radius θ_E of $0.5''$ to ensure the detectability of the arcs. With this procedure, we generate three samples of $\sim 100,000$ mock images with different Einstein radius distributions. The first sample represents a naturally distribution on θ_E, the second a flat distribution up to

$\sim 2''$, and the third sample has a lower limit of $\theta_E = 2''$. We present for all three samples specific CNNs through extensive tests on the architecture and hyper-parameter combination. In general, we find good performance of the networks, and advocate finally a flat distribution as it gives much better performance on the systems with larger image separation with an acceptable performance on the very small image separation systems. Moreover, we test in this chapter the improvement by training a network specifically for quads (four image configurations) and test as well the possibility of a network dedicated for doubles (two image configurations).

We further demonstrate in this chapter the possibility of using the CNN output to predict the lensed image position(s) and time delay(s), in case a short-lived transient such as a SN is additionally lensed in such a system. We find in general similar performance for all three different samples and no strong correlation on the lens/source brightness ratio as shown in Fig. 3.13. The accuracy is very good for the time delay, time-delay ratio and also image positions, while the precision is as expected for the adopted SIE profile and ground-based images.

In Chapter 4 we built upon this work and show the ability of predicting the SIE+γ_{ext} parameters with corresponding 1σ uncertainties through a residual neural network. For this, we upgrade the simulation pipeline by adding the external shear component, but also improve by including the Poisson noise on the arcs. We further refine the lens center and ellipticity estimation of the first and second brightness moments by including a proper mask of the lens. We then recenter the lens cutout and finally shift the mock image by up to ±3 pixel to enforce the network to learn the lens mass center instead of the cutout center. With the upgraded pipeline, we generate again $\sim 100,000$ mock images with a flat distribution of the Einstein radius up to $\theta_E \sim 2''$ for better performance over the full range. The most difficult parameters to infer are those of the external shear, possibly as the influence is very minor and only encoded in the arcs. However, our tests with a simplified setup suggest that the network is in general able to extract the external shear and blending with the arcs are not the main issue. Providing the network further information though including the FWHM values or the PSF directly as image did not result in an improvement on the external shear. Although we carried out extensive tests, investigations are needed to further test possible reasons for the difficulty and options to overcome it.

Even if the light distribution of the lens galaxy and the unlensed source galaxy are real, the presented networks are trained on mock images. Therefore we present in Chapter 5 a direct comparison on real lenses. The comparison sample consists of 32 grade A galaxy-scale lenses found in the SuGOHI program (Sonnenfeld et al., 2018a; Wong et al., 2018; Sonnenfeld et al., 2019; Chan et al., 2020; Jaelani et al., 2020a; Sonnenfeld et al., 2020; Jaelani et al., 2021). We apply the trained network from Chapter 4 to that sample, and also model them with GLEE & GLaD, state-of-the-art software packages with Bayesian inference. Since these techniques are very time and resource consuming, we develop a dedicated procedure to model galaxy-galaxy lenses observed with ground based facilities to reduce the required user input time. After modeling all 32 lenses uniformly with this code, we further refine some models to further improve the fit for the comparison. For this, we present *gleeauto.py*, a software package that allows the user to specify an optimization sequence which then the code performs without any further user input. Using both presented GLEE & GLaD automation codes, we obtained good models for our 32 lenses, which we then compared to our network predictions. We find overall very good performance of the network compared to the performance on the test set. This highlights how realistic our mocks are.

Since the main focus was on the model comparison rather than performing many dedicated tests with our automated modeling pipeline, there are several possibilities to improve the code further both in terms of fitting and computational time. For instance the second part, the lens light fit, is currently performed with GLEE and for the fourth part, the arc light modeling, transferred to GLaD. By performing directly the lens light modeling in GLaD, the first optimization step (m=1) would be unnecessary, resulting in a clear reduction of computational time. Moreover, the sampling sequences can be better tested and then optimized. Beside these technical improvements mainly to reduce the computational time, also investigations in automating the masking and parameter estimate would be helpful such that this code can fully autonomously model lenses. A possibility is to use auto encoders to deblend source and lens as suggested e.g., in Rojas et al. (2021) and Savary et al. (2021), or by making use of color information from different filters as done in CHITAH (Chan et al., 2020). This would allow to apply the code to a larger set of lenses and, for instance, use the obtained model to rule out some false-positives. This would increase the purity of lenses in our sample and save spectroscopic follow-up resources. We demonstrate this procedure in Taubenberger et al. (in prep.) with a dozen lens candidates and confirm with spectroscopic observations.

Given the higher uncertainties predicted by the network compared to GLEE & GLaD, a complete new project would be to combine DL and state-of-the-art optimization directly in one modeling pipeline in analogy to Pearson et al. (2021). This means the code first predict the SIE+γ_{ext} values through an optimized network, and use then these estimates as starting values for an MCMC sampling for refinement. One could also use the network output to select peculiar lensing systems with e.g., high image separation and just analyze them further. A direct combination of network and GLEE & GLaD, however, need definitively an automation in the masking or automated image position detection for an source/image position optimization.

Despite these possibilities of further developments, combining the modeling tools developed and presented in this thesis, ranging from machine-learning CNNs and ResNets to automated stat-of-the-art modeling tools, we are able to efficiently model the expected number of $\sim 100,000$ lenses from current and upcoming wide-field imaging surveys. This enables statistical analysis of thousands of lenses, fast and also detailed modeling without much user input for follow-up observations of lensed SNe allowing us to study their progenitor systems as well as time-delay cosmography to measure H_0.

7.2 Redshift estimation with NetZ

Beside the lensing aspect, we develop and present in Chapter 6 a novel approach to systematically predict photometric redshifts for galaxies. For this the network is trained directly on the image cutouts instead of extracted quantities such as color-magnitude or size-compactness information of a galaxy. The performance of the network has been demonstrated with HSC images through a detailed comparison to DEmP, another photo-z code that gave the best accuracies on HSC data. We find that our main network performs especially well in the high-redshift range without drastic cuts on e.g., the brightness, which will be very important for upcoming deep imaging surveys such as LSST. One of the main advantages of our technique is the possibility of data augmentation in the training process to obtain a more uniformly distributed training

sample. With this network we have predicted and published > 34 million photometric redshifts. Moreover, we present dedicated networks for the low end of the redshift range, which resulted in NetZ$_{\text{lowz}}$, and also for LRGs, as specific type of galaxies, resulting in NetZ$_{\text{LRG}}$. Both networks are also trained on HSC data and show very good results. Given the similarity of HSC images to the expected image quality and filters of LSST, this method will be immediately applicable with similar performance to LSST and thus enables a variety of science applications in the next decade.

The presented network demonstrate the possibility of using image cutouts directly and we are able to achieve encouraging performance. Therefore, a straight forward improvement of NetZ would be to include an uncertainty prediction similar as we have done for the modeling network in Chapter 4. A possibility here is to test also for NetZ whether a residual neural network provide an additional improvement. Another extension would be a hybrid network, which means the input data are a combination of image cutouts and extracted catalog entries such as color, magnitude, or size of the galaxy. This might result in an even better performance. In that case, one need to be careful if data augmentation like our image rotation and flipping is still possible given that the catalog entries are not changing at all. Although there are many options to test and possibly improve the performance, the presented results are very good for a novel technique.

List of Figures

List of Tables

Abbreviations

DM	Dark Matter. v, 2–5, 20, 45, 46, 53, 171
eBOSS	extended BOSS. 8, 28, 137
ELT	Extremely Large Telescope. 20
ESA	European Space Agency. 18, 20, 54
ESO	European Southern Observatory. 20, 28
ETG	Early Type Galaxy. 89
FC	Fully Connected. 31, 34, 61–63, 80, 91, 92, 94, 98–101, 140, 141, 143, 154
FPR	False Positive Rate. 17
FWHM	Full Width at Half Maximum. 101, 106, 158, 172
Gaia	Global Astrometric Interferometer for Astrophysics. 18
GAMA	Galaxy And Mass Assembly. 16
GLaD	Gravitational Lensing and Dynamics. 41, 42, 45, 47, 106–109, 111–118, 120, 122, 124–130, 158, 159, 162, 165, 197–206
GLEE	Gravitational Lens Efficient Explorer. 18, 41, 42, 45, 47, 51, 57, 86, 106–109, 111–120, 122, 125–130, 158, 159, 162, 165, 197–206
GR	Theory of General Relativity. 1, 2, 26
GW	Gravitational Wave. 10
H0LiCOW	H_0 Lenses in COSMOGRAIL's Wellspring. 6, 9
HOLISMOKES	Highly Optimized Lensing Investigations of Supernovae, Microlensing Objects, and Kinematics of Ellipticals and Spirals. 13, 55
HSC	Hyper Suprime-Cam. v–xi, 14, 16–18, 21, 23–25, 31, 32, 51, 52, 54–56, 58, 59, 61, 64, 80, 82, 83, 85, 86, 89, 90, 102, 105–108, 110–113, 122, 130, 131, 133, 134, 136, 138, 139, 144, 146, 147, 152, 154–157, 159–162, 165, 170, 197–199, 202, 204–206
HST	Hubble Space Telescope. 4, 8, 11, 12, 14, 18, 19, 42, 45, 51, 55, 80, 96, 147
HUDF	Hubble Ultra Deep Field. x, 52, 55, 58, 61, 81, 86, 90, 102, 105, 157
IFU	Integrale Field Unit. 47
IR	Infrared. 6, 8, 19, 169
JVAS	Jodrell Bank VLA Astrometric Survey. 17
JWST	James Webb Space Telescope. 20, 47

KiDS	Kilo Degree Survey. 15, 16, 19
LAMOST	Large sky Area Multi-Object fiber Spectroscopic Telescope. 28
LMC	Large Maggellanic Cloud. 3
LOS	Line-of-sight. v, 1, 6, 12, 44, 48, 89, 107, 116, 172
LRG	Luminous Red Galaxy. xi, 13–16, 43, 56, 86, 89, 105, 133, 134, 136, 137, 148, 149, 151–154, 160, 162
LSS	Large Scale Structure. 3, 4, 20, 28
LSST	Rubin Observatory Legacy Survey of Space and Time. v–viii, xi, 15, 17, 18, 20–23, 25, 29–32, 51, 53, 54, 56, 79–83, 90, 98, 106, 107, 122, 131, 133–136, 146, 148, 152, 155, 156, 159–161
MACHO	Massive Astrophysical Compact Halo Object. 3
MAD	Median Absolute Deviation. 145
maser	microwave amplification by stimulated emission of radiation. 8, 10
MCMC	Markov chain Monte Carlo. xi, 49, 50, 54, 78, 79, 81, 107, 116, 119–123, 125, 130, 159
ML	Machine Learning. vi, ix, 15, 16, 31–33, 35, 36, 51, 135, 152
MSD	Mass-Sheet Degeneracy. 46–48, 172
MSE	Mean Square Error. 62, 90, 105, 141
MW	Milky Way. 10
NFW	Navarro Frenk and White. 45, 46, 172, 173
NIR	Near-IR. 19–21, 28, 29, 45, 152
NN	Neural Network. v, vii, 16, 32–35, 51, 55, 61, 106–108, 140, 142
OLF	Outlier fraction. 147
Pan-STARRS	Panoramic Survey Telescope And Rapid Response System. 16, 18, 19, 21, 51, 54
PDF	Probability Density Function. 31, 135, 148
PDR	Public Data Release. 17, 23, 56, 59, 106, 136–138, 155, 157, 162, 165
PSF	Point Spread Function. 14, 44, 47, 58, 59, 101, 105, 106, 120, 123, 124, 143, 158
ReLU	Rectified Linear Unit. 33, 34, 61, 92
ResNet	Residual Network. v, vii–ix, 17, 35, 85, 91, 92, 96–99, 102, 107, 109, 111–115, 125–129, 159, 162

RMS	Root Mean Square. 30, 59, 88
SB	Surface Brightness. x, 5, 10, 39, 41, 45, 47, 116, 122, 125, 130
SDSS	Sloan Digital Sky Survey. 14, 18, 19, 23, 28, 29, 32, 55, 56, 83, 86, 105, 135, 137, 157, 170
SED	Spectral Energy Distribution. 23, 30, 31, 135
SFR	Star Formation Rate. 5, 28, 30
SH0ES	Supernovae, H_0, for the Equation of State. 6, 8–10
SIE	Singular Isothermal Ellipsoid. v, vii, x, 14, 36, 42, 43, 51, 52, 54–58, 60–62, 64, 66, 68, 70–72, 74–76, 78–82, 85, 86, 90, 96–98, 100, 102, 104–108, 113–116, 119, 123, 125, 128, 157, 158, 165, 171, 172, 197–199
SIE+γ_{ext}	Singular Isothermal Ellipsoid with external shear. x, 16, 36, 85, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 125–127, 129, 130, 158, 159, 162
SIS	Singular Isothermal Sphere. 43
SL2S	Strong Lensing Legacy Survey. 13, 14
SLACS	Sloan Lens ACS. 14
SN	Supernova. v, vii, ix, 8, 10–13, 18–21, 40, 48, 51, 53, 54, 75, 76, 81, 157–159, 161
SOM	Self Organized Maps. 31, 32
SPEMD	Singular Power-Law Elliptical Mass Distribution. 42, 171
SPT	South Pole Telescope. 8
SQLS	SDSS Quasar Lens Search. 18
SuGOHI	Survey of Gravitationally-lensed Objects in HSC Imaging. xii, 14, 15, 36, 107–113, 122, 124, 125, 158, 162, 165, 197–200, 202, 204–206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238
TRGB	Tip of the Red Giant Branch. 8, 10
UV	Ultraviolet. 28, 29, 137
VLBI	Very Large Baseline Interferometry. 10
WD	White Dwarf. 11, 53
WMAP	Wilkinson Microwave Anisotropy Probe. 8
ZTF	Zwicky Transient Facility. 13, 19, 21

Nomenclature

A	Amplitude of the SIE profile.	41, 43, 44, 57, 117, 120, 200, 201, 204–206
D	Angular diameter distance (with subscript d (deflector) or s (source)).	1, 5, 6, 10, 26, 37, 38, 41, 57
E	Amplitude of the SPEMD profile.	42
F	Flux.	39
G	Gravitational constant.	1, 2, 38
H_0	Hubble constant.	v, vii, 2, 6, 8–12, 26, 46–48, 53, 55, 85, 107, 157, 159, 161, 168, 170
H	Hubble parameter.	26
I	Intensity.	39, 43, 44, 68, 70, 74, 76, 77, 79, 88, 89
L	Loss function.	35, 62, 90, 141, 143
$M_\odot$	Mass of the sun.	1, 3
M	Mass of the lens.	1
N_{bin}	Number of galaxies in bin for f_{outlier}.	144, 145
N	Batch size.	62, 65, 90, 141
R	Training set.	35, 141
S	Stride (denoted as s in Chapter 6).	62, 141
T'	Temperature (scaling parameter in optimization algorithms).	49, 50
T	Test set.	35, 36, 141, 142, 144, 162
V	Validation set.	35, 141
Δt	Time delay.	5, 6, 46, 48, 76
Λ	Cosmological constant.	6, 26, 48, 53, 55, 85, 107
Ω_Λ	Cosmological constant for the energy content.	48, 55, 85, 107
Ω_M	Cosmological constant for the DM content.	3, 48, 55, 85, 107
Ψ	Lens potential.	5, 6, 39, 40
Σ_{crit}	Critical Surface mass density.	38
Σ	Surface mass density.	38
$\alpha_{\mathrm{poisson}}$	Scaling factor of Poisson noise map.	88, 89
$\boldsymbol{A}$	Jacobian matrix.	40
$\boldsymbol{\alpha}$	Scaled deflection angle.	37–39, 41, 46
$\boldsymbol{\beta}$	Source position (unlensed).	5, 37–41, 46, 172
$\boldsymbol{\theta}$	Image Position on the lens plane.	5, 37–41, 46

θ	Position angle of Sérsic profile (with subscript light/mass of lens/source, component). 42, 56, 57, 86, 117, 121, 198, 199, 202–206
a	Scale factor. 26, 27
c	Speed of light. 1, 2, 5, 38, 57
e_{c}	Ellipticity, in complex notation. 57
e_{x}	Ellipticity, x component in complex notation. 51, 52, 57, 61, 62, 66–68, 71, 73, 82, 92, 94, 96–98, 108, 113–115, 172
e_{y}	Ellipticity, y component in complex notation. 51, 52, 57, 61, 62, 66–68, 71, 73, 82, 92, 94, 96–98, 108, 113–115, 129, 172
f_{outlier}	Outlier fraction. 30, 144, 171
f	Kernel size (of a pooling layer). 61, 62
g	Reduced shear. 40
n	Sérsic index (with subscript light/mass of lens/source, component). 43, 44, 117, 121, 124, 202–206
p	Dropout rate. 63
q	Axis ratio (with subscript light/mass of lens/source, component). 42–44, 46, 56, 57, 66, 68, 86, 117, 120, 121, 124, 198, 199, 202–206
$r_{\odot}$	Radius of the sun. 1
r_{c}'	Elliptical core radius. 42
r_{c}	Core radius. 42
r_{eff}	Effective radius, also called half-light radius (with subscript light/mass of lens/source, component). 43, 44, 117, 120, 202–206
r_{learn}	learning rate. 63, 65, 93
r_{s}	Scale radius in the NFW profile. 45, 46
r	Radius. 38, 42, 43, 45, 46, 56, 57, 78
s	Source intensity. 45
v_{disp}	velocity dispersion. 56, 57, 66, 80, 161
v	variance map. 26, 88
w	Weighting factor. 62, 65, 67, 69, 70, 75, 82, 90, 93, 161
x	Coordinate along the x-axis with respect to the cutout center (with subscript lens/source, component). 51, 52, 56, 57, 61, 62, 67, 68, 71, 73, 76, 77, 82, 86, 92, 94, 96, 105, 108, 113–115, 117, 120, 121, 126, 128, 129, 172, 198, 199, 202, 203
y	Coordinate along the y-axis with respect to the cutout center (with subscript lens/source, component). 51, 52, 56, 57, 61, 62, 67, 68, 71, 73, 76, 77, 82, 86, 92, 94, 96, 105, 108, 113–115, 117, 120, 121, 126, 128, 129, 172, 198, 199, 202, 203

z Redshift, spectroscopic (spec-z) or photometric (photo-z); subscript: deflector (lens) d, source s. iii, ix, xi, 1, 3, 5, 6, 11, 12, 27–32, 56, 58, 66, 80, 90, 109, 133–155, 157, 159, 161, 162, 172, 174

Bibliography

Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018a, PRD, 98, 043526, [arXiv:1708.01530]

Abbott, T. M. C., Abdalla, F. B., Allam, S., et al. 2018b, ApJS, 239, 18, [arXiv:1801.03181]

Abdalla, F. B., Banerji, M., Lahav, O., & Rashkov, V. 2011, MNRAS, 417, 1891, [arXiv:0812.3831]

Abolfathi, B., Aguado, D. S., Aguilar, G., et al. 2018, ApJS, 235, 42, [arXiv:1707.09322]

Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, ApJS, 249, 3, [arXiv:1912.02905]

Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, PASJ, 106, [arXiv:1905.12221]

Aihara, H., AlSayyad, Y., Ando, M., et al. 2021, arXiv e-prints, arXiv:2108.13045, [arXiv:2108.13045]

Aihara, H., Arimoto, N., Armstrong, R., et al. 2018a, PASJ, 70, S4, [arXiv:1704.05858]

Aihara, H., Armstrong, R., Bickerton, S., et al. 2018b, PASJ, 70, S8, [arXiv:1702.08449]

Aiola, S., Calabrese, E., Maurin, L., et al. 2020, JCAP, 2020, 047, [arXiv:2007.07288]

Alam, S., Albareti, F. D., Prieto, C. A., et al. 2015, The Astrophysical Journal Supplement Series, 219, 12

Alam, S., Aubert, M., Avila, S., et al. 2021, PRD, 103, 083533, [arXiv:2007.08991]

Alcock, C., Akerlof, C. W., Allsman, R. A., et al. 1993, Nature, 365, 621, [arXiv:astro-ph/9309052]

Ali-Dib, M., Menou, K., Jackson, A. P., Zhu, C., & Hammond, N. 2020, icarus, 345, 113749, [arXiv:1906.08826]

Almosallam, I. A., Jarvis, M. J., & Roberts, S. J. 2016a, MNRAS, 462, 726, [arXiv:1604.03593]

Almosallam, I. A., Lindsay, S. N., Jarvis, M. J., & Roberts, S. J. 2016b, MNRAS, 455, 2387, [arXiv:1505.05489]

Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS, 310, 540, [arXiv:astro-ph/9902290]

Assef, R. J., Kochanek, C. S., & The AGES Collaboration. 2010, arXiv e-prints, arXiv:1004.5415, [arXiv:1004.5415]

Aubourg, E., Bareyre, P., Bréhin, S., et al. 1993, Nature, 365, 623

Auger, M. W., Treu, T., Bolton, A. S., et al. 2010, ApJ, 724, 511, [arXiv:1007.2880]

Bag, S., Kim, A. G., Linder, E. V., & Shafieloo, A. 2021, ApJ, 910, 65, [arXiv:2010.03774]

Balakrishnan, V., Champion, D., Barr, E., et al. 2021, MNRAS, 505, 1180, [arXiv:2010.07457]

Balkenhol, L., Dutcher, D., Ade, P. A. R., et al. 2021, PRD, 104, 083509, [arXiv:2103.13618]

Ball, N. M., Brunner, R. J., Myers, A. D., et al. 2008, ApJ, 683, 12, [arXiv:0804.3413]

Ball, N. M., Brunner, R. J., Myers, A. D., et al. 2007, ApJ, 663, 774, [arXiv:astro-ph/0612471]

Barkana, R. 1998, ApJ, 502, 531, [arXiv:astro-ph/9802002]

Barnabè, M., Czoske, O., Koopmans, L. V. E., Treu, T., & Bolton, A. S. 2011, MNRAS, 415, 2215, [arXiv:1102.2261]

Barnabè, M., Dutton, A. A., Marshall, P. J., et al. 2012, MNRAS, 423, 1073, [arXiv:1201.1692]

Baron, D. & Poznanski, D. 2017, MNRAS, 465, 4530, [arXiv:1611.07526]

Barrena, R., Biviano, A., Ramella, M., Falco, E. E., & Seitz, S. 2002, A&A, 386, 816, [arXiv:astro-ph/0202323]

Bartelmann, M. & Schneider, P. 2001, PhysRep, 340, 291, [arXiv:astro-ph/9912508]

Baum, W. A. 1962, in Problems of Extra-Galactic Research, ed. G. C. McVittie, Vol. 15, 390

Baxter, E. J. & Sherwin, B. D. 2021, MNRAS, 501, 1823, [arXiv:2007.04007]

Bayer, D., Chatterjee, S., Koopmans, L. V. E., et al. 2018, arXiv e-prints, arXiv:1803.05952, [arXiv:1803.05952]

Bayer, J., Huber, S., Vogl, C., et al. 2021, A&A, 653, A29, [arXiv:2101.06229]

Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, AJ, 132, 1729

Bellagamba, F., Tessore, N., & Metcalf, R. B. 2017, MNRAS, 464, 4823, [arXiv:1610.06003]

Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, PASP, 131, 018002, [arXiv:1902.01932]

Belokurov, V., Evans, N. W., Moiseev, A., et al. 2007, ApJL, 671, L9, [arXiv:0706.2326]

Benítez, N. 2000, ApJ, 536, 571, [arXiv:astro-ph/9811189]

Bernardi, M., Sheth, R. K., Dominguez-Sanchez, H., et al. 2018, MNRAS, 477, 2560, [arXiv:1712.05414]

Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393

Birrer, S., Amara, A., & Refregier, A. 2016, JCAP, 2016, 020, [arXiv:1511.03662]

Birrer, S., Shajib, A. J., Galan, A., et al. 2020, A&A, 643, A165, [arXiv:2007.02941]

Birrer, S., Treu, T., Rusu, C. E., et al. 2019, MNRAS, 484, 4726, [arXiv:1809.01274]

Blake, C., Collister, A., Bridle, S., & Lahav, O. 2007, MNRAS, 374, 1527, [arXiv:astro-ph/0605303]

Blakeslee, J. P., Jensen, J. B., Ma, C.-P., Milne, P. A., & Greene, J. E. 2021, ApJ, 911, 65, [arXiv:2101.02221]

Bolton, A. S., Burles, S., Koopmans, L. V. E., et al. 2008, ApJ, 682, 964, [arXiv:0805.1931]

Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., & Moustakas, L. A. 2006, ApJ, 638, 703, [arXiv:astro-ph/0511453]

Bolzonella, M., Miralles, J. M., & Pelló, R. 2000, A&A, 363, 476, [arXiv:astro-ph/0003380]

Bom, C., Poh, J., Nord, B., Blanco-Valentin, M., & Dias, L. 2019, arXiv e-prints, arXiv:1911.06341, [arXiv:1911.06341]

Bom, C. R., Makler, M., Albuquerque, M. P., & Brand t, C. H. 2017, A&A, 597, A135, [arXiv:1607.04644]

Bond, I. A., Udalski, A., Jaroszyński, M., et al. 2004, ApJL, 606, L155, [arXiv:astro-ph/0404309]

Bonfield, D. G., Sun, Y., Davey, N., et al. 2010, MNRAS, 405, 987, [arXiv:0910.4393]

Bonnett, C. 2015, MNRAS, 449, 1043, [arXiv:1312.1287]

Bonnett, C., Troxel, M. A., Hartley, W., et al. 2016, PRD, 94, 042005, [arXiv:1507.05909]

Bonvin, V., Courbin, F., Suyu, S. H., et al. 2017, MNRAS, 465, 4914, [arXiv:1607.01790]

Bosch, J., Armstrong, R., Bickerton, S., et al. 2018, PASJ, 70, S5, [arXiv:1705.06766]

Boulade, O., Charlot, X., Abbon, P., et al. 2003, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes, ed. M. Iye & A. F. M. Moorwood, 72–81

Bozinowski, S. 1981, UM-CS, 81

Bradshaw, E. J., Almaini, O., Hartley, W. G., et al. 2013, Monthly Notices of the Royal Astronomical Society, 433, 194–208

Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503, [arXiv:0807.1533]

Brehmer, J., Mishra-Sharma, S., Hermans, J., Louppe, G., & Cranmer, K. 2019, ApJ, 886, 49, [arXiv:1909.02005]

Brescia, M., Cavuoti, S., Longo, G., & De Stefano, V. 2014, A&A, 568, A126, [arXiv:1407.2527]

Breuval, L., Kervella, P., Anderson, R. I., et al. 2020, A&A, 643, A115, [arXiv:2006.08763]

Browne, I. W. A., Wilkinson, P. N., Jackson, N. J. F., et al. 2003, MNRAS, 341, 13, [arXiv:astro-ph/0211069]

Brownstein, J. R., Bolton, A. S., Schlegel, D. J., et al. 2012, ApJ, 744, 41, [arXiv:1112.3683]

Brunner, R. J., Connolly, A. J., Szalay, A. S., & Bershady, M. A. 1997, ApJL, 482, L21, [arXiv:astro-ph/9703058]

Cañameras, R., Schuldt, S., Shu, Y., et al. 2021, arXiv e-prints, arXiv:2107.07829, [arXiv:2107.07829]

Cañameras, R., Schuldt, S., Suyu, S. H., et al. 2020, arXiv e-prints, arXiv:2004.13048, [arXiv:2004.13048]

Cañameras et al., R. in prep.

Cabanac, R. A., Alard, C., Dantel-Fort, M., et al. 2007, A&A, 461, 813, [arXiv:astro-ph/0610362]

Camarena, D. & Marra, V. 2020, Physical Review Research, 2, 013028, [arXiv:1906.11814]

Campagne, J.-E. 2020, arXiv e-prints, arXiv:2002.10154, [arXiv:2002.10154]

Cardona, W., Kunz, M., & Pettorino, V. 2017, JCAP, 2017, 056, [arXiv:1611.06088]

Cardone, V. F. 2004, 415, 839, [astro-ph/0311559]

Carliles, S., Budavári, T., Heinis, S., Priebe, C., & Szalay, A. S. 2010, ApJ, 712, 511

Carrasco Kind, M. & Brunner, R. J. 2013, MNRAS, 432, 1483, [arXiv:1303.7269]

Cavuoti, S., Amaro, V., Brescia, M., et al. 2017, MNRAS, 465, 1959, [arXiv:1611.02162]

Cavuoti, S., Brescia, M., Longo, G., & Mercurio, A. 2012, A&A, 546, A13, [arXiv:1206.0876]

Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2019, The Pan-STARRS1 Surveys, [arXiv:1612.05560]

Chan, J. H. H., Rojas, K., Millon, M., et al. 2021, A&A, 647, A115, [arXiv:2007.14416]

Chan, J. H. H., Suyu, S. H., Sonnenfeld, A., et al. 2020, A&A, 636, A87, [arXiv:1911.02587]

Chao, D. C. Y., Chan, J. H. H., Suyu, S. H., et al. 2020a, A&A, 640, A88, [arXiv:1910.01140]

Chao, D. C. Y., Chan, J. H. H., Suyu, S. H., et al. 2020b, arXiv e-prints, arXiv:2009.07854, [arXiv:2009.07854]

Chen, B. H., Hashimoto, T., Goto, T., et al. 2022, MNRAS, 509, 1227, [arXiv:2110.09440]

Chen, G. C. F., Fassnacht, C. D., Suyu, S. H., et al. 2021a, arXiv e-prints, arXiv:2107.10304, [arXiv:2107.10304]

Chen, G. C. F., Fassnacht, C. D., Suyu, S. H., et al. 2019, MNRAS, 490, 1743, [arXiv:1907.02533]

Chen, G. C. F., Suyu, S. H., Wong, K. C., et al. 2016, MNRAS, 462, 3457, [arXiv:1601.01321]

Chen, G. C. F., Treu, T., Fassnacht, C. D., et al. 2021b, MNRAS, 508, 755, [arXiv:2106.11060]

Chen, Y., Steidel, C. C., Erb, D. K., et al. 2021c, MNRAS, 508, 19, [arXiv:2104.10173]

Cheng, T.-Y., Li, N., Conselice, C. J., et al. 2020, MNRAS, 494, 3750, [arXiv:1911.04320]

Chirivì, G., Yıldırım, A., Suyu, S. H., & Halkola, A. 2020, arXiv e-prints, arXiv:2003.08404, [arXiv:2003.08404]

Ciarcelluti, P. 2005a, International Journal of Modern Physics D, 14, 187, [arXiv:astro-ph/0409630]

Ciarcelluti, P. 2005b, International Journal of Modern Physics D, 14, 223, [arXiv:astro-ph/0409633]

Ciotti, L. & Bertin, G. 1999, 352, 447, [astro-ph/9911078]

Clowe, D., Bradač, M., Gonzalez, A. H., et al. 2006, ApJL, 648, L109, [arXiv:astro-ph/0608407]

Coil, A. L., Blanton, M. R., Burles, S. M., et al. 2011, The Astrophysical Journal, 741, 8

Colas, T., d'Amico, G., Senatore, L., Zhang, P., & Beutler, F. 2020, JCAP, 2020, 001, [arXiv:1909.07951]

Collett, T., Montanari, F., & Räsänen, S. 2019, Phys. Rev. Lett., 123, 231101, [arXiv:1905.09781]

Collett, T. E. 2015, ApJ, 811, 20, [arXiv:1507.02657]

Collett, T. E., Oldham, L. J., Smith, R. J., et al. 2018, Science, 360, 1342, [arXiv:1806.08300]

Collier, W. P., Smith, R. J., & Lucey, J. R. 2018, MNRAS, 473, 1103, [arXiv:1709.01931]

Collister, A. A. & Lahav, O. 2004, PASP, 116, 345, [arXiv:astro-ph/0311058]

Connolly, A. J., Csabai, I., Szalay, A. S., et al. 1995, AJ, 110, 2655, [arXiv:astro-ph/9508100]

Cool, R. J., Moustakas, J., Blanton, M. R., et al. 2013, The Astrophysical Journal, 767, 118

Cooper, M. C., Aird, J. A., Coil, A. L., et al. 2011, ApJS, 193, 14, [arXiv:1101.4018]

Cornachione, M. A., Bolton, A. S., Shu, Y., et al. 2018, ApJ, 853, 148, [arXiv:1708.08854]

Coupon, J., Ilbert, O., Kilbinger, M., et al. 2009, A&A, 500, 981, [arXiv:0811.3326]

Courbin, F., Bonvin, V., Buckley-Geer, E., et al. 2018, A&A, 609, A71

Courbin, F., Eigenbrod, A., Vuissoz, C., Meylan, G., & Magain, P. 2005, in Gravitational Lensing Impact on Cosmology, ed. Y. Mellier & G. Meylan, Vol. 225, 297–303

Craig, P., O'Connor, K., Chakrabarti, S., et al. 2021, A targeted search for strongly lensed supernovae and expectations for targeted searches in the Rubin era, [arXiv:2111.01680]

Cuillandre, J.-C. J., Withington, K., Hudelot, P., et al. 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8448, Observatory Operations: Strategies, Processes, and Systems IV, ed. A. B. Peck, R. L. Seaman, & F. Comeron, 84480M

Dahlen, T., Mobasher, B., Faber, S. M., et al. 2013, ApJ, 775, 93, [arXiv:1308.5353]

Dark Energy Survey Collaboration et al. 2005, arXiv e-prints

Davies, L. J. M., Robotham, A. S. G., Driver, S. P., et al. 2018, MNRAS, 480, 768, [arXiv:1806.05808]

Davis, M., Faber, S. M., Newman, J., et al. 2003, Discoveries and Research Prospects from 6- to 10-Meter-Class Telescopes II

Davis, M., Guhathakurta, P., Konidaris, N. P., et al. 2007, ApJL, 660, L1, [arXiv:astro-ph/0607355]

Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10, [arXiv:1208.0022]

de Diego, J. A., Nadolny, J., Bongiovanni, Á., et al. 2021, A&A, 655, A56

de Jaeger, T., Stahl, B. E., Zheng, W., et al. 2020, MNRAS, 496, 3402, [arXiv:2006.03412]

de Jong, J. T. A., Kuijken, K., Applegate, D., et al. 2013a, The Messenger, 154, 44

de Jong, J. T. A., Verdoes Kleijn, G. A., Boxhoorn, D. R., et al. 2015, A&A, 582, A62, [arXiv:1507.00742]

de Jong, J. T. A., Verdoes Kleijn, G. A., Erben, T., et al. 2017, A&A, 604, A134, [arXiv:1703.02991]

de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013b, Experimental Astronomy, 35, 25, [arXiv:1206.1254]

De Vaucouleurs, G. 1948, 227, 586

Deng, J., Dong, W., Socher, R., et al. 2009, in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248–255

Denzel, P., Coles, J. P., Saha, P., & Williams, L. L. R. 2021, MNRAS, 501, 784, [arXiv:2007.14398]

DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv e-prints, arXiv:1611.00036, [arXiv:1611.00036]

Dhawan, S., Jha, S. W., & Leibundgut, B. 2018, A&A, 609, A72, [arXiv:1707.00715]

Di Valentino, E., Mena, O., Pan, S., et al. 2021, Classical and Quantum Gravity, 38, 153001, [arXiv:2103.01183]

Diego, J. M., Broadhurst, T., Chen, C., et al. 2016, MNRAS, 456, 356, [arXiv:1504.05953]

Dieleman, S., Willett, K. W., & Dambre, J. 2015, MNRAS, 450, 1441, [arXiv:1503.07077]

D'Isanto, A. & Polsterer, K. L. 2018, A&A, 609, A111, [arXiv:1706.02467]

Drinkwater, M. J., Jurek, R. J., Blake, C., et al. 2010, Monthly Notices of the Royal Astronomical Society, 401, 1429–1452

Duncan, C. A. J., Harnois-Déraps, J., & Miller, L. 2021, arXiv e-prints, arXiv:2111.09867, [arXiv:2111.09867]

Duncan, K. J., Brown, M. J. I., Williams, W. L., et al. 2018a, MNRAS, 473, 2655, [arXiv:1709.09183]

Duncan, K. J., Jarvis, M. J., Brown, M. J. I., & Röttgering, H. J. A. 2018b, MNRAS, 477, 5177, [arXiv:1712.04476]

Dutcher, D., Balkenhol, L., Ade, P. A. R., et al. 2021, PRD, 104, 022003, [arXiv:2101.01684]

Dutton, A. A., Brewer, B. J., Marshall, P. J., et al. 2011, 417, 1621, [arXiv:1101.1622]

Dye, S., Evans, N. W., Belokurov, V., Warren, S. J., & Hewett, P. 2008, MNRAS, 388, 384, [arXiv:0804.4002]

Dye, S., Furlanetto, C., Dunne, L., et al. 2018, MNRAS, 476, 4383, [arXiv:1705.05413]

Dye, S. & Warren, S. J. 2005, ApJ, 623, 31, [arXiv:astro-ph/0411452]

Eales, S., Fullard, A., Allen, M., et al. 2015, MNRAS, 452, 3489, [arXiv:1506.05466]

Eddington, A. S. 1919, The Observatory, 42, 119

Einstein, A. 1908, Jahrbuch der Radioaktivität und Elektronik, 4, 411

Einstein, A. 1915, Sitzungsber. preuss.Akad. Wiss, 47, 831

Einstein, A. 1917, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), 142

Einstein, A. 1918, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 448-459.

Einstein, A. 1919, Naturwissenschaften, 7, 776

Einstein, A. 1936, Science, 84, 506

Eriksen, M., Alarcon, A., Cabayol, L., et al. 2020, arXiv e-prints, arXiv:2004.07979, [arXiv:2004.07979]

Ertl et al., S. in prep.

Evans, C., Puech, M., Afonso, J., et al. 2015, arXiv e-prints, arXiv:1501.04726, [arXiv:1501.04726]

Falco, E. E., Gorenstein, M. V., & Shapiro, I. I. 1985, ApJL, 289, L1

Faure, C., Anguita, T., Alloin, D., et al. 2011, A&A, 529, A72, [arXiv:1009.1545]

Feeney, S. M., Mortlock, D. J., & Dalmasso, N. 2018, MNRAS, 476, 3861, [arXiv:1707.00007]

Feldmann, R., Carollo, C. M., Porciani, C., et al. 2006, MNRAS, 372, 565, [arXiv:astro-ph/0609044]

Fernández Arenas, D., Terlevich, E., Terlevich, R., et al. 2018, MNRAS, 474, 1250, [arXiv:1710.05951]

Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, AJ, 150, 150, [arXiv:1504.02900]

Follin, B. & Knox, L. 2018, MNRAS, 477, 4534, [arXiv:1707.01175]

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306, [arXiv:1202.3665]

Fowlie, A., Handley, W., & Su, L. 2020, arXiv e-prints, arXiv:2006.03371, [arXiv:2006.03371]

Foxley-Marrable, M., Collett, T. E., Frohmaier, C., et al. 2020, MNRAS, 495, 4622, [arXiv:2003.14340]

Freedman, W. L., Madore, B. F., Hatt, D., et al. 2019, ApJ, 882, 34, [arXiv:1907.05922]

Freedman, W. L., Madore, B. F., Hoyt, T., et al. 2020, ApJ, 891, 57, [arXiv:2002.01550]

Freedman, W. L., Madore, B. F., Scowcroft, V., et al. 2012, ApJ, 758, 24, [arXiv:1208.3281]

Freeman, P. E., Newman, J. A., Lee, A. B., Richards, J. W., & Schafer, C. M. 2009, MNRAS, 398, 2012, [arXiv:0906.0995]

Friedmann, A. 1922, Zeitschrift fur Physik, 10, 377

Friedmann, A. 1924, Zeitschrift fur Physik, 21, 326

Fukushima, K. & Miyake, S. 1982, Competition and Cooperation in Neural Nets - Proceedings, Kyoto 1982 (Springer-Verlag), 267–285

Fuqiang, C., Yan, W., Yude, B., & Guodong, Z. 2014, PASA, 31, e001

Furusawa, H., Koike, M., Takata, T., et al. 2018, PASJ, 70, S3

Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, SSR, 123, 485, [arXiv:astro-ph/0606175]

Garilli, B., Guzzo, L., Scodeggio, M., et al. 2014, Astronomy & Astrophysics, 562, A23

Gavazzi, R., Marshall, P. J., Treu, T., & Sonnenfeld, A. 2014, ApJ, 785, 144, [arXiv:1403.1041]

Gavazzi, R., Treu, T., Koopmans, L. V. E., et al. 2008, ApJ, 677, 1046, [arXiv:0801.1555]

Gavazzi, R., Treu, T., Marshall, P. J., Brault, F., & Ruff, A. 2012, ApJ, 761, 170, [arXiv:1202.3852]

Gavazzi, R., Treu, T., Rhodes, J. D., et al. 2007, ApJ, 667, 176, [arXiv:astro-ph/0701589]

Gayathri, V., Healy, J., Lange, J., et al. 2020, arXiv e-prints, arXiv:2009.14247, [arXiv:2009.14247]

Geach, J. E. 2012, MNRAS, 419, 2633, [arXiv:1110.0005]

Gerdes, D. W., Sypniewski, A. J., McKay, T. A., et al. 2010, ApJ, 715, 823, [arXiv:0908.4085]

Goldstein, D. A. & Nugent, P. E. 2016, 834, L5

Goldstein, D. A., Nugent, P. E., & Goobar, A. 2019, ApJS, 243, 6, [arXiv:1809.10147]

Gomer, M. & Williams, L. L. R. 2020, JCAP, 2020, 045, [arXiv:1907.08638]

Gomes, Z., Jarvis, M. J., Almosallam, I. A., & Roberts, S. J. 2018, MNRAS, 475, 331, [arXiv:1712.02256]

Goobar, A., Amanullah, R., Kulkarni, S. R., et al. 2017, Science, 356, 291, [arXiv:1611.00014]

Goodman, J. & Weare, J. 2010, Communications in Applied Mathematics and Computational Science, 5, 65

Graff, P., Feroz, F., Hobson, M. P., & Lasenby, A. 2014, MNRAS, 441, 1741, [arXiv:1309.0790]

Graham, M. L., Connolly, A. J., Ivezić, Ž., et al. 2018, AJ, 155, 1, [arXiv:1706.09507]

Grillo, C., Christensen, L., Gallazzi, A., & Rasmussen, J. 2013, MNRAS, 433, 2604, [arXiv:1305.5844]

Grillo, C., Rosati, P., Suyu, S. H., et al. 2018, ApJ, 860, 94, [arXiv:1802.01584]

Grillo, C., Rosati, P., Suyu, S. H., et al. 2020, ApJ, 898, 87, [arXiv:2001.02232]

Grogin, N. A. & Narayan, R. 1996, ApJ, 473, 570, [arXiv:astro-ph/9512156]

Halkola, A., Seitz, S., & Pannella, M. 2006, 372, 1425, [astro-ph/0605470]

Han, B., Ding, H.-P., Zhang, Y.-X., & Zhao, Y.-H. 2016, Research in Astronomy and Astro-physics, 16, 74, [arXiv:1601.01739]

Hashim, N., De Laurentis, M., Zainal Abidin, Z., & Salucci, P. 2014, arXiv e-prints, arXiv:1407.0379, [arXiv:1407.0379]

Hasinoff, S. W. 2012

Hastings, W. K. 1970, Biometrika, 57, 97

Hatfield, P. W., Almosallam, I. A., Jarvis, M. J., et al. 2020, MNRAS, 498, 5498, [arXiv:2009.01952]

He, K., Zhang, X., Ren, S., & Sun, J. 2015, CoRR, abs/1512.03385, [arXiv:1512.03385]

He, K., Zhang, X., Ren, S., & Sun, J. 2016, CoRR, abs/1603.05027, [1603.05027]

Henghes, B., Pettitt, C., Thiyagalingam, J., Hey, T., & Lahav, O. 2021a, MNRAS, 505, 4847, [arXiv:2104.01875]

Henghes, B., Pettitt, C., Thiyagalingam, J., Hey, T., & Lahav, O. 2021b, arXiv e-prints, arXiv:2109.02503, [arXiv:2109.02503]

Herbst, W., Ford, C., Vinton, G., et al. 1992, JAAVSO, 21, 48

Hezaveh, Y. D., Dalal, N., Marrone, D. P., et al. 2016, ApJ, 823, 37, [arXiv:1601.01388]

Hezaveh, Y. D., Perreault Levasseur, L., & Marshall, P. J. 2017, Nature, 548, 555, [arXiv:1708.08842]

Hildebrandt, H., Arnouts, S., Capak, P., et al. 2010, A&A, 523, A31, [arXiv:1008.0658]

Hildebrandt, H., Erben, T., Kuijken, K., et al. 2012, MNRAS, 421, 2355, [arXiv:1111.4434]

Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, MNRAS, 465, 1454, [arXiv:1606.05338]

Hildebrandt, H., Wolf, C., & Benítez, N. 2008, A&A, 480, 703, [arXiv:0801.2975]

Hoekstra, H., Yee, H. K. C., & Gladders, M. D. 2002, ApJ, 577, 595, [arXiv:astro-ph/0204295]

Hoyle, B. 2016, Astronomy and Computing, 16, 34, [arXiv:1504.07255]

Hsieh, B. C. & Yee, H. K. C. 2014, ApJ, 792, 102, [arXiv:1407.5151]

Huang, J., Feng, L., Jin, K., Li, M., & Wang, G. 2022, Optics Communications, 505, 127515, [arXiv:2109.08826]

Huang, X., Storfer, C., Gu, A., et al. 2021, ApJ, 909, 27, [arXiv:2005.04730]

Huang, X., Storfer, C., Ravi, V., et al. 2020, ApJ, 894, 78, [arXiv:1906.00970]

Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168

Hubble, E. P. 1925a, Popular Astronomy, 33, 252

Hubble, E. P. 1925b, ApJ, 62, 409

Huber, S., Suyu, S. H., Ghoshdastidar, D., et al. 2021a, arXiv e-prints, arXiv:2108.02789, [arXiv:2108.02789]

Huber, S., Suyu, S. H., Noebauer, U. M., et al. 2019, A&A, 631, A161, [arXiv:1903.00510]

Huber, S., Suyu, S. H., Noebauer, U. M., et al. 2021b, A&A, 646, A110, [arXiv:2008.10393]

Hudelot, P., Cuillandre, J. C., Withington, K., et al. 2012, VizieR Online Data Catalog, II/317

Ibata, R. A., McConnachie, A., Cuillandre, J.-C., et al. 2017, ApJ, 848, 128, [arXiv:1708.06356]

Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841, [arXiv:astro-ph/0603217]

Inada, N., Oguri, M., Becker, R. H., et al. 2008, AJ, 135, 496, [arXiv:0708.0828]

Inada, N., Oguri, M., Shin, M.-S., et al. 2010, AJ, 140, 403, [arXiv:1005.5570]

Inada, N., Oguri, M., Shin, M.-S., et al. 2012, AJ, 143, 119, [arXiv:1203.1087]

Inami, H., Bacon, R., Brinchmann, J., et al. 2017, A&A, 608, A2, [arXiv:1710.03773]

Irwin, J. & Shmakova, M. 2005, New Astronomy Reviews, 49, 83

Ivanov, M. M., Simonović, M., & Zaldarriaga, M. 2020, JCAP, 2020, 042, [arXiv:1909.05277]

Ivezic, Z., Axelrod, T., Brandt, W. N., et al. 2008, Serbian Astronomical Journal, 176, 1

Izbicki, R., Lee, A. B., & Freeman, P. E. 2016, arXiv e-prints, arXiv:1604.01339, [arXiv:1604.01339]

Jacobs, C., Collett, T., Glazebrook, K., et al. 2019, ApJS, 243, 17, [arXiv:1905.10522]

Jacobs, C., Glazebrook, K., Collett, T., More, A., & McCarthy, C. 2017, MNRAS, 471, 167, [arXiv:1704.02744]

Jaelani, A. T., More, A., Oguri, M., et al. 2020a, MNRAS, 495, 1291, [arXiv:2002.01611]

Jaelani, A. T., More, A., Sonnenfeld, A., et al. 2020b, MNRAS, 494, 3156, [arXiv:1909.00120]

Jaelani, A. T., Rusu, C. E., Kayo, I., et al. 2021, MNRAS, 502, 1487, [arXiv:2006.16584]

Jaelani et al., A. T. in prep.

Jain, B. & Seljak, U. 1997, ApJ, 484, 560, [arXiv:astro-ph/9611077]

Jaroszynski, M., Park, C., Paczynski, B., & Gott, J. Richard, I. 1990, ApJ, 365, 22

Jee, M. J., White, R. L., Ford, H. C., et al. 2005, ApJ, 634, 813, [arXiv:astro-ph/0508044]

Johnston, D. C. & Welsh, B. M. 1994, Journal of the Optical Society of America A, 11, 394

Joye, W. A. & Mandel, E. 2003, in Astronomical Society of the Pacific Conference Series, Vol. 295, Astronomical Data Analysis Software and Systems XII, ed. H. E. Payne, R. I. Jedrzejewski, & R. N. Hook, 489

Jullo, E., Kneib, J. P., Limousin, M., et al. 2007, New Journal of Physics, 9, 447, [arXiv:0706.0048]

184

Kawanomoto, S., Uraguchi, F., Komiyama, Y., et al. 2018, PASJ, 70, 66

Kelly, P. L., Rodney, S. A., Treu, T., et al. 2015, Science, 347, 1123, [arXiv:1411.6009]

Kelly, P. L., Rodney, S. A., Treu, T., et al. 2016, ApJL, 819, L8, [arXiv:1512.04654]

King, L. J., Schneider, P., & Springel, V. 2001, A&A, 378, 748

Kinney, A. L. & Maran, S. P. 1991, PASP, 103, 1237

Knabel, S., Steele, R. L., Holwerda, B. W., et al. 2020, AJ, 160, 223, [arXiv:2009.09493]

Kochanek, C. S. 2004, ApJ, 605, 58, [arXiv:astro-ph/0307422]

Komiyama, Y., Obuchi, Y., Nakaya, H., et al. 2018, PASJ, 70, S2

Koo, D. C. 1985, AJ, 90, 418

Koopmans, L. V. E., Treu, T., Bolton, A. S., Burles, S., & Moustakas, L. A. 2006, ApJ, 649, 599, [arXiv:astro-ph/0601628]

Kormann, R., Schneider, P., & Bartelmann, M. 1994, A&A, 284, 285

Kourkchi, E., Tully, R. B., Anand, G. S., et al. 2020, ApJ, 896, 3, [arXiv:2004.14499]

Kriek, M., van Dokkum, P. G., Labbé, I., et al. 2009, ApJ, 700, 221, [arXiv:0905.1692]

Krone-Martins, A., Ishida, E. E. O., & de Souza, R. S. 2014, MNRAS, 443, L34, [arXiv:1308.4145]

Krywult, J., Tasca, L. A. M., Pollo, A., et al. 2017, A&A, 598, A120, [arXiv:1605.05502]

Kunz, M., Nesseris, S., & Sawicki, I. 2016, PRD, 94, 023510, [arXiv:1604.05701]

Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, ApJS, 224, 24, [arXiv:1604.02350]

Laigle, C., Pichon, C., Arnouts, S., et al. 2018, MNRAS, 474, 5437, [arXiv:1702.08810]

Lanusse, F., Ma, Q., Li, N., et al. 2018, MNRAS, 473, 3895, [arXiv:1703.02642]

Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, arXiv:1110.3193, [arXiv:1110.3193]

Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, Astronomy & Astrophysics, 559, A14

Le Fèvre, O., Saisse, M., Mancini, D., et al. 2003, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes, ed. M. Iye & A. F. M. Moorwood, 1670–1681

Leal-Taixé, L., Canton Ferrer, C., & Schindler, K. 2016, arXiv e-prints, arXiv:1604.07866, [arXiv:1604.07866]

Leavitt, H. S. 1908, Annals of Harvard College Observatory, 60, 87

Leavitt, H. S. & Pickering, E. C. 1912, Harvard College Observatory Circular, 173, 1

Lecun, Y., Bengio, Y., & Hinton, G. 2015, Nature, 521, 436

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, Proceedings of the IEEE, 86, 2278

Leistedt, B. & Hogg, D. W. 2017, ApJ, 838, 5, [arXiv:1612.00847]

Leistedt, B., Hogg, D. W., Wechsler, R. H., & DeRose, J. 2019, ApJ, 881, 80, [arXiv:1807.01391]

Lemaître, G. 1927, Annales de la Société Scientifique de Bruxelles, 47, 49

Lemon, C. A., Auger, M. W., & McMahon, R. G. 2019, MNRAS, 483, 4242, [arXiv:1810.04480]

Lemon, C. A., Auger, M. W., McMahon, R. G., & Koposov, S. E. 2017, MNRAS, 472, 5023, [arXiv:1709.08976]

Lemon, C. A., Auger, M. W., McMahon, R. G., & Ostrovski, F. 2018, MNRAS, 479, 5060, [arXiv:1803.07601]

Li, C., Zhang, Y., Cui, C., et al. 2021, MNRAS, [arXiv:2110.14951]

Li, L.-L., Zhang, Y.-X., Zhao, Y.-H., & Yang, D.-W. 2007, CJAA, 7, 448

Li, Z. & Scheraga, H. A. 1987, Proceedings of the National Academy of Sciences, 84, 6611, [https://www.pnas.org/content/84/19/6611.full.pdf]

Liao, K., Shafieloo, A., Keeley, R. E., & Linder, E. V. 2019, ApJL, 886, L23, [arXiv:1908.04967]

Liao, K., Shafieloo, A., Keeley, R. E., & Linder, E. V. 2020, ApJL, 895, L29, [arXiv:2002.10605]

Lilly, S. J., Le Brun, V., Maier, C., et al. 2009, ApJS, 184, 218

Lima, M., Cunha, C. E., Oyaizu, H., et al. 2008, MNRAS, 390, 118, [arXiv:0801.3822]

Liske, J., Baldry, I. K., Driver, S. P., et al. 2015, Monthly Notices of the Royal Astronomical Society, 452, 2087–2126

Loh, E. D. & Spillar, E. J. 1986, ApJ, 303, 154

Lombardi, M. & Bertin, G. 1998, A&A, 335, 1, [arXiv:astro-ph/9801244]

López Fune, E. 2018, MNRAS, 475, 2132, [arXiv:1712.03880]

Loubser, S. I., Babul, A., Hoekstra, H., et al. 2020, MNRAS, [arXiv:2006.05706]

LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv e-prints, arXiv:0912.0201, [arXiv:0912.0201]

Luo, A.-L., Zhao, Y.-H., Zhao, G., et al. 2015, Research in Astronomy and Astrophysics, 15, 1095–1124

Mackay, D. J. C. 1992

Mackay, D. J. C. 2003, Information Theory, Inference and Learning Algorithms

Madireddy, S., Li, N., Ramachandra, N., et al. 2019, arXiv e-prints, arXiv:1911.03867, [arXiv:1911.03867]

Maller, A. H., Simard, L., Guhathakurta, P., et al. 2000, ApJ, 533, 194, [arXiv:astro-ph/9910207]

Mannucci, F., Basile, F., Poggianti, B. M., et al. 2001, MNRAS, 326, 745, [arXiv:astro-ph/0104427]

Mao, S. & Paczynski, B. 1991, ApJL, 374, L37

Marianer, T., Poznanski, D., & Prochaska, J. X. 2021, MNRAS, 500, 5408, [arXiv:2010.11949]

Marshall, P. J., Verma, A., More, A., et al. 2016, MNRAS, 455, 1171, [arXiv:1504.06148]

Masters, D., Capak, P., Stern, D., et al. 2015, ApJ, 813, 53, [arXiv:1509.03318]

Maturi, M., Mizera, S., & Seidel, G. 2014, A&A, 567, A111, [arXiv:1305.3608]

McGaugh, S. S. 2016, ApJ, 816, 42, [arXiv:1511.06387]

McGreer, I. D., Clément, B., Mainali, R., et al. 2018, MNRAS, 479, 435, [arXiv:1706.09428]

McLure, R. J., Pearce, H. J., Dunlop, J. S., et al. 2012, Monthly Notices of the Royal Astronomical Society, 428, 1088–1106

Medling, A. M., Kewley, L. J., Calzetti, D., et al. 2021, arXiv e-prints, arXiv:2111.01025, [arXiv:2111.01025]

Medwadowski, S. J. 1989, APSS, 160, 33

Meneghetti, M., Bartelmann, M., & Moscardini, L. 2003, 340, 105, [astro-ph/0201501]

Meng, X.-L., Treu, T., Agnello, A., et al. 2015, JCAP, 2015, 059, [arXiv:1506.07640]

Metcalf, R. B., Meneghetti, M., Avestruz, C., et al. 2019, A&A, 625, A119, [arXiv:1802.03609]

Millon, M., Courbin, F., Bonvin, V., et al. 2020a, A&A, 640, A105, [arXiv:2002.05736]

Millon, M., Galan, A., Courbin, F., et al. 2020b, A&A, 639, A101, [arXiv:1912.08027]

Miyazaki, S., Komiyama, Y., Kawanomoto, S., et al. 2018, PASJ, 70, S1

Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, The Astrophysical Journal Supplement Series, 225, 27

Morales, M. D., Antelis, J. M., Moreno, C., & Nesterov, A. I. 2021, Sensors, 21, 3174, [arXiv:2009.04088]

More, A., Cabanac, R., More, S., et al. 2012, ApJ, 749, 38, [arXiv:1109.1821]

More, A., Verma, A., Marshall, P. J., et al. 2016, MNRAS, 455, 1191, [arXiv:1504.05587]

More et al., A. in prep.

Morgan, C. W., Hyer, G. E., Bonvin, V., et al. 2018, ApJ, 869, 106, [arXiv:1812.05639]

Morgan, C. W., Kochanek, C. S., Morgan, N. D., & Falco, E. E. 2010, ApJ, 712, 1129, [arXiv:1002.4160]

Morganson, E., Gruendl, R. A., Menanteau, F., et al. 2018, PASP, 130, 074501, [arXiv:1801.03177]

Morningstar, W. R., Hezaveh, Y. D., Perreault Levasseur, L., et al. 2018, arXiv e-prints, arXiv:1808.00011, [arXiv:1808.00011]

Morningstar, W. R., Perreault Levasseur, L., Hezaveh, Y. D., et al. 2019, ApJ, 883, 14, [arXiv:1901.01359]

Mortlock, D. J. & Webster, R. L. 2000, MNRAS, 319, 860, [arXiv:astro-ph/0008079]

Mu, Y.-H., Qiu, B., Zhang, J.-N., Ma, J.-C., & Fan, X.-D. 2020, Research in Astronomy and Astrophysics, 20, 089

Mukherjee, S., Ghosh, A., Graham, M. J., et al. 2020, arXiv e-prints, arXiv:2009.14199, [arXiv:2009.14199]

Mukherjee, S., Lavaux, G., Bouchet, F. R., et al. 2021, A&A, 646, A65, [arXiv:1909.08627]

Mullen, K. M. 2014, Journal of Statistical Software, Articles, 60, 1

Nair, V. & Hinton, G. E. 2010, in ICML, ed. J. Fürnkranz & T. Joachims (Omnipress), 807–814

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493, [arXiv:astro-ph/9611107]

Neichel, B., Rigaut, F., Vidal, F., et al. 2014, MNRAS, 440, 1002, [arXiv:1402.6906]

Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, The Astrophysical Journal Supplement Series, 208, 5

Nielsen, E. L., Rosa, R. J. D., Rameau, J., et al. 2017, AJ, 154, 218, [arXiv:1705.06851]

Nightingale, J., Hayes, R., & Griffiths, M. 2021a, The Journal of Open Source Software, 6, 2550, [arXiv:2102.04472]

Nightingale, J., Hayes, R., Kelly, A., et al. 2021b, The Journal of Open Source Software, 6, 2825, [arXiv:2106.01384]

Nightingale, J. W., Dye, S., & Massey, R. J. 2018, MNRAS, 478, 4738, [arXiv:1708.07377]

Nikolaus, K. S. & Hundertmark, M. 2018, arXiv e-prints, arXiv:1804.10136, [arXiv:1804.10136]

Nishizawa, A. J., Hsieh, B.-C., Tanaka, M., & Takata, T. 2020, arXiv e-prints, arXiv:2003.01511, [arXiv:2003.01511]

Nomoto, K. 1982, ApJ, 257, 780

Oguri, M. 2015, MNRAS, 449, L86, [arXiv:1411.6443]

Oguri, M. 2019, Reports on Progress in Physics, 82, 126901, [arXiv:1907.06830]

Oguri, M., Inada, N., Pindor, B., et al. 2006, AJ, 132, 999, [arXiv:astro-ph/0605571]

Oguri, M., Inada, N., Strauss, M. A., et al. 2008, AJ, 135, 512, [arXiv:0708.0825]

Ostrovski, F., McMahon, R. G., Connolly, A. J., et al. 2017, MNRAS, 465, 4325, [arXiv:1607.01391]

Oyaizu, H., Lima, M., Cunha, C. E., et al. 2008, ApJ, 674, 768, [arXiv:0708.0030]

Paczynski, B. 1986, ApJ, 304, 1

Paczynski, B. 1991, ApJL, 371, L63

Pakmor, R., Röpke, F., Hillebrandt, W., et al. 2010, in Progenitors and Environments of Stellar Explosions, 62

Pasquet, J., Bertin, E., Treyer, M., Arnouts, S., & Fouchez, D. 2019, A&A, 621, A26, [arXiv:1806.06607]

Pasquet-Itam, J. & Pasquet, J. 2018, A&A, 611, A97, [arXiv:1712.02777]

Pearson, J., Li, N., & Dye, S. 2019, MNRAS, 488, 991, [arXiv:1904.06199]

Pearson, J., Maresca, J., Li, N., & Dye, S. 2021, arXiv e-prints, arXiv:2103.03257, [arXiv:2103.03257]

Perreault Levasseur, L., Hezaveh, Y. D., & Wechsler, R. H. 2017, ApJL, 850, L7, [arXiv:1708.08843]

Pesce, D. W., Braatz, J. A., Reid, M. J., et al. 2020, ApJL, 891, L1, [arXiv:2001.09213]

Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2019a, MNRAS, 482, 807, [arXiv:1807.04764]

Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2017, MNRAS, 472, 1129, [arXiv:1702.07675]

Petrillo, C. E., Tortora, C., Vernardos, G., et al. 2019b, MNRAS, 484, 3879, [arXiv:1812.03168]

Petters, A. O. 1995a, Journal of Mathematical Physics, 36, 4263

Petters, A. O. 1995b, Journal of Mathematical Physics, 36, 4276

Philcox, O. H. E., Ivanov, M. M., Simonović, M., & Zaldarriaga, M. 2020, JCAP, 2020, 032, [arXiv:2002.04035]

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6, [arXiv:1807.06209]

Pogosian, L., Zhao, G.-B., & Jedamzik, K. 2020, ApJL, 904, L17, [arXiv:2009.08455]

Powell, D., Vegetti, S., McKean, J. P., et al. 2021, MNRAS, 501, 515, [arXiv:2005.03609]

Primack, J. R. 1997, arXiv e-prints, astro, [arXiv:astro-ph/9707285]

Qi, J.-Z., Zhao, J.-W., Cao, S., Biesiada, M., & Liu, Y. 2021, MNRAS, 503, 2179, [arXiv:2011.00713]

Quider, A. M., Pettini, M., Shapley, A. E., & Steidel, C. C. 2009, MNRAS, 398, 1263, [arXiv:0906.2412]

Razim, O., Cavuoti, S., Brescia, M., et al. 2021, MNRAS, 507, 5034, [arXiv:2108.04784]

Refsdal, S. 1964, MNRAS, 128, 307

Reid, M. J., Pesce, D. W., & Riess, A. G. 2019, ApJL, 886, L27, [arXiv:1908.05625]

Renzini, A. & Daddi, E. 2009, Journeying the Redshift Desert, [arXiv:0906.4662]

Richards, J. W., Homrighausen, D., Freeman, P. E., Schafer, C. M., & Poznanski, D. 2012, MNRAS, 419, 1121, [arXiv:1103.6034]

Riess, A. G., Casertano, S., Yuan, W., et al. 2021, ApJL, 908, L6, [arXiv:2012.08534]

Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., & Scolnic, D. 2019, ApJ, 876, 85, [arXiv:1903.07603]

Rigaut, F., Neichel, B., Boccas, M., et al. 2014, MNRAS, 437, 2361, [arXiv:1310.6199]

Ritondale, E., Vegetti, S., Despali, G., et al. 2019, MNRAS, 485, 2179, [arXiv:1811.03627]

Rizzo, F., Vegetti, S., Fraternali, F., & Di Teodoro, E. 2018, MNRAS, 481, 5606, [arXiv:1809.07340]

Robert, C. P. & Casella, G. 2004, 677

Rodney, S. A., Brammer, G. B., Pierel, J. D. R., et al. 2021, Nature Astronomy, [arXiv:2106.08935]

Rojas, K., Savary, E., Clément, B., et al. 2021, arXiv e-prints, arXiv:2109.00014, [arXiv:2109.00014]

Romanowsky, A. J. & Kochanek, C. S. 1999, ApJ, 516, 18, [arXiv:astro-ph/9805080]

Rousseeuw, P. J. & Croux, C. 1993, Journal of the American Statistical Association, 88, 1273

Rubin, D., Hayden, B., Huang, X., et al. 2018, ApJ, 866, 65, [arXiv:1707.04606]

Russakovsky, O., Deng, J., Su, H., et al. 2015, ImageNet Large Scale Visual Recognition Challenge, [arXiv:1409.0575]

Rusu, C. E., Wong, K. C., Bonvin, V., et al. 2020, MNRAS, [arXiv:1905.09338]

Sadeh, I., Abdalla, F. B., & Lahav, O. 2016, PASP, 128, 104502, [arXiv:1507.00490]

Salmon, B., Coe, D., Bradley, L., et al. 2018, ApJL, 864, L22, [arXiv:1801.03103]

Sarmiento, R., Huertas-Company, M., Knapen, J. H., et al. 2021, ApJ, 921, 177, [arXiv:2104.08292]

Savary, E., Rojas, K., Maus, M., et al. 2021, arXiv e-prints, arXiv:2110.11972, [arXiv:2110.11972]

Schaefer, C., Geiger, M., Kuntzer, T., & Kneib, J. P. 2018, A&A, 611, A2, [arXiv:1705.07132]

Schlegel, D., Finkbeiner, D., & Davis, M. 1998, in Wide Field Surveys in Cosmology, ed. S. Colombi, Y. Mellier, & B. Raban, Vol. 14, 297, [arXiv:astro-ph/9809230]

Schmidt, M. & Lipson, H. 2009, Science, 324, 81

Schmidt, S. J., Malz, A. I., Soo, J. Y. H., et al. 2020, arXiv e-prints, arXiv:2001.03621, [arXiv:2001.03621]

Schneider, P. 2005, arXiv e-prints, astro, [arXiv:astro-ph/0509252]

Schneider, P. 2019, A&A, 624, A54

Schneider, P., Kochanek, C., & Wambsganss, J. 2006, Gravitational Lensing: Strong, Weak and Micro - Saas-Fee Advanced Course 33 (Berlin Heidelberg: Springer Science & Business Media)

Schneider, P. & Seitz, C. 1995, A&A, 294, 411, [arXiv:astro-ph/9407032]

Schneider, P. & Sluse, D. 2013, A&A, 559, A37, [arXiv:1306.0901]

Schneider, P., van Waerbeke, L., Mellier, Y., et al. 1998, A&A, 333, 767, [arXiv:astro-ph/9705122]

Schombert, J. & McGaugh, S. 2014, PASA, 31, e036, [arXiv:1407.6778]

Schombert, J., McGaugh, S., & Lelli, F. 2020, AJ, 160, 71, [arXiv:2006.08615]

Schuldt, S. 2018

Schuldt, S., Chirivì, G., Suyu, S. H., et al. 2019, A&A, 631, A40, [arXiv:1901.02896]

Schuldt, S., Suyu, S. H., Cañameras, R., et al. 2021a, A&A, 651, A55, [arXiv:2011.12312]

Schuldt, S., Suyu, S. H., Meinhardt, T., et al. 2021b, A&A, 646, A126, [arXiv:2010.00602]

Schuldt et al. in prep a

Schuldt et al. in prep b

Sciortino, F., Howard, N. T., Marmar, E. S., et al. 2020, arXiv e-prints, arXiv:2006.06798, [arXiv:2006.06798]

Seidel, G. & Bartelmann, M. 2007, A&A, 472, 341, [arXiv:astro-ph/0607547]

Sérsic, J. L. 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41

Shajib, A. J., Birrer, S., Treu, T., et al. 2020, MNRAS, 494, 6072, [arXiv:1910.06306]

Shajib, A. J., Birrer, S., Treu, T., et al. 2019, MNRAS, 483, 5649, [arXiv:1807.09278]

Sharon, K. & Johnson, T. L. 2015, ApJL, 800, L26, [arXiv:1411.6933]

Shu, Y., Bolton, A. S., Kochanek, C. S., et al. 2016a, ApJ, 824, 86, [arXiv:1604.01842]

Shu, Y., Bolton, A. S., Mao, S., et al. 2016b, ApJ, 833, 264, [arXiv:1608.08707]

Shu, Y., Brownstein, J. R., Bolton, A. S., et al. 2017, ApJ, 851, 48, [arXiv:1711.00072]

Shu, Y., Marques-Chaves, R., Evans, N. W., & Pérez-Fournon, I. 2018, MNRAS, 481, L136, [arXiv:1809.07337]

Shu et al. in prep.

Sim, S. A., Röpke, F. K., Hillebrandt, W., et al. 2010, ApJL, 714, L52, [arXiv:1003.2917]

Singal, J., Shmakova, M., Gerke, B., Griffith, R. L., & Lotz, J. 2011, PASP, 123, 615, [arXiv:1101.4011]

Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, The Astrophysical Journal Supplement Series, 214, 24

Slipher, V. M. 1913, Lowell Observatory Bulletin, 1, 56

Slipher, V. M. 1915, Popular Astronomy, 23, 21

Slipher, V. M. 1917a, Proceedings of the American Philosophical Society, 56, 403

Slipher, V. M. 1917b, Lowell Observatory Bulletin, 3, 59

Slipher, V. M. 1918, PASP, 30, 63

Slipher, V. M. 1921, Popular Astronomy, 29, 128

Smith, K. W., Williams, R. D., Young, D. R., et al. 2019, Research Notes of the AAS, 3, 26

Soltis, J., Casertano, S., & Riess, A. G. 2021, ApJL, 908, L5, [arXiv:2012.09196]

Sonnenfeld, A., Chan, J. H. H., Shu, Y., et al. 2018a, PASJ, 70, S29, [arXiv:1704.01585]

Sonnenfeld, A., Gavazzi, R., Suyu, S. H., Treu, T., & Marshall, P. J. 2013, ApJ, 777, 97, [arXiv:1307.4764]

Sonnenfeld, A., Jaelani, A. T., Chan, J., et al. 2019, A&A, 630, A71, [arXiv:1904.10465]

Sonnenfeld, A., Leauthaud, A., Auger, M. W., et al. 2018b, MNRAS, 481, 164, [arXiv:1801.01883]

Sonnenfeld, A., Treu, T., Marshall, P. J., et al. 2015, ApJ, 800, 94, [arXiv:1410.1881]

Sonnenfeld, A., Verma, A., More, A., et al. 2020, A&A, 642, A148, [arXiv:2004.00634]

Soo, J. Y. H., Joachimi, B., Eriksen, M., et al. 2021, MNRAS, 503, 4118, [arXiv:2101.03723]

Soo, J. Y. H., Moraes, B., Joachimi, B., et al. 2018, MNRAS, 475, 3613, [arXiv:1707.03169]

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014, Journal of Machine Learning Research, 15, 1929

Stafford, S. G., McCarthy, I. G., Kwan, J., et al. 2021, MNRAS, 508, 2537, [arXiv:2109.11956]

Steidel, C. C., Shapley, A. E., Pettini, M., et al. 2004, ApJ, 604, 534, [arXiv:astro-ph/0401439]

Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, AJ, 123, 485

Strigari, L. E. 2013, PhysRep, 531, 1, [arXiv:1211.7090]

Suyu, S. H. 2012, MNRAS, 426, 868, [arXiv:1202.0287]

Suyu, S. H., Auger, M. W., Hilbert, S., et al. 2013, ApJ, 766, 70, [arXiv:1208.6010]

Suyu, S. H. & Blandford, R. D. 2006, MNRAS, 366, 39, [arXiv:astro-ph/0506629]

Suyu, S. H., Bonvin, V., Courbin, F., et al. 2017, MNRAS, 468, 2590, [arXiv:1607.00017]

Suyu, S. H. & Halkola, A. 2010, A&A, 524, A94, [arXiv:1007.4815]

Suyu, S. H., Hensel, S. W., McKean, J. P., et al. 2012, ApJ, 750, 10, [arXiv:1110.2536]

Suyu, S. H., Huber, S., Cañameras, R., et al. 2020, arXiv e-prints, arXiv:2002.08378, [arXiv:2002.08378]

Suyu, S. H., Marshall, P. J., Auger, M. W., et al. 2010, ApJ, 711, 201, [arXiv:0910.2773]

Suyu, S. H., Marshall, P. J., Hobson, M. P., & Blandford, R. D. 2006, MNRAS, 371, 983, [astro-ph/0601493]

Suyu, S. H., Treu, T., Hilbert, S., et al. 2014, ApJL, 788, L35, [arXiv:1306.4732]

Tagliaferri, R., Longo, G., Andreon, S., et al. 2003, Neural Networks for Photometric Redshifts Evaluation, Vol. 2859, 226–234

Talbot, M. S., Brownstein, J. R., Bolton, A. S., et al. 2018, MNRAS, 477, 195, [arXiv:1803.03604]

Talbot, M. S., Brownstein, J. R., Dawson, K. S., Kneib, J.-P., & Bautista, J. 2021, MNRAS, 502, 4617, [arXiv:2007.09006]

Tanaka, M., Coupon, J., Hsieh, B.-C., et al. 2018, PASJ, 70, S9, [arXiv:1704.05988]

Tanoglidis, D., Drlica-Wagner, A., Wei, K., et al. 2020, arXiv e-prints, arXiv:2006.04294, [arXiv:2006.04294]

Taubenberger, S., Suyu, S. H., Komatsu, E., et al. 2019, A&A, 628, L7, [arXiv:1905.12496]

Taubenberger et al. in prep.

Tihhonova, O., Courbin, F., Harvey, D., et al. 2020, MNRAS, 498, 1406, [arXiv:2005.12295]

Treu, T. 2010, ARA&A, 48, 87, [arXiv:1003.5567]

Treu, T., Brammer, G., Diego, J. M., et al. 2016, ApJ, 817, 60, [arXiv:1510.05750]

Treu, T., Dutton, A. A., Auger, M. W., et al. 2011, MNRAS, 417, 1601, [arXiv:1104.5663]

Treu, T. & Ellis, R. S. 2015, Contemporary Physics, 56, 17, [arXiv:1412.6916]

Treu, T. & Koopmans, L. V. E. 2002a, arXiv e-prints, astro, [arXiv:astro-ph/0205335]

Treu, T. & Koopmans, L. V. E. 2002b, MNRAS, 337, L6, [arXiv:astro-ph/0210002]

Treu, T. & Koopmans, L. V. E. 2004, ApJ, 611, 739, [arXiv:astro-ph/0401373]

Tsallis, C. 1988, J. Statist. Phys., 52, 479

Tsallis, C. & Stariolo, D. A. 1996, Physica A: Statistical Mechanics and its Applications, 233, 395–406

Tsapras, Y., Street, R. A., Horne, K., et al. 2001, MNRAS, 325, 1205, [arXiv:astro-ph/0103517]

Tucker, W., Blanco, P., Rappoport, S., et al. 1998, ApJL, 496, L5, [arXiv:astro-ph/9801120]

Turner, E. L., Ostriker, J. P., & Gott, J. R., I. 1984, ApJ, 284, 1

Tyson, J. A., Valdes, F., & Wenk, R. A. 1990, ApJL, 349, L1

Udalski, A., Szymanski, M., Kaluzny, J., et al. 1993, Acta Astronomica, 43, 289

van de Ven, G., Falcón-Barroso, J., McDermid, R. M., et al. 2010, ApJ, 719, 1481, [arXiv:0807.4175]

Vayner, A., Wright, S. A., Murray, N., et al. 2021, ApJ, 919, 122, [arXiv:2106.08337]

Vegetti, S. & Koopmans, L. V. E. 2009a, MNRAS, 392, 945, [arXiv:0805.0201]

Vegetti, S. & Koopmans, L. V. E. 2009b, MNRAS, 400, 1583, [arXiv:0903.4752]

Vegetti, S., Koopmans, L. V. E., Auger, M. W., Treu, T., & Bolton, A. S. 2014, MNRAS, 442, 2017, [arXiv:1405.3666]

Vegetti, S. & Vogelsberger, M. 2014, MNRAS, 442, 3598, [arXiv:1406.1170]

Villar, V. A., Cranmer, M., Berger, E., et al. 2021, ApJS, 255, 24, [arXiv:2103.12102]

Wadadekar, Y. 2005, PASP, 117, 79, [arXiv:astro-ph/0412005]

Wales, D. J. 2003, Energy Landscapes (Cambridge University Press)

Wales, D. J. & Doye, J. P. K. 1997, The Journal of Physical Chemistry A, 101, 5111–5116

Wales, D. J. & Scheraga, H. A. 1999, Science, 285, 1368, [https://www.science.org/doi/pdf/10.1126/science.285.5432.1368]

Walsh, D., Carswell, R. F., & Weymann, R. J. 1979, Nature, 279, 381

Wambsganss, J., Cen, R., Xu, G., & Ostriker, J. P. 1997, ApJL, 475, L81

Wang, C., Ma, N., Ming, Y., Wang, Q., & Xia, J. 2019, Advances in Space Research, 64, 886

Wang, D., Zhang, Y.-X., Liu, C., & Zhao, Y.-H. 2008, CJAA, 8, 119, [arXiv:0707.2250]

Wang, Y., Bahcall, N., & Turner, E. L. 1998, AJ, 116, 2081, [arXiv:astro-ph/9804195]

Warren, S. J. & Dye, S. 2003, ApJ, 590, 673, [arXiv:astro-ph/0302587]

Way, M. J., Foster, L. V., Gazis, P. R., & Srivastava, A. N. 2009, ApJ, 706, 623, [arXiv:0905.4081]

Way, M. J. & Klose, C. D. 2012, PASP, 124, 274, [arXiv:1201.1098]

Whelan, J. & Iben, Icko, J. 1973, ApJ, 186, 1007

Wilson, D., Nayyeri, H., Cooray, A., & Häußler, B. 2020, ApJ, 888, 83, [arXiv:1911.00210]

Wizinowich, P. L., Le Mignant, D., Bouchez, A. H., et al. 2006, PASP, 118, 297

Wojtak, R., Hjorth, J., & Gall, C. 2019, MNRAS, 487, 3342, [arXiv:1903.07687]

Wolf, C. 2009, MNRAS, 397, 520, [arXiv:0904.3438]

Wong, K. C., Keeton, C. R., Williams, K. A., Momcheva, I. G., & Zabludoff, A. I. 2011, ApJ, 726, 84, [arXiv:1011.2504]

Wong, K. C., Sonnenfeld, A., Chan, J. H. H., et al. 2018, ApJ, 867, 107, [arXiv:1809.07341]

Wong, K. C., Suyu, S. H., Chen, G. C. F., et al. 2020, MNRAS, [arXiv:1907.04869]

Wucknitz, O. 2002, MNRAS, 332, 951, [arXiv:astro-ph/0202376]

Xiang, Y. & Gong, X. G. 2000, Phys. Rev. E, 62, 4473

Xiang, Y., Gubian, S., Suomela, B., & Hoeng, J. 2013, The R Journal Volume 5(1):13-29, June 2013, 5

Xiang, Y., Sun, D. Y., Fan, W., & Gong, X. G. 1997, Physics Letters A, 233, 216

Yahalomi, D. A., Schechter, P. L., & Wambsganss, J. 2017, arXiv e-prints, arXiv:1711.07919, [arXiv:1711.07919]

Yıldırım, A., Suyu, S. H., Chen, G. C. F., & Komatsu, E. 2021, arXiv e-prints, arXiv:2109.14615, [arXiv:2109.14615]

Yıldırım, A., Suyu, S. H., & Halkola, A. 2020, MNRAS, 493, 4783, [arXiv:1904.07237]

York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, AJ, 120, 1579, [arXiv:astro-ph/0006396]

Yuan, W., Riess, A. G., Macri, L. M., Casertano, S., & Scolnic, D. M. 2019, ApJ, 886, 61, [arXiv:1908.00993]

Zettl, A. 2010, Sturm-Liouville Theory (American Mathematical Society)

Zhang, X. & Huang, Q.-G. 2019, Communications in Theoretical Physics, 71, 826

Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, Research in Astronomy and Astrophysics, 12, 723

Zhou, R., Cooper, M. C., Newman, J. A., et al. 2019, MNRAS, 488, 4565, [arXiv:1903.08174]

Zwicky, F. 1933, Helvetica Physica Acta, 6, 110

Zwicky, F. 1937, Phys. Rev., 51, 290

Zwicky, F. 1937, ApJ, 86, 217

Zwicky, F. 1937, Phys. Rev., 51, 679

A Model details of the 32 SuGOHI lens systems

In this chapter we provide the best-fit parameter values for the 32 HSC SuGOHI lenses obtained through our modeling with GLEE & GLaDas described in Sect. 5.4. In detail, we give in Tab. A.1 the obtained χ^2 and $\chi^2_{\rm red}$ values of the final model. Moreover, we give the best.fit values for the external shear, and in Tab. A.2 the best-fit values of the corresponding SIE profile describing the mass distribution of the lens. The obtained best-fit values for the source light distribution are given in Tab. A.3 (amplitudes) and Tab. A.4 (structural parameters).We provide in Tab. A.5 the obtained parameter values of the Sérsic profiles describing the lens light distribution.

We further show here figures (Fig. A.1 - Fig. A.32) of all the 32 lenses to visualize the models. Each figure show one lens system, from left to right the observed, the model, and the normalized residuals, and from top to bottom the four different filters g, r, i, and z.

Table A.1: Achieved χ^2 values and best fit values for the external shear by modeling the 32 HSC SuGOHI lenses with GLEE & GLaD. Column 2 and 3 give, respectively, the χ^2 and the reduced χ^2, which is defined as χ^2 divided by the degrees of freedom (i.e. number of modeled pixel minus varying parameters). This is followed by the external shear strength $\gamma_{\rm ext}$ and its orientation $\theta_{\rm ext}$.

Name	χ^2	$\chi^2_{\rm red}$	$\gamma_{\rm ext}$	$\theta_{\rm ext}$ [rad]
HSCJ015618-010747	1.68×10^4	1.19	$0.084^{+0.005}_{-0.005}$	$-0.47^{+0.09}_{-0.08}$
HSCJ020141-030946	1.97×10^4	1.55	$0.073^{+0.003}_{-0.003}$	$0.76^{+0.02}_{-0.02}$
HSCJ020241-064611	2.08×10^4	1.35	$0.12^{+0.01}_{-0.03}$	$-0.5^{+0.1}_{-0.1}$
HSCJ020955-024442	1.70×10^4	1.14	$0.033^{+0.005}_{-0.005}$	$1.62^{+0.07}_{-0.07}$
HSCJ021737-051329	1.55×10^4	1.18	$0.067^{+0.02}_{-0.008}$	$1.61^{+0.02}_{-0.03}$
HSCJ022346-053418	1.66×10^4	1.22	$0.010^{+0.003}_{-0.002}$	$-0.5^{+0.1}_{-0.1}$
HSCJ022610-042011	1.90×10^4	1.46	$0.07^{+0.01}_{-0.02}$	$3.0^{+0.1}_{-0.1}$
HSCJ023217-021703	1.72×10^4	1.15	$0.040^{+0.005}_{-0.005}$	$2.18^{+0.09}_{-0.09}$
HSCJ023322-020530	2.52×10^4	1.87	$0.108^{+0.001}_{-0.001}$	$0.8214^{+0.005}_{-0.005}$
HSCJ085046+003905	1.97×10^4	1.36	$0.058^{+0.01}_{-0.005}$	$3.14^{+0.01}_{-0.04}$
HSCJ085855-010208	1.84×10^4	1.30	$0.074^{+0.003}_{-0.003}$	$1.12^{+0.02}_{-0.02}$
HSCJ090429-010228	1.71×10^4	1.22	$0.053^{+0.002}_{-0.002}$	$1.622^{+0.006}_{-0.01}$
HSCJ094427-014742	1.58×10^4	1.12	$0.085^{+0.005}_{-0.01}$	$-2.47^{+0.3}_{-0.07}$

Table A.1 Continued: Achieved χ^2 values and best fit values for the external shear by modeling the 32 HSC SuGOHI lense with GLEE & GLaD.

Name	χ^2	χ^2_{red}	γ_{ext}	θ_{ext} [rad]
HSCJ120623+001507	1.61×10^4	1.14	$0.120^{+0.009}_{-0.02}$	$0.98^{+0.09}_{-0.09}$
HSCJ121052-011905	1.92×10^4	1.29	$0.040^{+0.005}_{-0.005}$	$2.218^{+0.007}_{-0.007}$
HSCJ121504+004726	1.91×10^4	1.38	$0.080^{+0.002}_{-0.004}$	$1.96^{+0.07}_{-0.08}$
HSCJ124320-004517	1.58×10^4	1.17	$0.086^{+0.003}_{-0.07}$	$-2.02^{+0.03}_{-0.02}$
HSCJ125254+004356	1.70×10^4	1.24	$0.087^{+0.003}_{-0.006}$	$-2.98^{+0.04}_{-0.04}$
HSCJ135138+002839	1.82×10^4	1.65	$0.048^{+0.001}_{-0.001}$	$0.87^{+0.01}_{-0.01}$
HSCJ141136-010215	1.82×10^4	1.27	$0.032^{+0.006}_{-0.006}$	$3.05^{+0.05}_{-0.06}$
HSCJ141815+015832	1.62×10^4	1.17	$0.09^{+0.01}_{-0.01}$	$2.85^{+0.04}_{-0.06}$
HSCJ142720+001916	1.60×10^4	1.14	$0.090^{+0.003}_{-0.005}$	$1.06^{+0.05}_{-0.05}$
HSCJ144132-005358	1.97×10^4	1.69	$0.06^{+0.02}_{-0.01}$	$2.89^{+0.06}_{-0.02}$
HSCJ144320-012537	1.86×10^4	1.30	$0.058^{+0.001}_{-0.001}$	$2.27^{+0.02}_{-0.02}$
HSCJ145242+425731	1.56×10^4	1.24	$0.060^{+0.006}_{-0.005}$	$0.60^{+0.03}_{-0.03}$
HSCJ150021-004936	1.78×10^4	1.30	$0.102^{+0.001}_{-0.003}$	$3.045^{+0.006}_{-0.01}$
HSCJ150112+422113	2.03×10^4	1.63	$0.134^{+0.001}_{-0.001}$	$1.557^{+0.009}_{-0.008}$
HSCJ223733+005015	2.01×10^4	1.36	$0.114^{+0.002}_{-0.005}$	$3.126^{+0.007}_{-0.007}$
HSCJ230335+003703	1.91×10^4	1.26	$0.027^{+0.003}_{-0.003}$	$-0.59^{+0.03}_{-0.03}$
HSCJ230521-000211	2.08×10^4	1.53	$0.049^{+0.001}_{-0.001}$	$-0.59^{+0.02}_{-0.02}$
HSCJ233130+003733	2.10×10^4	1.45	$0.075^{+0.001}_{-0.001}$	$0.45^{+0.01}_{-0.01}$
HSCJ233146+013845	1.76×10^4	1.21	$0.084^{+0.003}_{-0.003}$	$1.78^{+0.01}_{-0.01}$

Table A.2: Obtained best fit parameter values for the SIE mass profile by modeling the 32 HSC SuGOHI lense with GLEE & GLaD. The SIE mass parameters are the lens center x_l and y_l, the lens mass axis ratio q_{lm}, its orientation θ_{lm}, and the Einstein radius θ_E.

Name	$x_l['']$	$y_l['']$	q_{lm}	θ_{lm} [rad]	$\theta_E['']$
HSCJ015618-010747	$5.3^{+0.001}_{-0.001}$	$5.446^{+0.001}_{-0.001}$	$0.42^{+0.02}_{-0.01}$	$0.32^{+0.05}_{-0.04}$	$0.99^{+0.02}_{-0.01}$
HSCJ020141-030946	$5.371^{+0.001}_{-0.001}$	$5.324^{+0.001}_{-0.001}$	$0.79^{+0.01}_{-0.01}$	$1.77^{+0.04}_{-0.04}$	$1.406^{+0.007}_{-0.007}$
HSCJ020241-064611	$5.369^{+0.001}_{-0.001}$	$5.437^{+0.001}_{-0.001}$	$0.78^{+0.08}_{-0.04}$	$2.7^{+0.2}_{-0.2}$	$1.26^{+0.03}_{-0.02}$
HSCJ020955-024442	$5.305^{+0.001}_{-0.001}$	$5.397^{+0.002}_{-0.002}$	$0.78^{+0.02}_{-0.02}$	$1.74^{+0.03}_{-0.03}$	$1.043^{+0.003}_{-0.003}$
HSCJ021737-051329	$5.323^{+0.001}_{-0.001}$	$5.448^{+0.001}_{-0.001}$	$0.793^{+0.02}_{-0.02}$	$0.08^{+0.08}_{-0.02}$	$1.252^{+0.007}_{-0.02}$
HSCJ022346-053418	$5.287^{+0.001}_{-0.001}$	$5.356^{+0.001}_{-0.001}$	$0.719^{+0.004}_{-0.004}$	$2.499^{+0.01}_{-0.008}$	$1.397^{+0.004}_{-0.003}$

Table A.2 Continued: Obtained best fit parameter values for the SIE mass profile by modeling the 32 HSC SuGOHI lense with GLEE & GLaD.

Name	$x_l['']$	$y_l['']$	q_{lm}	θ_{lm} [rad]	$\theta_E['']$
HSCJ022610-042011	$5.435^{+0.001}_{-0.001}$	$5.353^{+0.001}_{-0.001}$	$0.90^{+0.06}_{-0.05}$	$2.1^{+0.4}_{-0.3}$	$1.16^{+0.02}_{-0.02}$
HSCJ023217-021703	$5.318^{+0.001}_{-0.001}$	$5.331^{+0.001}_{-0.001}$	$0.65^{+0.02}_{-0.02}$	$1.03^{+0.03}_{-0.03}$	$1.345^{+0.009}_{-0.009}$
HSCJ023322-020530	$5.343^{+0.002}_{-0.002}$	$5.433^{+0.001}_{-0.001}$	$0.74^{+0.003}_{-0.001}$	$-2.284^{+0.006}_{-0.006}$	$1.669^{+0.001}_{-0.001}$
HSCJ085046+003905	$5.284^{+0.001}_{-0.001}$	$5.383^{+0.001}_{-0.001}$	$0.70^{+0.03}_{-0.02}$	$4.59^{+0.02}_{-0.02}$	$1.750^{+0.006}_{-0.01}$
HSCJ085855-010208	$5.369^{+0.001}_{-0.001}$	$5.391^{+0.001}_{-0.001}$	$0.769^{+0.006}_{-0.007}$	$1.04^{+0.02}_{-0.01}$	$1.065^{+0.002}_{-0.002}$
HSCJ090429-010228	$5.335^{+0.001}_{-0.001}$	$5.269^{+0.001}_{-0.001}$	$0.81^{+0.005}_{-0.005}$	$0.08^{+0.005}_{-0.007}$	$1.232^{+0.002}_{-0.002}$
HSCJ094427-014742	$5.371^{+0.001}_{-0.001}$	$5.458^{+0.001}_{-0.001}$	$0.49^{+0.02}_{-0.08}$	$1.59^{+0.06}_{-0.03}$	$1.08^{+0.05}_{-0.01}$
HSCJ120623+001507	$5.315^{+0.001}_{-0.001}$	$5.36^{+0.001}_{-0.001}$	$0.73^{+0.06}_{-0.04}$	$-1.83^{+0.08}_{-0.1}$	$1.07^{+0.01}_{-0.02}$
HSCJ121052-011905	$5.405^{+0.001}_{-0.001}$	$5.281^{+0.001}_{-0.001}$	$0.733^{+0.001}_{-0.001}$	$1.046^{+0.005}_{-0.005}$	$1.529^{+0.008}_{-0.008}$
HSCJ121504+004726	$5.324^{+0.001}_{-0.001}$	$5.399^{+0.001}_{-0.001}$	$0.96^{+0.03}_{-0.03}$	$1.2^{+0.2}_{-0.1}$	$1.398^{+0.008}_{-0.007}$
HSCJ124320-004517	$5.410^{+0.001}_{-0.001}$	$5.406^{+0.001}_{-0.001}$	$0.86^{+0.02}_{-0.02}$	$3.99^{+0.06}_{-0.08}$	$1.506^{+0.004}_{-0.005}$
HSCJ125254+004356	$5.336^{+0.001}_{-0.001}$	$5.442^{+0.001}_{-0.001}$	$0.91^{+0.01}_{-0.02}$	$1.8^{+0.1}_{-0.1}$	$1.899^{+0.007}_{-0.005}$
HSCJ135138+002839	$5.444^{+0.001}_{-0.001}$	$5.359^{+0.001}_{-0.001}$	$0.777^{+0.002}_{-0.002}$	$0.706^{+0.008}_{-0.008}$	$2.216^{+0.001}_{-0.001}$
HSCJ141136-010215	$5.32^{+0.001}_{-0.001}$	$5.442^{+0.001}_{-0.001}$	$0.77^{+0.01}_{-0.01}$	$0.54^{+0.02}_{-0.03}$	$1.081^{+0.003}_{-0.004}$
HSCJ141815+015832	$5.283^{+0.001}_{-0.001}$	$5.438^{+0.001}_{-0.001}$	$0.81^{+0.04}_{-0.04}$	$1.82^{+0.08}_{-0.07}$	$1.34^{+0.01}_{-0.01}$
HSCJ142720+001916	$5.253^{+0.001}_{-0.001}$	$5.285^{+0.001}_{-0.001}$	$0.73^{+0.03}_{-0.03}$	$1.23^{+0.04}_{-0.04}$	$1.491^{+0.009}_{-0.010}$
HSCJ144132-005358	$5.377^{+0.001}_{-0.001}$	$5.335^{+0.001}_{-0.001}$	$0.608^{+0.005}_{-0.007}$	$0.41^{+0.01}_{-0.01}$	$1.078^{+0.02}_{-0.006}$
HSCJ144320-012537	$5.410^{+0.001}_{-0.001}$	$5.349^{+0.001}_{-0.001}$	$0.665^{+0.004}_{-0.004}$	$0.580^{+0.008}_{-0.008}$	$1.206^{+0.001}_{-0.001}$
HSCJ145242+425731	$5.436^{+0.001}_{-0.001}$	$5.456^{+0.001}_{-0.001}$	$0.663^{+0.005}_{-0.007}$	$-2.75^{+0.02}_{-0.01}$	$1.99^{+0.01}_{-0.01}$
HSCJ150021-004936	$5.465^{+0.001}_{-0.001}$	$5.47^{+0.001}_{-0.001}$	$0.716^{+0.003}_{-0.004}$	$2.56^{+0.01}_{-0.02}$	$3.063^{+0.01}_{-0.008}$
HSCJ150112+422113	$5.31^{+0.001}_{-0.001}$	$5.384^{+0.001}_{-0.001}$	$0.531^{+0.008}_{-0.008}$	$0.255^{+0.005}_{-0.005}$	$1.117^{+0.002}_{-0.002}$
HSCJ223733+005015	$5.391^{+0.001}_{-0.001}$	$5.461^{+0.001}_{-0.001}$	$0.622^{+0.003}_{-0.003}$	$0.565^{+0.001}_{-0.001}$	$1.58^{+0.006}_{-0.004}$
HSCJ230335+003703	$5.359^{+0.001}_{-0.001}$	$5.441^{+0.001}_{-0.001}$	$0.992^{+0.002}_{-0.002}$	$3.0^{+0.1}_{-0.2}$	$1.020^{+0.003}_{-0.003}$
HSCJ230521-000211	$5.452^{+0.001}_{-0.001}$	$5.456^{+0.001}_{-0.001}$	$0.618^{+0.005}_{-0.005}$	$-0.085^{+0.005}_{-0.005}$	$1.777^{+0.003}_{-0.003}$
HSCJ233130+003733	$5.324^{+0.001}_{-0.001}$	$5.349^{+0.001}_{-0.001}$	$0.688^{+0.004}_{-0.004}$	$0.961^{+0.005}_{-0.005}$	$1.502^{+0.001}_{-0.001}$
HSCJ233146+013845	$5.380^{+0.001}_{-0.001}$	$5.346^{+0.001}_{-0.001}$	$0.882^{+0.006}_{-0.005}$	$-0.50^{+0.03}_{-0.03}$	$1.481^{+0.001}_{-0.001}$

Table A.3: Reconstructed amplitudes of the adopted Sérsic profiles representing the source light obtained with GLEE & GLaD. We give for the different components (column 2) the amplitudes A for all four modeled filters g, r, i and z.

Name	component	$A_{g,s,comp}$	$A_{r,s,comp}$	$A_{i,s,comp}$	$A_{z,s,comp}$
HSCJ015618-010747	1	93^{+18}_{-30}	116^{+22}_{-37}	83^{+28}_{-30}	238^{+47}_{-76}
HSCJ020141-030946	1	$2.205^{+0.09}_{-0.09}$	$3.0^{+0.1}_{-0.1}$	$3.4^{+0.1}_{-0.1}$	$3.6^{+0.2}_{-0.2}$
HSCJ020241-064611	1	$0.87^{+0.2}_{-0.2}$	$1.1^{+0.2}_{-0.2}$	$0.85^{+0.2}_{-0.2}$	$0.97^{+0.2}_{-0.2}$
HSCJ020955-024442	1	$0.71^{+0.04}_{-0.04}$	$2.6^{+0.1}_{-0.1}$	$1.92^{+0.1}_{-0.09}$	$1.71^{+0.1}_{-0.09}$
HSCJ021737-051329	1	9.2^{+2}_{-3}	10^{+2}_{-3}	11^{+2}_{-4}	12^{+2}_{-4}
HSCJ022346-053418	1	25^{+1}_{-1}	35^{+2}_{-2}	48^{+1}_{-3}	75^{+3}_{-4}
HSCJ022610-042011	1	$2.6^{+2}_{-0.7}$	$2.8^{+2}_{-0.8}$	4^{+2}_{-1}	6^{+4}_{-2}
HSCJ023217-021703	1	$0.48^{+0.09}_{-0.06}$	$0.9^{+0.2}_{-0.1}$	$0.8^{+0.2}_{-0.1}$	$0.8^{+0.1}_{-0.1}$
HSCJ023322-020530	1	$146.6^{+0.5}_{-0.7}$	$262.6^{+0.4}_{-0.4}$	$513.7^{+0.5}_{-0.6}$	772^{+1}_{-2}
	2	$11.5^{+0.3}_{-0.3}$	$7.76^{+0.2}_{-0.2}$	$16.4^{+0.4}_{-0.4}$	$9.76^{+0.5}_{-0.6}$
	3	$1.24^{+0.05}_{-0.05}$	$0.674^{+0.03}_{-0.03}$	$3.43^{+0.1}_{-0.1}$	$0.90^{+0.08}_{-0.07}$
HSCJ085046+003905	1	$3.31^{+0.08}_{-0.09}$	$3.85^{+0.1}_{-0.1}$	$3.8^{+0.1}_{-0.1}$	$3.66^{+0.06}_{-0.08}$
HSCJ085855-010208	1	$1.24^{+0.06}_{-0.06}$	$1.67^{+0.08}_{-0.08}$	$2.5^{+0.1}_{-0.1}$	$3.8^{+0.2}_{-0.2}$
HSCJ090429-010228	1	$7.16^{+0.09}_{-0.1}$	$16.4^{+0.1}_{-0.2}$	$18.9^{+0.2}_{-0.3}$	$18.67^{+0.09}_{-0.06}$
HSCJ094427-014742	1	11^{+1}_{-2}	20^{+2}_{-3}	28^{+3}_{-4}	45^{+4}_{-6}
HSCJ120623+001507	1	41^{+6}_{-10}	40^{+6}_{-10}	40^{+6}_{-10}	37^{+6}_{-9}
HSCJ121052-011905	1	24^{+2}_{-3}	33^{+2}_{-4}	36^{+3}_{-5}	46^{+3}_{-6}
HSCJ121504+004726	1	$11.2^{+0.2}_{-0.2}$	$11.3^{+0.2}_{-0.2}$	$11.6^{+0.2}_{-0.2}$	$17.1^{+0.3}_{-0.3}$
HSCJ124320-004517	1	$1.43^{+0.09}_{-0.1}$	$2.1^{+0.1}_{-0.1}$	$2.3^{+0.1}_{-0.2}$	$2.5^{+0.2}_{-0.2}$
HSCJ125254+004356	1	$0.28^{+0.02}_{-0.02}$	$0.37^{+0.02}_{-0.02}$	$0.66^{+0.04}_{-0.04}$	$0.87^{+0.05}_{-0.05}$
HSCJ135138+002839	1	$0.64^{+0.06}_{-0.05}$	$0.68^{+0.06}_{-0.06}$	$0.73^{+0.06}_{-0.06}$	$0.82^{+0.07}_{-0.07}$
HSCJ141136-010215	1	39^{+4}_{-5}	44^{+4}_{-6}	42^{+4}_{-5}	39^{+4}_{-5}
HSCJ141815+015832	1	$1.6^{+0.1}_{-0.1}$	$2.2^{+0.2}_{-0.2}$	$2.1^{+0.2}_{-0.2}$	$2.4^{+0.2}_{-0.2}$
HSCJ142720+001916	1	$1.6^{+0.1}_{-0.1}$	$1.8^{+0.2}_{-0.2}$	$2.2^{+0.2}_{-0.2}$	$3.2^{+0.3}_{-0.3}$
HSCJ144132-005358	1	39^{+6}_{-13}	36^{+5}_{-12}	30^{+4}_{-10}	41^{+6}_{-15}
HSCJ144320-012537	1	$4.62^{+0.08}_{-0.08}$	$5.08^{+0.08}_{-0.08}$	$5.25^{+0.08}_{-0.08}$	$5.36^{+0.09}_{-0.1}$
HSCJ145242+425731	1	$0.7^{+0.6}_{-0.3}$	$1.04^{+0.8}_{-0.4}$	$1.5^{+1.2}_{-0.6}$	$1.6^{+1}_{-0.6}$
HSCJ150021-004936	1	10^{+8}_{-5}	12^{+9}_{-6}	8^{+6}_{-4}	13^{+11}_{-7}
HSCJ150112+422113	1	$15.4^{+0.4}_{-0.4}$	$19.9^{+0.5}_{-0.4}$	$18.4^{+0.4}_{-0.4}$	$23^{+0.6}_{-0.6}$

Table A.3 Continued: Reconstructed amplitudes of the adopted Sérsic profiles representing the source light obtained with GLEE & GLaD.

Name	component	$A_{\mathrm{g,s,comp}}$	$A_{\mathrm{r,s,comp}}$	$A_{\mathrm{i,s,comp}}$	$A_{\mathrm{z,s,comp}}$
HSCJ223733+005015	1	44^{+3}_{-4}	44^{+3}_{-4}	39^{+3}_{-4}	46^{+3}_{-5}
HSCJ230335+003703	1	40^{+4}_{-7}	44^{+4}_{-8}	50^{+5}_{-8}	49^{+5}_{-8}
HSCJ230521-000211	1	$12.5^{+0.2}_{-0.2}$	$12.3^{+0.2}_{-0.2}$	$11.2^{+0.2}_{-0.2}$	$10.8^{+0.2}_{-0.2}$
HSCJ233130+003733	1	$10^{+0.9}_{-0.7}$	13^{+1}_{-1}	15^{+1}_{-1}	18^{+2}_{-1}
HSCJ233146+013845	1	$2.3^{+0.2}_{-0.2}$	$4.0^{+0.4}_{-0.3}$	$4.8^{+0.5}_{-0.4}$	$5.6^{+0.6}_{-0.4}$

Table A.4: Reconstructed structural parameter values describing the source light distribution of the 32 HSC SuGOHI lenses obtained with GLEE & GLaD. We give, respectively for the Sérsic profile(s) specified in column 2, the unlensed source center coordinates x_s and y_s, the axis ratio q_s, its orientation θ_s, the effective radius $r_{\text{eff,s}}$, and the Sérsic index n_s.

Name	component	$x_{s,\text{comp}}['']$	$y_{s,\text{comp}}['']$	$q_{s,\text{comp}}$	$\theta_{s,\text{comp}}$ [rad]	$r_{\text{eff,s,comp}}$	$n_{s,\text{comp}}$
HSCJ015618-010747	1	$6.0^{+0.1}_{-0.1}$	$6.23^{+0.04}_{-0.05}$	$0.25^{+0.04}_{-0.04}$	$-1.36^{+0.04}_{-0.04}$	$0.06^{+0.02}_{-0.01}$	$2.61^{+0.85}_{-0.87}$
HSCJ020141-030946	1	$5.02^{+0.03}_{-0.01}$	$5.35^{+0.03}_{-0.03}$	$0.140^{+0.005}_{-0.005}$	$-2.889^{+0.009}_{-0.009}$	$0.240^{+0.006}_{-0.006}$	$0.52^{+0.03}_{-0.02}$
HSCJ020241-064611	1	$5.6^{+0.2}_{-0.2}$	$6.08^{+0.1}_{-0.2}$	$0.38^{+0.08}_{-0.08}$	$1.0^{+0.1}_{-0.1}$	$0.13^{+0.01}_{-0.01}$	$1.03^{+0.3}_{-0.2}$
HSCJ020955-024442	1	$5.5^{+0.044}_{-0.043}$	$5.24^{+0.03}_{-0.03}$	$0.27^{+0.02}_{-0.02}$	$0.02^{+0.02}_{-0.03}$	$0.082^{+0.003}_{-0.003}$	$0.51^{+0.03}_{-0.01}$
HSCJ021737-051329	1	$5.45^{+0.04}_{-0.04}$	$5.11^{+0.04}_{-0.1}$	$0.38^{+0.04}_{-0.04}$	$0.07^{+0.1}_{-0.06}$	$0.059^{+0.009}_{-0.005}$	$0.6^{+0.2}_{-0.06}$
HSCJ022346-053418	1	$5.08^{+0.02}_{-0.02}$	$5.60^{+0.02}_{-0.02}$	$0.068^{+0.006}_{-0.005}$	$1.761^{+0.007}_{-0.01}$	$0.096^{+0.007}_{-0.006}$	$0.512^{+0.02}_{-0.009}$
HSCJ022610-042011	1	$5.1^{+0.1}_{-0.1}$	$5.1^{+0.1}_{-0.1}$	$0.20^{+0.06}_{-0.04}$	$2.26^{+0.05}_{-0.04}$	$0.18^{+0.02}_{-0.03}$	$1.0^{+0.3}_{-0.2}$
HSCJ023217-021703	1	$5.75^{+0.05}_{-0.05}$	$5.35^{+0.05}_{-0.05}$	$0.13^{+0.01}_{-0.01}$	$-1.58^{+0.02}_{-0.02}$	$0.30^{+0.02}_{-0.02}$	$1.82^{+0.30}_{-0.24}$
HSCJ023322-020530	1	$5.477^{+0.008}_{-0.007}$	$4.650^{+0.007}_{-0.006}$	$0.302^{+0.004}_{-0.005}$	$-3.098^{+0.004}_{-0.004}$	$0.064^{+0.001}_{-0.001}$	$0.503^{+0.005}_{-0.002}$
	2	$\equiv x_{s1}$	$\equiv y_{s1}$	$0.128^{+0.004}_{-0.004}$	$2.092^{+0.002}_{-0.003}$	$0.201^{+0.002}_{-0.003}$	$0.502^{+0.003}_{-0.001}$
	3	$4.839^{+0.009}_{-0.007}$	$4.660^{+0.007}_{-0.006}$	$0.202^{+0.003}_{-0.002}$	$3.10^{+0.02}_{-0.02}$	$0.179^{+0.005}_{-0.005}$	$0.506^{+0.01}_{-0.005}$
HSCJ085046+003905	1	$4.97^{+0.04}_{-0.06}$	$5.93^{+0.04}_{-0.02}$	$0.53^{+0.04}_{-0.04}$	$0.29^{+0.1}_{-0.1}$	$0.06^{+0.003}_{-0.002}$	$2.39^{+0.2}_{-0.5}$
HSCJ085855-010208	1	$5.34^{+0.02}_{-0.02}$	$4.76^{+0.02}_{-0.02}$	$0.29^{+0.01}_{-0.01}$	$-2.21^{+0.01}_{-0.01}$	$0.208^{+0.005}_{-0.005}$	$1.00^{+0.04}_{-0.04}$
HSCJ090429-010228	1	$5.79^{+0.01}_{-0.02}$	$4.954^{+0.006}_{-0.004}$	$0.065^{+0.008}_{-0.007}$	$0.498^{+0.003}_{-0.003}$	$0.064^{+0.002}_{-0.002}$	$0.53^{+0.04}_{-0.02}$
HSCJ094427-014742	1	$4.7^{+0.3}_{-0.06}$	$4.78^{+0.06}_{-0.2}$	$0.20^{+0.03}_{-0.06}$	$-0.38^{+0.05}_{-0.03}$	$0.096^{+0.01}_{-0.009}$	$0.55^{+0.07}_{-0.04}$
HSCJ120623+001507	1	$4.66^{+0.09}_{-0.1}$	$4.41^{+0.2}_{-0.09}$	$0.6^{+0.1}_{-0.1}$	$3.04^{+0.07}_{-0.09}$	$0.031^{+0.005}_{-0.003}$	$1.3^{+0.3}_{-0.3}$
HSCJ121052-011905	1	$5.46^{+0.04}_{-0.04}$	$5.64^{+0.02}_{-0.02}$	$0.01^{+0.001}_{-0.001}$	$2.269^{+0.001}_{-0.001}$	$0.16^{+0.01}_{-0.01}$	$1.9^{+0.4}_{-0.3}$
HSCJ121504+004726	1	$6.62^{+0.01}_{-0.03}$	$5.27^{+0.09}_{-0.09}$	$0.60^{+0.02}_{-0.02}$	$-0.05^{+0.03}_{-0.03}$	$0.128^{+0.002}_{-0.002}$	$0.88^{+0.04}_{-0.03}$
HSCJ124320-004517	1	$5.66^{+0.04}_{-0.03}$	$4.96^{+0.05}_{-0.03}$	$0.15^{+0.01}_{-0.01}$	$-1.399^{+0.01}_{-0.01}$	$0.28^{+0.008}_{-0.007}$	$0.52^{+0.02}_{-0.01}$

Table A.4 Continued: Reconstructed structual source light parameter values obtained with GLEE & GLaD.

Name	component	$x_{\text{s,comp}}[''] $	$y_{\text{s,comp}}['']$	$q_{\text{s,comp}}$	$\theta_{\text{s,comp}}$ [rad]	$r_{\text{eff,s,comp}}$	$n_{\text{s,comp}}$
HSCJ125254+004356	1	$4.34^{+0.04}_{-0.04}$	$5.73^{+0.05}_{-0.05}$	$0.58^{+0.03}_{-0.03}$	$0.64^{+0.05}_{-0.05}$	$0.272^{+0.008}_{-0.008}$	$1.49^{+0.1}_{-0.09}$
HSCJ135138+002839	1	$5.280^{+0.005}_{-0.005}$	$5.02^{+0.01}_{-0.01}$	$0.27^{+0.02}_{-0.02}$	$-0.03^{+0.02}_{-0.02}$	$0.157^{+0.008}_{-0.008}$	$1.9^{+0.2}_{-0.1}$
HSCJ141136-010215	1	$5.05^{+0.04}_{-0.04}$	$5.60^{+0.04}_{-0.02}$	$0.22^{+0.01}_{-0.02}$	$-2.94^{+0.03}_{-0.03}$	$0.033^{+0.004}_{-0.004}$	$1.5^{+0.5}_{-0.4}$
HSCJ141815+015832	1	$5.38^{+0.09}_{-0.08}$	$6.01^{+0.08}_{-0.08}$	$0.19^{+0.01}_{-0.01}$	$2.97^{+0.02}_{-0.02}$	$0.20^{+0.01}_{-0.01}$	$0.89^{+0.1}_{-0.08}$
HSCJ142720+001916	1	$5.56^{+0.07}_{-0.07}$	$4.28^{+0.04}_{-0.02}$	$0.133^{+0.009}_{-0.008}$	$0.5^{+0.02}_{-0.01}$	$0.28^{+0.01}_{-0.01}$	$0.78^{+0.1}_{-0.09}$
HSCJ144132-005358	1	$5.17^{+0.09}_{-0.09}$	$5.93^{+0.1}_{-0.06}$	$0.011^{+0.001}_{-0.001}$	$3.112^{+0.005}_{-0.005}$	$0.31^{+0.07}_{-0.04}$	$0.7^{+0.3}_{-0.1}$
HSCJ144320-012537	1	$5.748^{+0.008}_{-0.009}$	$5.39^{+0.01}_{-0.01}$	$0.586^{+0.008}_{-0.008}$	$-3.11^{+0.02}_{-0.02}$	$0.108^{+0.001}_{-0.001}$	$0.69^{+0.03}_{-0.03}$
HSCJ145242+425731	1	$4.72^{+0.05}_{-0.05}$	$5.42^{+0.01}_{-0.01}$	$0.026^{+0.004}_{-0.003}$	$-1.221^{+0.003}_{-0.003}$	$0.5^{+0.1}_{-0.1}$	$5.4^{+0.5}_{-0.9}$
HSCJ150021-004936	1	$4.87^{+0.02}_{-0.02}$	$5.715^{+0.009}_{-0.01}$	$0.041^{+0.006}_{-0.005}$	$0.192^{+0.004}_{-0.006}$	$0.12^{+0.04}_{-0.02}$	$5.7^{+0.2}_{-0.5}$
HSCJ150112+422113	1	$5.78^{+0.01}_{-0.01}$	$4.64^{+0.01}_{-0.01}$	$0.72^{+0.02}_{-0.02}$	$-3.137^{+0.005}_{-0.003}$	$0.081^{+0.002}_{-0.001}$	$0.501^{+0.001}_{-0.001}$
HSCJ223733+005015	1	$5.04^{+0.03}_{-0.02}$	$6.56^{+0.02}_{-0.02}$	$0.016^{+0.001}_{-0.001}$	$1.53^{+0.003}_{-0.003}$	$0.134^{+0.009}_{-0.007}$	$0.52^{+0.02}_{-0.01}$
HSCJ230335+003703	1	$5.26^{+0.02}_{-0.02}$	$5.856^{+0.01}_{-0.01}$	$0.011^{+0.001}_{-0.001}$	$-0.851^{+0.001}_{-0.001}$	$0.159^{+0.006}_{-0.005}$	$0.55^{+0.07}_{-0.04}$
HSCJ230521-000211	1	$5.79^{+0.02}_{-0.02}$	$5.882^{+0.007}_{-0.008}$	$0.51^{+0.01}_{-0.01}$	$2.51^{+0.02}_{-0.02}$	$0.07^{+0.001}_{-0.001}$	$0.60^{+0.01}_{-0.01}$
HSCJ233130+003733	1	$4.627^{+0.002}_{-0.002}$	$5.04^{+0.02}_{-0.02}$	$0.76^{+0.02}_{-0.02}$	$0.40^{+0.05}_{-0.05}$	$0.049^{+0.002}_{-0.002}$	$2.6^{+0.2}_{-0.2}$
HSCJ233146+013845	1	$5.84^{+0.02}_{-0.02}$	$5.16^{+0.02}_{-0.02}$	$0.209^{+0.01}_{-0.01}$	$-2.971^{+0.009}_{-0.009}$	$0.086^{+0.002}_{-0.003}$	$1.1^{+0.1}_{-0.1}$

Table A.5: Lens light parameter values obtained with GLEE & GLaD for the 32 HSC SuGOHI lenses. We give, respectively for the first and second Sérsic profile the axis ratio q_{ll}, its orientation θ_{ll}, the effective radius $r_{eff,ll}$, Sérsic index n_{ll}, the amplitudes A_{ll} for the four different filters $g, r, i,$ and z.

Name	$q_{ll,1}$ / $q_{ll,2}$	$\theta_{ll,1}$ [rad] / $\theta_{ll,2}$ [rad]	$r_{eff,ll,1}$ / $r_{eff,ll,2}$	$n_{ll,1}$ / $n_{ll,2}$	$A_{g,ll,1}$ / $A_{g,ll,2}$	$A_{r,ll,1}$ / $A_{r,ll,2}$	$A_{i,ll,1}$ / $A_{i,ll,2}$	$A_{z,ll,1}$ / $A_{z,ll,2}$
HSCJ015618-010747	$0.62^{+0.02}_{-0.03}$	$0.42^{+0.04}_{-0.05}$	$2.5^{+0.1}_{-0.1}$	$0.53^{+0.04}_{-0.02}$	$0.09^{+0.01}_{-0.01}$	$0.044^{+0.004}_{-0.004}$	$0.49^{+0.06}_{-0.05}$	$0.29^{+0.05}_{-0.05}$
	$0.645^{+0.009}_{-0.009}$	$0.69^{+0.02}_{-0.02}$	$0.63^{+0.05}_{-0.05}$	$3.3^{+0.2}_{-0.3}$	$1.7^{+0.2}_{-0.2}$	$0.37^{+0.05}_{-0.04}$	$3.9^{+0.5}_{-0.4}$	$6.2^{+0.8}_{-0.7}$
HSCJ020141-030946	$0.563^{+0.003}_{-0.003}$	$1.615^{+0.003}_{-0.003}$	$0.71^{+0.03}_{-0.03}$	$4.6^{+0.1}_{-0.1}$	$2.7^{+0.2}_{-0.2}$	$0.45^{+0.03}_{-0.03}$	$5.7^{+0.4}_{-0.4}$	$8.4^{+0.6}_{-0.5}$
	$0.80^{+0.02}_{-0.02}$	$2.60^{+0.05}_{-0.05}$	$2.37^{+0.03}_{-0.03}$	$0.76^{+0.04}_{-0.04}$	$0.167^{+0.006}_{-0.005}$	$0.059^{+0.001}_{-0.001}$	$0.25^{+0.01}_{-0.01}$	$0.30^{+0.02}_{-0.02}$
HSCJ020241-064611	$0.946^{+0.005}_{-0.006}$	$1.61^{+0.05}_{-0.05}$	$0.541^{+0.003}_{-0.003}$	$1.79^{+0.02}_{-0.02}$	$2.98^{+0.05}_{-0.06}$	$0.709^{+0.009}_{-0.01}$	$6.90^{+0.07}_{-0.07}$	$5.9^{+0.4}_{-0.3}$
	$0.97^{+0.02}_{-0.03}$	$-0.5^{+0.5}_{-0.6}$	$0.69^{+0.02}_{-0.02}$	$2.9^{+0.1}_{-0.2}$	$0.03^{+0.03}_{-0.02}$	$0.002^{+0.004}_{-0.002}$	$0.01^{+0.02}_{-0.01}$	$2.3^{+0.3}_{-0.3}$
HSCJ020955-024442	$0.88^{+0.01}_{-0.01}$	$1.15^{+0.05}_{-0.06}$	$0.98^{+0.01}_{-0.01}$	$1.14^{+0.07}_{-0.06}$	$0.38^{+0.01}_{-0.01}$	$0.121^{+0.005}_{-0.005}$	$0.78^{+0.02}_{-0.02}$	$1.06^{+0.03}_{-0.04}$
	$0.78^{+0.08}_{-0.06}$	$1.787^{+0.2}_{-0.2}$	$0.115^{+0.007}_{-0.007}$	$0.58^{+0.1}_{-0.06}$	$5.9^{+0.4}_{-0.5}$	$0.3^{+0.2}_{-0.2}$	28^{+1}_{-3}	46^{+2}_{-4}
HSCJ021737-051329	$0.85^{+0.01}_{-0.01}$	$2.87^{+0.04}_{-0.05}$	$3.7^{+0.3}_{-0.3}$	$3.4^{+0.3}_{-0.3}$	$0.044^{+0.006}_{-0.005}$	$0.017^{+0.002}_{-0.002}$	$0.076^{+0.01}_{-0.009}$	$0.11^{+0.02}_{-0.01}$
	$0.97^{+0.02}_{-0.02}$	$-0.1^{+0.2}_{-0.2}$	$0.20^{+0.01}_{-0.02}$	$5.90^{+0.08}_{-0.2}$	$3.29^{+0.48}_{-0.33}$	$0.36^{+0.06}_{-0.06}$	13^{+2}_{-1}	21^{+4}_{-2}
HSCJ022346-053418	$0.683^{+0.006}_{-0.006}$	$2.98^{+0.02}_{-0.02}$	$0.36^{+0.02}_{-0.02}$	$2.8^{+0.2}_{-0.1}$	$4.7^{+0.3}_{-0.4}$	$0.75^{+0.05}_{-0.05}$	$11.5^{+0.8}_{-0.9}$	17^{+1}_{-1}
	$0.553^{+0.005}_{-0.005}$	$2.639^{+0.006}_{-0.007}$	$2.92^{+0.08}_{-0.07}$	$1.82^{+0.08}_{-0.08}$	$0.24^{+0.02}_{-0.02}$	$0.071^{+0.003}_{-0.003}$	$0.48^{+0.03}_{-0.03}$	$0.66^{+0.04}_{-0.04}$
HSCJ022610-042011	$0.802^{+0.004}_{-0.004}$	$2.64^{+0.02}_{-0.02}$	$0.63^{+0.08}_{-0.05}$	$5.98^{+0.02}_{-0.03}$	$1.8^{+0.2}_{-0.3}$	$0.36^{+0.05}_{-0.05}$	$4.1^{+0.5}_{-0.6}$	$6.0^{+0.8}_{-0.9}$
	$0.831^{+0.01}_{-0.009}$	$-0.95^{+0.03}_{-0.05}$	$1.69^{+0.01}_{-0.01}$	$0.9^{+0.1}_{-0.1}$	$0.27^{+0.02}_{-0.02}$	$0.074^{+0.004}_{-0.004}$	$0.50^{+0.04}_{-0.05}$	$0.75^{+0.05}_{-0.08}$
HSCJ023217-021703	$0.845^{+0.006}_{-0.007}$	$0.72^{+0.02}_{-0.02}$	$1.18^{+0.02}_{-0.02}$	$1.16^{+0.05}_{-0.05}$	$0.44^{+0.01}_{-0.01}$	$0.112^{+0.004}_{-0.004}$	$0.99^{+0.03}_{-0.03}$	$1.43^{+0.04}_{-0.04}$
	$0.84^{+0.02}_{-0.03}$	$1.36^{+0.08}_{-0.08}$	$0.152^{+0.006}_{-0.005}$	$0.9^{+0.1}_{-0.1}$	$12.4^{+0.7}_{-0.7}$	$2.3^{+0.1}_{-0.1}$	30^{+2}_{-2}	44^{+3}_{-2}
HSCJ023322-020530	$0.636^{+0.008}_{-0.008}$	$0.30^{+0.01}_{-0.01}$	$0.499^{+0.004}_{-0.004}$	$0.504^{+0.006}_{-0.003}$	$1.66^{+0.02}_{-0.02}$	$0.22^{+0.02}_{-0.02}$	$6.63^{+0.06}_{-0.07}$	$11.1^{+0.1}_{-0.1}$
	$0.873^{+0.01}_{-0.01}$	$0.18^{+0.05}_{-0.05}$	$2.19^{+0.03}_{-0.03}$	$0.92^{+0.05}_{-0.05}$	$0.118^{+0.003}_{-0.003}$	$0.079^{+0.003}_{-0.002}$	$0.283^{+0.009}_{-0.009}$	$0.46^{+0.01}_{-0.01}$
HSCJ085046+003905	$0.78^{+0.06}_{-0.06}$	$2.4^{+0.2}_{-0.2}$	$0.028^{+0.006}_{-0.005}$	$1.3^{+0.7}_{-0.4}$	3^{+1}_{-2}	$20.7^{+0.5}_{-0.4}$	48^{+1}_{-2}	47^{+2}_{-2}
	$0.839^{+0.009}_{-0.01}$	$1.58^{+0.03}_{-0.03}$	$0.71^{+0.01}_{-0.01}$	$3.3^{+0.1}_{-0.1}$	$0.255^{+0.008}_{-0.009}$	$0.046^{+0.003}_{-0.003}$	$0.83^{+0.03}_{-0.03}$	$1.85^{+0.05}_{-0.07}$

Table A.5 Continued: Lens light parameter values obtained with GLEE & GLaD for the 32 HSC SuG-OHI lenses.

Name	$q_{ll,1}$ / $q_{ll,2}$	$\theta_{ll,1}$ [rad] / $\theta_{ll,2}$ [rad]	$r_{eff},ll,1$ / $r_{eff,ll,2}$	$n_{ll,1}$ / $n_{ll,2}$	$A_{g,ll,1}$ / $A_{g,ll,2}$	$A_{r,ll,1}$ / $A_{r,ll,2}$	$A_{i,ll,1}$ / $A_{i,ll,2}$	$A_{z,ll,1}$ / $A_{z,ll,2}$
HSCJ085855-010208	$0.78^{+0.02}_{-0.02}$	$0.43^{+0.04}_{-0.04}$	$1.52^{+0.05}_{-0.07}$	$4.9^{+0.1}_{-0.2}$	$0.21^{+0.03}_{-0.02}$	$0.044^{+0.006}_{-0.003}$	$0.50^{+0.06}_{-0.04}$	$0.73^{+0.09}_{-0.05}$
	$0.77^{+0.01}_{-0.01}$	$0.76^{+0.02}_{-0.02}$	$0.998^{+0.007}_{-0.006}$	$0.71^{+0.02}_{-0.02}$	$1.31^{+0.02}_{-0.03}$	$0.342^{+0.008}_{-0.009}$	$2.41^{+0.04}_{-0.06}$	$3.30^{+0.06}_{-0.08}$
HSCJ090429-010228	$0.37^{+0.02}_{-0.02}$	$3.41^{+0.02}_{-0.02}$	$0.19^{+0.03}_{-0.02}$	$5.7^{+0.7}_{-0.8}$	$0.05^{+0.05}_{-0.03}$	$0.016^{+0.008}_{-0.004}$	6^{+2}_{-2}	19^{+6}_{-5}
	$0.92^{+0.02}_{-0.02}$	$1.64^{+0.1}_{-0.1}$	$1.30^{+0.02}_{-0.02}$	$0.501^{+0.001}_{-0.001}$	$0.207^{+0.004}_{-0.004}$	$0.05^{+0.002}_{-0.002}$	$0.25^{+0.006}_{-0.006}$	$0.35^{+0.01}_{-0.01}$
HSCJ094427-014742	$0.81^{+0.02}_{-0.02}$	$3.11^{+0.03}_{-0.04}$	$0.27^{+0.04}_{-0.02}$	$5.88^{+0.09}_{-0.2}$	$3.1^{+0.6}_{-0.7}$	$0.60^{+0.09}_{-0.09}$	9^{+2}_{-2}	14^{+2}_{-3}
	$0.90^{+0.03}_{-0.03}$	$-1.3^{+0.1}_{-0.1}$	$2.8^{+0.3}_{-0.4}$	$3.0^{+0.5}_{-0.4}$	$0.06^{+0.02}_{-0.01}$	$0.020^{+0.006}_{-0.003}$	$0.10^{+0.03}_{-0.02}$	$0.15^{+0.05}_{-0.03}$
HSCJ120623+001507	$0.751^{+0.009}_{-0.009}$	$2.47^{+0.02}_{-0.02}$	$0.159^{+0.002}_{-0.002}$	$1.3^{+0.1}_{-0.1}$	$11.7^{+0.2}_{-0.3}$	$2.17^{+0.05}_{-0.05}$	$33.3^{+0.5}_{-0.7}$	$48.9^{+0.7}_{-1}$
	$0.877^{+0.004}_{-0.004}$	$2.25^{+0.02}_{-0.02}$	$1.58^{+0.02}_{-0.01}$	$1.64^{+0.05}_{-0.05}$	$0.305^{+0.005}_{-0.006}$	$0.078^{+0.002}_{-0.002}$	$0.78^{+0.01}_{-0.02}$	$1.12^{+0.02}_{-0.02}$
HSCJ121052-011905	$0.91^{+0.01}_{-0.01}$	$-2.03^{+0.2}_{-0.07}$	$0.148^{+0.004}_{-0.003}$	$5.9^{+0.1}_{-0.3}$	$4.2^{+0.2}_{-0.3}$	$0.02^{+0.03}_{-0.009}$	$16.1^{+0.5}_{-0.9}$	$24.8^{+0.8}_{-1}$
	$0.651^{+0.007}_{-0.006}$	$1.418^{+0.009}_{-0.01}$	$1.99^{+0.02}_{-0.03}$	$2.07^{+0.05}_{-0.06}$	$0.126^{+0.004}_{-0.003}$	$0.062^{+0.002}_{-0.001}$	$0.352^{+0.008}_{-0.007}$	$0.56^{+0.01}_{-0.01}$
HSCJ121504+004726	$0.602^{+0.008}_{-0.01}$	$1.14^{+0.01}_{-0.01}$	$0.586^{+0.008}_{-0.007}$	$5.94^{+0.04}_{-0.07}$	$0.77^{+0.01}_{-0.01}$	$0.017^{+0.007}_{-0.005}$	$3.04^{+0.05}_{-0.03}$	$4.55^{+0.07}_{-0.07}$
	$0.64^{+0.01}_{-0.01}$	$-2.84^{+0.02}_{-0.02}$	$1.43^{+0.03}_{-0.03}$	$2.93^{+0.1}_{-0.09}$	$0.243^{+0.01}_{-0.009}$	$0.126^{+0.005}_{-0.004}$	$0.48^{+0.02}_{-0.02}$	$0.8^{+0.03}_{-0.03}$
HSCJ124320-004517	$0.706^{+0.006}_{-0.006}$	$0.01^{+0.01}_{-0.01}$	$1.47^{+0.04}_{-0.07}$	$5.96^{+0.03}_{-0.07}$	$0.197^{+0.009}_{-0.008}$	$0.035^{+0.002}_{-0.002}$	$0.66^{+0.03}_{-0.02}$	$0.99^{+0.04}_{-0.04}$
	$0.64^{+0.02}_{-0.02}$	$2.69^{+0.03}_{-0.03}$	$1.64^{+0.03}_{-0.03}$	$0.506^{+0.01}_{-0.004}$	$0.146^{+0.006}_{-0.006}$	$0.063^{+0.003}_{-0.003}$	$0.22^{+0.01}_{-0.01}$	$0.31^{+0.02}_{-0.02}$
HSCJ125254+004356	$0.738^{+0.001}_{-0.001}$	$1.804^{+0.008}_{-0.003}$	$0.94^{+0.01}_{-0.01}$	$5.989^{+0.008}_{-0.007}$	$0.494^{+0.009}_{-0.01}$	$0.078^{+0.002}_{-0.003}$	$1.77^{+0.03}_{-0.04}$	$2.79^{+0.06}_{-0.06}$
	$0.9^{+0.001}_{-0.001}$	$2.149^{+0.02}_{-0.008}$	$1.979^{+0.02}_{-0.008}$	$0.502^{+0.003}_{-0.001}$	$0.119^{+0.002}_{-0.002}$	$0.070^{+0.002}_{-0.002}$	$0.216^{+0.009}_{-0.008}$	$0.269^{+0.007}_{-0.008}$
HSCJ135138+002839	$0.768^{+0.001}_{-0.002}$	$0.626^{+0.006}_{-0.006}$	$0.424^{+0.002}_{-0.002}$	$3.24^{+0.01}_{-0.01}$	$4.30^{+0.03}_{-0.03}$	$0.728^{+0.009}_{-0.008}$	$10.25^{+0.06}_{-0.07}$	$14.44^{+0.09}_{-0.09}$
	$0.87^{+0.001}_{-0.001}$	$2.986^{+0.007}_{-0.008}$	$2.047^{+0.005}_{-0.005}$	$0.501^{+0.001}_{-0.001}$	$0.343^{+0.002}_{-0.002}$	$0.128^{+0.001}_{-0.001}$	$0.666^{+0.004}_{-0.005}$	$0.904^{+0.007}_{-0.008}$
HSCJ141136-010215	$0.38^{+0.02}_{-0.02}$	$0.84^{+0.01}_{-0.01}$	$0.136^{+0.005}_{-0.004}$	$2.2^{+0.4}_{-0.3}$	$4.8^{+0.2}_{-0.3}$	$0.03^{+0.04}_{-0.02}$	$20.7^{+0.5}_{-1}$	49^{+1}_{-2}
	$0.56^{+0.02}_{-0.01}$	$0.694^{+0.02}_{-0.02}$	$1.29^{+0.02}_{-0.03}$	$0.503^{+0.005}_{-0.002}$	$0.117^{+0.007}_{-0.007}$	$0.069^{+0.003}_{-0.003}$	$0.24^{+0.02}_{-0.02}$	$0.55^{+0.04}_{-0.04}$
HSCJ141815+015832	$0.81^{+0.02}_{-0.02}$	$1.18^{+0.05}_{-0.06}$	$0.38^{+0.06}_{-0.03}$	$5.6^{+0.3}_{-0.4}$	$2.4^{+0.3}_{-0.4}$	$0.35^{+0.05}_{-0.06}$	$6.5^{+0.8}_{-1}$	10^{+1}_{-2}
	$0.71^{+0.03}_{-0.03}$	$1.66^{+0.04}_{-0.04}$	$2.4^{+0.2}_{-0.1}$	$2.0^{+0.2}_{-0.2}$	$0.09^{+0.02}_{-0.01}$	$0.057^{+0.006}_{-0.006}$	$0.22^{+0.03}_{-0.04}$	$0.32^{+0.05}_{-0.05}$

Table A.5 Continued: Lens light parameter values obtained with GLEE & GLaD for the 32 HSC SuG-OHI lenses.

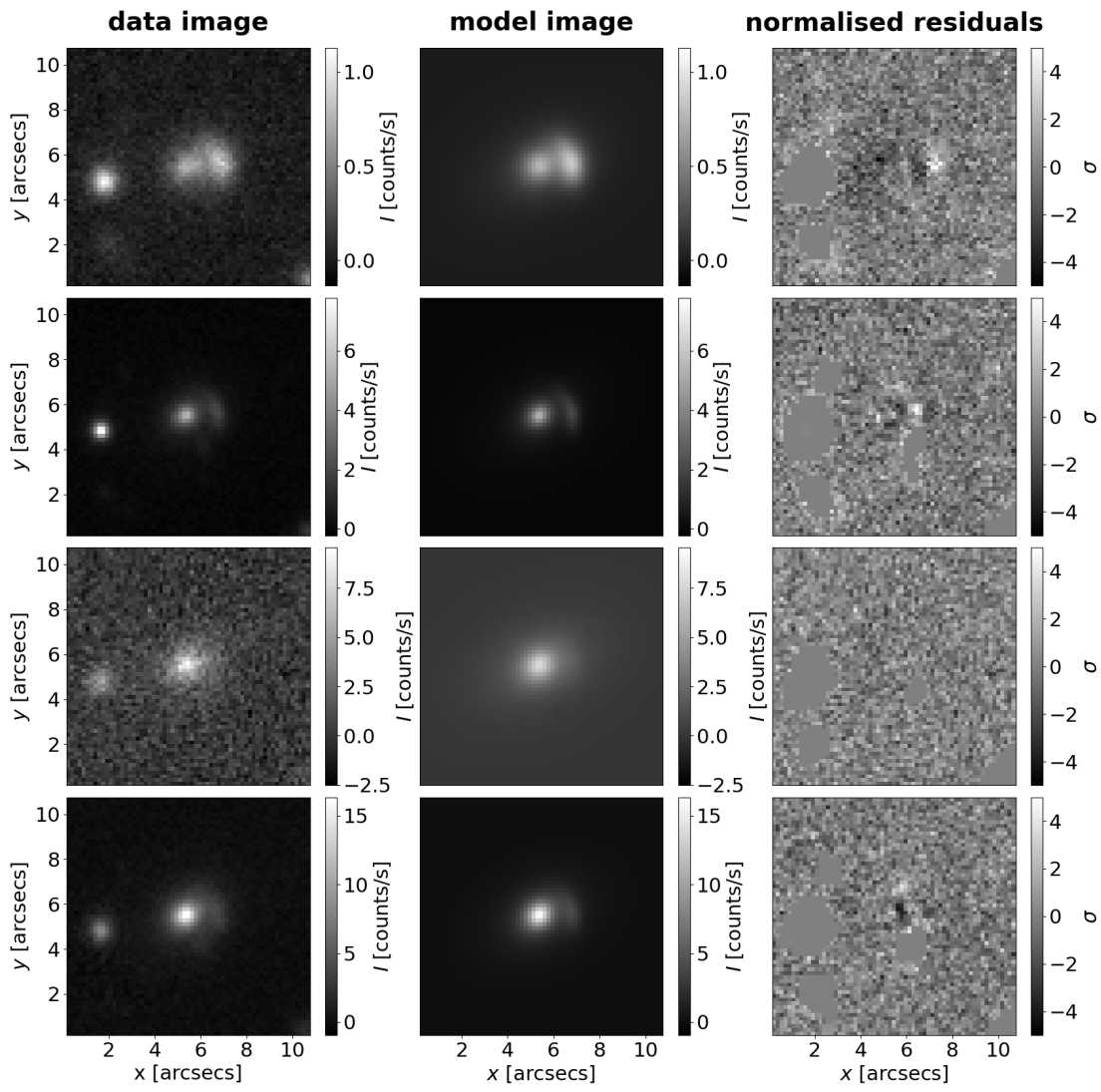| Name	$q_{\mathrm{ll},1}$ $q_{\mathrm{ll},2}$	$\theta_{\mathrm{ll},1}$ [rad] $\theta_{\mathrm{ll},2}$ [rad]	$r_{\mathrm{eff},\mathrm{ll},1}$ $r_{\mathrm{eff},\mathrm{ll},2}$	$n_{\mathrm{ll},1}$ $n_{\mathrm{ll},2}$	$A_{\mathrm{g},\mathrm{ll},1}$ $A_{\mathrm{g},\mathrm{ll},2}$	$A_{\mathrm{r},\mathrm{ll},1}$ $A_{\mathrm{r},\mathrm{ll},2}$	$A_{\mathrm{i},\mathrm{ll},1}$ $A_{\mathrm{i},\mathrm{ll},2}$	$A_{\mathrm{z},\mathrm{ll},1}$ $A_{\mathrm{z},\mathrm{ll},2}$
HSCJ142720+001916	$0.829^{+0.006}_{-0.007}$ $0.752^{+0.006}_{-0.007}$	$-0.13^{+0.02}_{-0.02}$ $0.21^{+0.02}_{-0.02}$	$0.35^{+0.02}_{-0.008}$ $1.44^{+0.06}_{-0.03}$	$5.7^{+0.2}_{-0.5}$ $2.14^{+0.07}_{-0.1}$	$2.7^{+0.1}_{-0.1}$ $0.26^{+0.01}_{-0.02}$	$0.54^{+0.03}_{-0.02}$ $0.062^{+0.005}_{-0.005}$	$8.07^{+0.3}_{-0.4}$ $0.58^{+0.04}_{-0.04}$	$12.2^{+0.5}_{-0.6}$ $0.91^{+0.06}_{-0.07}$
HSCJ144132-005358	$0.678^{+0.005}_{-0.004}$ $0.509^{+0.002}_{-0.002}$	$0.414^{+0.005}_{-0.005}$ $0.471^{+0.002}_{-0.002}$	$0.55^{+0.02}_{-0.02}$ $3.96^{+0.04}_{-0.04}$	$1.71^{+0.03}_{-0.03}$ $1.33^{+0.03}_{-0.04}$	$2.62^{+0.07}_{-0.07}$ $0.281^{+0.005}_{-0.005}$	$0.60^{+0.02}_{-0.02}$ $0.037^{+0.001}_{-0.001}$	$7.2^{+0.2}_{-0.2}$ $0.76^{+0.01}_{-0.01}$	$10.9^{+0.3}_{-0.3}$ $1.13^{+0.02}_{-0.02}$
HSCJ144320-012537	$0.52^{+0.01}_{-0.01}$ $0.87^{+0.02}_{-0.02}$	$0.57^{+0.01}_{-0.01}$ $-1.19^{+0.06}_{-0.06}$	$1.18^{+0.04}_{-0.04}$ $1.33^{+0.02}_{-0.02}$	$4.0^{+0.2}_{-0.1}$ $0.502^{+0.004}_{-0.002}$	$0.094^{+0.006}_{-0.006}$ $0.179^{+0.006}_{-0.006}$	$0.010^{+0.001}_{-0.001}$ $0.134^{+0.003}_{-0.003}$	$0.38^{+0.02}_{-0.02}$ $0.208^{+0.009}_{-0.009}$	$0.99^{+0.05}_{-0.05}$ $0.29^{+0.02}_{-0.02}$
HSCJ145242+425731	$0.77^{+0.02}_{-0.02}$ $0.65^{+0.01}_{-0.01}$	$2.73^{+0.04}_{-0.05}$ $2.99^{+0.02}_{-0.02}$	$2.42^{+0.06}_{-0.06}$ $0.252^{+0.004}_{-0.004}$	$0.86^{+0.07}_{-0.06}$ $5.8^{+0.1}_{-0.3}$	$0.067^{+0.004}_{-0.004}$ $3.90^{+0.09}_{-0.08}$	$0.03^{+0.002}_{-0.002}$ $0.58^{+0.03}_{-0.03}$	$0.20^{+0.01}_{-0.01}$ $16.0^{+0.4}_{-0.3}$	$0.3^{+0.02}_{-0.02}$ $24.4^{+0.5}_{-0.5}$
HSCJ150021-004936	$0.878^{+0.004}_{-0.004}$ $0.88^{+0.003}_{-0.003}$	$2.74^{+0.03}_{-0.04}$ $2.08^{+0.01}_{-0.01}$	$0.281^{+0.003}_{-0.004}$ $1.80^{+0.02}_{-0.03}$	$2.08^{+0.06}_{-0.06}$ $2.49^{+0.1}_{-0.08}$	$11.8^{+0.2}_{-0.3}$ $0.52^{+0.02}_{-0.01}$	$2.06^{+0.05}_{-0.07}$ $0.128^{+0.004}_{-0.004}$	$26.6^{+0.5}_{-0.6}$ $0.99^{+0.04}_{-0.03}$	$37.9^{+0.7}_{-0.9}$ $1.40^{+0.06}_{-0.04}$
HSCJ150112+422113	$0.756^{+0.006}_{-0.005}$ $0.711^{+0.005}_{-0.005}$	$0.23^{+0.01}_{-0.01}$ $0.308^{+0.009}_{-0.008}$	$0.684^{+0.008}_{-0.007}$ $0.59^{+0.01}_{-0.01}$	$2.90^{+0.07}_{-0.2}$ $5.8^{+0.1}_{-0.2}$	$3.7^{+0.1}_{-0.3}$ $0.2^{+0.2}_{-0.1}$	$0.37^{+0.06}_{-0.06}$ $0.78^{+0.06}_{-0.06}$	$0.8^{+0.2}_{-0.3}$ $6.3^{+0.2}_{-0.2}$	$2.2^{+0.3}_{-0.4}$ $7.1^{+0.3}_{-0.3}$
HSCJ223733+005015	$0.841^{+0.005}_{-0.005}$ $0.909^{+0.006}_{-0.005}$	$-2.59^{+0.02}_{-0.02}$ $2.40^{+0.03}_{-0.03}$	$0.114^{+0.001}_{-0.001}$ $1.277^{+0.007}_{-0.008}$	$5.6^{+0.3}_{-0.4}$ $0.8^{+0.02}_{-0.01}$	$16.5^{+0.2}_{-0.4}$ $0.310^{+0.006}_{-0.005}$	$3.9^{+0.1}_{-1}$ $0.091^{+0.003}_{-0.003}$	$49.3^{+0.5}_{-2}$ $0.81^{+0.02}_{-0.01}$	$77.5^{+0.8}_{-2}$ $1.25^{+0.03}_{-0.02}$
HSCJ230335+003703	$0.672^{+0.004}_{-0.003}$ $0.847^{+0.004}_{-0.004}$	$3.005^{+0.009}_{-0.009}$ $-3.10^{+0.01}_{-0.01}$	$0.335^{+0.006}_{-0.006}$ $1.89^{+0.02}_{-0.01}$	$3.29^{+0.06}_{-0.05}$ $2.04^{+0.04}_{-0.03}$	$4.59^{+0.1}_{-0.09}$ $0.329^{+0.006}_{-0.006}$	$0.74^{+0.02}_{-0.02}$ $0.094^{+0.002}_{-0.002}$	$10.3^{+0.3}_{-0.2}$ $0.68^{+0.01}_{-0.01}$	$15.0^{+0.4}_{-0.3}$ $0.91^{+0.02}_{-0.02}$
HSCJ230521-000211	$0.86^{+0.005}_{-0.005}$ $0.777^{+0.005}_{-0.005}$	$0.04^{+0.02}_{-0.02}$ $3.08^{+0.01}_{-0.01}$	$0.183^{+0.002}_{-0.001}$ $1.497^{+0.008}_{-0.008}$	$5.94^{+0.05}_{-0.1}$ $0.98^{+0.02}_{-0.02}$	$13.15^{+0.1}_{-0.2}$ $0.418^{+0.005}_{-0.005}$	$1.97^{+0.05}_{-0.05}$ $0.134^{+0.003}_{-0.003}$	$33.5^{+0.3}_{-0.6}$ $0.867^{+0.008}_{-0.008}$	$49.4^{+0.4}_{-0.9}$ $1.16^{+0.01}_{-0.01}$
HSCJ233130+003733	$0.402^{+0.001}_{-0.001}$ $0.445^{+0.001}_{-0.001}$	$0.850^{+0.003}_{-0.002}$ $0.797^{+0.007}_{-0.006}$	$1.60^{+0.02}_{-0.02}$ $2.25^{+0.02}_{-0.02}$	$2.508^{+0.02}_{-0.009}$ $0.518^{+0.004}_{-0.003}$	$0.351^{+0.006}_{-0.004}$ $0.191^{+0.005}_{-0.005}$	$0.052^{+0.002}_{-0.002}$ $0.102^{+0.003}_{-0.003}$	$1.054^{+0.02}_{-0.007}$ $0.324^{+0.009}_{-0.01}$	$1.68^{+0.03}_{-0.02}$ $0.38^{+0.01}_{-0.02}$
HSCJ233146+013845	$0.71^{+0.01}_{-0.01}$ $0.893^{+0.004}_{-0.004}$	$2.63^{+0.02}_{-0.02}$ $2.38^{+0.02}_{-0.03}$	$1.71^{+0.02}_{-0.02}$ $0.88^{+0.02}_{-0.02}$	$0.52^{+0.03}_{-0.02}$ $3.32^{+0.04}_{-0.04}$	$0.193^{+0.007}_{-0.006}$ $0.99^{+0.03}_{-0.03}$	$0.057^{+0.003}_{-0.003}$ $0.195^{+0.007}_{-0.007}$	$0.37^{+0.01}_{-0.01}$ $2.24^{+0.07}_{-0.07}$	$0.61^{+0.02}_{-0.02}$ $3.10^{+0.1}_{-0.09}$

Figure A.1: Fit of lens HSCJ015618-010747. Top to bottom: g, r, i, and z filters.
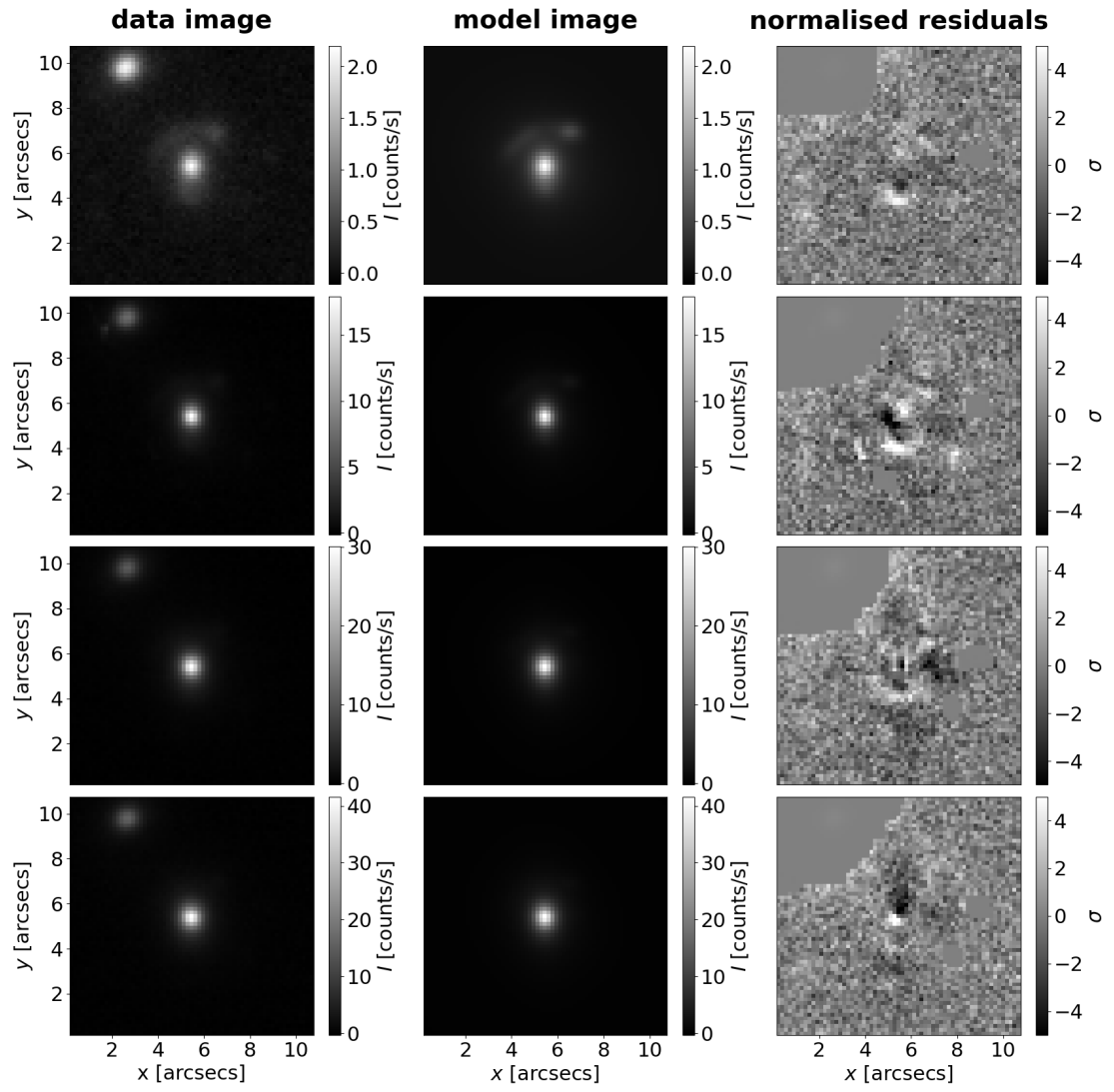
Figure A.2: Fit of lens HSCJ020141-030946. Top to bottom: $g, r, i,$ and z filters.
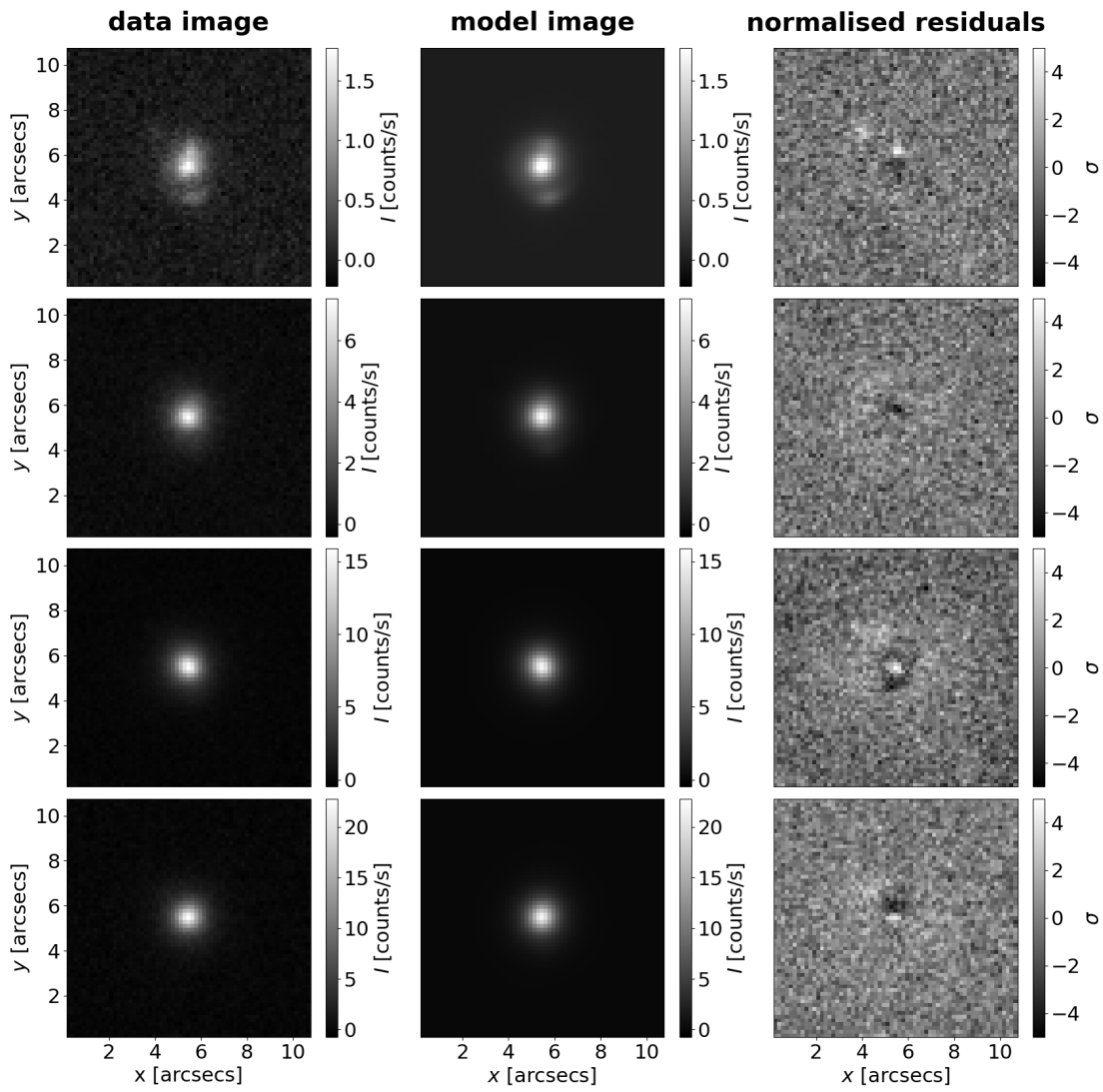
Figure A.3: Fit of lens HSCJ020241-064611. Top to bottom: g, r, i, and z filters.
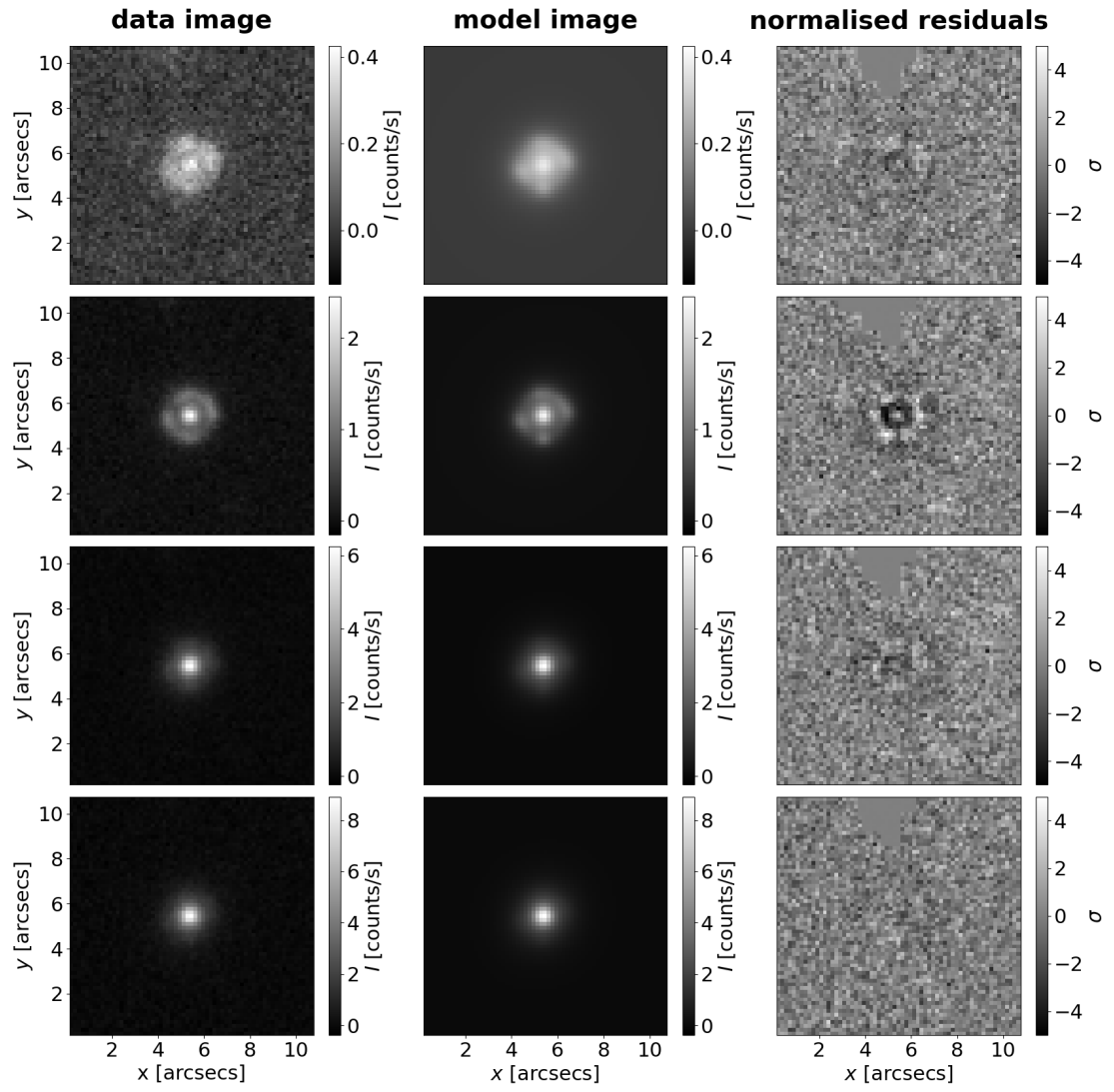
Figure A.4: Fit of lens HSCJ020955-024442. Top to bottom: *g, r, i*, and *z* filters.
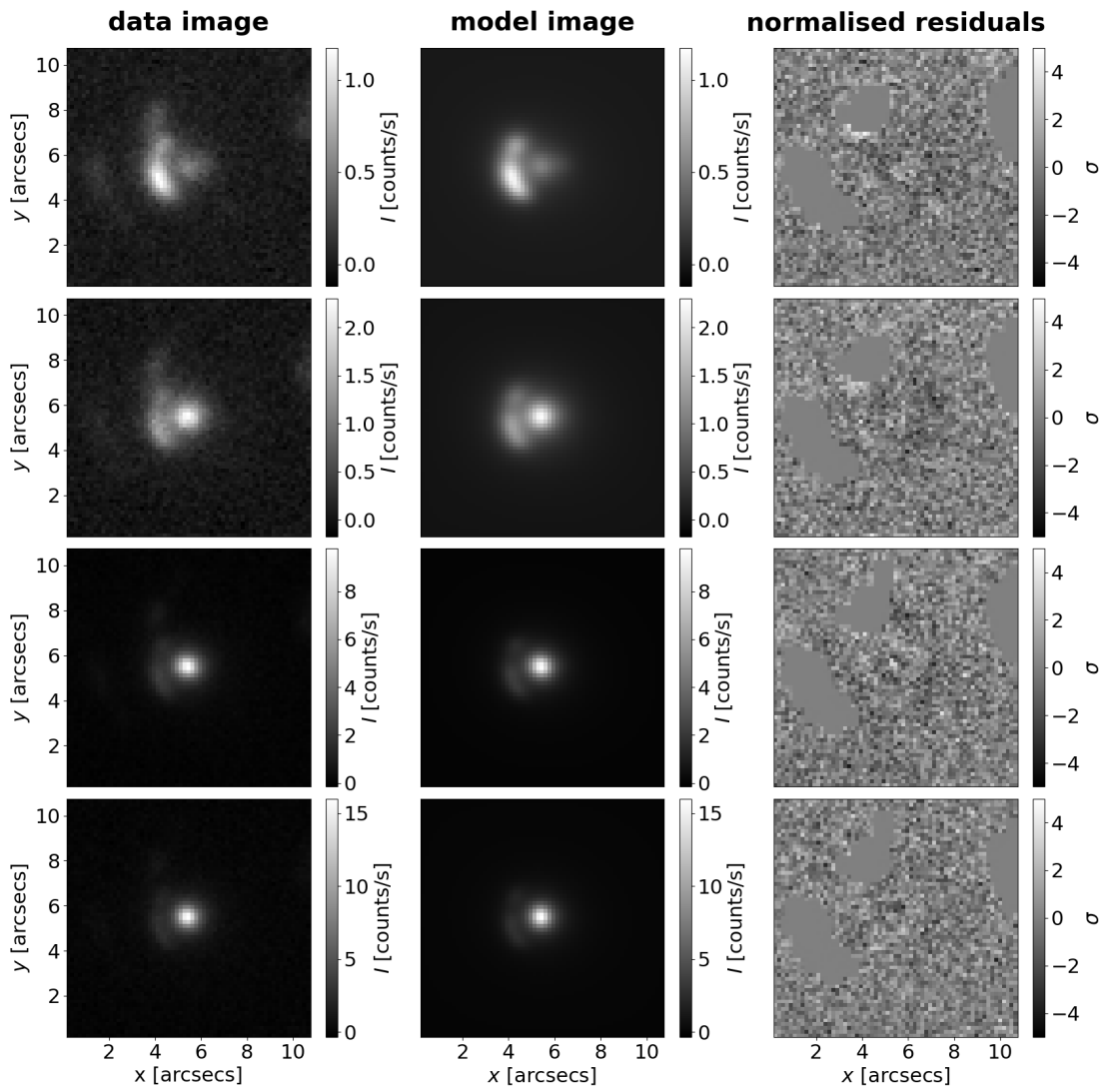
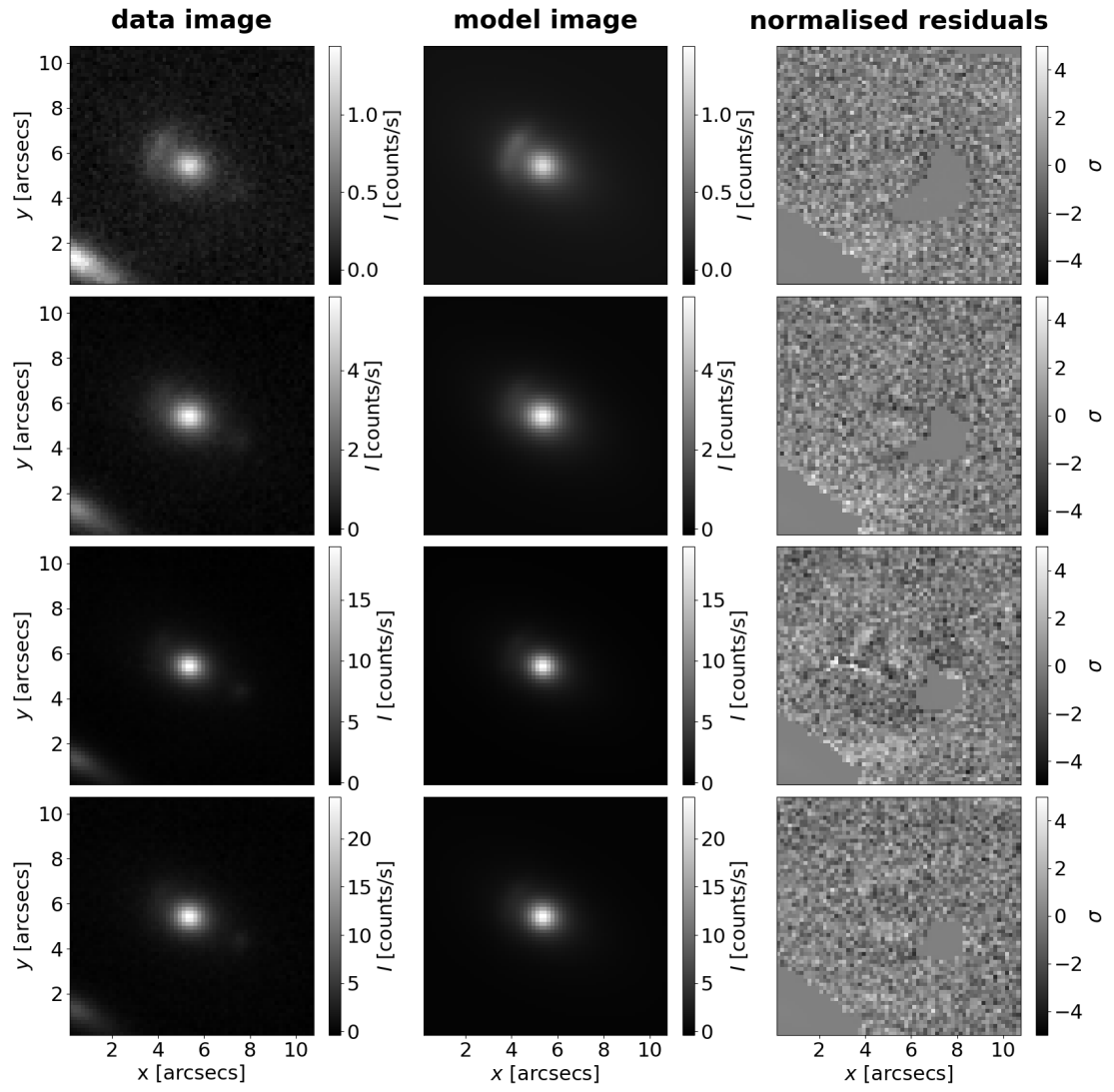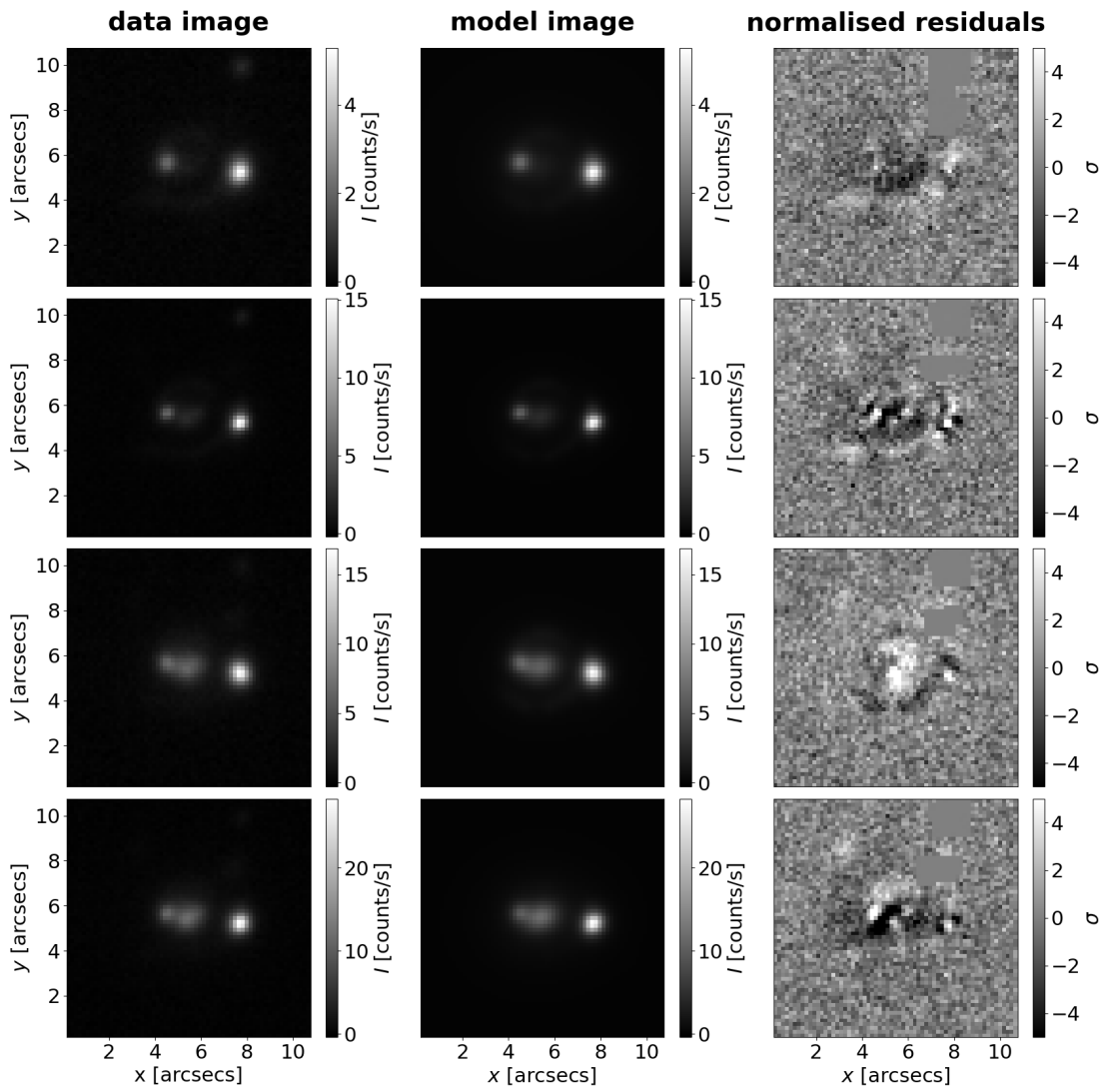**Figure A.5:** Fit of lens HSCJ021737-051329. Top to bottom: $g, r, i,$ and z filters.

Figure A.6: Fit of lens HSCJ022346-053418. Top to bottom: *g*, *r*, *i*, and *z* filters.
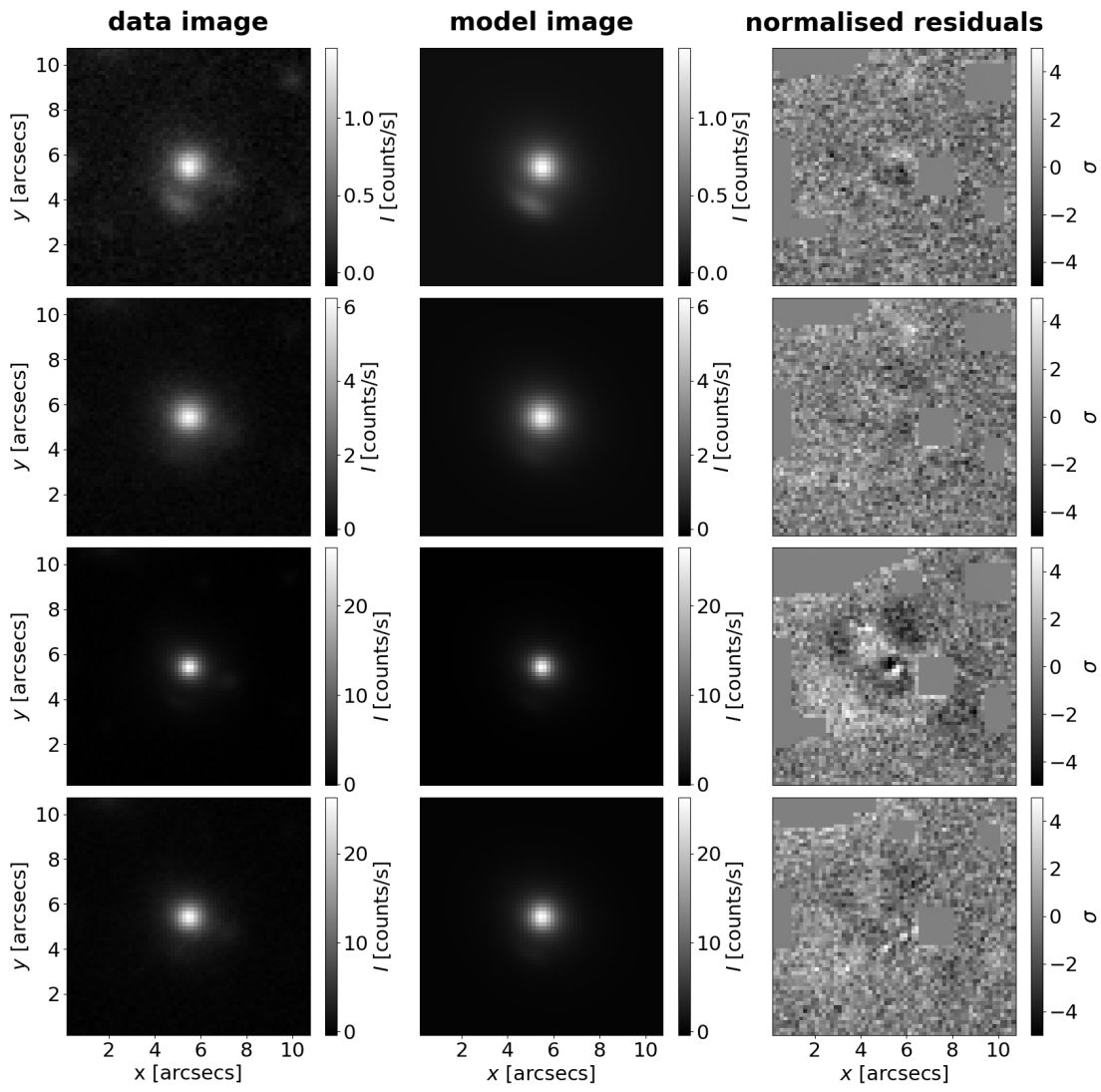
Figure A.7: Fit of lens HSCJ022610-042011. Top to bottom: g, r, i, and z filters.
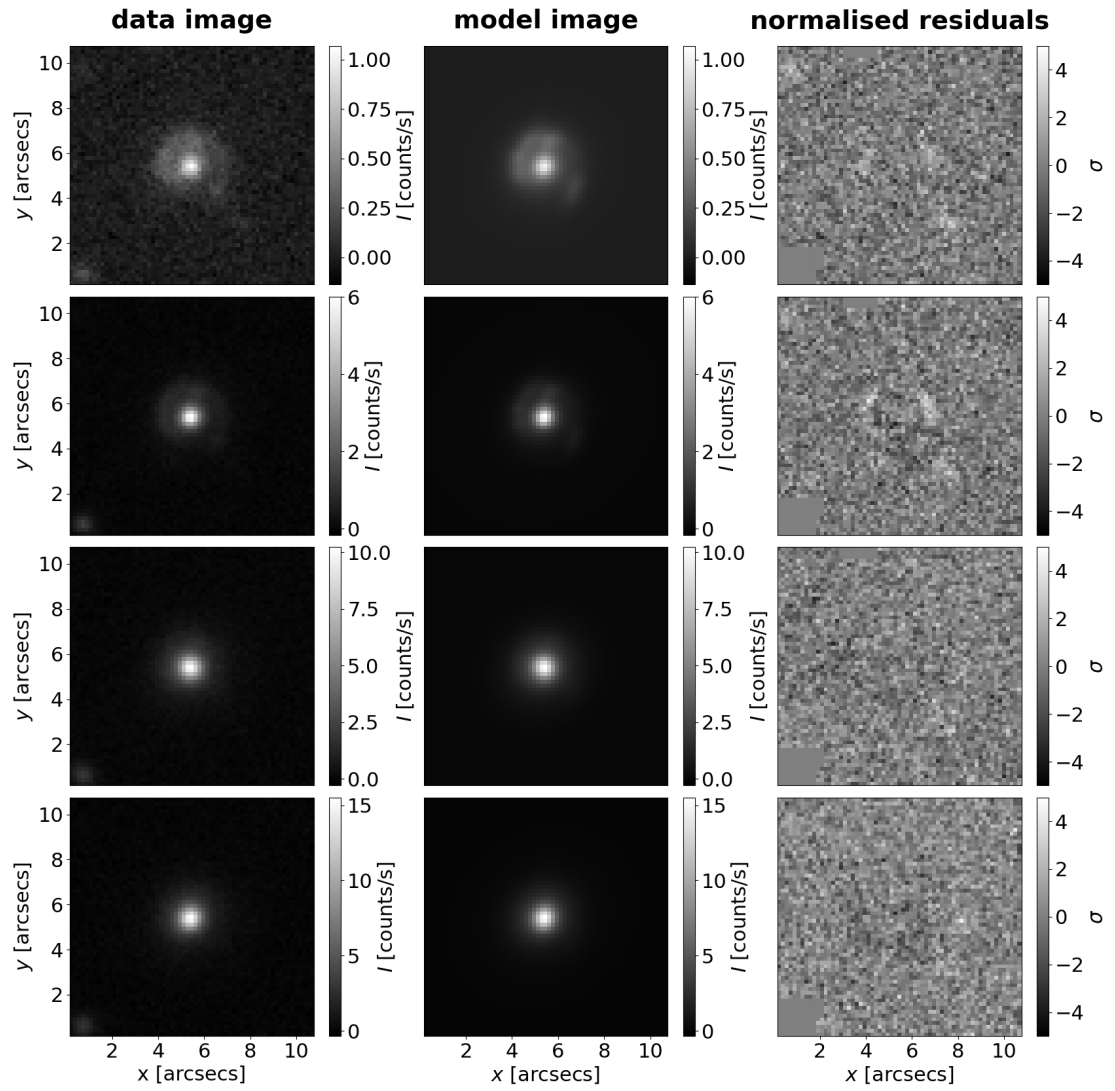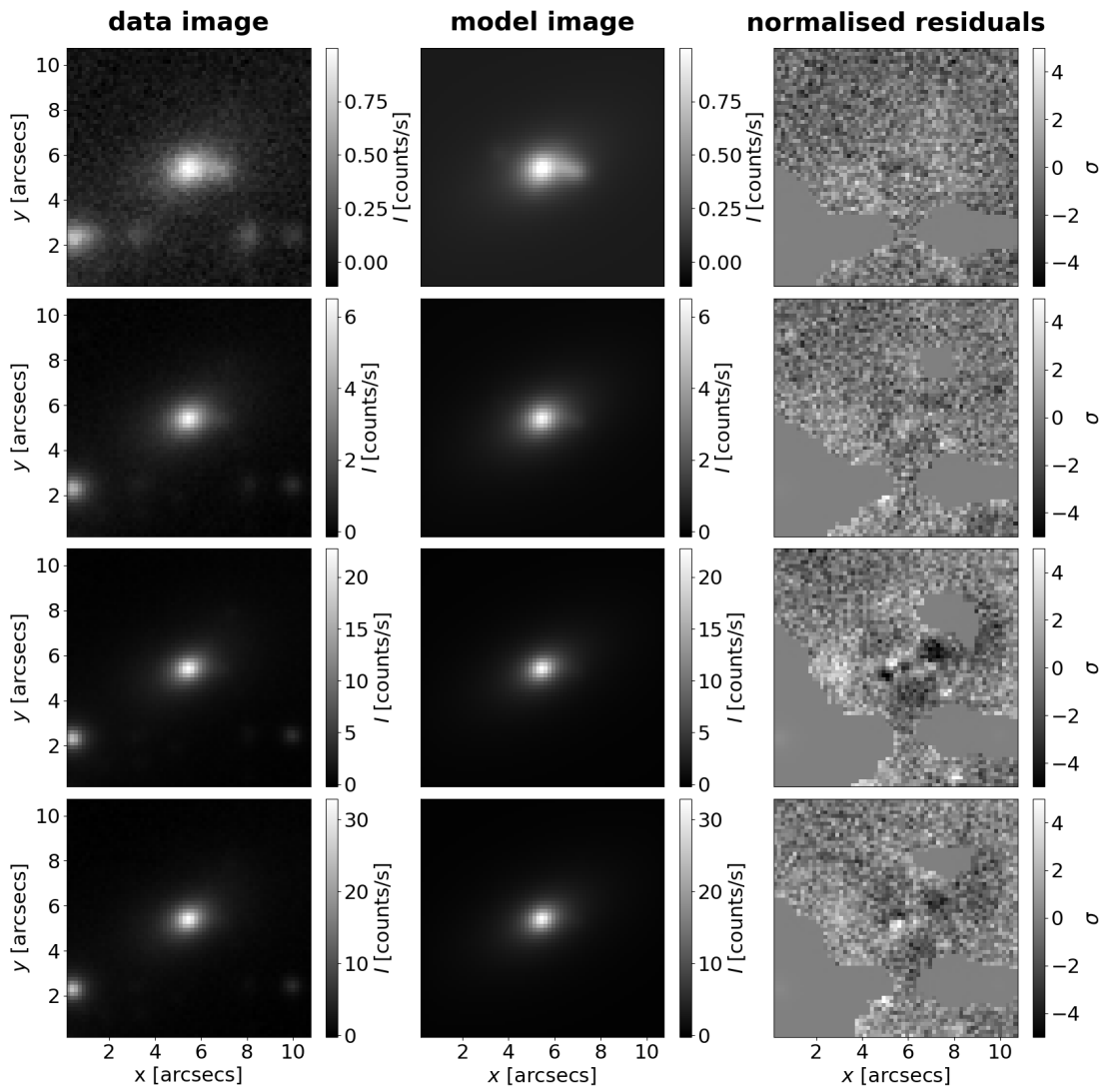
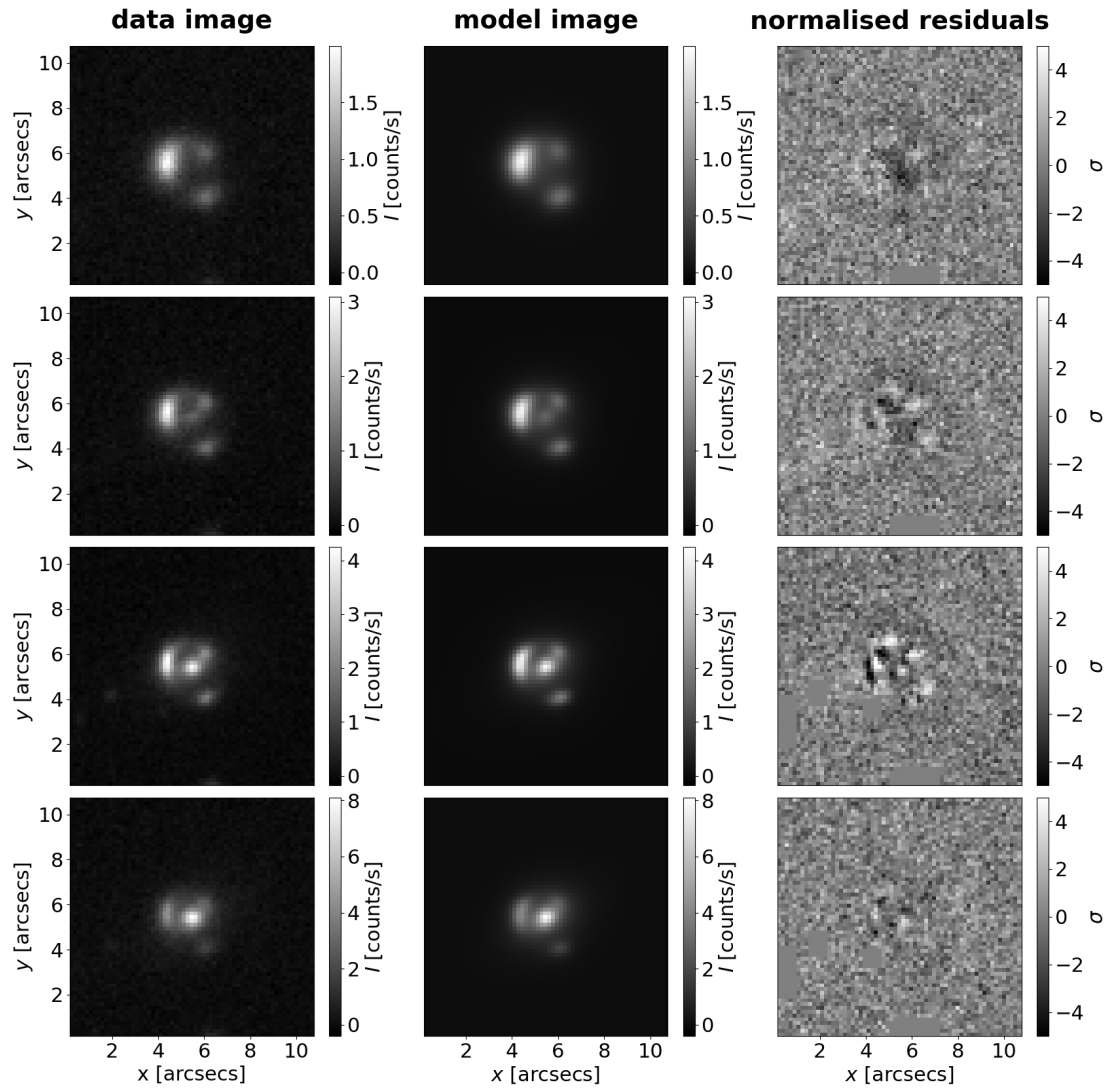**Figure A.8:** Fit of lens HSCJ023217-021703. Top to bottom: *g*, *r*, *i*, and *z* filters.

Figure A.9: Fit of lens HSCJ023322-020530. Top to bottom: g, r, i, and z filters.
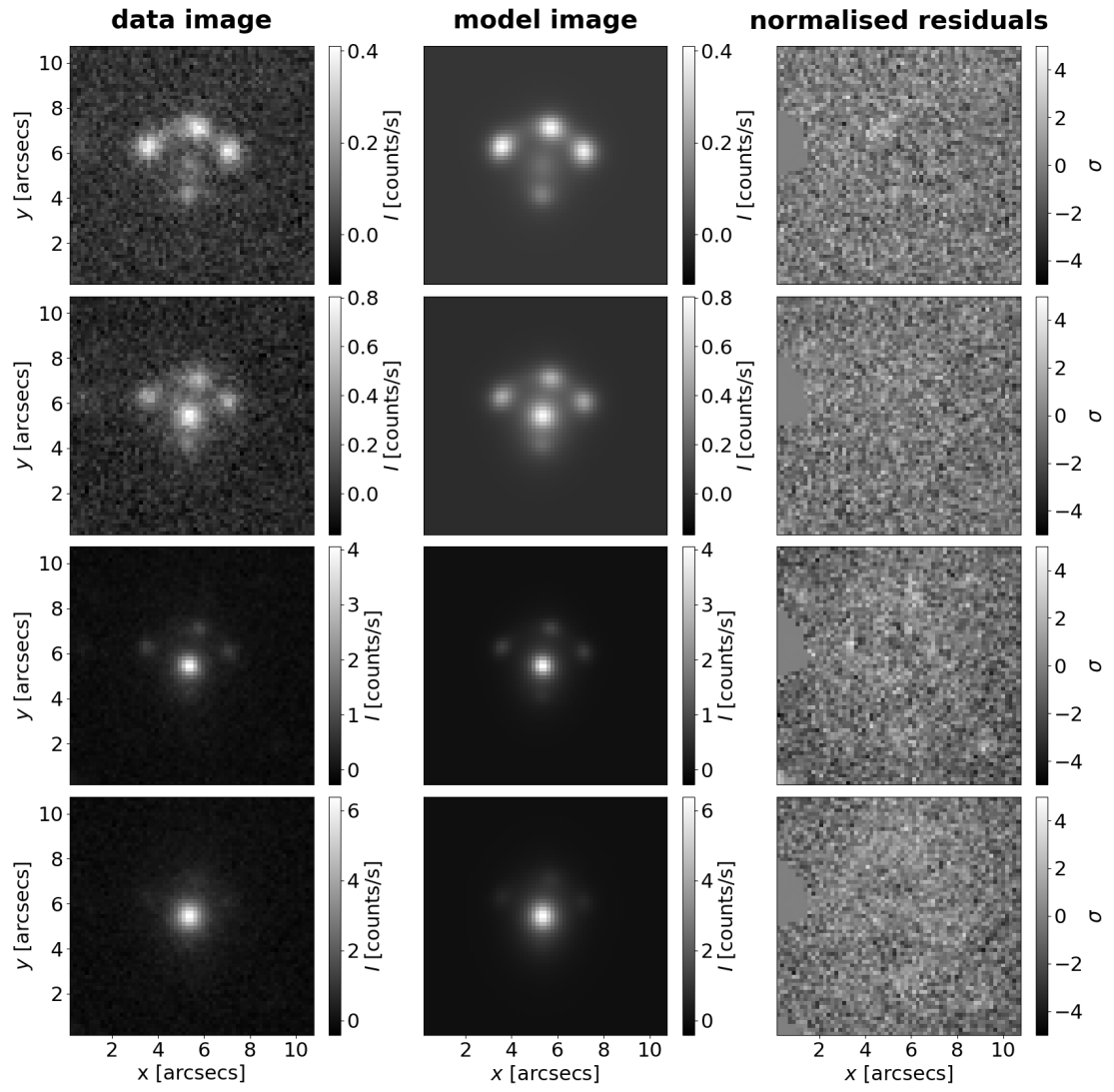
Figure A.10: Fit of lens HSCJ085046+00390. Top to bottom: $g, r, i,$ and z filters.
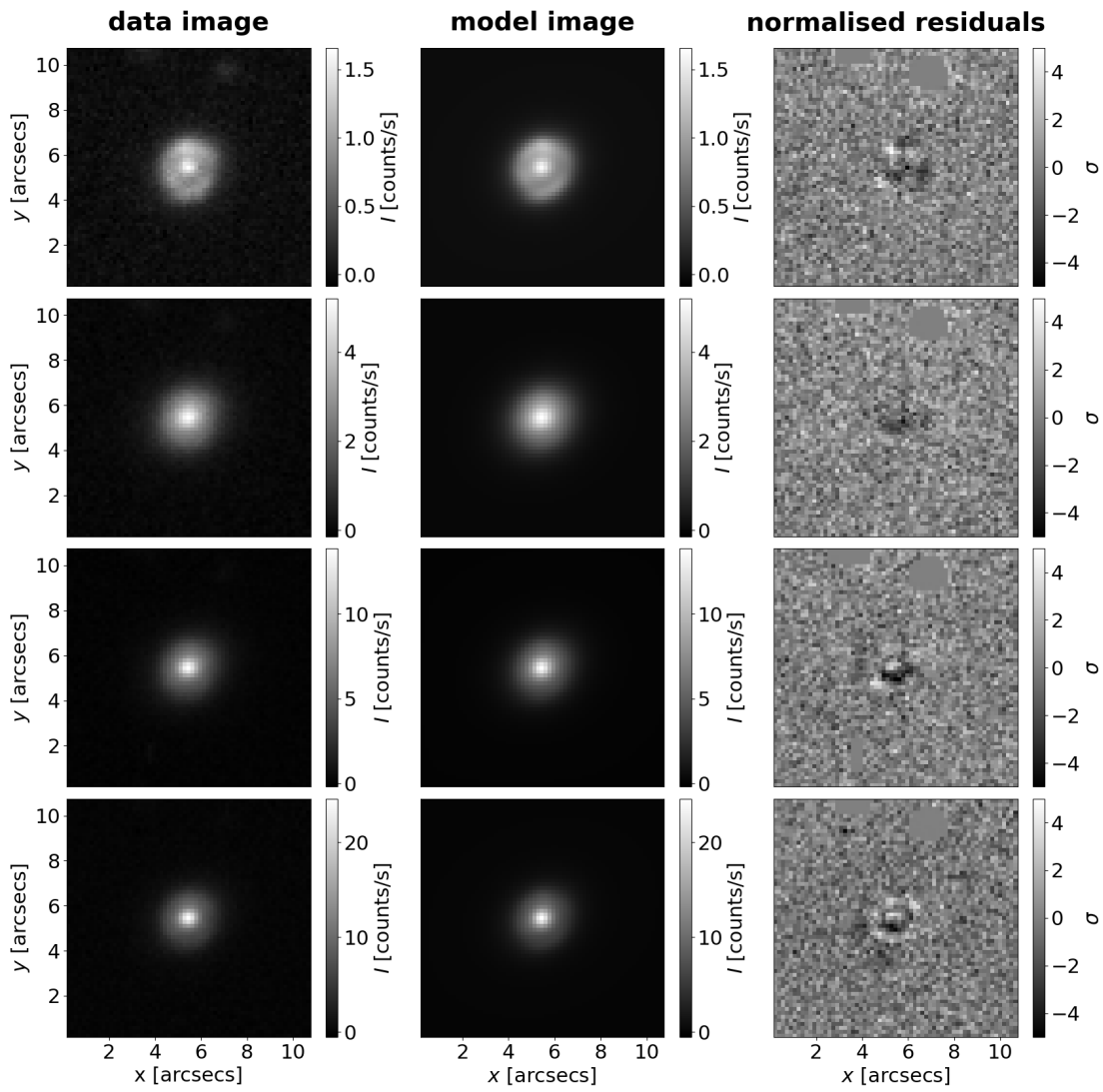
Figure A.11: Fit of lens HSCJ085855-010208. Top to bottom: g, r, i, and z filters.
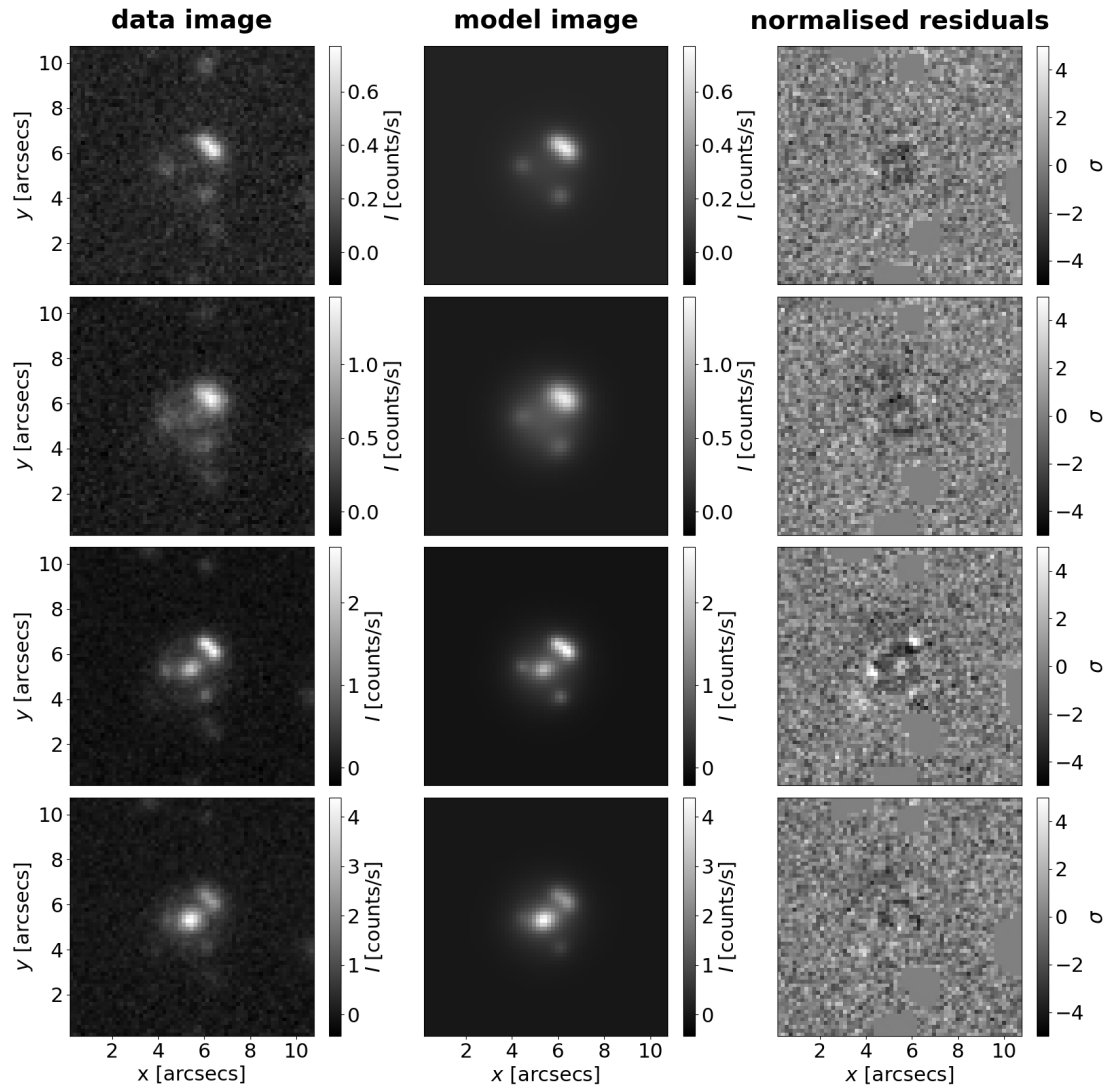
Figure A.12: Fit of lens HSCJ090429-010228. Top to bottom: g, r, i, and z filters.
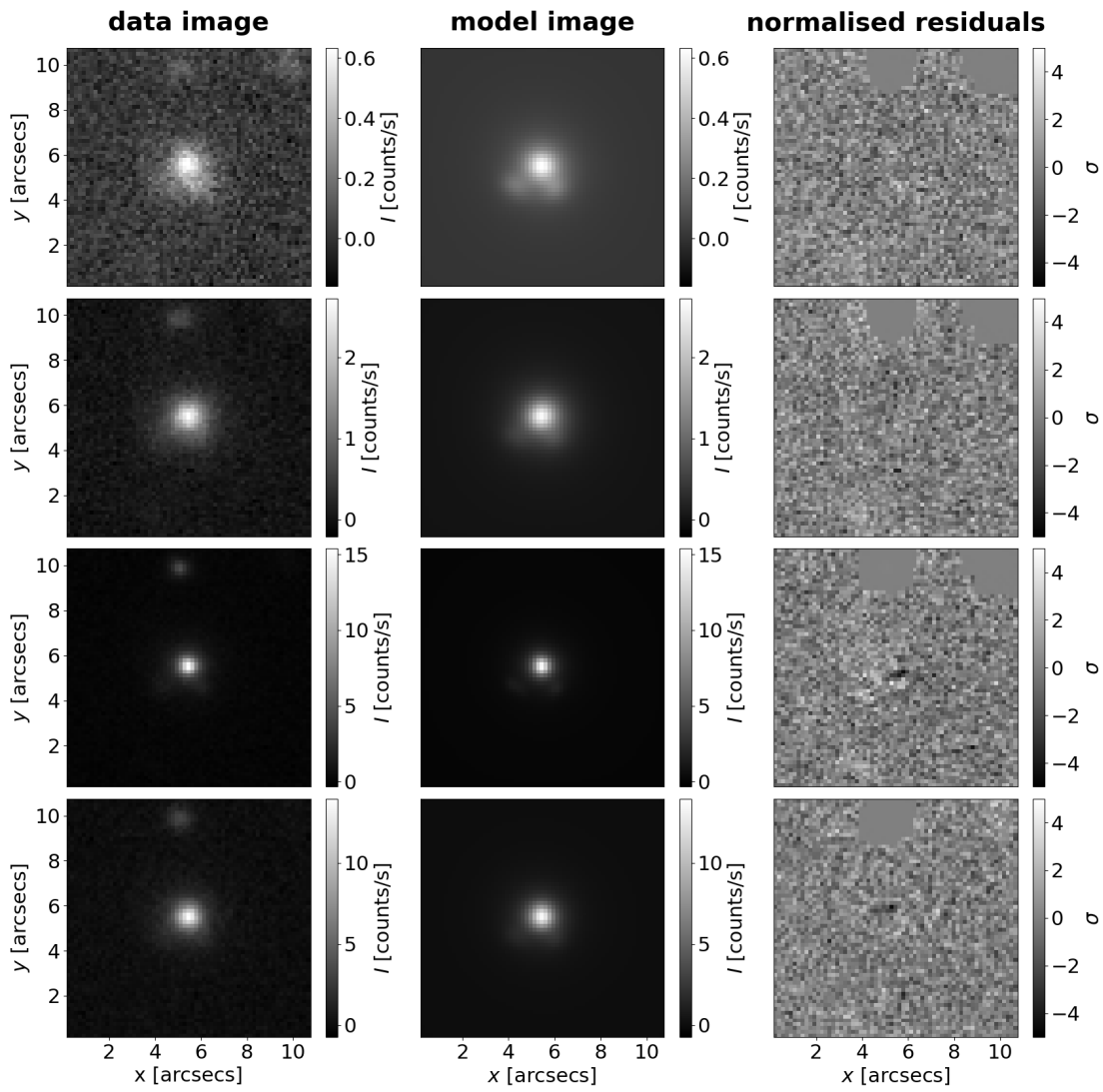
Figure A.13: Fit of lens HSCJ094427-014742. Top to bottom: g, r, i, and z filters.
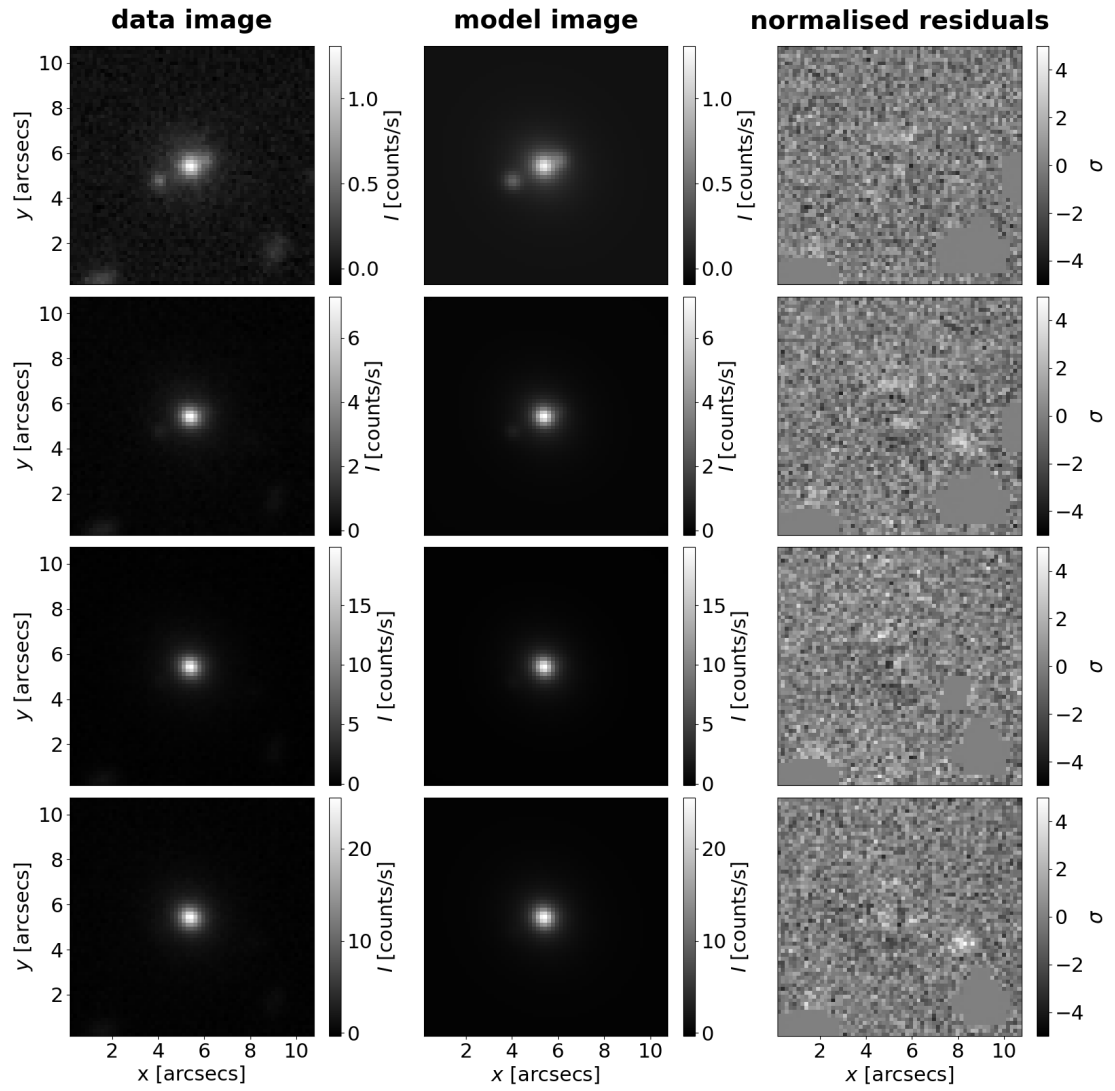
Figure A.14: Fit of lens HSCJ120623+001507. Top to bottom: $g, r, i,$ and z filters.
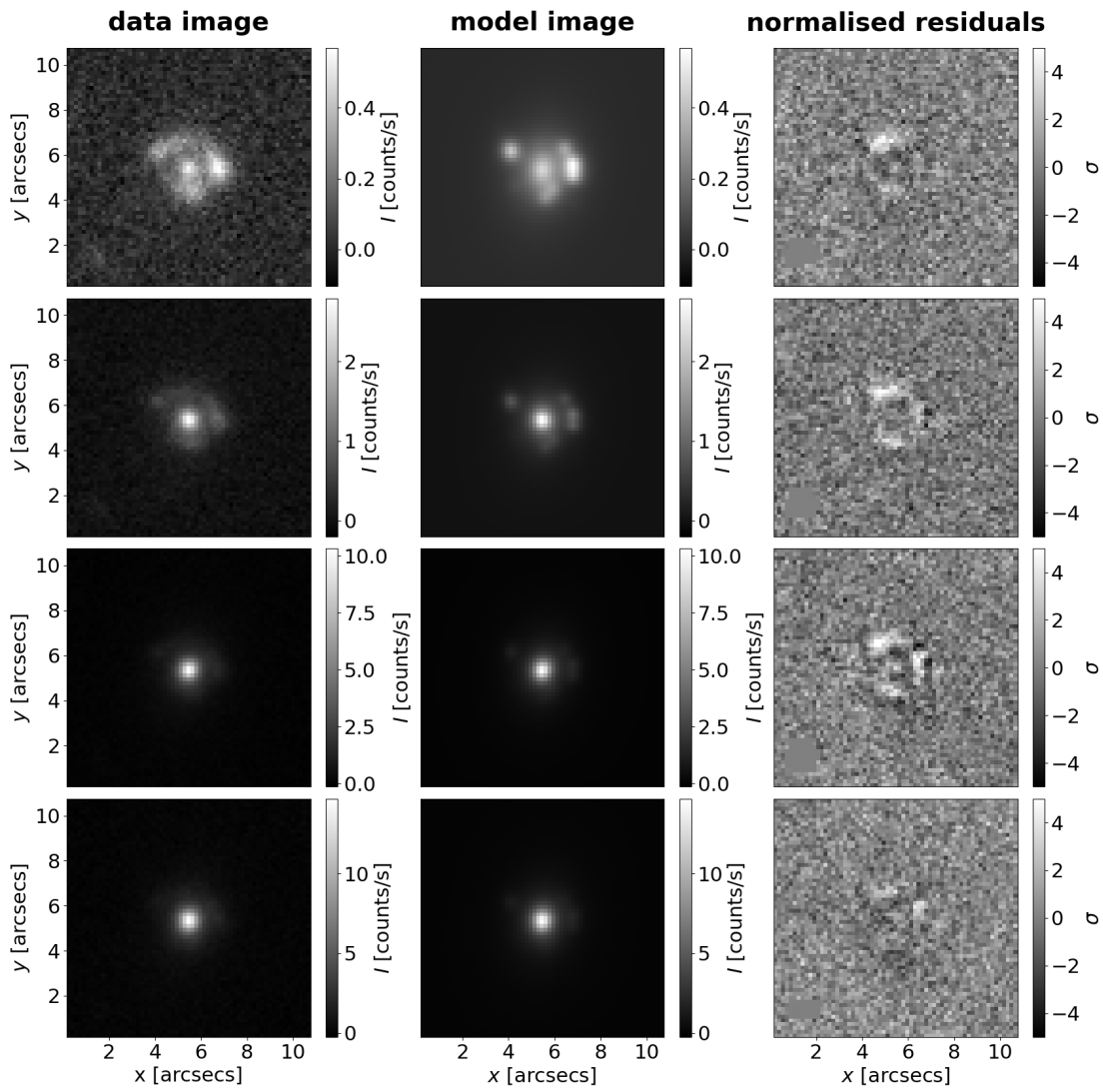
Figure A.15: Fit of lens HSCJ121052-011905. Top to bottom: g, r, i, and z filters.
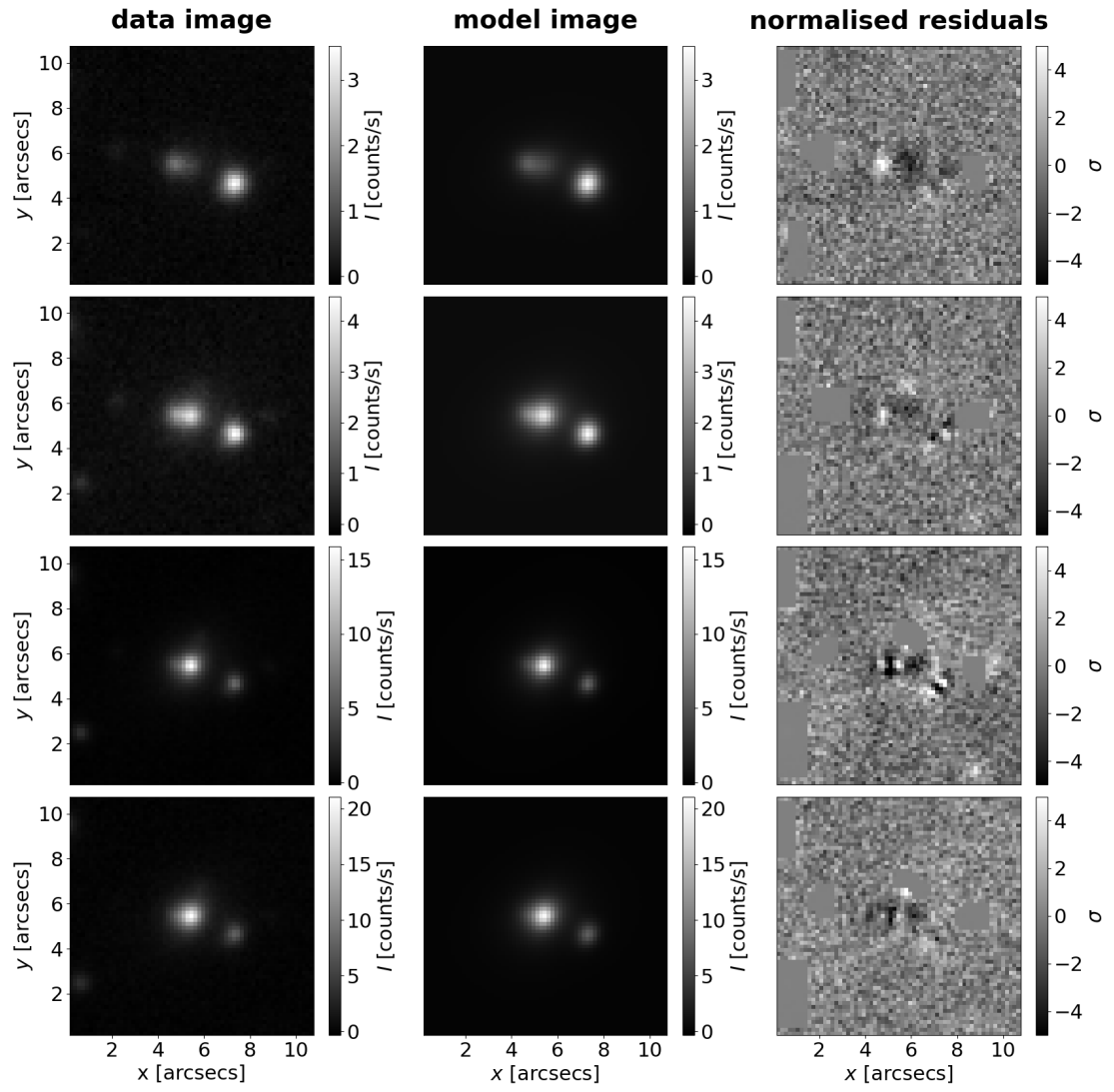
Figure A.16: Fit of lens HSCJ121504+004726. Top to bottom: g, r, i, and z filters.
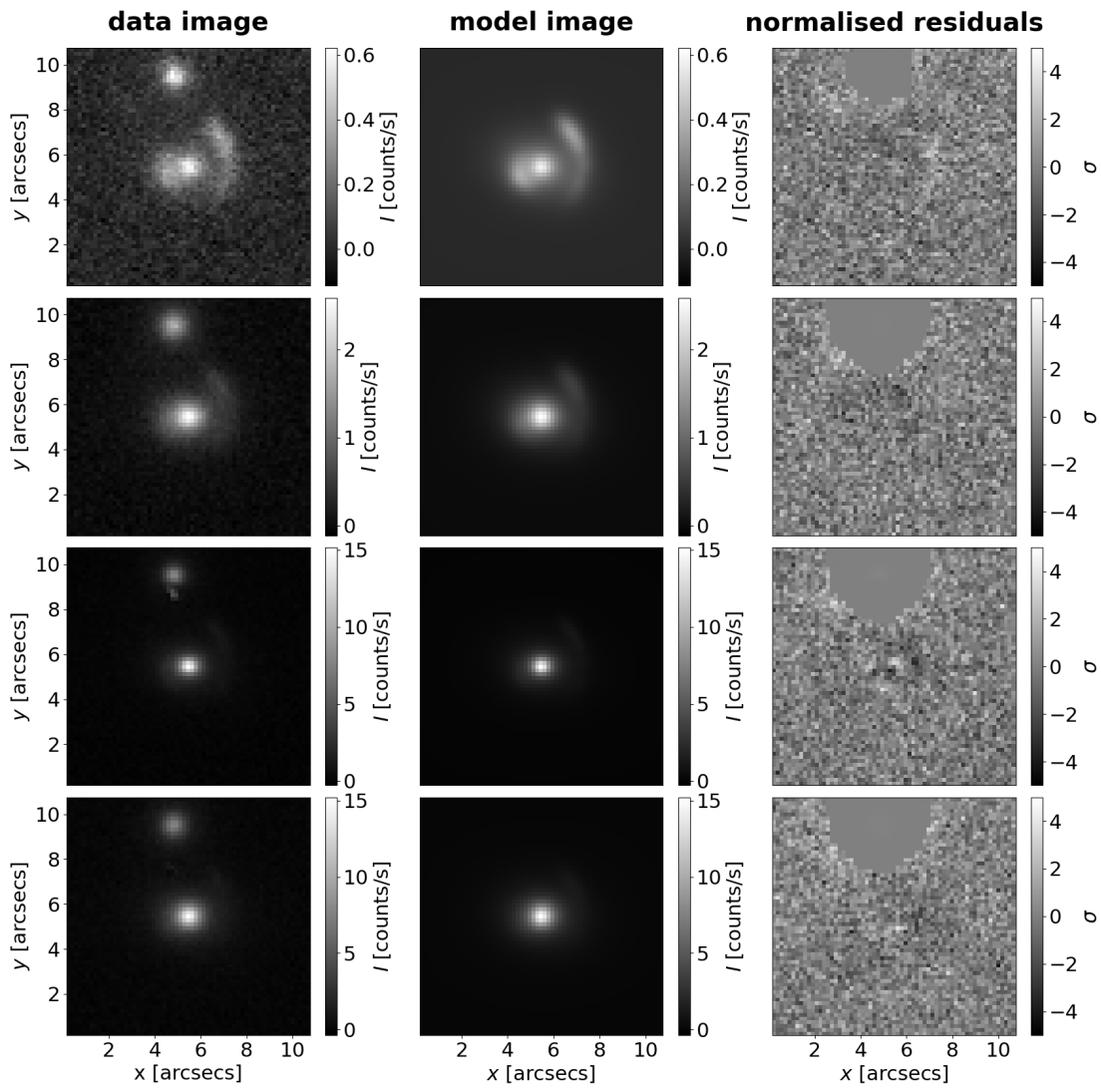
Figure A.17: Fit of lens HSCJ124320-004517. Top to bottom: g, r, i, and z filters.
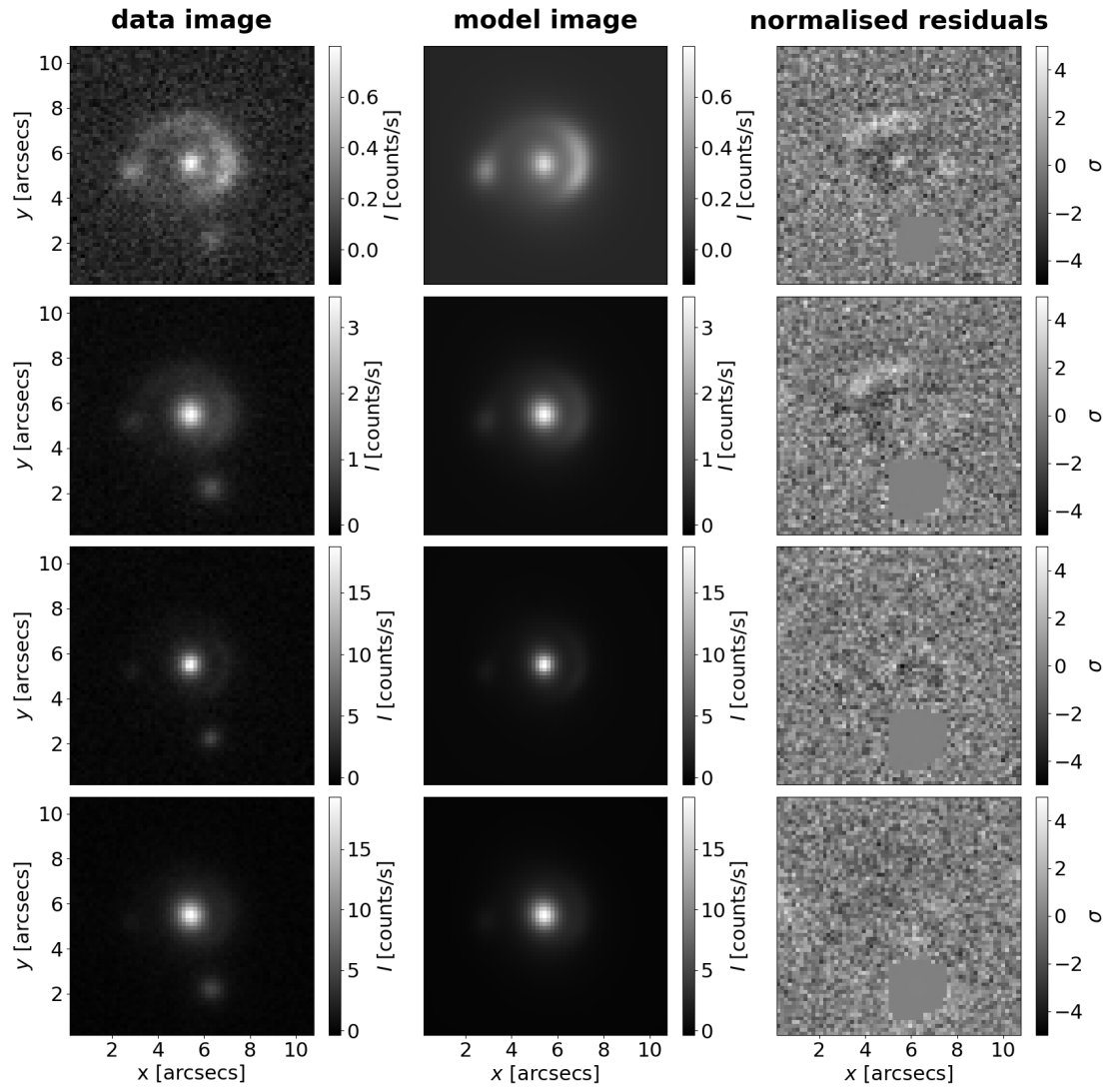
Figure A.18: Fit of lens HSCJ125254+004356. Top to bottom: $g, r, i,$ and z filters.
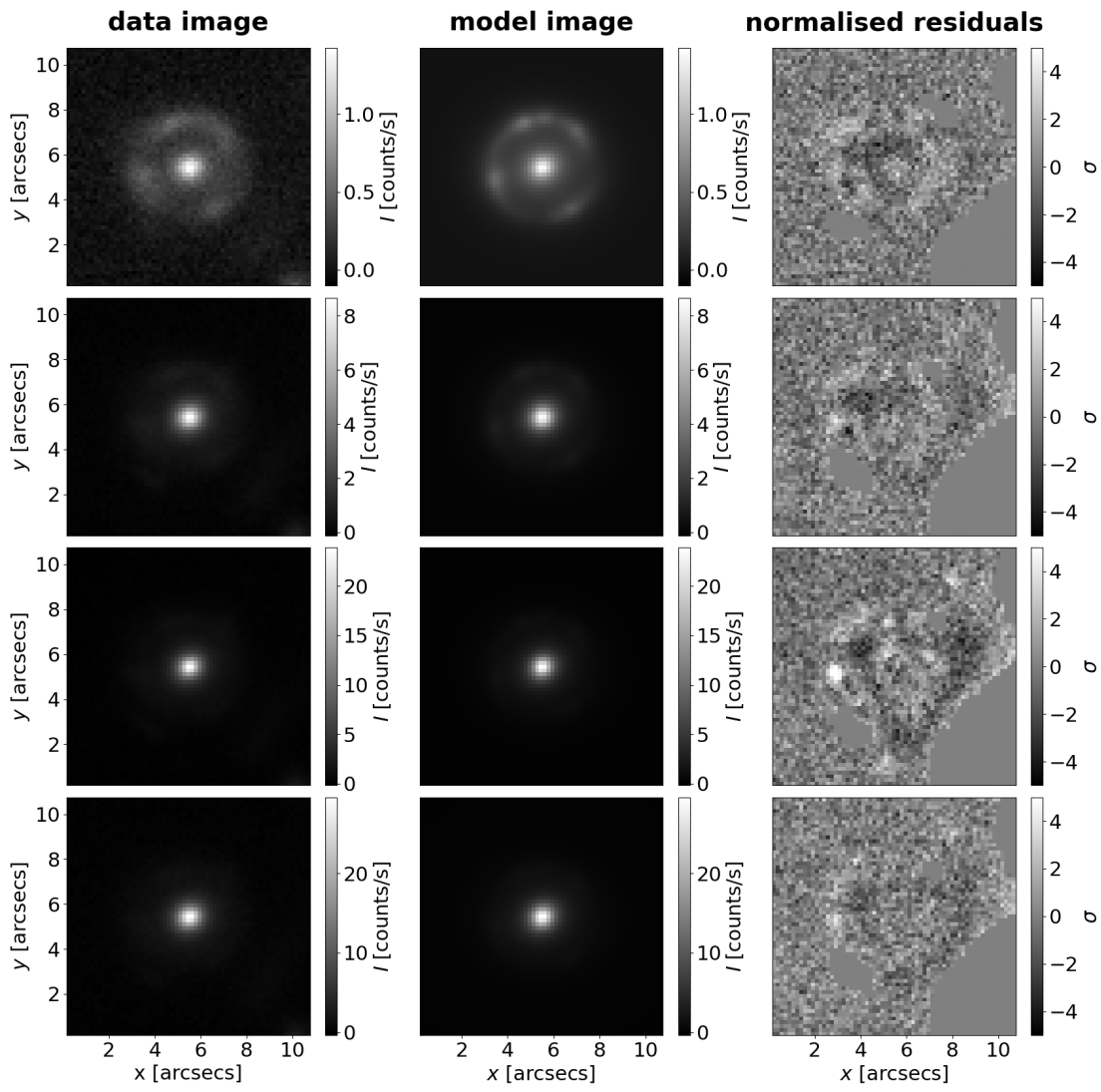
Figure A.19: Fit of lens HSCJ135138+002839. Top to bottom: g, r, i, and z filters.
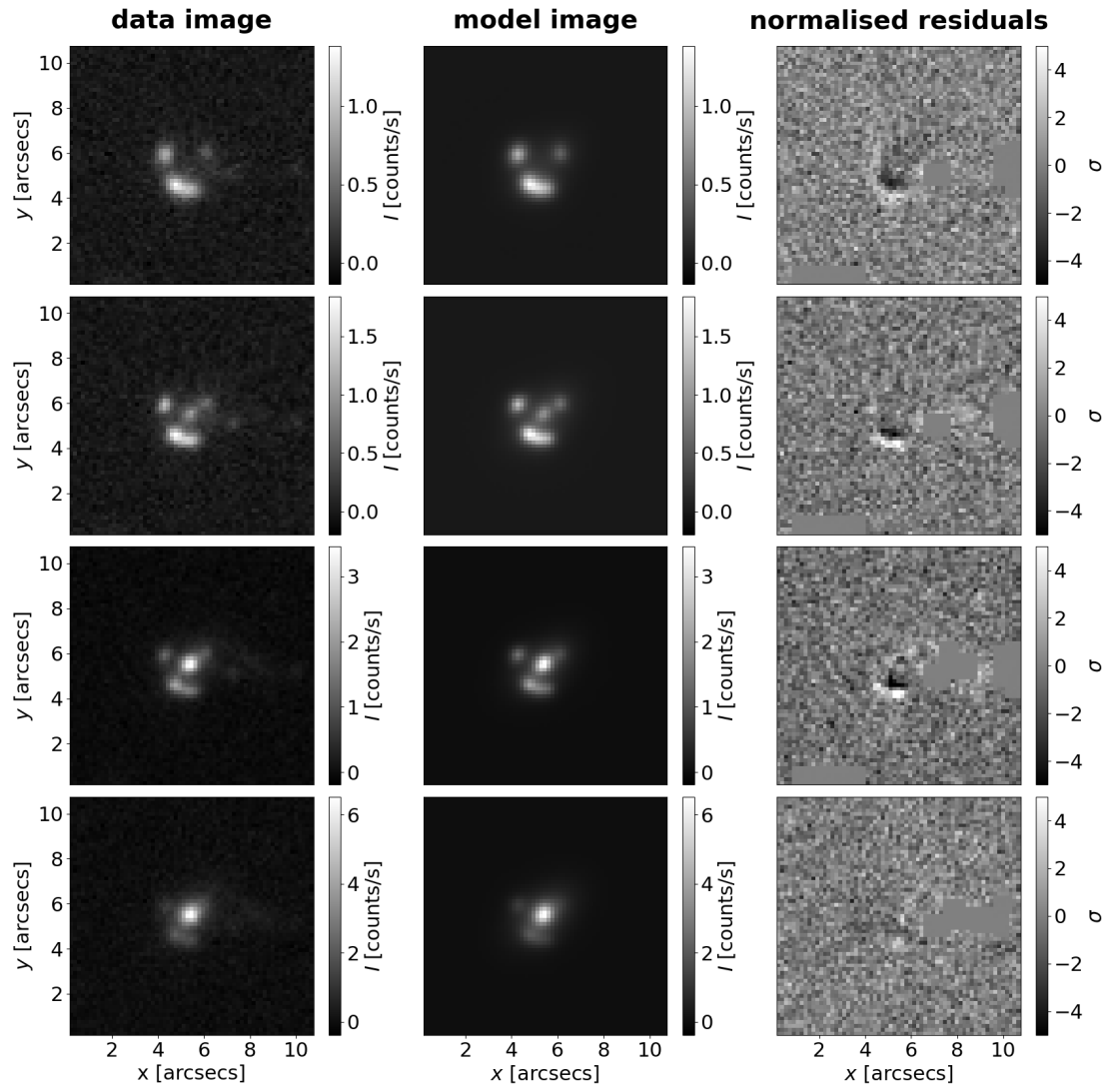
Figure A.20: Fit of lens HSCJ141136-010215. Top to bottom: g, r, i, and z filters.
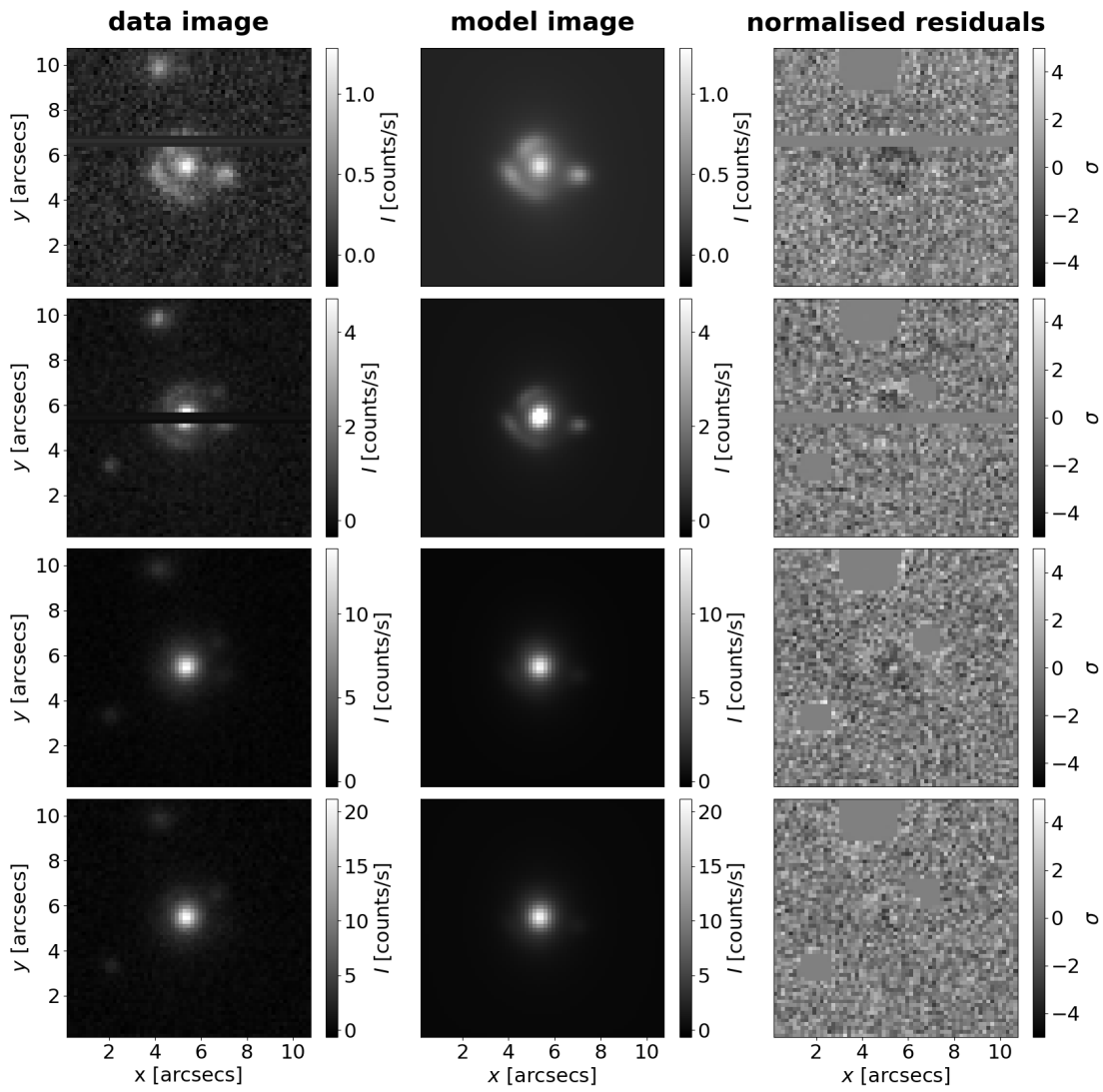
Figure A.21: Fit of lens HSCJ141815+015832. Top to bottom: g, r, i, and z filters.
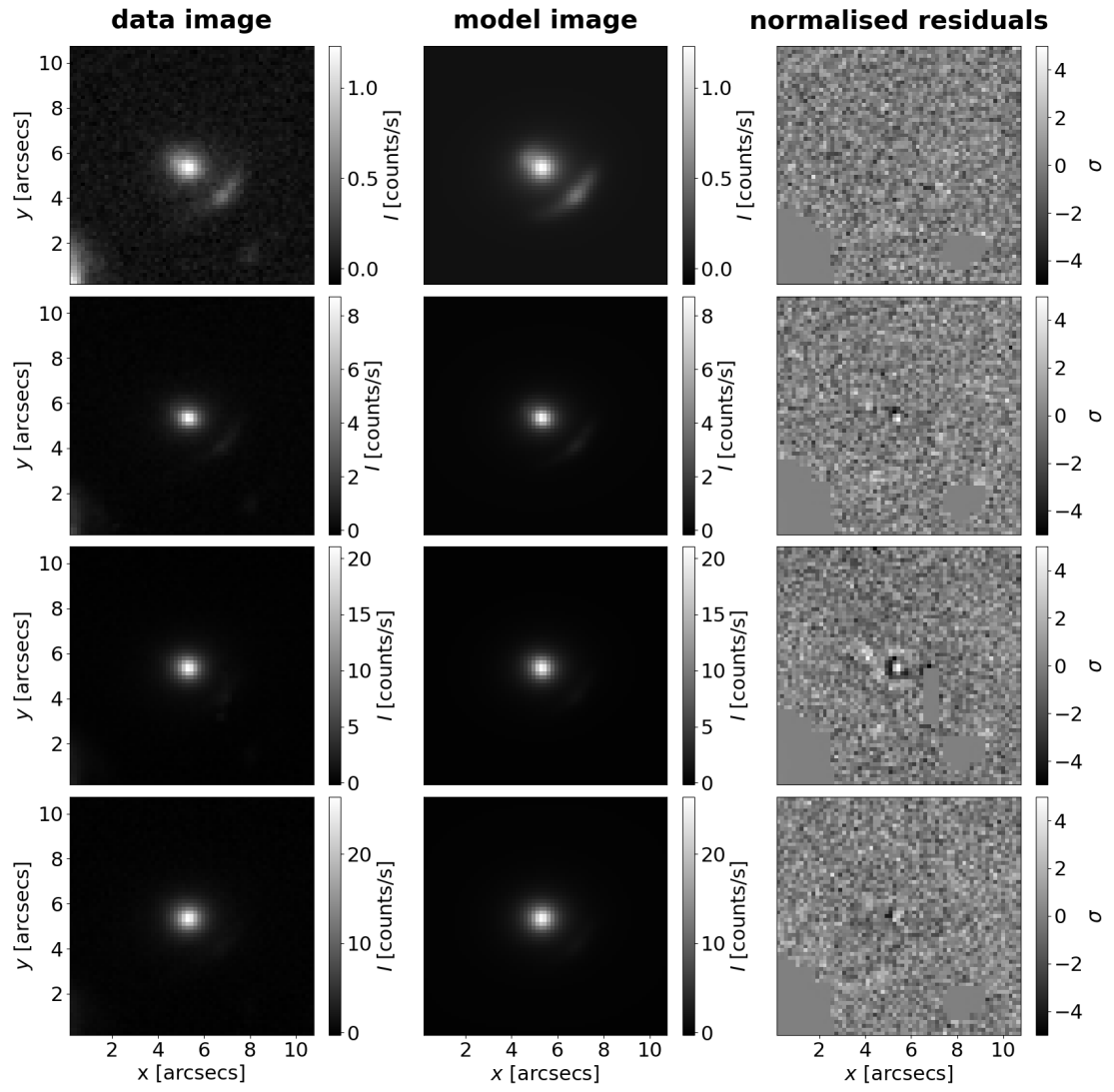
Figure A.22: Fit of lens HSCJ142720+001916. Top to bottom: g, r, i, and z filters.

Figure A.23: Fit of lens HSCJ144132-005358. Top to bottom: $g, r, i,$ and z filters.

Figure A.24: Fit of lens HSCJ144320-012537. Top to bottom: $g, r, i,$ and z filters.
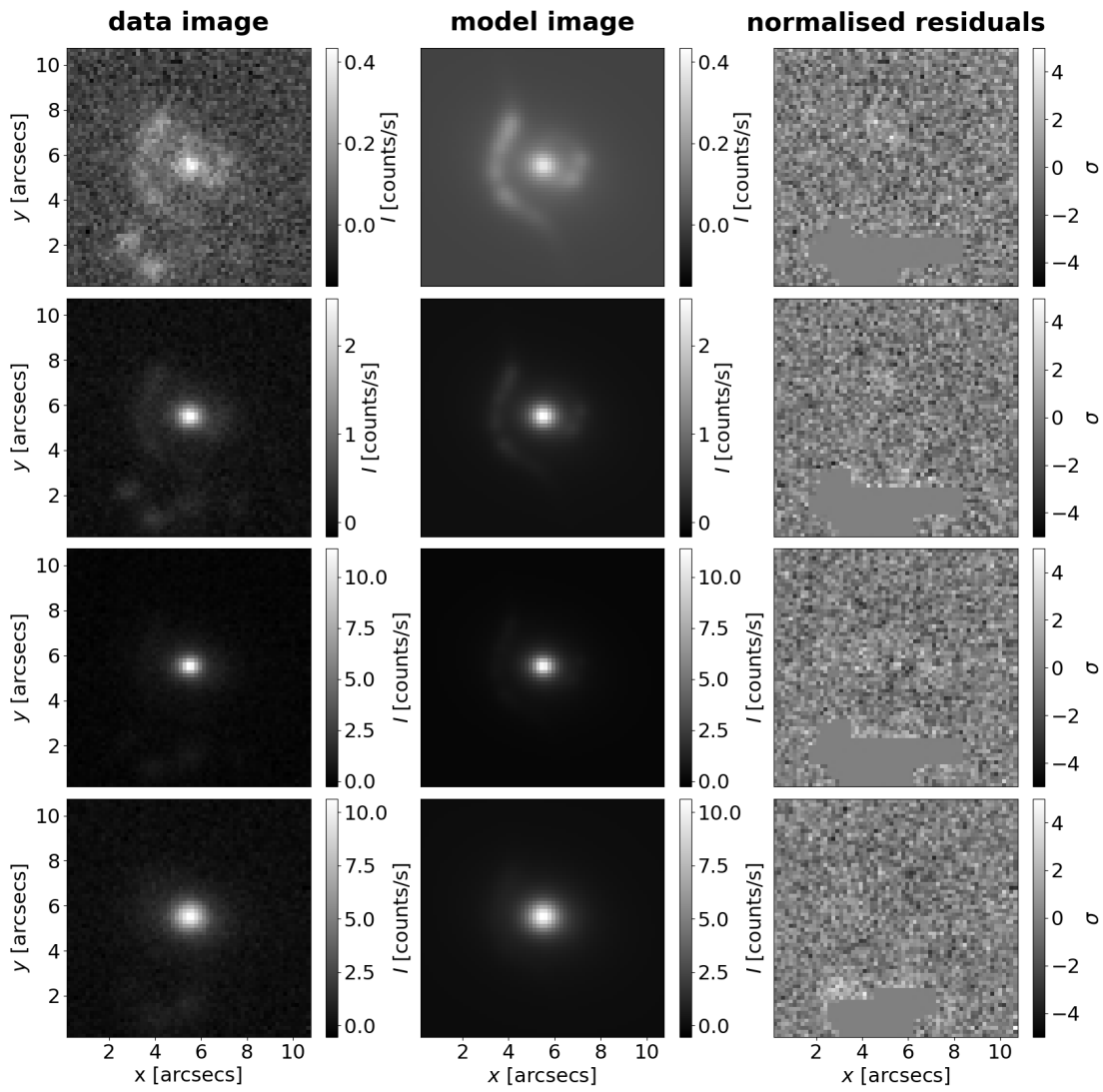
Figure A.25: Fit of lens HSCJ145242+425731. Top to bottom: g, r, i, and z filters.
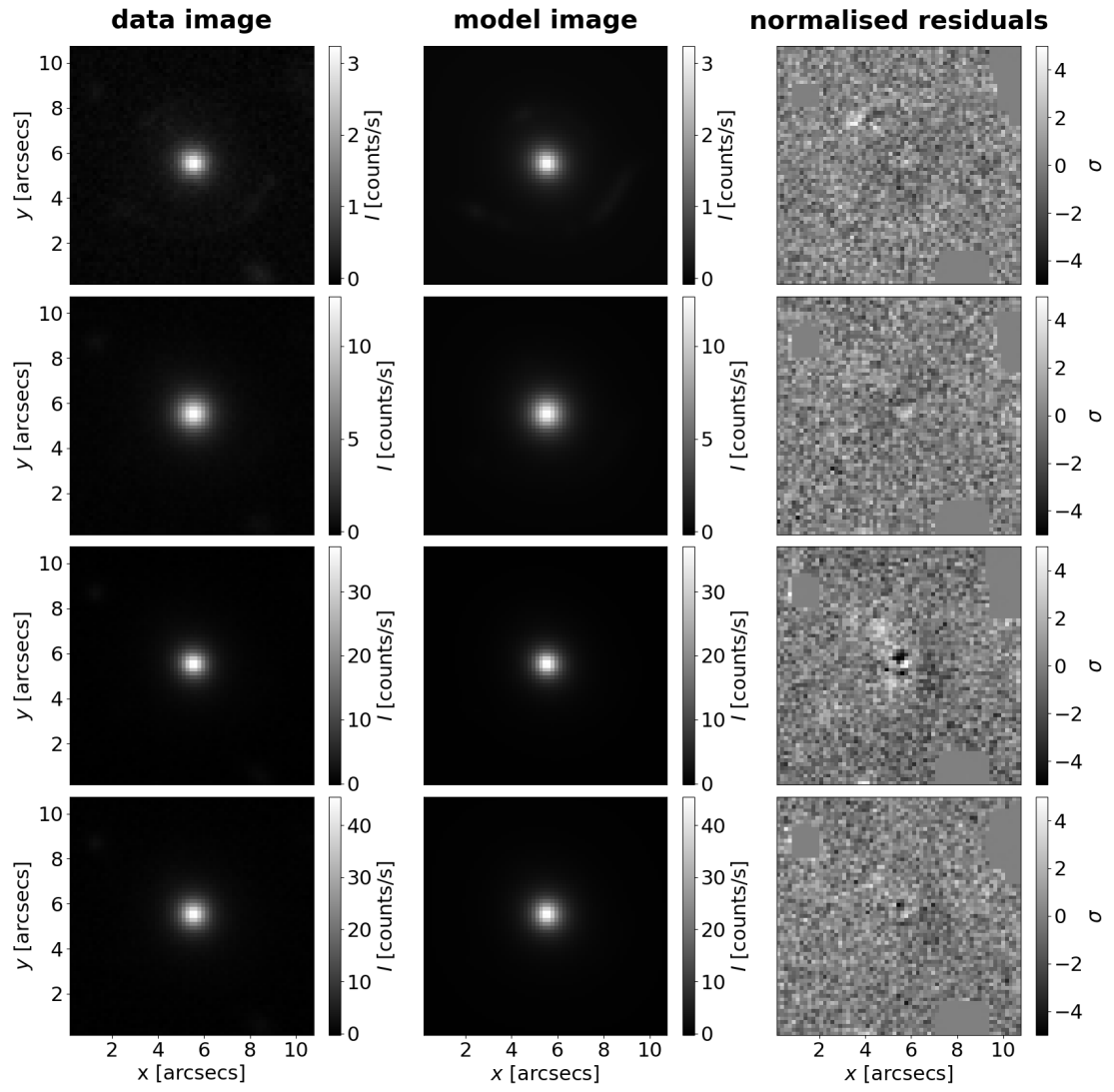
Figure A.26: Fit of lens HSCJ150021-004936. Top to bottom: *g*, *r*, *i*, and *z* filters.
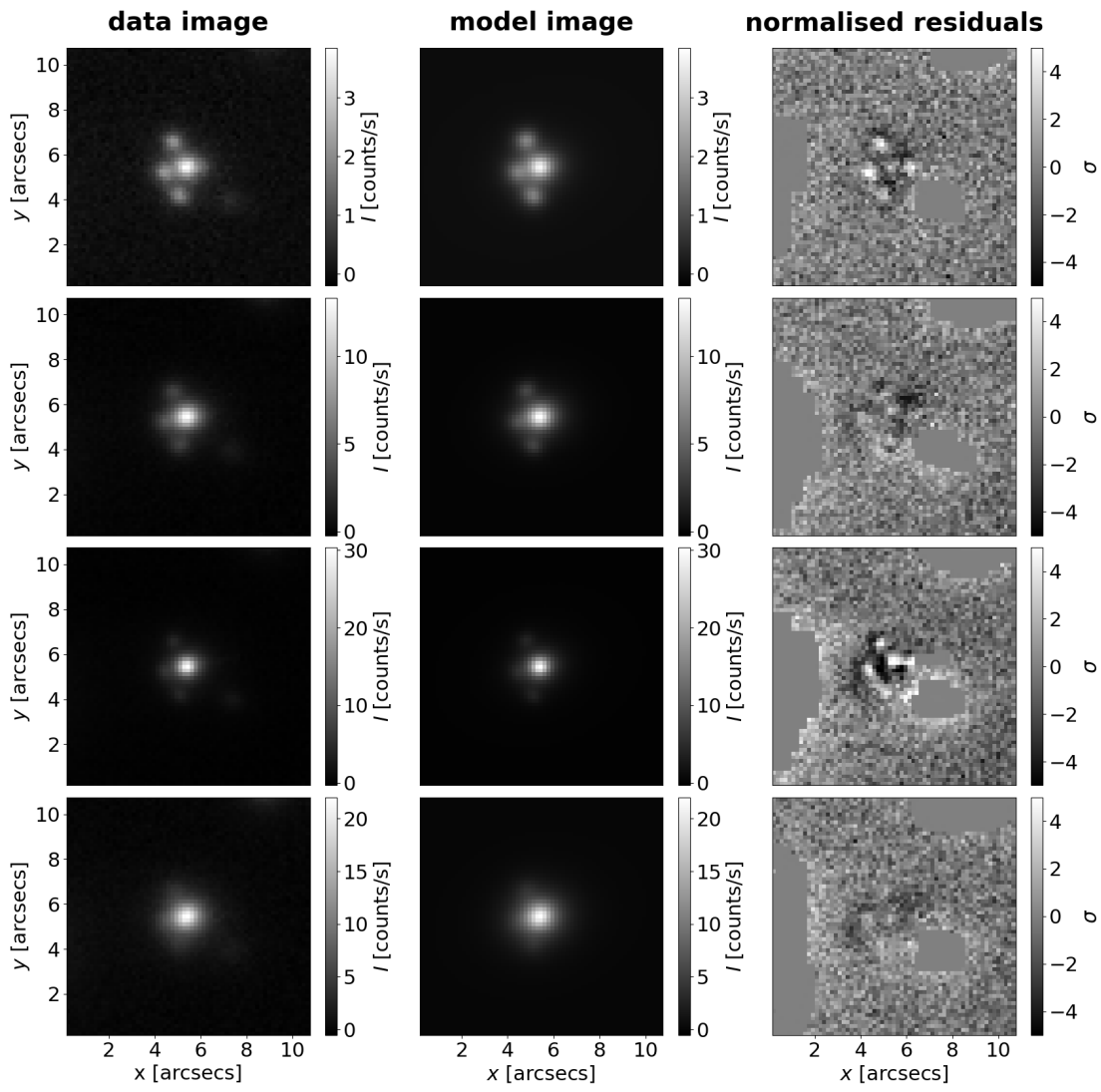
Figure A.27: Fit of lens HSCJ150112+422113. Top to bottom: g, r, i, and z filters.
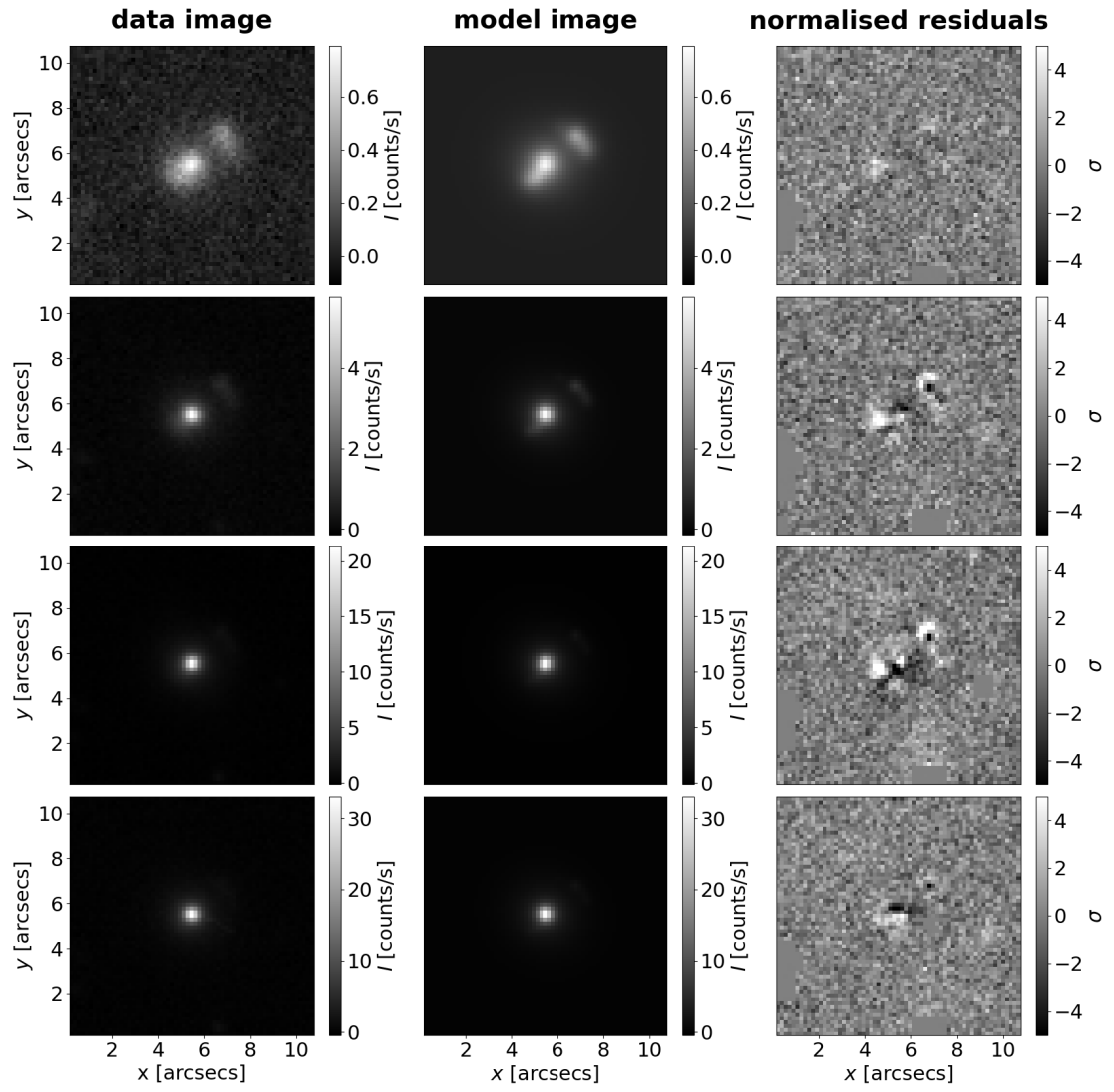
Figure A.28: Fit of lens HSCJ223733+005015. Top to bottom: $g, r, i,$ and z filters.
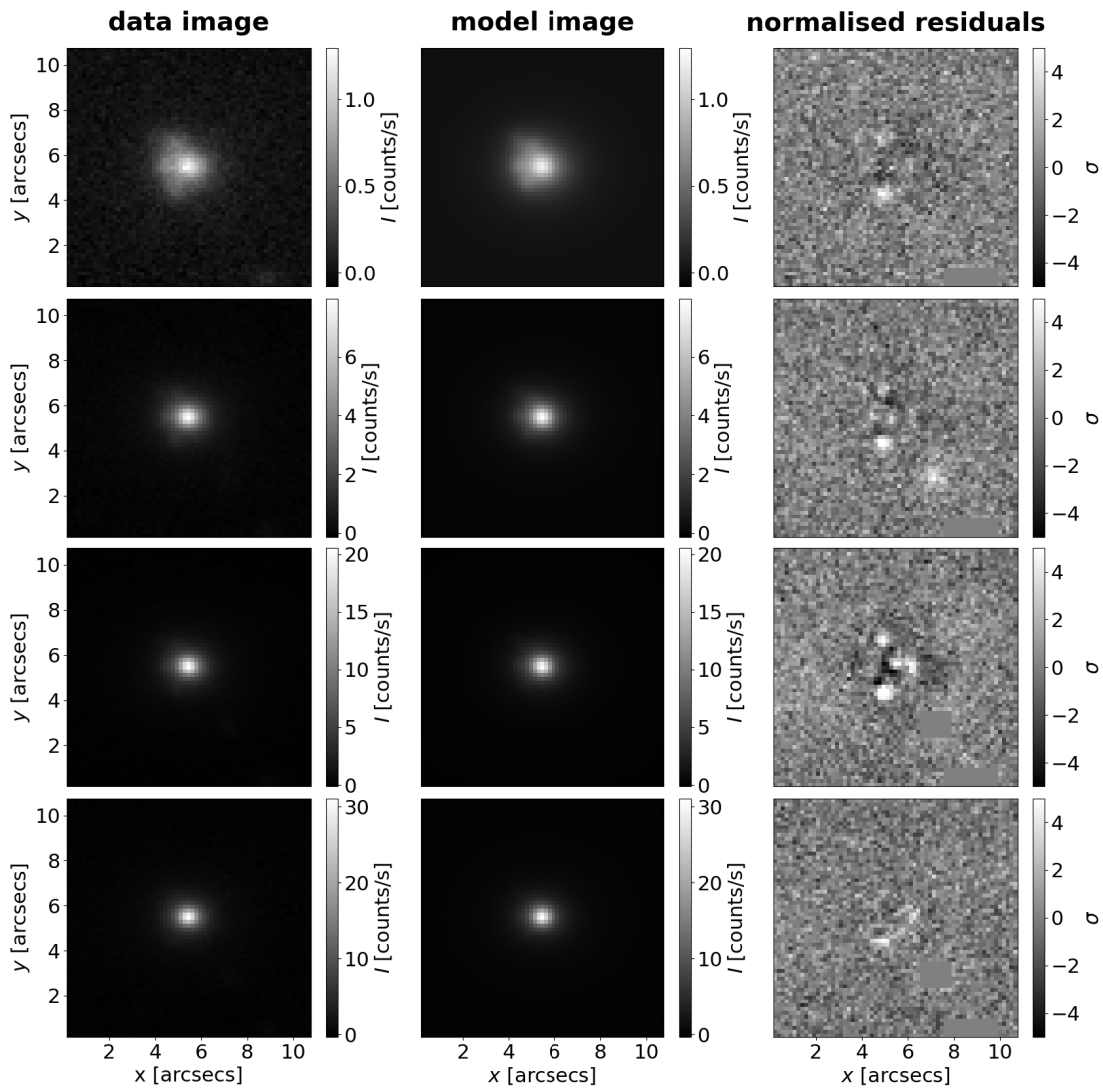
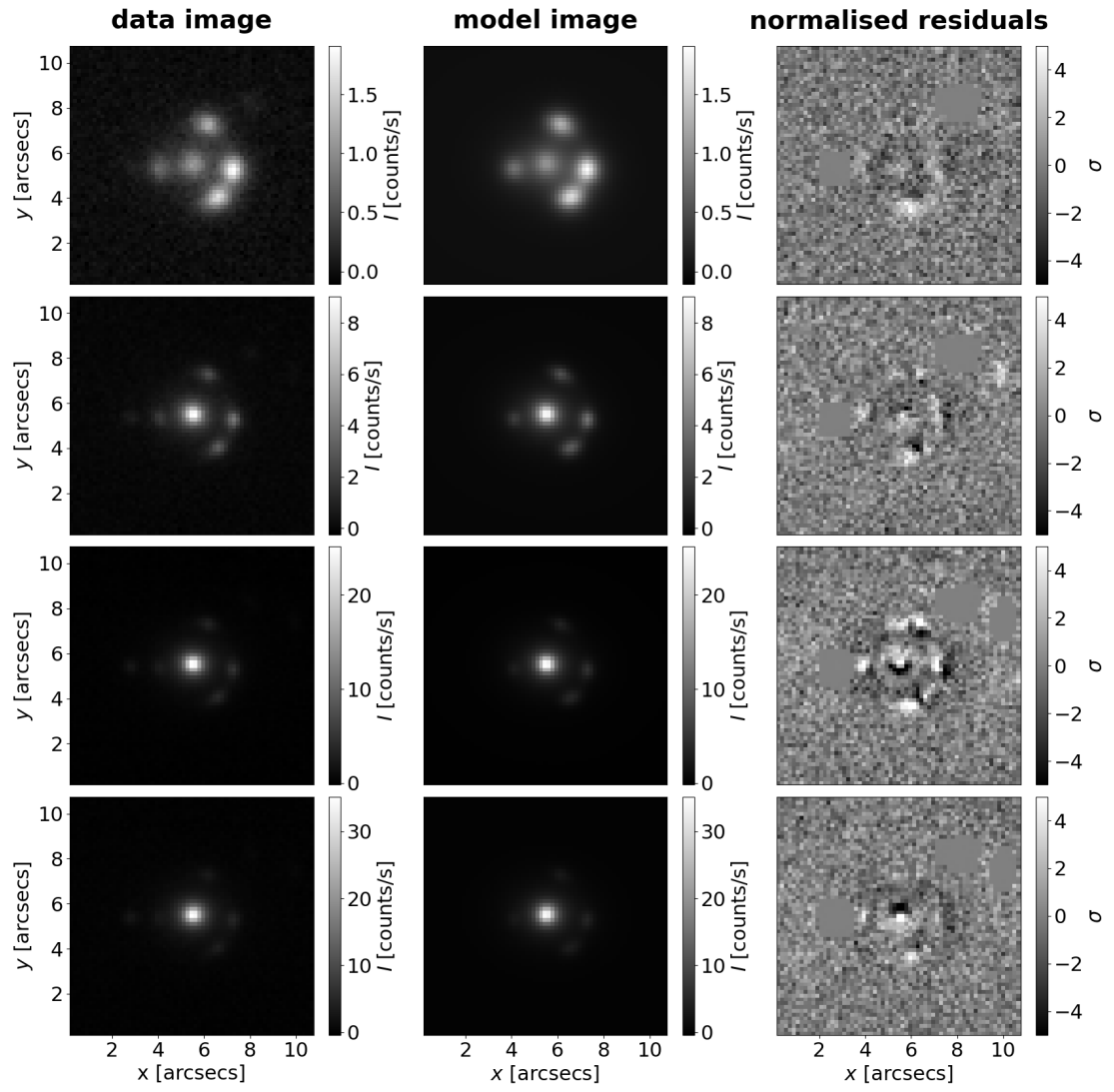**Figure A.29:** Fit of lens HSCJ230335+003703. Top to bottom: g, r, i, and z filters.

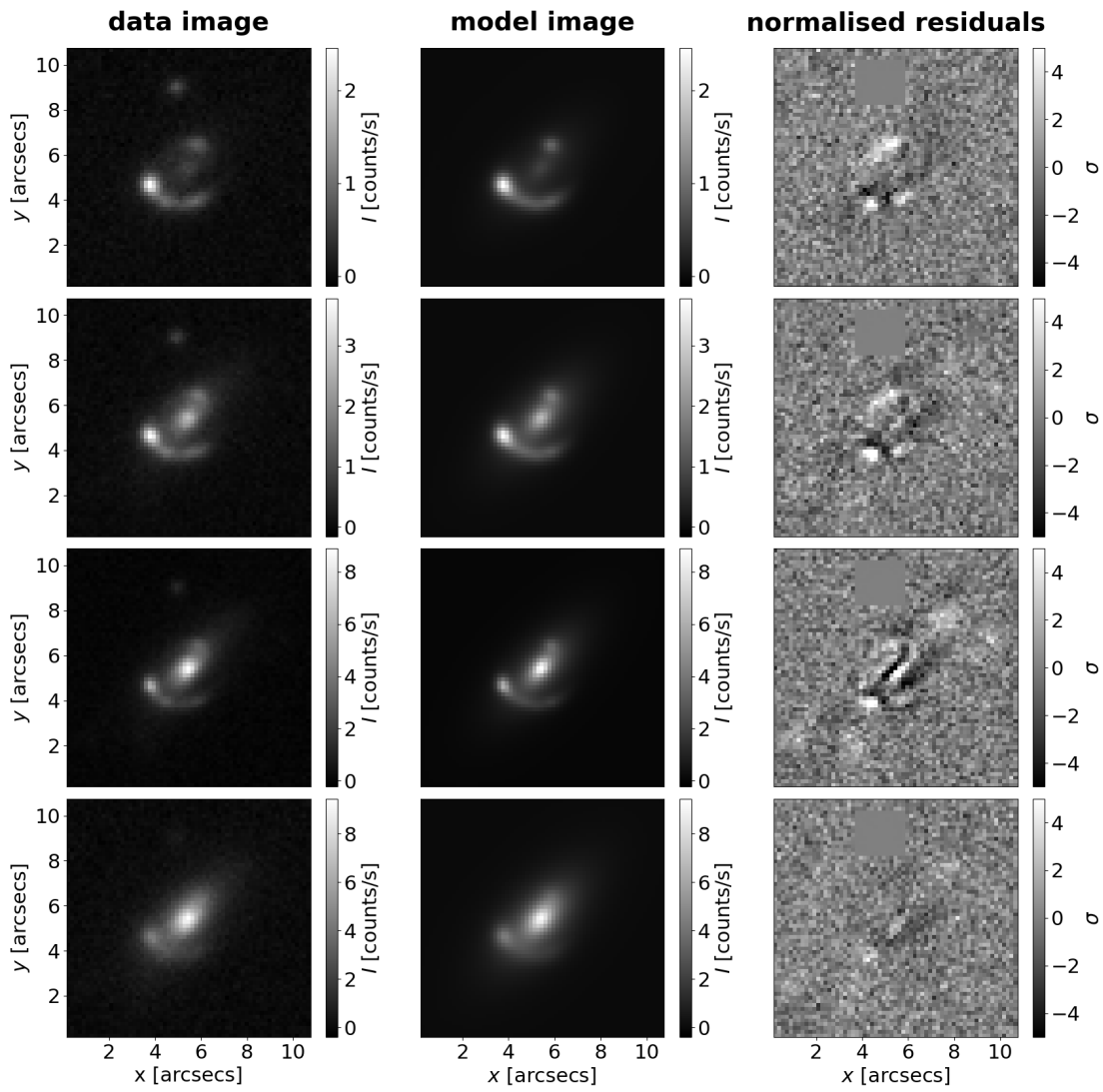**Figure A.30:** Fit of lens HSCJ230521-000211. Top to bottom: $g, r, i,$ and z filters.

Figure A.31: Fit of lens HSCJ233130+003733. Top to bottom: g, r, i, and z filters.
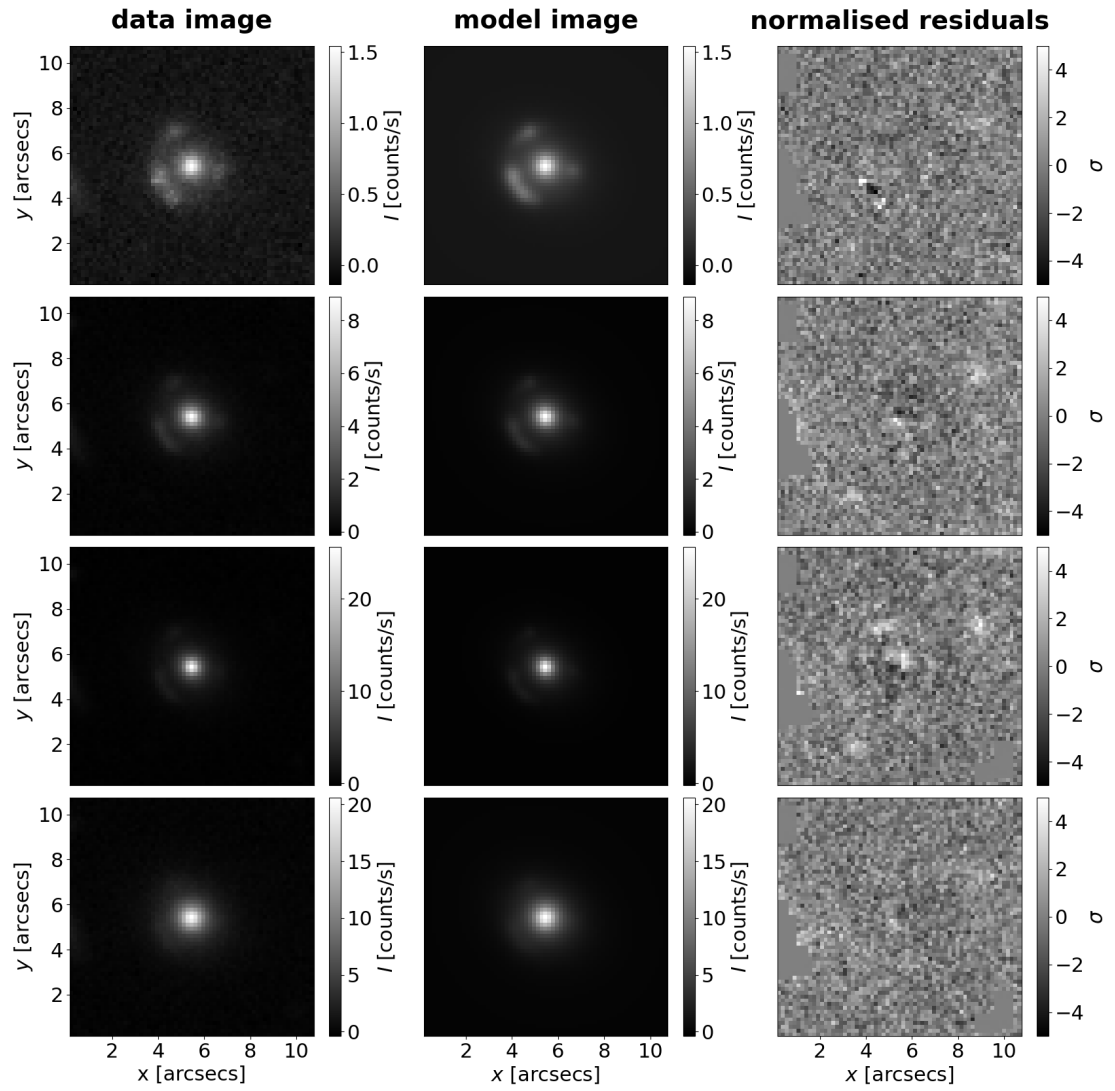
Figure A.32: Fit of lens HSCJ233146+013845. Top to bottom: g, r, i, and z filters.