

Teaching machines how to find strongly lensed galaxies in cosmological sky surveys

Présentée le 15 novembre 2022

Faculté des sciences de base
Laboratoire d'astrophysique
Programme doctoral en physique

pour l'obtention du grade de Docteur ès Sciences

par

Elodie Marie Charlène SAVARY

Acceptée sur proposition du jury

Prof. V. Savona, président du jury
Prof. F. Courbin, directeur de thèse
Dr R. Gavazzi, rapporteur
Dr T. Colett, rapporteur
Prof. L. Zdeborova, rapporteuse

For things to reveal themselves to us, we need to be ready to abandon our views about them.

— Thich Nhat Hanh, *Being Peace*

To my friends and family

Acknowledgements

The years of my PhD were undoubtedly not the easiest part of my life. Despite that, I enjoyed every bit of this experience. This would certainly not have been possible without the support of people I was lucky enough to meet on my journey toward this PhD.

I would like first to thank my PhD advisor, Prof. Frédéric Courbin, who, despite his busy schedule, always found the time to provide me with advice and guidance. Thanks for offering me the opportunity to live this adventure at the LASTRO.

I collaborated intensively with Karina Rojas and Benjamin Clément during my entire thesis. I wish to thank them for their support and invaluable help in understanding astrophysics and the academic world.

These years of research would have been bland without the amazing colleagues I met at the LASTRO and Unige, Javier, Mark, Loic, Damien, Cameron, Christoph, James, Yoan, Martin, Amélie, Aymeric, Andrei, Eric, Thibault, Sophie, and all the others. Thanks for the good moments and not always science-related discussions that made my work days brighter and the commuting hours much shorter.

To make it to a PhD in astrophysics, I had to go through the first years of physics at EPFL. This would not have been possible without the help of friends who helped me fill the knowledge gap between an apprenticeship in mechanics and the first year at EPFL. I take this occasion to thank Patrick, Nathalie, Florian, and Joël for their time explaining basic mathematics to me.

I want to thank my close friends from EPFL, Laetitia, Romain, Emily, Olivier, David, Patrik, Coralie, Jeanne, Baptiste, Cindy, Boris, Morgane, Mario, Alexandre, Claudio, Mohamed, and all the others for the moral support during my thesis, all the good moments spent during our studies and for some the help proofreading this thesis.

Je tiens également à remercier mes vieux amis et les membres de ma famille qui m'ont encouragé le long de ce parcours, en particulier mes parents, qui m'ont soutenu malgré le défi que représentait cette thèse.

Last but not least, I would like to thank Alexandre for his love, comprehension, and unlimited support during my PhD.

Ecublens, 30 Juillet 2022

E. S.

Abstract

Strong gravitational lensing results from the space-time distortions induced by massive objects. The trajectory of light rays coming from a background source is deflected if they pass near a massive body before reaching the observer. Strong lensing happens when the distortion is significant enough to cause the appearance of multiple images of the source. Analyzing gravitationally lensed objects enables a wide range of physical and cosmological applications, such as probing the dark matter content in galaxies and clusters or measuring the Universe's expansion rate. The precision of these applications can be improved drastically with a larger number of lenses. For this reason, finding new strong gravitational lensing systems is crucial. Traditionally, gravitational lenses were found with meticulous visual inspections of imaging datasets. However, the next generation of large-scale imaging surveys will produce such a large amount of data that visual inspection will be impractical, thus motivating the development of efficient automated detection methods to handle large datasets and improve the accuracy of the different applications. The main topic of this thesis is to develop and improve our lens-finding algorithms and to study new strong gravitational lenses in large-scale imaging surveys.

The first part of this thesis has been to develop a new tool that allows the production of lens simulations. Deep learning algorithms, particularly convolutional neural networks (CNNs), have recently proven their efficiency in detecting lensing systems. In particular, CNNs can be adapted to find a large variety of lenses at once, making them especially suited to be part of automated detection pipelines. However, convolutional neural networks usually require large sets of images to be trained. Unfortunately, the number of known lenses is at this time too low to constitute a sufficient training set. For this reason, robust and flexible tools to simulate realistic lenses have to be developed to generate training sets. Lens simulations must, however, be as realistic as possible to avoid biases. In this thesis, we propose a simulation tool that enables the production of large sets of realistic lens simulations. This tool has been designed to be flexible and, thus, enables the production of lens simulations with different types of deflectors for any imaging survey.

The second part consists in separating the light of a lensed source from the foreground object. Indeed, in some cases, the lens features are hidden by the light of the foreground object and deblending might help to identify the lenses with small image separation and enables photometric redshift measurements or the initialization of lens models. This thesis presents a

new data-driven method for deblending strong gravitational lenses based on neural networks. Finally, due to the low occurrence rate of strong lensing, the false positive rate of detection algorithms is still a significant challenge. Therefore, we present different tools that enable the inspection of large sets of candidate images. In addition, evaluating the quality of lens candidates is somewhat subjective since the features that define a lens may differ for different experts. Therefore, we propose a set of grading guidelines that can be used for the subsequent ground-based imaging lens searches in this thesis.

The CNN-based classifier, the deblending algorithm, and the simulation and visualization tools are part of an automated lens-finding pipeline. This flexible pipeline can be adapted to future large-scale surveys. We discuss in this thesis its first applications to the Canada-France Imaging Survey and the Dark Energy Survey, which led to the discovery of 133 and 403 high-quality lens candidates, respectively.

Key words: Strong gravitational lensing – Deep Learning – galaxy lensing – galaxy surveys.

Résumé

L'effet de lentille gravitationnelle intervient lorsque des rayons lumineux émis par une source distante passent à proximité d'un objet massif. Lorsque que cela se produit, les rayons lumineux sont déviés par la courbure de l'espace-temps engendrée par la présence de cet objet. Quand cette courbure est suffisamment importante pour que plusieurs images de la même source de lumière apparaisse, on parle de lentille gravitationnelle forte.

L'analyse des lentilles fortes est utile à un large éventail d'applications physiques et cosmologiques allant de l'estimation du contenu en matière noire des galaxies à la mesure de la vitesse d'expansion de l'Univers. La précision de ces applications pourrait être grandement améliorée grâce à la découverte de nouvelles lentilles. Pour cette raison, la recherche de nouvelles lentilles est actuellement un sujet important en astronomie. Dans le passé, les images de lentilles gravitationnelles étaient découvertes grâce à une inspection méticuleuse des images provenant de différents relevés astronomiques. Cependant, la prochaine génération de relevés d'images astronomiques à grande échelle produira une telle quantité de données qu'une inspection visuelle de chaque objet serait fastidieuse. Il est donc essentiel de développer des méthodes de détection automatisées efficaces. Cette thèse est dédiée au développement et à l'amélioration d'algorithmes pour trouver et étudier de nouvelles lentilles gravitationnelles fortes dans les programmes de relevé d'images astronomiques à grande échelle.

Les algorithmes d'apprentissage automatique profonds, plus particulièrement les réseaux de neurones convolutifs, se sont récemment distingués dans le domaine de la détection de lentilles dans les images astronomiques. Leur flexibilité les rend particulièrement adaptés lors d'une utilisation à l'intérieur de pipelines de détection de lentilles. Cependant, l'entraînement de réseaux de neurones convolutifs nécessite un grand nombre de données. Or, le nombre de lentilles connues est encore très restreint et ne suffit pas pour entraîner des réseaux de neurones. Pour cette raison, il est également important de développer des outils robustes et flexibles pour simuler des lentilles fortes en parallèle de la recherche de lentilles. Les simulations doivent cependant être aussi réalistes que possible pour éviter la création d'artefacts pouvant perturber les réseaux de neurones. Dans cette thèse, je présente un outil de simulation permettant de produire des ensembles de lentilles simulées dans le but d'entraîner des réseaux de neurones. Cet outil est conçu pour être flexible et permettre de réaliser des simulations de différents types de déflecteurs ou de différents télescopes.

Dans certains cas, les images de la source passant par la lentille sont perturbées par la lumière de l'objet occupant le premier plan. Distinguer ces deux flux lumineux permet d'aider à la détection des lentilles pour lesquelles la séparation entre les images est très faible. La distinction entre les images rend également possible l'estimation du décalage vers le rouge des composants de la lentille ainsi que l'initialisation des paramètres pour la modélisation des lentilles. Cette thèse présente une nouvelle méthode basée sur les réseaux de neurones pour séparer les composants de lentilles gravitationnelles fortes.

En raison de la rareté des lentilles gravitationnelles fortes, le taux de faux positifs reste très important avec les algorithmes de détection actuels. Pour cette raison, une partie de cette thèse est consacrée à l'élaboration d'outils permettant l'inspection de grands ensembles de candidats lentilles. De plus, l'évaluation de la qualité des candidats identifiés par les algorithmes de détection est subjective. En effet, les caractéristiques qui définissent une lentille peuvent varier d'un expert à l'autre. Par conséquent, j'introduis dans cette thèse un ensemble de lignes directrices qui pourront être utilisées lors des prochaines recherches de lentilles.

Le détecteur de lentille basé sur les réseaux de neurones convolutifs, l'algorithme de séparation et les outils de simulation et de visualisation font partie d'un pipeline automatisé dédié à la recherche de lentilles fortes. Ce pipeline est très flexible et peut être adapté à de nombreux types de données. Je présente dans cette thèse la première application de ce pipeline sur les données de deux relevés d'images astronomiques : "Canada-France Imaging Survey" et "Dark Energy Survey" qui a conduit à la découverte de 133 et 405 nouvelles potentielles lentilles.

Mots clefs : Lentillage gravitationnel fort – Apprentissage automatique profond – lentillage de galaxies – relevé de galaxies.

Contents

Acknowledgements	i
Abstract (English/Français/Deutsch)	iii
List of figures	xi
List of tables	xiii
1 Introduction	1
1.1 Gravitation from Newton to Einstein	1
1.2 Gravitational lensing	2
2 Gravitational lensing theoretical framework	7
2.1 Formalism	7
2.1.1 The lens equation	7
2.1.2 Properties of the lensed source images	9
2.1.3 Einstein radius	12
2.2 Regimes of gravitational lensing	14
2.3 Strong lensing applications	15
2.3.1 Natural telescope	16
2.3.2 Dark matter content of clusters and galaxies	16
2.3.3 Hubble constant measurement	17
3 Machine learning methods	21
3.1 A simple introduction to machine learning	21
3.1.1 Learning in the context of machine learning	22
3.1.2 Model parameter versus hyper-parameter	23
3.1.3 Overfitting, underfitting	23
3.1.4 The curse of dimensionality	23
3.1.5 Principal component analysis	25
3.2 Neural networks	28
3.2.1 Artificial neural networks: Theoretical introduction	29
3.2.2 Convolutional neural networks	36
3.2.3 Auto-encoders	38
3.3 Performance metrics	40
	vii

3.3.1	Binary classification	41
3.3.2	Image Generation	45
4	Finding galaxy-galaxy lenses in imaging data	47
4.1	Citizen science	47
4.2	Automated methods	49
4.2.1	Early methods	49
4.2.2	Neural networks based methods	50
4.3	Gravitational lens finding challenge 1.0 and 2.0	53
4.3.1	Challenge 1.0	53
4.3.2	Challenge 2.0	54
4.4	Limitations of CNN-based methods	58
4.4.1	Low occurrence rate of lenses	58
4.4.2	High false positive rate	59
4.4.3	The problem of using simulations in training sets	60
4.5	Summary	61
5	Lens simulations	63
5.1	Comparison of the two versions of the simulation tool	65
6	Visual inspection of the lens candidates	69
6.1	Visual inspection tool	69
6.1.1	Detailed inspection tool	69
6.1.2	Classification guidelines	72
6.2	Method	73
6.3	Discussion	74
7	Lens deblending	77
7.1	Multi-output auto-encoder	78
7.1.1	Training	79
7.1.2	Evaluation of the deblended images	80
7.2	Comparison of the two-output auto-encoder with two separated auto-encoders	82
7.3	Application on real data	84
7.4	Summary	88
8	First application of an end-to-end lens finding pipeline in CFIS and DES	89
8.1	Lens finding in CFIS and DES	90
8.1.1	Statement of contribution	90
8.2	Discussion	160
9	Conclusion	163
A	Galaxy Image Simulation Using Progressive GANs	165
	Bibliography	186

Curriculum Vitae**187**

List of Figures

2.1	Illustration of the geometry of a gravitational lens.	8
2.2	Illustration of the effect of the convergence κ and shear γ_1, γ_2 on a circular source (in black).	11
2.3	Scheme illustrating examples of the three typical lensing configurations formed with an elliptical deflector approximated as an SIE.	13
2.4	Image of the lensed quasar J1131–1231, combined image from NASA's Chandra X-ray Observatory (X-rays domain) and HST (optical domain).	14
3.1	Orange with altered colors.	22
3.2	Example of underfitting and overfitting with linear regression.	24
3.3	Illustration of the curse of dimensionality problem	26
3.4	Example of PCA for a set of 12 images with 2 pixels	27
3.5	Percentage of explained variance for the 24 first component of the PCA decomposition of the MNIST dataset.	28
3.6	Reconstruction after PCA of images from the MNIST test set.	29
3.7	Scheme of an artificial neuron with four inputs and one output.	30
3.8	Example of a neural network composed of three dense layers.	32
3.9	Illustration of the convolution operation taking place in convolutional layers	37
3.10	Illustration of a residual block.	39
3.11	Sketch of an autoencoder.	40
3.12	Reconstruction of MNIST images using an auto-encoder with one hidden layer and sigmoid activation in the output layer.	41
3.13	Confusion matrix representing the four possible outcomes of a binary classification.	42
3.14	Comparison of precision-recall curve and ROC curve.	44
4.1	Architecture of our modified Efficientnet B0.	56
4.2	Examples of Euclid lens simulations (VIS band) matching the criteria to be included in the positive examples of our training set.	62
5.1	Illustration of the two different parts of the simulation pipeline.	66
5.2	Comparison of CFIS lens simulations produced with the two versions of the simulation pipeline.	67

6.1	Screenshot of the one-band version of the detailed inspection tool.	70
6.2	Screenshot of the color version of the mosaic tool.	71
6.3	Color-map depicting the mean number of clicks in each cell of the Mosaic tool during the visual inspection of the candidates of the lens searches (Savary et al., 2021) and (Rojas et al., 2021).	72
6.4	Illustrative examples for the four categories of the grading guidelines.	73
6.5	Legacy Survey images of the 33 SL candidates of the CFIS lens search	76
7.1	Comparison of the CAS statistics for 1000 lens simulations and 1000 lenses reconstructed with the multi-output auto-encoder	81
7.2	Difference in magnitudes between lensed sources and deflectors reconstructed with the multi-output deblending auto-encoder and their ground truth.	82
7.3	Scheme of two possible deblending models.	83
7.4	Difference in magnitudes between lensed sources and deflectors reconstructed with two different auto-encoders and their ground truth.	84
7.5	Comparison of deblending results on simulated lenses for two different architectures	85
7.6	Result of the auto-encoder deblending for 4 spirals extracted from the false positive catalog of Savary et al. (2021)	86
7.7	Result of the auto-encoder deblending for 16 ML candidates of Savary et al. (2021) .	87

List of Tables

4.1	Results of our three architectures after training with the modified set of lenses on VIS band data.	58
6.1	Percentage of objects classified as SL , ML, F, and NL for each visual inspector after the detailed visual inspection step of Savary et al. (2021)	74

1 Introduction

Gravitational lensing is a visible consequence of the curvature of space-time by massive objects. Multiple images of the source may appear when a background source is located behind a massive object such as a galaxy or a cluster. This phenomenon is called strong lensing. This thesis aims to improve the search for strong lenses in large-scale imaging surveys. More precisely, we focus on galaxies lensed by other galaxies (galaxy-galaxy lenses).

1.1 Gravitation from Newton to Einstein

To describe the chain of events that led to the development of General Relativity, I follow the historical introduction from S. Weinberg (1972), to which the interested reader can refer for more details. In the era of Newtonian mechanics, scientists understood gravitation as an attractive force proportional to the masses of the objects involved in the interaction and inversely proportional to the squared distance between the two objects. Einstein, however, revolutionized the description of gravitation at the beginning of the nineteen-hundreds with a new theory, General Relativity, in which gravity is described as a curvature of space-time so that the concept of gravitational force is no longer needed.

In developing this theory, Einstein was guided by the observation that the motion of an object in an inertial reference frame is indistinguishable from the motion of the same object in free space subjected to a suitable uniform acceleration. This finding led him to formulate the so-called strong equivalence principle: "Locally, and at any point of space-time physics follows the special relativity laws and is not affected by the presence of a gravitational field." Einstein already described in his previous theory, Special Relativity, how space and time can be represented as intertwined dimensions of a four-dimensional continuum.

However, the combination of this vision of space, time, and gravity with his equivalence principle was only possible thanks to the development of non-Euclidean geometry. The idea of non-Euclidean geometry arose from the efforts of numerous mathematicians to prove the fifth postulate of Euclidean geometry: "If a straight line falling on two straight lines makes

the interior angles on the same side of the lines taken together smaller than two right angles, then the two straight lines, if produced indefinitely, meet on the side where the sum of the angles is smaller than two right angles," can be derived from the other postulates, definitions, and axioms described by Euclid in his treatise "Elements." All attempts to convert the fifth postulate into a simple theorem failed. However, this prompted some mathematicians to consider the possibility of a geometry framework in which the fifth postulate would not be satisfied. The first variant of non-Euclidean geometry, called hyperbolic geometry, was discovered independently by Carl Friedrich Gauss (1777-1855), János Bolyai (1802-1860), and Nikolai Lobachevsky (1793-1856). The space described by this geometry has fascinating properties and consequences, such as constant negative curvature and triangles where the sum of the angles is less than 180 degrees. The real breakthrough here was not in the odd properties of such a space but in the fact that a geometry with curvature opened up the possibility of describing curved spaces as a function of their inner properties, such as their metrics, without having to refer to conventional Euclidean space. This opened the possibility of describing our physical world as a curved space and incidentally gave Einstein the necessary mathematical background to develop the General Theory of Relativity, in which gravity is described as a curvature of the space-time metric.

1.2 Gravitational lensing

Gravitational lensing is an observable consequence of General Relativity. Massive bodies such as stars, black holes, or galaxies cause distortions in the space-time metric. The trajectory of photons coincides with the geodesics of this metric. Gravitational lensing occurs when a photon has to pass through the gravitational potential of a massive object before reaching an observer. In this case, its trajectory is deflected. Therefore, the observer sees the photon as if it had been emitted from another location. Although gravitational lensing was already described in the framework of Newtonian mechanics, in particular by Johan von Soldner in 1804, the correct calculation of the deflection angle was only possible in the framework of General Relativity. It should also be noted that the Newtonian calculation of the deflection angle requires a mathematical trick to assign a mass to the photons. This fact already seemed to unsettle Soldner when he wrote, "Hopefully no one will find it objectionable that I treat a ray of light completely like a heavy body. Because those light rays have all fundamental properties of matter, one sees the phenomenon of aberration, which is only possible if light rays are material. - And nothing can be imagined to exist and act on our senses without having the properties of matter."

After achieving the development of General Relativity, Einstein recomputed the value of the deflection angle and found a value two times greater than the Newtonian value. He was proven right on the occasion of a solar eclipse in 1919. On this occasion, Arthur Stanley Eddington and Frank Watson Dyson conducted an experiment where they measured the displacement of the stars during the eclipse and found that the values matched the prediction of General Relativity. Not long after the eclipse, in 1924, Orest Khvolson suggested for the first time

that gravitational lensing may produce multiple images of the same source or even ring-like features (Chwolson, 1924). Einstein was, however, very pessimistic about the possibility of observing this phenomenon as he considered lensing by stars. He concluded in Einstein (1936) that the separation between the images would be on the order of the milli- arcsecond, and the multiple images could therefore not be resolved. Fritz Zwicky was, however, more optimistic on the subject. Instead of considering lensing by stars, he assessed the possibility of observing lensing by galaxies (even though, at this time, galaxies were not completely understood and called "extra-galactic nebulae"). He concluded that the separation between the images is sufficiently large for such objects to be resolved by ground-based telescopes. Therefore, he thought such an event is very likely to be observed. He gathered his thoughts in two publications Zwicky (1937a) and Zwicky (1937b), where he also described for the first time the possibility of using gravitational lensing as a way to probe the mass of galaxies.

As we detail in Sect. 2.3, gravitational lensing has numerous applications in astrophysics, going from constraining cosmological models to probing the fraction of dark matter in foreground galaxies. The precision of these measurements depends dramatically on the number of lenses.

The first galaxy acting as a gravitational lens was discovered fortuitously by Walsh et al. (1979). Here, the lensed source was a quasar. Similar discoveries of quasars lensed by galaxies followed shortly after this observation (e.g. Huchra et al., 1985; Weymann et al., 1980).

Gravitational lensing not only changes the observed position of the background source but can also magnify it. Consequently, in a magnitude-limited sample, the probability of an object being magnified is higher than for a random field source (Narayan & Wallington, 1993). The first systematic searches of gravitational lenses used this effect, called the magnification bias. They looked for bright sources within populations of quasars or radio sources. The candidates obtained with this method were then reinspected with high-quality imaging data to confirm their lens nature.

The first lensed quasars found in the context of systematic searches in the optical domain are described in Magain et al.; Surdej et al. (1988, 1987). They were followed by larger searches performed for example in the Hamburg-ESO bright quasars catalogue (Blackburne et al., 2008; Wisotzki et al., 2004; Wisotzki et al., 2003; Wisotzki et al., 2002; Wisotzki et al., 2000; Wisotzki et al., 1993) or in the Sloan Digital Sky Survey (SDSS) (Inada et al., 2012; Oguri et al., 2006). Similar strong lens searches were also performed in the radio domain, for example, in the CLASS survey (Myers et al., 2003). In this type of search, the pre-selection of candidates is carried on the properties of the lensed sources. They are therefore often referred to as "source-selected" searches. Source-selected searches enable us to identify very bright objects as quasars.

Galaxies lensed by galaxies can also produce visible lensing features. Since there are much more galaxies than quasars in the visible universe, we expect that galaxy-galaxy lenses are more frequent than quasar-galaxy lenses. The main difference here is that even considering the magnification effect, the difference between the luminosity of the foreground objects and

the lensed features is small. Therefore such lenses cannot be found via the magnification bias. The searches for galaxy-galaxy lenses are usually performed by pre-selecting samples of galaxies that have a high probability of acting as lenses and are therefore called "lens-selected". Luminous red galaxies (LRG) are the most frequent deflectors targeted in lens-selected searches due to their large masses. However, other types of galaxies, such as spirals, may sometimes produce visible lenses.

One of the first examples of lens-selected search is the Sloan Lens ACS (SLACS) survey (Bolton et al., 2006). It was performed by looking for the presence of potential high redshift emission lines in the spectra of a sample of LRGs. Other examples using a similar method are the BOSS Emission-Line Lens Survey (BELLS) (Brownstein et al., 2012) which targeted LRG acting as lenses, and the Sloan WFC Edge-on Late-type Lens Survey (SWELLS) (Treu et al., 2011) which focused on edge-on spiral galaxy deflectors. Spectroscopic searches are limited in redshift since they require visible emissions lines. By contrast, searches on imaging data allow us to find lensed sources at high redshift, provided that they display visible lensing features. However, some of the candidates found with spectroscopic searches may not be found with imaging searches if the separation between the source images is too small, especially with low-resolution data. The two methods are therefore complementary and find different candidates. Recently, lens-selected searches have been performed in different surveys such as the Kilo Degree Survey (KiDS) (de Jong et al., 2013), the Dark Energy Survey (DES) (Dark Energy Survey Collaboration et al., 2016), Panoramic Survey Telescope And Rapid Response System (PanSTARRS) (Chambers et al., 2016), the Hyper Suprime-Cam Subaru Strategic Program (HSC) (Aihara et al., 2019) and in the Ultraviolet Near- Infrared Optical Northern Survey (UNIONS) (Guinot et al. in prep.).

Future and ongoing large-scale imaging surveys, for example Euclid (e.g. Amiaux et al., 2012; Laureijs et al., 2011), the Rubin Observatory Legacy Survey of Space and Time (LSST) (e.g. Ivezić et al., 2019) and Roman (Spergel et al., 2015) will provide a large amount of high-quality imaging data and consequently numerous gravitational lenses.

There exist several methods to find lenses in imaging data, such as clustering algorithms using pixel intensities to detect the presence of elongated and curved objects in an image (Cabanac et al., 2007; Lenzen et al., 2004) or analysis of the residuals after the subtraction of the foreground galaxy (Gavazzi et al., 2014; Joseph et al., 2014). Nevertheless, recently the use of machine learning, in particular convolutional neural networks (CNN), to classify images has become prevalent due to the significant improvement of the graphics processing units. It has already been used successfully to find candidates in the surveys mentioned above (e.g. Jacobs et al., 2019a; Rojas et al., 2021; Savary et al., 2021; Cañameras et al., 2020; Petrillo et al., 2018; Jacobs et al., 2017; Petrillo et al., 2017).

The long-term goal of this thesis is to enable astrophysical and cosmological applications of strong gravitational lensing at galaxy, group, and cluster scales, given present and future wide-field surveys like DES, CFIS, Euclid, LSST. The focus of my thesis was principally on

enabling the detection of galaxy-galaxy lenses. This involves developing a reliable set of tools to find and classify these rare objects and understanding the limitations of the algorithms used for this purpose. In this thesis, the focus is on the use of neural networks to find galaxy-scale lenses. It will first require to design tools to produce the training sets for Machine Learning methods and then the design of a fully flexible but robust lens-finding pipeline.

In this thesis, I detail my contributions to the realization of the lens-finding pipeline. It is organized as follows. Chapter 2 introduces the theoretical aspects of gravitational lensing. It also summarizes the main applications of strong lensing. Chapter 3 summarizes the theoretical concepts and machine learning methods used throughout this work. Chapter 4 summarizes the methods used to search galaxy-galaxy lenses in imaging data and the main challenges encountered with machine learning methods. I also discuss our participation in the Lens finding challenge 2.0. Chapter 5 presents our tool to simulate lenses, which serves to produce training sets to train the classifiers of the lens-finding pipeline. In Chap. 6, I present the interface I developed to inspect large sets of candidates as well as the guidelines we built to standardize the visual inspection. In Chap. 7, the algorithm I developed to deblend the light of the background source from the light of the deflector in imaging data is presented. Finally, Chap. 8 describes the two first applications of the lens finding pipeline to real data.

2 Gravitational lensing theoretical framework

Gravitational lensing is one of the observable consequences of General Relativity. It is also a powerful tool for a variety of astrophysical applications. In this Chapter, I summarize the theoretical aspects of gravitational lensing and present the different regimes of gravitational lensing: strong, weak, and micro. At the end of this chapter, I detail some applications of strong lensing at galaxy scale, which is the main focus of this thesis.

2.1 Formalism

Here, I summarize the theoretical aspects of gravitational lensing. A more complete description can be found, for example, in Narayan and Bartelmann (1996) , Schneider (2006) and Meneghetti (2021) from which this section is widely inspired.

2.1.1 The lens equation

We present a simplified scheme of a gravitational lens in Fig. 2.1. Let us consider a source and lens placed respectively at a distance D_S and D_L from the observer. The distance between the source and the lens is D_{LS} . It should be emphasized that the distances used here are angular distances and thus, depends on the choice of the cosmological model. In particular, $D_{LS} = D_L + D_S$ is not necessarily true. We hypothesize that the lens and source plane are much thinner than D_L and D_S and that the deflection angle is very small, such that we can approximate the source and the lens as 2D objects lying in the source and lens plane, respectively. In this model, called the thin lens approximation, the deflection effect takes place within the lens plane, and therefore, the bending of the light is approximated as a sharp angle. Let's further assume that all the angles depicted in Fig. 2.1 are very small. Consequently, we can express the impact parameter as $\xi = D_L \theta'$ and the real angular position of the source $\vec{\eta} = D_{LS} \vec{\beta}$. We define the optical axis as the axis perpendicular to the source and lens planes and passing through the center of the deflector.

Due to the presence of a massive deflector in the line of sight, the light emitted from the

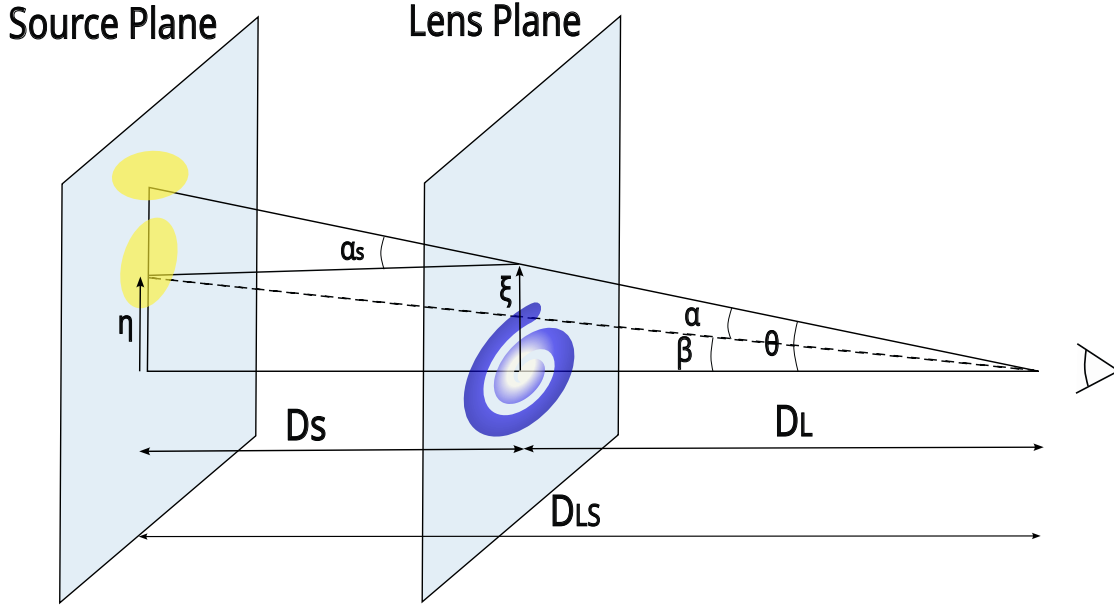


Figure 2.1: Illustration of the geometry of a gravitational lens. θ and β are the angular position of one image and the source, respectively. Due to the gravitational lensing effect, the observer sees the source light as if it came from the angular position θ .

angular position β is seen as if coming from the angular position θ . The deflection angle $\vec{\alpha}_s$ is always expressed for a particular angular position θ such that we write it $\vec{\alpha}_s(\theta)$. It was the parameter measured in the first observations of gravitational lensing. However, we use another quantity here, the scaled deflection angle $\vec{\alpha}(\theta)$, which directly expresses the difference between the real angular position of the source β and the position of the images seen by the observer α .

The scaled deflection angle can be related to the deflection angle as follows:

$$\vec{\alpha}(\theta) = \frac{D_{LS}}{D_S} \vec{\alpha}_s(\theta), \quad (2.1)$$

where $\alpha_s(\theta)$ is the deflection angle. The relation between the angular position of one of the lensed source image θ and the real angular position of the source β can then be derived using simple geometry:

$$\beta = \theta - \vec{\alpha}(\theta), \quad (2.2)$$

where $\vec{\alpha}(\theta)$ is the scaled deflection angle.

Equation 2.2 is called the lens equation. If more than one angular position θ verifies the lens equation for a fixed β , the observer sees multiple images of the source.

In analogy with classical mechanics, the scaled deflection angle can be expressed as the gradient of a potential, called the deflection potential, $\vec{\alpha}(\theta) = \nabla \Phi(\theta)$. It is defined as follows:

$$\Phi(\theta) = \frac{1}{4} \int \kappa(\theta') \ln(|\theta - \theta'|) d\theta', \quad (2.3)$$

where $\kappa(\theta)$ is the dimensionless surface mass density, also called convergence for a given angular position of the source. It can be written:

$$\kappa(\theta) = \frac{\Sigma(D_L \theta)}{\Sigma_c}, \quad (2.4)$$

where the critical density Σ_c corresponds to:

$$\Sigma_c = \frac{c^2}{4\pi G} \frac{D_s}{D_l D_{LS}}. \quad (2.5)$$

The lens equation, in combination with the expression of the deflection potential, links the angular position of the observed images with the surface mass density of the lens and the different distances involved in the problem.

With the help of these two equations, we will now derive some properties of the lensed source images.

2.1.2 Properties of the lensed source images

Gravitational lensing is achromatic, meaning that all photons are deflected similarly independently of their wavelength, and the surface brightness of the source is conserved because of Liouville's theorem. However, gravitational lensing affects the source images' size, shape, and total flux.

To understand better the distortion of the source images, let us consider the mapping $f : \theta \rightarrow \beta$ which converts the angular position of the observed images of the source θ into the real angular position of the source β . With the help of the lens equation 2.2, we can express the Jacobian matrix of this transformation, also called the amplification matrix, as follows:

$$A(\theta) = \frac{\partial \beta}{\partial \theta} = \frac{\partial \theta}{\partial \theta} - \frac{\partial \vec{\alpha}}{\partial \theta} = \mathbb{I}_2 - \frac{\partial^2 \Phi}{\partial \theta_i \partial \theta_j} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}, \quad (2.6)$$

where κ is the convergence defined in Eq. 2.4 and γ_1 and γ_2 correspond to the following terms:

$$\gamma_1 = \frac{1}{2} \left(\frac{\partial^2 \Phi}{\partial^2 \theta_x} - \frac{\partial^2 \Phi}{\partial^2 \theta_y} \right), \quad (2.7)$$

$$\gamma_2 = \frac{1}{2} \left(\frac{\partial^2 \Phi}{\partial \theta_x \partial \theta_y} \right). \quad (2.8)$$

The amplification matrix $A(\theta)$ give use access to a first order approximation of the mapping $f : \mathcal{S} \rightarrow \mathcal{B}$. To highlight better the transformations happening to the source $f : \mathcal{B} \rightarrow \mathcal{S}$ we will look at the inverse of the amplification matrix, the magnification matrix:

$$M(\theta) = \mu \begin{pmatrix} 1 - \kappa + \gamma_1 & \gamma_2 \\ \gamma_2 & 1 - \kappa - \gamma_1 \end{pmatrix} = (1 - \kappa) \mu \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \mu \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{pmatrix}, \quad (2.9)$$

where $\mu = \det A(\theta)^{-1}$ is named magnification since it corresponds to the ratio of the fluxes of the lensed image of the source and the real source. The existence of the magnification matrix $M(\theta)$ is conditioned by the fact that $\det A(\theta)$ must be non null. In fact, all the positions of the lens plane θ where $\det A(\theta) = 0$ form a boundary called critical curve where the magnification is theoretically infinite. The magnification remains however finite in practice, even if very high, since the geometrical optics approximation does not hold in this case, and wave optics predicts a finite magnification. The mapping of the critical curve to the source plane is called the caustic lines. Each time the source crosses a caustic from the outside to the inside, a new pair of images of the source either appears. Conversely, an image pair disappears when the source crosses the caustic in the reverse direction. The first term of the magnification matrix

$$(1 - \kappa) \mu \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad (2.10)$$

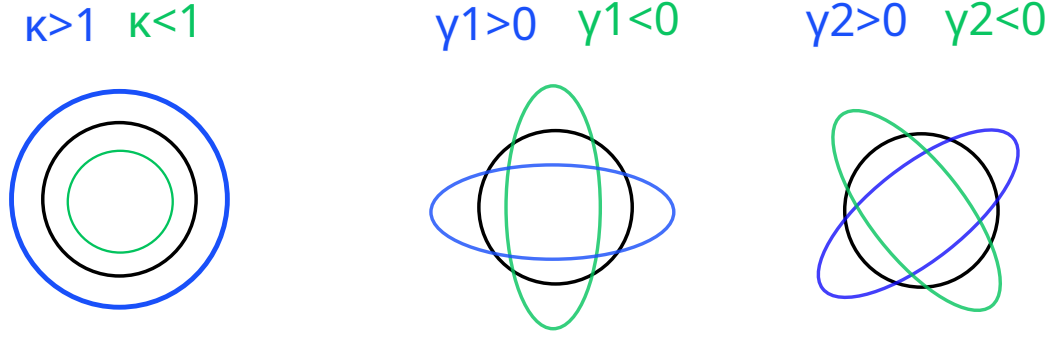
operates an isotropic re-scaling of the coordinates. The convergence κ acts therefore on the size of the images but does not change the shape. The second part of the magnification matrix

$$\mu \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{pmatrix}, \quad (2.11)$$

stretches the coordinates along the axis $\theta_x = \theta_y$ and either $\theta_x = 0$ or $\theta_y = 0$. For this reason γ_1 and γ_2 are called shear components. The effect of κ , γ_1 , and γ_2 on the source images is represented in Fig. 2.2.

Gravitational lensing preserves the surface brightness of the source, provided that there is no absorption or emission of light. The apparent solid angle of the lensed source $d\Omega_\theta$ is, however, different from the real one $d\Omega_\beta$ because of the deformations induced by lensing.

We mentioned earlier that the lens equation could have multiple solutions, and thus the multiple images of the source can be visible. We will now examine the formation of multiple



Convergence only

Shear only

Figure 2.2: Illustration of the effect of the convergence κ and shear γ_1, γ_2 on a circular source (in black).

images and their positions.

By using the mathematical identity $\beta - \theta = \frac{1}{2} \nabla (\beta - \theta)^2$, the lens equation Eq.2.2 can be rewritten as follows:

$$\nabla \left(\frac{1}{2} (\beta - \theta)^2 - \Phi(\theta) \right) = \nabla \tau(\beta, \theta) = 0, \quad (2.12)$$

where we introduced the Fermat potential $\tau(\beta, \theta) = \frac{1}{2} (\beta - \theta)^2 - \Phi(\theta)$. Therefore, the lens equations' solutions correspond to the Fermat potential's stationary points. As we will detail in Sect. 2.3.3, this equation also reflects Fermat's principle, which states that light follows trajectories for which the travel time is stationary.

The condition $\nabla \tau(\beta, \theta) = 0$ forces images to appear only at the extrema of the Fermat potential. Provided the source is not located on a caustic, the images' positions correspond to the Fermat potential's minima, maxima, or saddle points. Since the amplification matrix is the Hessian matrix of the Fermat potential, the type of extrema on which images are located will be reflected by the signs of the eigenvalues of the amplification matrix $A(\theta)$ and hence with the sign of the magnification μ . The trace of the amplification matrix can be expressed:

$$\text{tr } A(\theta) = 2(1 - \kappa). \quad (2.13)$$

We can therefore have one of the following three situations:

- The image is positioned at a minimum of the Fermat potential. The eigenvalues of the amplification matrix are positive, as well as the magnification μ . The trace of the amplification matrix is positive. Therefore $2(1 - \kappa) > 0$ implies $\kappa < 1$.
- The image is positioned at a maximum of the Fermat potential. The eigenvalues of the amplification matrix are all negative. Thus, the magnification μ is positive, and the

convergence $\kappa > 1$.

- The image is positioned at a saddle point of the Fermat potential. The eigenvalues of the amplification matrix have opposite signs; hence the magnification is negative. $\text{tr } A$ can be either negative or positive. Therefore, we cannot conclude on the sign of the convergence κ . In this situation, the image is seen as flipped by a mirror effect.

The number of images for a fixed source position β is predicted by the odd-number theorem (Burke, 1981), which states that the number of multiple images produced by a bounded and smooth lens is odd and equal to the number of saddle points + 1. For this, we can deduce the conditions for multiple images to happen. As we said earlier, each time the source crosses a caustic, i.e., a point where $\det A(\theta) = 0$, two images appear or disappear. We also know that if the source is located far from the optical axis, only one image should be visible, and this image should be located at a minimum of the Fermat potential. Therefore because of the odd-number theorem, a lens can produce multiple images if and only if there exist positions θ for which $\det A(\theta) < 0$. Similarly, $\kappa > 1$ ensures that multiple images are produced. This is, however, not a necessary condition.

The shape of the observable lensing features depends on the geometry of the deflector's gravitational potential and the source's position. We show in Fig. 2.3 the three main possible configurations for an elliptical deflector approximated with a singular isothermal ellipsoid (SIE) (Kormann et al., 1994; Kassiola & Kovner, 1993). More complex profiles may lead to a larger number of images and possible configurations (e.g. Orban de Xivry & Marshall, 2009). It is, therefore, difficult to concisely word the visual aspect of a gravitational lens.

2.1.3 Einstein radius

The Einstein radius estimates the order of magnitude of the separation between the images. We will derive its expression here in the context of a simple model of lens called the Schwarzschild lens. This model approximates the lens and the source as point masses. We can, therefore, use the Schwarzschild metric, which is an exact solution to the Einstein field equations and describes the gravitational field outside a spherical mass. Using the Schwarzschild metric, the deflection angle can be expressed as follow:

$$\alpha_S(\theta) = \frac{4GM}{c^2\xi}, \quad (2.14)$$

where G is the gravitational constant, and M is the mass of the lens. Note that we expressed all the angular positions as scalar since the problem can be projected in a 2D plane. If the lens and the source are perfectly aligned, the observer will see a ring around the lens. The radius of this ring is what we call the Einstein radius. We can derive its expression by inserting Eq. 2.14

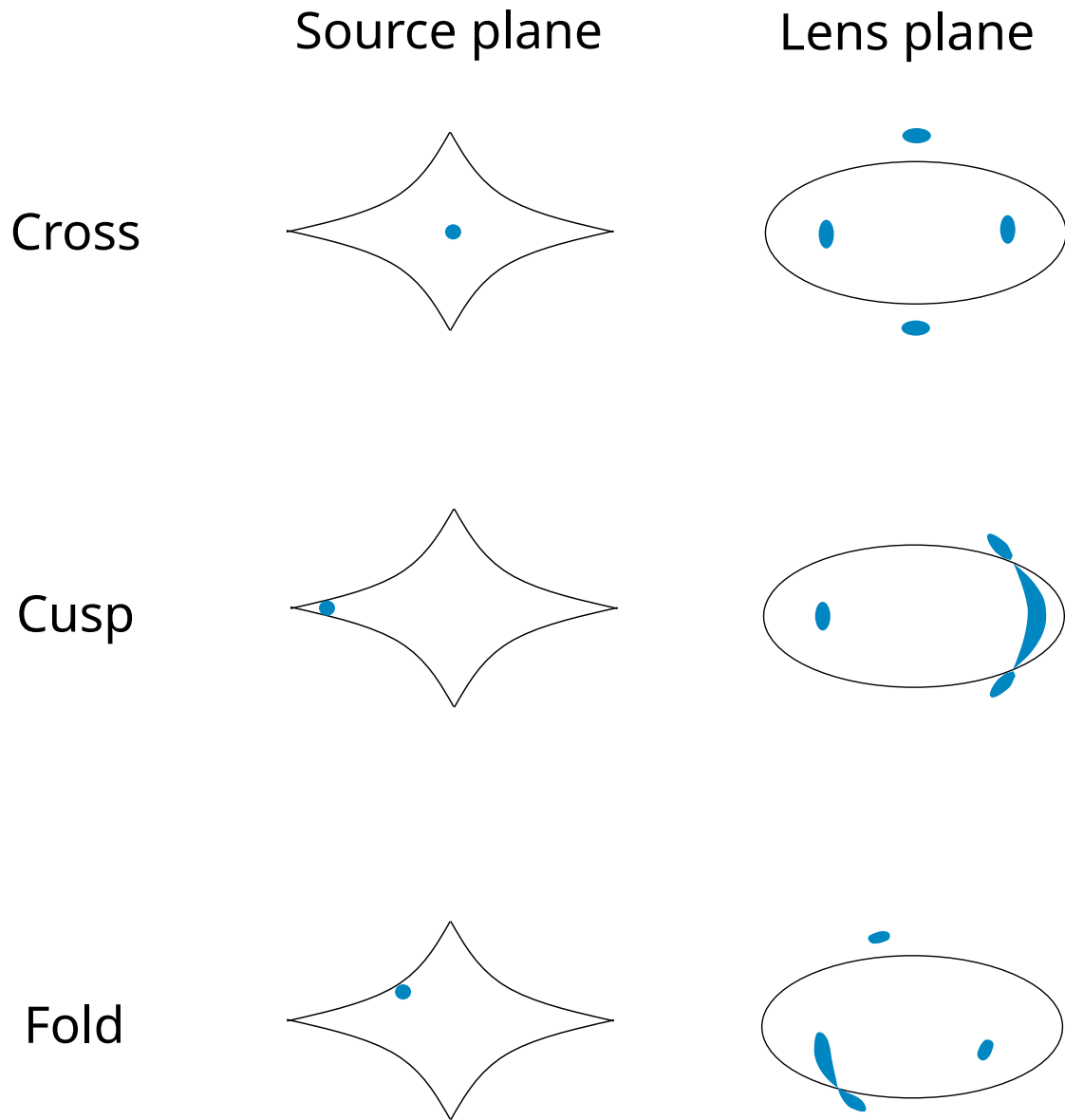


Figure 2.3: Scheme illustrating examples of the three typical lensing configurations formed with an elliptical deflector approximated as an SIE. The left column displays the source positions (in blue) in the source plane with respect to the caustics (in black). The right column shows the shape position with respect to the critical lines (in black) and the shape of the observed lensing features (in blue).



Figure 2.4: Image of the lensed quasar J1131–1231, combined image from NASA’s Chandra X-ray Observatory (X-rays domain) and HST (optical domain). It displays the effect of strong lensing on punctual and extended objects since the source is the addition of a point source (the quasar in pink) and an extended source (the galaxy hosting the quasar, in orange). The substructure of the host galaxy is also visible in the Einstein ring. Credit: adapted from Chandra X-ray observatory (2018)

into the lens equation Eq. 2.2:

$$\theta_E = 2\sqrt{\frac{GM}{c^2} \frac{D_{LS}}{D_S D_L}}. \quad (2.15)$$

Cluster scale lenses with masses that can be higher than $5 \times 10^{14} M_\odot$ can give rise to Einstein radii in the order of magnitude of 10 arcseconds (Bartelmann & Schneider, 2001). For galaxy-scale lenses (early type deflector), the Einstein radii are in the order of magnitude of 1 arcsecond (e.g. Gavazzi et al., 2012).

2.2 Regimes of gravitational lensing

Gravitational lensing events can usually fit into one of the following three categories: strong, weak, or micro, depending on the mass of the lens and the relative alignment between all

bodies taking part in the lens system.

We use the expression strong or macro lensing when multiple images or arclets are noticeable around the lens. Since the image separation is proportional to the lens mass, it requires very massive foreground bodies. For this reason, the foreground objects involved in strong lensing events are usually galaxies or galaxy clusters. However, the sources can be galaxies, quasars, or, less frequently, supernovae. The aspect of the lensing features changes drastically depending on the nature of the source. If the source is an extended object, such as a galaxy, the lensing features take the form of arclets or a ring. More precisely, the lensed source forms a ring, named the Einstein ring, around the deflector when the alignment between the source, the lens, and the observer is perfect. If the lensing system departs from this position, the ring splits into arclets until the alignment becomes too off-axis, and the multiple images disappear. In this case, the lensing system enters the weak lensing regime. In the case of a point source, the Einstein ring is replaced with an Einstein cross. In this configuration, there is also a fifth image in the center of the Einstein cross. However, it is usually not visible as it is demagnified and blended into the lens light. When the source moves away from the configuration producing the Einstein cross, the number and relative position of the images varies based on the position of the source with respect to the caustics. We show in Fig. 2.4 an example of a gravitational lens that displays both point-like and extended lensing features.

Weak lensing encompasses the cases where only distortions are visible in the lensed source image. It generally occurs when the lens is very massive, but the source is not sufficiently well aligned with the lens and the observer to cause a strong lensing event. In such a case, there are no multiple images. Weak lensing is challenging to observe on a unique lensed source since the distortions can be minimal. However, it is visible in the spatial distribution of ellipticities and the alignment of objects close to the lens. In general, the galaxies affected by weak lensing seem slightly aligned and elongated in the direction tangential to the lens.

Microlensing is essentially analogous to strong lensing since it also leads to multiple images of the lensed source. The difference is that the mass of the lens is much smaller than for strong lenses, generally in the range of $10^{-6} \leq m/M_{\odot} \leq 10^6 m/M_{\odot}$ (Schneider, 2006). This induces very small image separation, on the order of the milli- arcsecond for distances $10 \text{ kpc} \leq D_{LS} \leq 1 \text{ Gpc}$. Therefore, the different images cannot be observed separately with a telescope. The overlap of the images boosts the observed luminosity of the lensed source. The apparent magnification of the source corresponds then with the sum of the individual magnification of the images.

2.3 Strong lensing applications

By forming and magnifying multiple images, strong lensing can provide invaluable help in an extensive range of astrophysical and cosmological problems. We will present here a brief overview of the leading applications.

2.3.1 Natural telescope

Thanks to the magnification effect, strong lensing makes the images of distant background sources appear brighter and larger than if they were observed directly. This allows us to observe galaxies that are too faint or too small to be resolved by usual instruments, for example, in Welch et al. (2022), who report a strongly lensed star at redshift 6.2. This also enables us to probe the small scale structures of distant galaxies (Marshall et al., 2007) and in particular to study their properties, for example, their ages or star formation (e.g. Vanzella et al., 2022; Khullar et al., 2021; Johnson et al., 2017; Wuyts et al., 2014). This application is not limited to the visible domain. It is also used to detect high redshift sub-millimeter galaxies to estimate their contribution to the cosmic background radiation (e.g. Blain & Phillips, 2002; Blain et al., 1999).

Cluster scales lenses allow us to unveil the furthest and fainter objects. However, galaxy-galaxy lenses enable a detailed study of particular sources. Examples of such sources are the Cloverleaf Quasar (Magain et al., 1988), a quasar at redshift 6.5 (Fan et al., 2019), SPT-S J031132-5823.4 (Vieira et al., 2013) and a quasar at redshift 6 in a compound lens (Collett & Smith, 2020). Deep and high-resolution imaging with the new generation of telescopes, such as the James Webb Space Telescope (JWST), will enable a larger use of this application with galaxy-galaxy lensing.

2.3.2 Dark matter content of clusters and galaxies

The first evidence of the existence of an invisible but massive matter populating the universe comes from Zwicky (1933). He noted that the velocity dispersion of galaxies within the Coma cluster was significantly larger than expected from the luminous mass. He derived the cluster's mass from the Virial Theorem and noted a large discrepancy between the luminous and gravitational masses. The existence of an unknown, invisible matter left most astronomers skeptical until the seminal work of Vera Rubin (Rubin et al., 1980, 1978), who established that the rotation curves of galaxies (i.e., the orbital velocities of stars and gas relative to their radial distance from the center of the galaxy) remain flat beyond the half-light radius. From this, we know that galaxies are surrounded by a halo of invisible matter called dark matter. The nature of dark matter is still unknown. We know, however, that it only interacts with baryonic matter through gravitational force.

Strong lensing is linked to the mass distribution of the deflector. In particular, it is sensitive not only to the luminous mass but also to the dark matter component and can be used to study galaxy mass profiles (Bellagamba et al., 2017; Sonnenfeld et al., 2015; Koopmans & Treu, 2003).

For example, this property can be used to confirm the presence of dark matter halos around galaxies. Indeed, strong gravitational lensing combined with stellar kinematics of the deflector galaxy allows us to probe the mass distribution at radii beyond the effective radius (e.g. Barnabè

et al., 2009; Treu & Koopmans, 2004). Gravitational lensing can also be used to probe the dark matter distribution, which could help to understand the interplay between baryon and dark matter (e.g. Jiang & Kochanek, 2007). Numerical simulations of cold dark matter predict a cusp in the center of the density profile of galaxies. However, this prediction contradicts observations of the rotation curve of cold dark matter-dominated dwarf galaxies, which indicate a flat profile in the center. Because of its ability to probe the dark matter, strong lensing can help to solve the contradiction (e.g. N. Li & Chen, 2009).

The CDM model predicts the existence of substructures of the dark matter halos. However, there is a discrepancy between the abundance of luminous satellites observed around galaxies and the theoretical prediction obtained from N-body simulations (e.g. Kravtsov, 2010). One possibility to explain this discrepancy is that the substructures are constituted with dark matter only and thus are not visible. In the case that non-visible substructures could not explain the discrepancy, this would involve an alternative model for dark matter and galaxy formation (e.g. Miranda & Macciò, 2007). Gravitational lensing is currently the only method able to observe dark matter substructures (e.g. Kravtsov, 2010).

Small massive bodies belonging to the deflector galaxies, such as stars or dark matter clumps, cause small luminosity variations between the lensed source's different images. These changes in the luminosity of the images depend on the velocity and frequency of the small deflectors and their nature and are visible as perturbation of the magnification pattern. In the case of point sources, this effect produces important differences in the flux ratio between the different images to what could be expected without the substructures. It can therefore be used to detect their presence (e.g. Ritondale et al., 2019; Chatterjee & Koopmans, 2018; Dalal & Kochanek, 2002; Metcalf & Zhao, 2002; Mao & Schneider, 1998).

The presence of substructures can also reveal itself in astrometric and time-delay measurements. Indeed the substructures affect geometry of the lensing system, and thus, the lensing observables (e.g. Keeton & Moustakas, 2009; Chen et al., 2007). In the case of galaxy-galaxy lenses, the perturbations can be observed in systems with a complete or almost complete Einstein ring (Vegetti et al., 2018; Vegetti et al., 2014; Vegetti et al., 2012; Vegetti et al., 2010; Vegetti & Koopmans, 2009; Koopmans, 2005).

2.3.3 Hubble constant measurement

The current expansion rate of the Universe or Hubble constant H_0 is one of the parameters of the Λ CDM model. Finding its value would represent an essential step toward validating and constraining the cosmological model of the Universe (D. H. Weinberg et al., 2013).

Latterly, a discrepancy emerged between the value of H_0 obtained from "early" universe measurements and "late" universe measurements (Freedman, 2017). The value derived from the temperature fluctuations of the cosmic microwave background (Planck Collaboration et al., 2020; Planck Collaboration et al., 2016) is in tension with the value obtained from local

measurements of supernovae type Ia and Cepheid distances from the Supernova H0 for the Equation of State (SH0ES) project (Riess et al., 2019; Riess et al., 2016). The measurements of H_0 using the cosmological microwave background fluctuations (CMB) strongly depend on the cosmological model since they depend on the fraction of the different Universe components. However, local measurements of H_0 are almost independent of the cosmological model. A discrepancy between these two values could therefore call into question the validity of the Λ CDM model. This necessitates independent measurements of H_0 , which gravitational lensing can provide via so-called 'time delay' measurements (e.g. Linder, 2011; Treu et al., 2011).

The Fermat potential is related to the time delay between a photon emitted from a source affected by the gravitational potential of the lens and a photon emitted from the same position but in the absence of the lens in the following way:

$$t(\beta, \theta) = \frac{1}{c} \frac{D_L D_S}{D_{LS}} (1 + z_s) \tau(\beta, \theta), \quad (2.16)$$

where z_s is the redshift of the source.

This quantity is not directly measurable because we do not have access to the travel time of the unlensed light ray. However, it is possible to observe the difference in time delays between pairs of images.

The difference of time delay between images i and j is given by:

$$\Delta t_{i,j} = \frac{1}{c} \frac{D_L D_S}{D_{SL}} (1 + z_s) (\tau_i(\beta_i, \theta_i) - \tau_j(\beta_j, \theta_j)). \quad (2.17)$$

The different angular distances mentioned in Eq.2.17 are inversely proportional to H_0 . Since the other terms do not depend on H_0 , $\Delta t_{i,j}$ is proportional to H_0^{-1} . Time delay differences also depend on the cosmological model since the angular distances depend on the matter, dark energy, and curvature density. Nevertheless, this dependence is weak (only to the power $\frac{1}{2}$). Thanks to this weak dependency, time delay measurements can be used as an independent probe to solve the H_0 tension. This method requires using a varying source, i.e., a source whose luminosity is not constant in time. To estimate the time delay, we measure the shift between the light curves (luminosity as a function of time) of the different images of the source. This method was first proposed by Refsdal (1964) using type Ia supernovae (SNIa) as lensed sources. However, this method has not yet been used with this type of source as the first resolved lensed SNIa was only recently discovered by Goobar et al. (2017). Instead, the first measurements were performed with lensed quasars.

Measurements of time delays of strongly lensed quasars have been conducted by the H_0 Lenses in COSMOGRAIL's Wellspring collaboration (H0LiCOW) (Suyu et al., 2017). Recently, they achieved 2.4% accuracy in the measurement of H_0 (Wong et al., 2020) using joint analysis of the time delays of six lensed quasars.

However, using quasars for this type of measurement has some drawbacks compared to SNIa. First, degeneracies that can affect the measurement of time delays arise during the reconstruction of the lens potential. The so-called source-position transformations of the lens mass model leave the positions of the images and the flux ratios unchanged. They are known as the source-position transformations expounded in Schneider and Sluse, 2013 (2014, 2013) , of which the mass sheet degeneracy is a particular case. SNIa can overcome these degeneracies since they are standard candles. Indeed, the knowledge of the intrinsic luminosity gives the possibility to estimate the absolute magnification of the images (e.g. Schneider & Sluse, 2014, 2013) which is impossible for quasars. In addition, measurements of SNIa time delays require much shorter monitoring than quasars (weeks compared to years) (Bonvin et al., 2019).

In the future, the precision H_0 measurements with lensed SNIa will be possible thanks to future large-scale surveys. For example, around 500 lensed SNIa are expected in LSST; among them, 100 would be suitable for measurement time delays (Suyu et al., 2020). In this context, the Highly Optimized Lensing Investigations of Supernovae, Microlensing Objects, and Kinematics of Ellipticals and Spirals (HOLISMOKES) (Suyu et al., 2020) was launched to measure H_0 with SNIa. In the context of their project, the sample of lensed SNIa will be extended by monitoring the position of known galaxy-galaxy lenses in search of transient events with the Zwicky Transient Facility (ZTF) (Bellm et al., 2019; Masci et al., 2019). The lens candidates found in the context of this thesis in the CFIS survey are included in the sample of known lenses that HOLISMOKES currently monitors.

3 Machine learning methods

In the previous chapters, we described the formalism of gravitational lensing and the importance of finding large sets of new lenses. With the advent of large-scale surveys, the available data will grow exponentially. This requires the development of automated algorithms to detect or classify astronomical objects. Because of their ability to make inferences directly from data, machine learning algorithms are especially suited for such tasks. In this thesis, the focus is on automated methods to detect galaxy-galaxy lenses in imaging data. We present here the basic concepts of machine learning and a set of techniques and metrics especially suited to the field of lens finding. This chapter is inspired from Goodfellow et al. (2016) to which we refer the interested reader for a more detailed description.

3.1 A simple introduction to machine learning

In the early day of computer algorithms, computers were used for tasks that could be solved with a list of instructions. The era of big data called for another type of algorithm to deal with tasks that are not easily translatable into sets of instructions. One good example is image classification. If one wants to build an algorithm to separate images of oranges from pears images, one can write instructions about the fruits' shape and color. For example: "if the object is orange and round: classify the image as an orange. If the object is green or red and not round: classify the fruit as a pear". However, an image as Fig. 3.1 will be impossible to classify with this algorithm.

In contrast, it will be straightforward for a human to identify the image as an orange even if the color is altered. The algorithm may be improved by adding a set of more complex instructions but will fail to classify images that do not match precisely the instructions. A cleverer approach would be to mimic how a child would learn to differentiate oranges from apples, i.e., by inferring the important features after being exposed to a large number of examples of the two different kinds of fruits. It is the idea behind machine learning. Machine learning encompasses a large set of algorithms able to infer information directly from the data.



Figure 3.1: Orange with altered colors. In the context of classification of fruits images, this type of image is difficult to categorize with an algorithm based on a list of instructions. A machine learning algorithm will identify the presence of features typical to orange and may classify it correctly despite its unusual color.

3.1.1 Learning in the context of machine learning

The term "learning" within the framework of machine learning has to be understood as learning by experience, as stressed in the definition by Mitchell (1997) : "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E ."

It is very important to realize that, in machine learning, it is not only the algorithm that plays an essential role. The choice of the experience and the performance metrics are also crucial to solving a task. For a machine learning algorithm, the experience corresponds to the data given to the algorithm. These data can be of different nature, for example, images, numerical measurements, text, and sound. The performance is measured with a function called "objective function" or "loss function" that the algorithm aims to optimize. The definition of learning provided by Mitchell (1997) highlights that learning is a step-by-step process. After each iteration, the performance on the task is quantified by performance measurement. Then, the algorithm adjusts its parameters to maximize the performance.

Machine learning algorithms can be separated into two categories: "supervised" or "unsupervised," depending on the type of data they are exposed to. In supervised learning, the algorithm learns a function or a model that maps a set of inputs with a set of desired outputs which are called the labels. These outputs are provided to the algorithm during the training. In supervised learning, the performance measures the difference between the labels and the algorithm's output. By contrast, in unsupervised learning, the algorithm has to learn a representation of the set of inputs without having any feedback from the desired output. In unsupervised learning, the loss function penalizes the difference between the output of the algorithm and the input itself, i.e., the inputs serve as labels.

3.1.2 Model parameter versus hyper-parameter

Most machine learning algorithms analyze the data through statistics, i.e., at each step, they update the parameters of a statistical model based on the data and the performance metric. Machine learning models can have two different types of parameters. The first category encompasses variables external to the model. They cannot be estimated directly from data and must be adjusted by the programmer. The second category contains inner variables. The algorithm estimates these from data during the learning phase. To make a clear difference between the two, we name the variable belonging to the first category "hyper-parameter" and the second type "parameter."

3.1.3 Overfitting, underfitting

Following the definition of Mitchell (1997), we can understand machine learning as an optimization problem. However, in machine learning, we are generally interested in applying the model to new data. The generalization error measures the algorithm's error when applied to unseen data. Two situations can lead to a large generalization error. They are illustrated in 3.2 for an example with a linear regression algorithm. On the one hand, the model can adapt "too well" to the training data and thus, may not be able to generalize when applied to new data. This situation is called overfitting and usually happens when a model learns the details and noise of the training data. This is a problem because these details will not necessarily be the same or present in the new data. On the contrary, underfitting refers to a model that can neither explain well the training data nor generalize to new datasets. It happens when the model cannot capture all the essential features of the training data.

For this reason, we usually split the data into three different sets: the training, validation, and test set. The instances of the different sets must be independent and identically distributed. Before reaching the optimal model, not only must the parameters be optimized, but also the hyper-parameters. The validation set serves to optimize the hyperparameters. After the training, the performance is evaluated on the validation set, and the hyperparameters are adapted if necessary. The reason for not using the test set to this end is that in this case, the error on the test set would not correspond anymore to the generalization error as the hyper-parameter would have been chosen to maximize the performance on the test set. The test set is used only after the hyperparameters and parameters are optimized to estimate the generalization error correctly. The validation set can also be used for early stopping, which consists of stopping the training when the error on the validation set increases. This limits the risk of overfitting, (e.g. Caruana et al., 2000).

3.1.4 The curse of dimensionality

When the number of dimensions of the input data is high, many machine learning problems become incredibly difficult. The problem called "the curse of dimensionality" arises when the

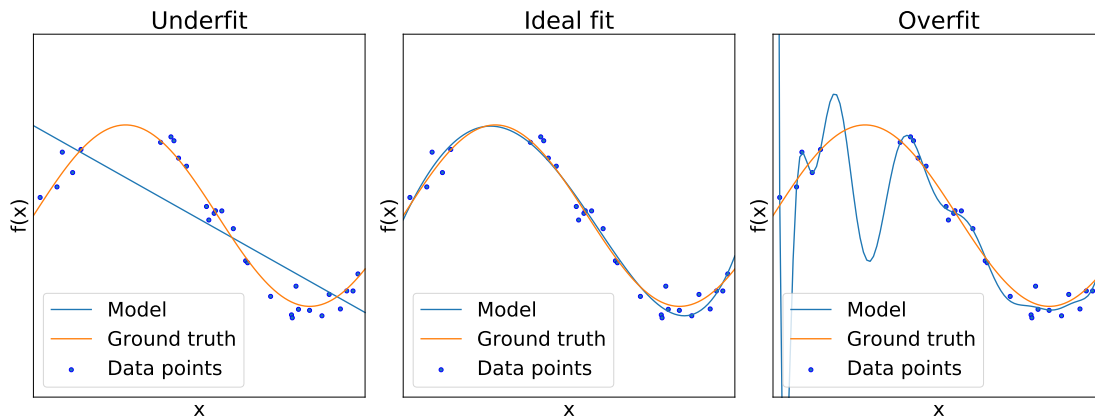


Figure 3.2: Example of underfitting and overfitting with linear regression: The true function is shown in orange. The data points correspond to this function plus noise. the left panel shows a situation of underfitting. The function is fitted with a degree 1 polynomial, which is insufficient to fit the data. The middle panel shows a fit with a degree 4 polynomial, which is sufficient to approximate correctly $f(x)$. The right panel displays a fit with a degree 15 polynomial. In this case, the polynomial has too many degrees of freedom, the fit captures the noise present in the data leading to overfitting. Credit: adapted from the documentation of scikit-learn (Pedregosa et al., 2011).

number of possible configurations in the data (the dimension of the problem) is much larger than the number of training samples. The curse of dimensionality, which is a widespread issue in actual machine learning applications, was first described in (Bellman, 1957). Since the number of possible configurations grows exponentially with the dimension of the problem D_p ($\sim \exp(D_p)$), this issue may occur when the dimension of data is increased if the number of training examples is not adapted.

For example, if one chooses to work on the generation of 256 by 256 greyscale images of faces, the dimension of the pixel space is already 65536. This problem can be visualized by taking an example of binary classification. Let us consider the problem of classifying cats and dogs. We have at our disposal numerical values corresponding to specific features such as size, ear shape, and the number of paws for each animal. We can first consider only one feature, for instance, the size, and project all the data on the subspace spanned by all the possibilities for this particular feature. The situation is displayed schematically in the top panel of Fig. 3.3. However, as we can see, this feature is insufficient to separate the two classes well. A model considering only this feature will under-fit the training data. If we add a second feature, for instance, the ear shape as presented in the middle panel of Fig. 3.3, we can easily classify the data by drawing a line in $y = 11$. We may be tempted to consider more features, for example, the number of paws, to refine the classification (i.e., to augment the dimension of the problem). However, if we keep the same number of training data, there will not be enough representative training data for each possible configuration to enable the algorithm to generalize. This is what we call "the curse of dimensionality."

In Fig. 3.3 we can observe that the 3D plot becomes already very sparse (i.e., some configurations are not or are hardly represented in the training dataset). The sparsity grows exponentially as a function of the dimension of the problem if the number of data stays constant. Therefore if we add more dimensions without augmenting the number of examples in the training set, the model may overfit the training data. When this happens, the model will have trouble drawing the boundary between the two classes in the parts of the feature space that are less represented in the training set. This will translate into a discrepancy between the training and generalization errors. One solution to this problem is to select only the "most meaningful" features. For example, if we aim to build a face recognition algorithm, not all the pixels of the images contain relevant information. The pixels in the nose and eyes are more valuable to identify a person than those in the picture's edges containing only the background.

Theoretically, all supervised machine learning models must be limited by the curse of dimensionality. Consequently, the number of necessary training examples grows exponentially with the dimension of the data. However, deep neural networks can, in practice, overcome this effect (e.g. Mallat, 2016). We present them in Sect. 3.2.1

3.1.5 Principal component analysis

Principal component analysis (PCA) (F.R.S., 1901) is a standard method to reduce the dimensionality of the data by finding the most meaningful features. Since it is also a simple illustrative example of an unsupervised algorithm, we quickly summarize the different steps of the algorithm. Let us consider the problem of reducing the dimension of a set of n images containing k pixels each. The dataset is represented as a matrix x with k rows and n columns. Before applying the PCA, the data must be standardized and mean-subtracted. The space spanned by all the possible values each pixel can take is called the feature space. Each coordinate axis of the feature space represents the value of one of the K pixels.

PCA aims at finding the directions that explain the most of the data, i.e., where the variance among the data is maximal, and project the data on a meaningful basis. Finding the direction of maximal variance \vec{c} , called the first principal component, can be reformulated as maximizing $\vec{c}_1^T X^T X \vec{c}_1$ under the constraint $\vec{c}_1^T \vec{c}_1 = 1$. Solving this problem is equivalent to computing the eigenvalues of the covariance matrix $Cov(X, X)$. \vec{c} is then the eigenvector associated with the largest eigenvalue.

Therefore, the first step of PCA is to compute the covariance matrix. The eigenvector of the covariance matrix (principal component) forms a basis of the feature space. We then sort the principal components such that the first corresponds to the eigenvector with the largest eigenvalue, the second to the second largest eigenvalue, and so forth. We can then project the data on the new basis formed by the principal components. Fig. 3.4 displays a simple example in 2D of the first steps of PCA.

The total variance corresponds to the sum of the variance of the data along the direction

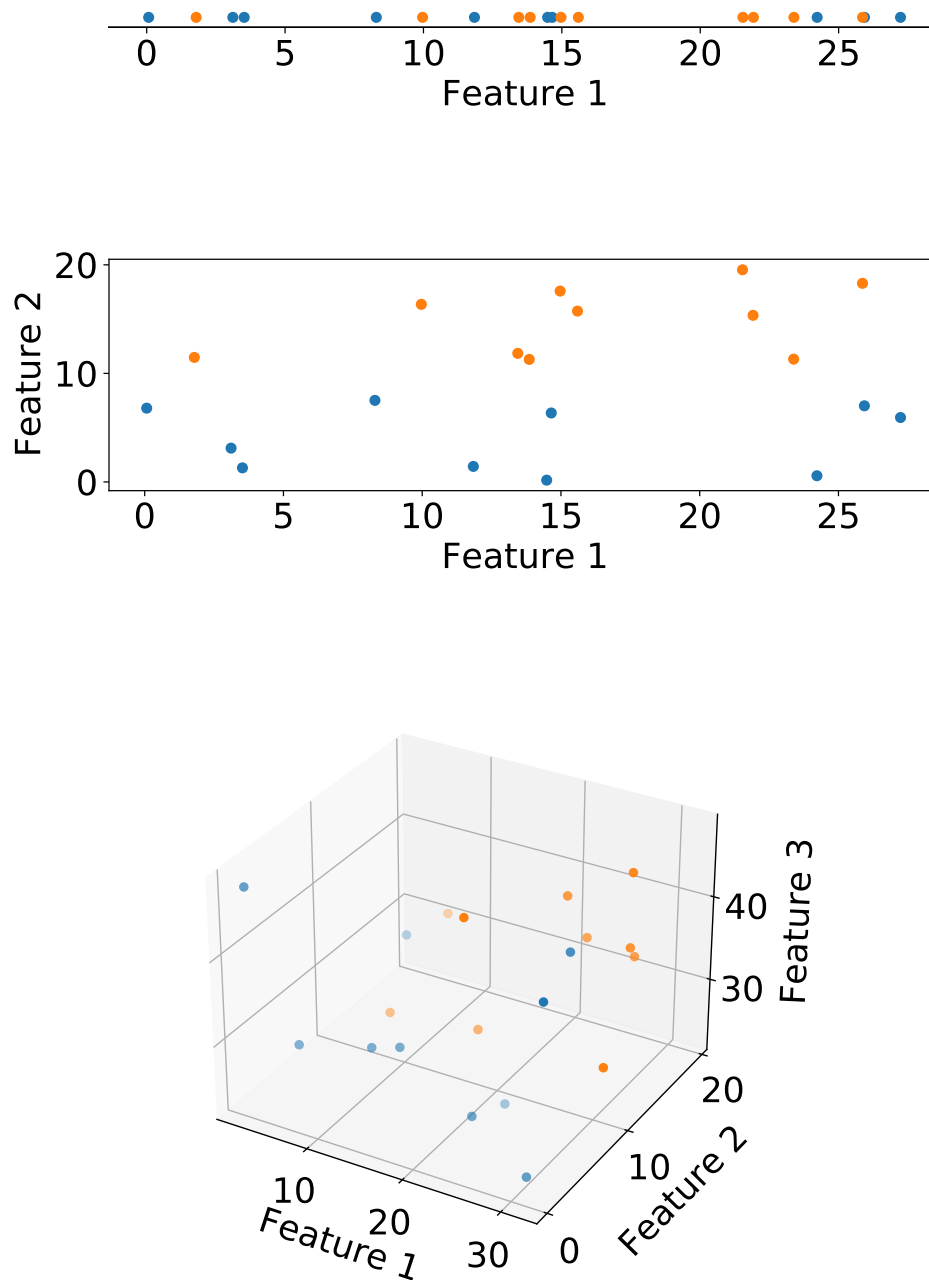


Figure 3.3: Illustration of the curse of dimensionality problem. Here, the aim is to find a classification boundary between the 20 orange dots and the 20 blue dots. The upper panel represents the two data-sets projected on the dimension spanned by one of the features. In the middle panel, we added the dimension spanned by a second feature. The lower panel represents the data-set as a function of the two features used in the previous plots and a third feature.

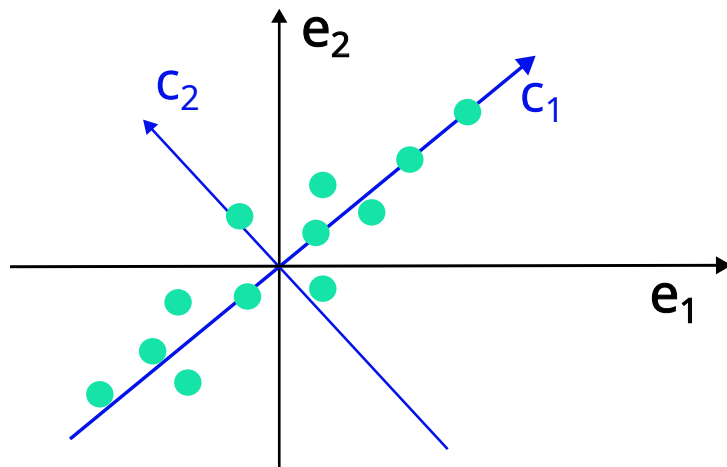


Figure 3.4: Example of PCA for a set of 12 images with 2 pixels. The direction e_1 and e_2 represent all possible values for pixels 1 and 2. The first component of the PCA c_1 is aligned with the direction where the variance among the data is maximal. The second axis of the PCA c_2 correspond to the direction where the variance is maximal after c_1 with the constrain to be perpendicular to c_1 .

of each individual principal component. We define the fraction of variance explained by a principal component as the ratio between the variance of the data along the direction of that principal component and the total variance. By construction, the eigenvalues are directly related to the fraction of variance explained in the axis of the corresponding eigenvector. The variance explained by each principal component can be visualized on a scree plot, We show in Fig. 3.5 an example of scree plot, which displays the percentage of explained variance for each principal component for the PCA decomposition of the test set of Modified National Institute of Standards and Technology (MNIST) dataset of handwritten digits (L. Deng, 2012), which consists of 28x28 grey-scale images of handwritten digits. The last step of the PCA is to discard the last principal components that explained the less variance and project the data on the first p principal components such that the dimension will be reduced from k to p but most of the essential information is kept. This corresponds to the matrix operation:

$$Y = A_p X, \text{ where } A_p = \begin{pmatrix} \bar{c}_1^T \\ \vdots \\ \bar{c}_p^T \end{pmatrix}. \quad (3.1)$$

The number p must be adjusted as a function of the desired quality of the reconstruction. The scree plot can help finding the right number of dimensions to keep. We display in Fig. 3.6 an example of the reconstruction of the MNIST images after PCA for different p . We can see that using only two principal components is insufficient to obtain recognizable numbers. This could have been guessed from the scree plot 3.5 where we see that these two components explain only 9.8 % of the variance. Starting from 60 principal components, the shapes are well

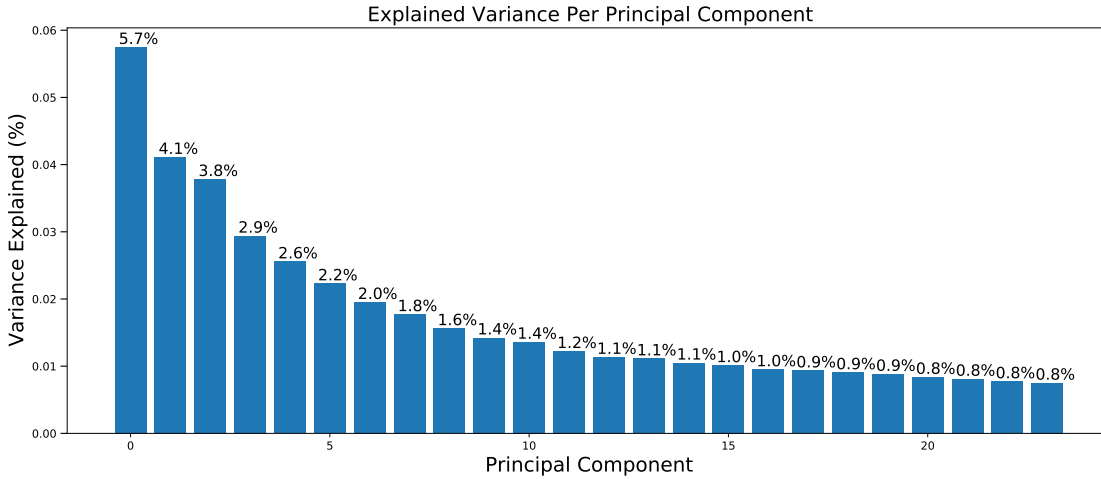


Figure 3.5: Percentage of explained variance for the 24 first component of the PCA decomposition of the MNIST dataset.

reconstructed. Increasing to 100 principal components does not significantly improve the quality of the reconstruction.

PCA has already been used in gravitational lensing, for example, to find lensed sources in the quasar spectra (Courbin et al., 2012; Boroson & Lauer, 2010) or to deblend lenses (Paraficz et al., 2016; Joseph et al., 2014). It has, however, a few limitations. First, PCA can apply only linear mapping. Since it uses a covariance matrix, PCA assumes that the data follows a multidimensional normal distribution. There are currently multiple generalizations of PCA able to overcome these issues. An alternative to PCA for dimension reduction is auto-encoders which are described in Sect. 3.2.3. We present one of them in Sect. 3.2.3, the auto-encoder, which can be seen as a generalization of PCA.

3.2 Neural networks

The idea of artificial neural networks arose at the end of the 1950s with the need to imitate the brain's functioning for research in the field of psychology. To this end, Frank Rosenblatt developed the idea of the Perceptron (Rosenblatt, 1957) which was the first version of artificial neural networks. It consisted of one layer of artificial neurons.

The output of each of these neurons was binary, 0 or 1. Although this perceptron could perform simple classification, it was only able to learn linear functions. At this time, the stacking of layers was not considered since it was thought they would take an infinite time to be trained. Rumelhart et al. (1986) overcame this limit by discovering error backpropagation.

Not long after, Cybenko (1989) proved the first version of the Universal Approximation Theorem which states that a network with at least one hidden layer can approximate any continuous functions on compact subsets of \mathbb{R}_n . Consequently, multi-layer perceptrons are potent tools

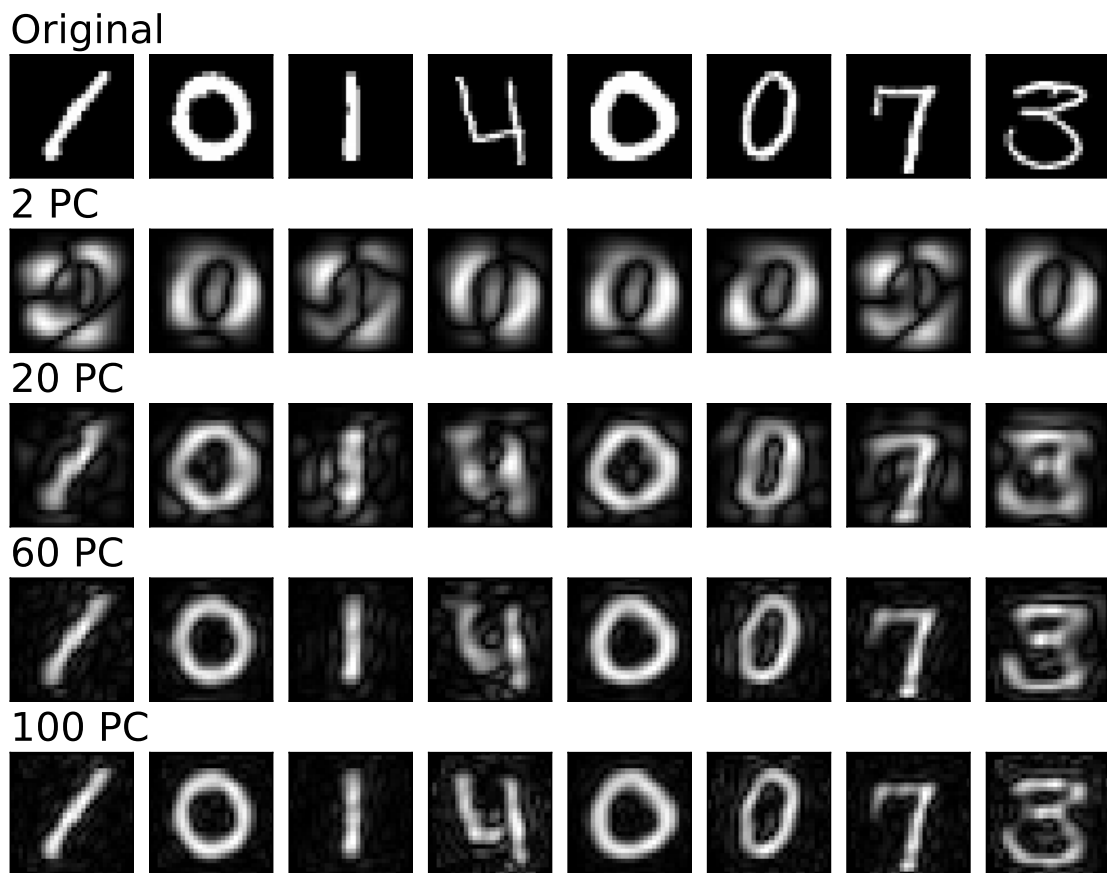


Figure 3.6: Reconstruction after PCA of images from the MNIST test set. The first row displays the original images and the following rows the reconstruction using respectively 2, 20, 60 and 100 principal components.

since they are not limited to approximating linear functions as one-layer perceptrons. However, these algorithms were extremely demanding in computational resources. Neural network research was, therefore, slowed until the advent of graphics processing units (GPU) in the 2010s. The interest in neural network methods took off in particular after the outstanding performance of "ALexNet", a convolutional neural network, in the Large Scale Visual Recognition Challenge (Krizhevsky et al., 2012a).

3.2.1 Artificial neural networks: Theoretical introduction

The neurons are the basic building blocks of an artificial neural network (ANN). They are developed to mimic the behavior of biological neurons. Biological neurons have dendrites to receive signals, a cell body to process their electrical inputs, and an axon to send signals to other neurons. Similarly, artificial neurons have inputs, a processing stage, and a unique output that is transferred to other artificial neurons. The processing stage can be summarized in three steps schematized in Fig. 3.7.

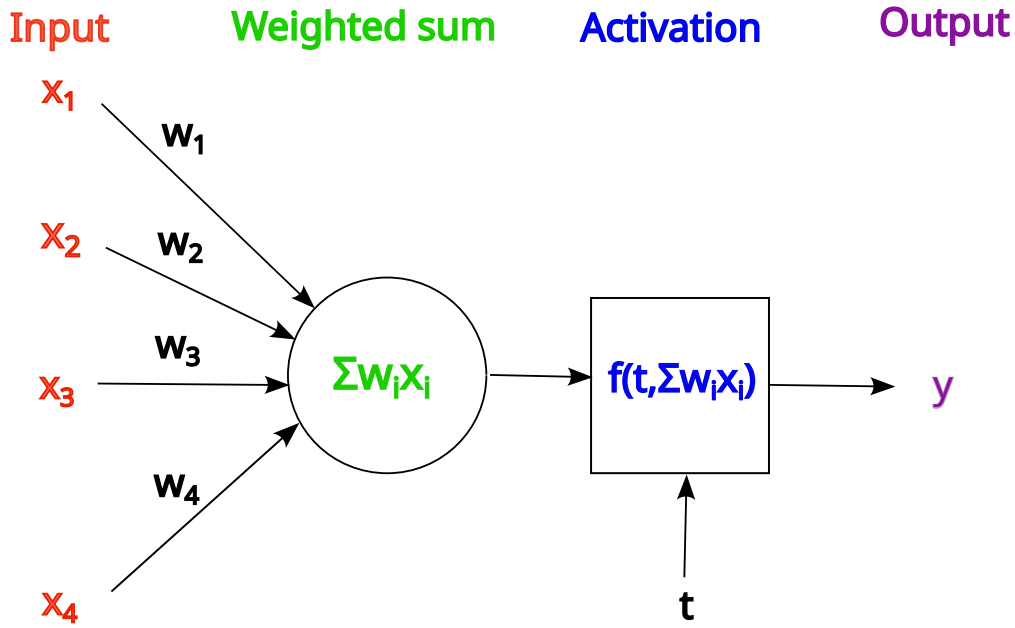


Figure 3.7: Scheme of an artificial neuron with four inputs and one output.

First, the input data are re-scaled with multiplication by a specific weight assigned to each input component. During the training phase, the neurons adjust the weights to reduce the final output error. All the weighted inputs are summed in the second step, and an offset is added to this sum. This offset is called bias and is also adjusted during the training phase.

Finally, the results of the neuron calculation are passed to the activation function. This function decides the neuron's output and introduces non-linearity in neural networks. One of the most simple activation functions is the Heavyside step function, which takes the following form:

$$\sigma(x) = \begin{cases} 1 & \text{if } w \cdot x + b > 0 \\ 0 & \text{otherwise} \end{cases}, \quad (3.2)$$

where x is the vector of the inputs, w is the vector containing all the weights, and b is the vector of the biases.

Using this activation function allows the network to be only in two states, activated when the output is equal to one or inactive when the output is equal to zero. This behavior mimics very well the behavior of biological neurons. However, this function is not differentiable at zero and differentiates to zero at all other points. As we will see later, this causes problems during the training phase since gradient descent is used to optimize the weights and biases. A suitable alternative to the Heavyside step function this activation is the sigmoid activation:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (3.3)$$

Sigmoid activation is commonly used in binary classification problems. Its shape is close to the one of the Heavyside step function. However, it is differentiable in all its domain of definition. In the case of multi-class classification (more than two categories), a generalization of the sigmoid function called the softmax function can be used. It takes the following form:

$$\sigma(x_i) = \frac{e^{-x_i}}{\sum_{j=1}^k e^{-x_j}}, \quad (3.4)$$

where k is the number of classes in the problem. The softmax and sigmoid function output continuously span the interval between 0 and 1. Thus, if used in the model's output neurons, they reflect the probability for x to belong to one of the classes.

The main flaw of the sigmoid function is its saturation at the tails. The gradient in these regions is very close to zero and may increase the training time or prevent the model from converging if the output falls here. The sigmoid function was one of the most widely used activation function until the emergence of rectified linear units (ReLU) (Fukushima, 1975):

$$\sigma(x) = \max(x, 0) \quad (3.5)$$

ReLU activations are the most simple nonlinear activations functions. They allow the usage of shorter training time than with sigmoid activation Krizhevsky et al. (2012b) . They also ease the obtainment of a sparse representation with the network. For example, if the weights are uniformly initialized, the output of roughly 50 percent of hidden units will be zero due to the shape of the function below zero. This property is, however, double-edged. The gradient of the ReLU is canceled for values below zero, which may cause the ReLU units to freeze during the training. Indeed, the weights and biases are not updated anymore during gradient descent if the gradient is zero. This issue is called the dying ReLU problem (e.g Lu et al., 2019). To avoid this but keep some benefits, it is possible to use slightly modified ReLU functions as an exponential linear unit or leaky ReLU.

A single neuron is already able to perform some simple tasks. For example, a neuron with the Heavyside activation can perform an easy linear classification. It is necessary to connect multiple neurons for more complex tasks. Such an ensemble of neurons is called a neural network. The neurons in a neural network are organized in layers, as depicted in the example in Fig. 3.8. The first layer is called the input layer, and the last is the output layer. The inner layers of the network are called hidden layers. The neurons belonging to a same layer

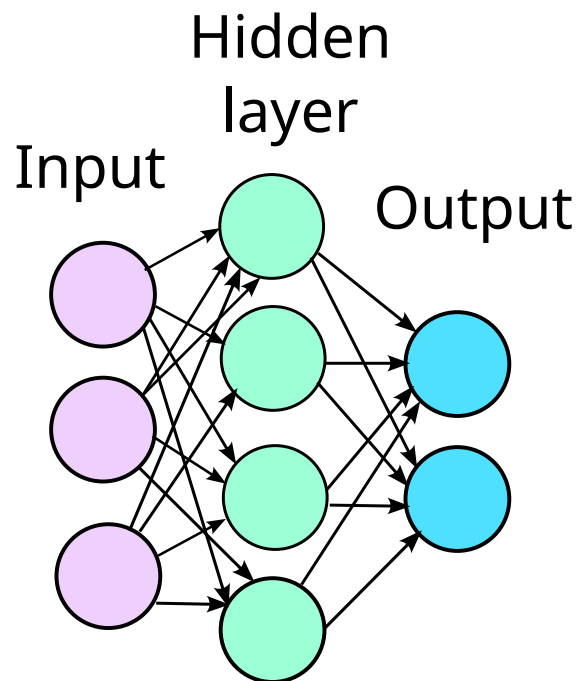


Figure 3.8: Example of a neural network composed of three dense layers. This network has only one four-dimensional hidden layer (In light green). The input dimension is three (violet) and the dimension output is two (blue). The arrows depict the connections between the neurons and the direction of propagation of the information through the network.

are not connected together. It allows them to react differently to identical input data and, consequently, makes them sensitive to different data features. The number of neurons per layer is only constrained in the input and output layers, where the number of neurons has to match the input and output dimensions. However, the number of neurons in the hidden layers and, similarly, the number of hidden layers are part of the hyper-parameters for a machine learning model based on a neural network. The user, therefore, settles them.

Learning with gradient descent and error back propagation

The non-linearity of neural networks is one of their main assets since it allows them to approximate easily non-linear function or non-linear decision boundaries in the case of a classification problem. However, the non-linearity transforms the training of neural networks into a non-convex optimization problem. For this reason, neural networks are generally trained with a variation of the stochastic gradient descent algorithm, which can be used either for convex or non-convex optimization. We must, however, keep in mind that in the case of non-convex problems, the convergence is not certain and depends strongly on the choice of initial parameters.

During the learning phase, the parameters of neural networks are optimized to minimize the value of the loss function. There are many valid options for the choice of the loss function

used for neural network optimization. One example is the mean square error (MSE) which is widely used for linear regression or image reconstruction:

$$J(\theta) = \frac{1}{N} \sum_{i=1}^N (f^{(i)}(x_i, \theta) - y^{(i)})^2, \quad (3.6)$$

where N is the number of data, x_i are the input data, θ the model parameters, $f^{(i)}(x_i, \theta)$ the output of the model and $y^{(i)}$ the truth vector which contains the target values the model aims to recover.

If the activation functions are sigmoids, this loss function may cause slow convergence toward the target since, as detailed below, during gradient descent, the gradient of the loss function is used to update the parameters. This gradient takes the form:

$$\frac{\partial J}{\partial \theta} = -(y - f(\vec{x}, \theta)) f'(\vec{x}, \theta) \vec{x} \quad (3.7)$$

The particular shape of the sigmoid function implies that when f tends to 0 or 1, f' is close to 0. Therefore, if the difference between the output function f and the truth vector y is large, the derivative f' will approach 0, leading to a decrease in the convergence speed. This behavior is critical since we expect the convergence speed to be higher when the error is significant. An alternative can be used for binary classification problems (assuming the target values are 0 or 1), the cross entropy function:

$$J(\theta) = \frac{1}{N} \sum_{i=1}^N (y^{(i)} \log(f^{(i)}(x_i, \theta)) + (1 - y^{(i)}) \log(1 - f^{(i)}(x_i, \theta))) \quad (3.8)$$

The cross-entropy measures the divergence between two probability distributions. If the cross entropy is large, the difference between the two distributions will be significant too. On the other hand, if the cross entropy is small, this implies that the two distributions are almost similar. Cross entropy can be easily generalized for multi classes problems in the following way:

$$J(\theta) = \sum_{i=1}^N \sum_{k=1}^K -(y^{(i)(k)} \log(f^{(i)(k)}(x_i, \theta))), \quad (3.9)$$

where K is the total number of classes.

Generally, loss functions penalize the reconstruction error, i.e., the error related to the model's output. To help the model generalize on new data and limit the chance of overfitting is possible to add an additional term in the loss function, called the regularization function. The regularization function often takes the form:

$$\lambda R(\theta), \quad (3.10)$$

where R is a function that becomes large when the parameters θ are large. The constant λ has

to be tuned according to the importance we want to put on generalization. Choosing a large λ may lead to underfitting. On the contrary, a too small λ may not inhibit enough overfitting. One of the most common regularization functions is the Ridge regression, also named L2 regularization. It takes the following form:

$$R(\theta) = \frac{\lambda}{2} \|\theta\|^2. \quad (3.11)$$

Intuitively, the L2 regularization encourages the network to use all of its inputs rather than focusing only on a few inputs by penalizing "very large" weight vectors (Krogh & Hertz, 1991). Another standard regularization function is the Lasso regression or L1 regularization function:

$$R(\theta) = \lambda \|\theta\| \quad (3.12)$$

Contrary to L2 regression, L1 regression encourages the network to use only a small subset of the inputs containing the most important features. This regularization tends to yield a sparse representation of the inputs and to determine the most relevant features (Ng, 2004).

In addition to adding a regularization term to the loss function, other methods exist to limit overfitting. One example is dropout, which consists of randomly deleting some connections in neural networks (Hinton et al., n.d.; Srivastava et al., 2014).

In order to be used to train a neural network, the loss function has to fulfill two hypotheses. First, the loss function can be decomposed as $J = \sum_{i=1}^N \frac{J_i}{N}$ where N is the number of sample. Secondly, it can be expressed as a function of the outputs from the neural network.

As mentioned earlier, the training phase aims to find the global minimum of the loss function with respect to the weights and biases of the network. To illustrate how backpropagation and gradient descent are used to optimize the network, let us consider a neural network containing L layers. The backpropagation algorithm from Rumelhart et al. (1986) can be decomposed into a few simple steps. The steps are written only for one training sample to simplify the notation.

1. Weights and biases are initialized to a chosen value.
2. An input and the related trust vector are fed into the neural network
3. Weighted sums z and activations a are computed by propagating forward the input.
4. The error for the neurons of the last layer is computed as follows:

$$\partial_{jL} = \frac{\partial J}{\partial a_{jL}} \frac{\partial a_{jL}}{\partial z_{jL}} \quad (3.13)$$

5. The error on the last layer is back-propagated to the next to the last layer. The error for

the neurons belonging to this layer is computed as follows:

$$\partial_{jL-1} = \sum_{jL} \frac{\partial J}{\partial z_{jL}} \frac{\partial z_{jL}}{\partial a_{jL-1}} \frac{\partial a_{jL-1}}{\partial z_{jL-1}} = \sum_{jL} \partial_{jL} \frac{\partial z_{jL}}{\partial a_{jL-1}} \frac{\partial a_{jL-1}}{\partial z_{jL-1}} \quad (3.14)$$

6. The error for the neurons located in the remaining layers is computed in the same way as (3.14).
7. The parameters of the network are updated. For the parameters related to the j^{st} neuron in the 1^{st} layer, the update step is:

$$\theta^{(t+1)} = \theta^{(t)} - \alpha \frac{\partial J}{\partial \theta_{jl-1jl}} = \theta^{(t)} - \alpha \partial_{jL} \frac{\partial z_{jl}}{\partial \theta_{jl-1jl}}, \quad (3.15)$$

where α is a hyper-parameter called the learning rate. The update procedure we just described is called stochastic gradient descent (SGD). However, updating the parameters after each input makes the training procedure unstable and sensitive to outliers. For this reason, in practice, we prefer to feed many inputs (batches) forward to the network before updating the parameters. In this case, the error is computed by summing the error obtained for all the individual inputs in a batch. The parameters are updated after all batch inputs are passed through the network. Choosing the value of the learning rate α is crucial. A too low value for α may lead the system to get stuck in a local minimum or increases dramatically the number of steps necessary to reach the global minimum. On the contrary, a too high value leads to substantial and noisy parameter changes. In this case, the system may even miss the local minimum.

Batch gradient descent does not guarantee the convergence of the model because of the non-convexity of the loss function. In addition, finding the optimal learning rate for the problem is complex, and applying the same learning rate on all the parameters may not be optimal for some cases. In particular, when the data are sparse or if the features are not standardized, updating the parameters at different rates may drastically improve the chances for the algorithm to find the global minimum. Many improvements to the gradient descent algorithm have been implemented. In this work, we use the Adaptive Moments Estimation (Adam) optimization (D. P. Kingma & Ba, 2014), which is an extension of stochastic gradient descent. The Adam algorithm computes adaptive learning rates for each parameter. At the end of each complete pass of the training dataset through the algorithm, an exponentially decaying average of past squared gradients v_t and exponentially decaying average of past gradients m_t are saved. These two values are computed as follows:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \text{ and } v_t = \beta_2 m_{t-1} + (1 - \beta_2) g_t^2 \quad (3.16)$$

where m_t and v_t are estimates of the first moment (mean) and the second moment (non centered variance) of the gradient g_t and β_1 and β_2 are the decay rates applied to the moments. The moments tend to be biased toward zero when they are initialized with null vectors. To

avoid this D. P. Kingma and Ba (2014) propose to use their unbiased equivalents:

$$\hat{m}_t = \frac{m_t}{(1 - \beta_1^t)} \text{ and } \hat{v}_t = \frac{v_t}{(1 - \beta_2^t)} \quad (3.17)$$

The parameters are then updated in this way:

$$\theta^{(t+1)} = \theta^{(t)} - \frac{\alpha}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t, \quad (3.18)$$

where ϵ is a small smoothing term used to avoid division by zero.

3.2.2 Convolutional neural networks

To identify the content of an image, we usually do not scan the entire image at once, but we look for the presence of known features. Indeed, in image recognition, the spatial correlation between pixels is more important than their values. This idea lies behind the concept of convolutional neural networks (CNNs), which were introduced by Lecun and Bengio (1995). Unlike fully connected neural networks, CNNs can account for spatial structures in the data.

Convolutional layers are the most important building blocks of CNNs as they act as feature detectors. The neurons of such layers are arranged along three dimensions: width, height, and depth. The width and height correspond to the dimensions of the images in the first layer or the dimension of the previous feature maps for deeper layers. The depth matches the number of channels of the images (i.e., in astronomy, the number of bands). In convolutional layers, neurons are only connected to a small number of neurons of the previous layer by contrast to a fully connected layer where they would receive inputs from all the neurons in the previous layer. If the convolutional layer receives an image as input, it will treat a small area of the image separately. We call the sets of neurons connected to a particular area of the receptive image units. These units can extract spatial features such as edges, corners, or more complex patterns. In image classification, the features specific to a class must be detected independently of their position in the image (translation invariance). To implement this in convolutional layers, the receptive units tracing the same feature at different positions are forced to share the same biases and weights and span the entire image, hence the name. This process can be visualized as a convolution with a similar kernel or filter across the entire image. We display an example of this process in Fig. 3.9. The output of all the convolutions with the same kernel is called a feature map. It is important to emphasize that the filters spread along a defined width and height but always operate on the entire image depth.

The number of filters, their dimensions, and the stride (i.e., the distance the filter moves over the input image between two convolutions) are the hyper-parameters of CNNs. Therefore, they must be decided by the programmer. The kernels are learned during the training through the weights and biases of the receptive units. It allows CNNs to select the most appropriate filters for each task. In some cases, the dimension of the input image is not a multiple of the

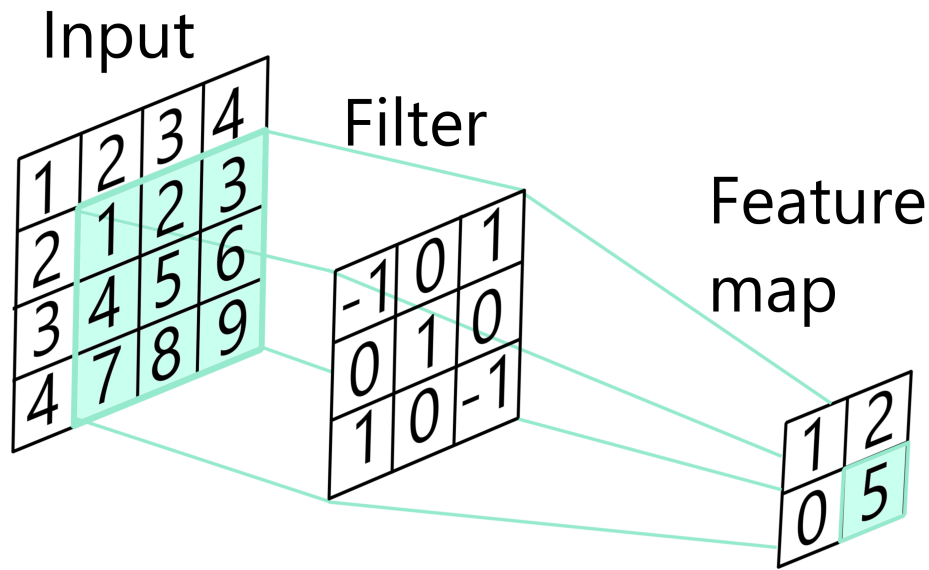


Figure 3.9: Illustration of the convolution operation taking place in convolutional layers. The behavior of the neurons belonging to convolutional layers can be represented by a convolution of the input image by filters that are learned during the training.

dimension of the filters. Thus we must decide what happens when the filter falls partially outside the input image. There exist two options of padding to expand the original image:

- Valid padding or no padding: In this case, the convolutions in which parts of the filters are located outside the image are not considered. This results in a feature map smaller than the input image.
- Same padding: The input image is expanded by adding pixels of a given value (usually zero) to the input image so that the last convolution does not fall outside the image. If the stride is equal to one, this produces a feature map with the same dimensions as the input image.

For a square kernel, the width of the feature map can be calculated as follows:

$$\frac{W_1 - F + 2P}{S}, \quad (3.19)$$

where W_1 is the width of the input image, F the filter size, P the amount of padding, and S the stride. Similarly, the height of the feature map is expressed:

$$\frac{H_1 - F + 2P}{S}, \quad (3.20)$$

where H_1 is the height of the input image.

A minor drawback of convolutional layers is that the result of the convolution with the filter depends strongly on the position of the features with respect to the filter. The results can be very different for small shifts in the original image. Adding pooling layers after convolutional layers is common to reduce this effect. Like convolutional layers, pooling layers can be represented by a filter that scans all locations of the input image according to its stride and returns a single output number for each time the filter stops. However, pooling layers have no parameters since the filters are not learnable. Instead, they return the result of a predefined operation, for example, the maximum pixel value over the region of the input image spanned by the filter (maximum pooling). Pooling layers have the additional advantage of reducing the dimension of the learned feature maps, decreasing the number of parameters and, consequently, the computational cost.

Deep convolutional neural networks were designed to capture more abstract features and potentially more efficiently classify complex images. Indeed, when CNNs are used to classify images, the first convolutional layer generally learns to detect edges, the second layer the textures, and the third layer more complex features and similar for the last layers (Mallat, 2016). However, in practice, there is a limited number of layers above which the performance of the CNNs starts to drop compared to shallower architectures. This limitation was overcome with residual networks (ResNet) (He et al., 2015).

The idea of the residual blocks, which are the basic building blocks of ResNet, stemmed from the fact that there exists a theoretical solution allowing a deeper network to reach at least the same accuracy as shallower networks. Indeed, if we consider a shallow CNN and add layers to it to obtain a deeper CNN, then the added layers could learn the identity mapping, and the original layers could take the same values as in the shallower CNN. The performance of the two networks would therefore be the same. However, as the weights of convolutional layers are usually initialized with random values or zeros, finding the identity function is as challenging to learn as any other function. To solve this, residual blocks include a new type of connection, the shortcut connections or skip connections, that allow them to learn the difference between the input and the output instead of the whole mapping. More precisely in convolutional block the output takes the form $H(x) = f(wx + b)$, where x is the input, w the weights and b the biases. In a residual block, the output is changed to $H(x) = f(wx + b) + x$. By using this type of connection, the network can "skip" the training of the useless layers. In addition, it avoids the vanishing gradient problem that can happen with deep CNNs. We present in Fig. 3.10 a scheme of a residual block. Residual networks are built by stacking residual blocks. The ImageNet Large Scale Visual Recognition Challenge 2015 (Russakovsky et al., 2015) was won by a Resnet. Since then, they have been widely used in image classification.

3.2.3 Auto-encoders

Neural networks are often used for classification or regression tasks. That is to say, they determine a mapping $y = f(x)$ between the input X and the labels y . Auto-encoders are

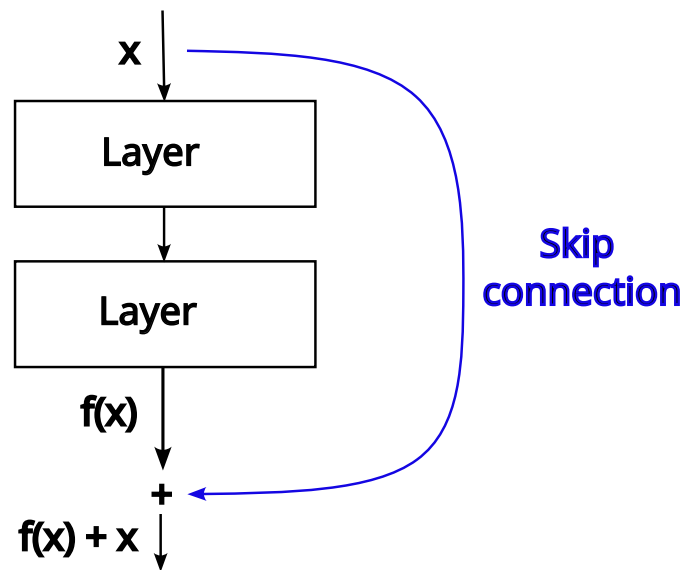


Figure 3.10: Illustration of a residual block. The skip connection, which is the particularity of ResNets, is depicted in blue. Credit: adapted from He et al. (2015) .

noteworthy in this respect because the inputs serve as labels, i.e., the mapping inferred by the auto-encoder takes the form $X = f(X)$. The interest in such mapping lies in the ability of the hidden layers to learn meaningful representations of the data. Thus, auto-encoders can be used for many applications such as dimensionality reduction and denoising (e.g. Vincent et al., 2010; Vincent et al., 2008). Auto-encoder-derived architectures such as variational auto-encoders (VAEs) may be used as generative models (e.g. D. Kingma & Welling, 2014).

We show a simple example of an auto-encoder architecture in Fig. 3.11. We can separate the architecture in two parts: the encoder and the decoder. The decoder is symmetrical to the encoder, meaning they have the same number of layers and neurons per layer. The central layer, called the bottleneck layer, connects the two parts of the autoencoder. In general, the number of neurons decreases from the input and output layers to the bottleneck layer. This peculiar architecture allows the network to learn a lower-dimensional representation of the original input.

In the context of denoising and dimensionality reduction, auto-encoders can be understood as a generalization of PCA (Hinton & Salakhutdinov, 2006). An auto-encoder with only one hidden layer of dimension p and linear activation is closely related to PCA. When trained with a mean-square cost function, they find the same optimal feature subspace as the PCA for p principal components (Bourlard & Kamp, 1988). However, unlike the eigenvectors of the PCA, the basis vectors of the optimal subspace found by the auto-encoder, may be correlated (not orthogonal) and are not sorted as a function of the explained variance. With non-linear activations, auto-encoders can encode the essential information in fewer dimensions than the PCA (Hinton & Salakhutdinov, 2006).

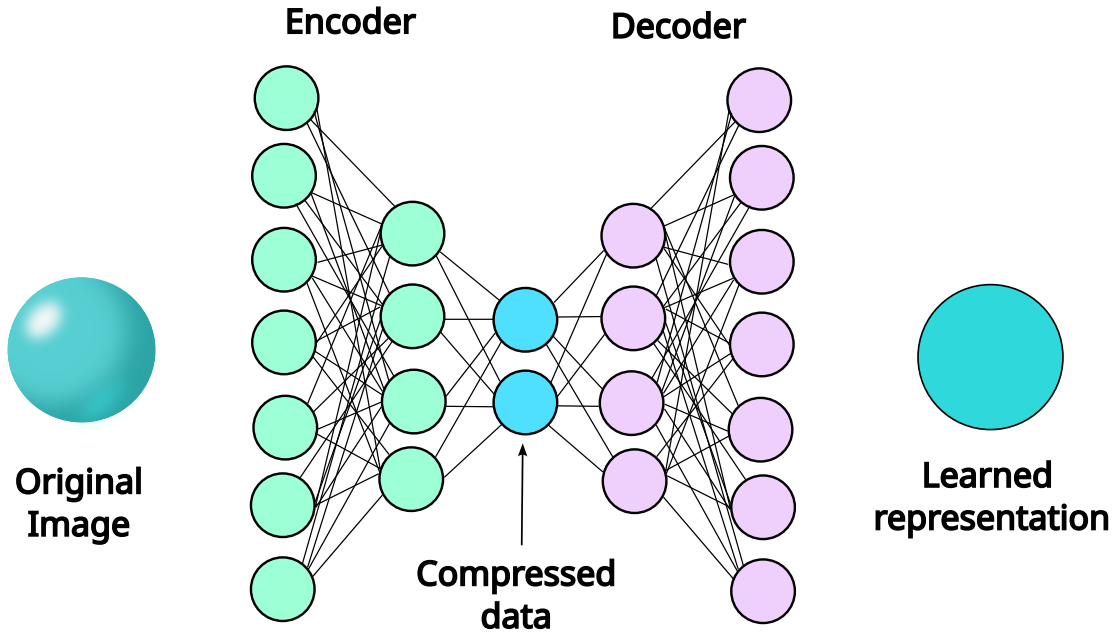


Figure 3.11: Sketch of an autoencoder. The architecture contains a bottleneck in its center that allows the network to learn a simplified representation of the original image

To highlight the advantage of the non-linear activation of auto-encoders, we present in Fig. 3.12 the results obtained on the MNIST dataset using an auto-encoder with one hidden layer only and sigmoid activation. We present the results for hidden layers of different dimensions. We can compare these results to those in Fig. 3.6, which displays the reconstruction of the same MNIST images with PCA. We see that the reconstruction with an auto-encoder is more accurate for an optimal subspace with an equal number of dimensions. The numbers are recognizable with a 20D optimal subspace in the case of the auto-encoder. In contrast, at least 60 principal components are necessary for the PCA reconstruction to reach an equivalent result.

In astrophysics, auto-encoders have found use in numerous applications, e.g., spectral energy distribution denoising (Frontera-Pons et al., 2017), galaxy deblending (Arcelin et al., 2021), gravitational wave detection (Moreno et al., 2022; Morawski et al., 2021), or galaxy image generation (e.g. Lanusse et al., 2021).

In Sect. 7, we present a new application of auto-encoder in the field of gravitational lensing. We use a modified auto-encoder architecture to separate the source and the lens light in images of gravitational lenses.

3.3 Performance metrics

During training, performance is approximated by the loss function that the algorithm aims to optimize. Although important for optimization, the loss function may not be the optimal

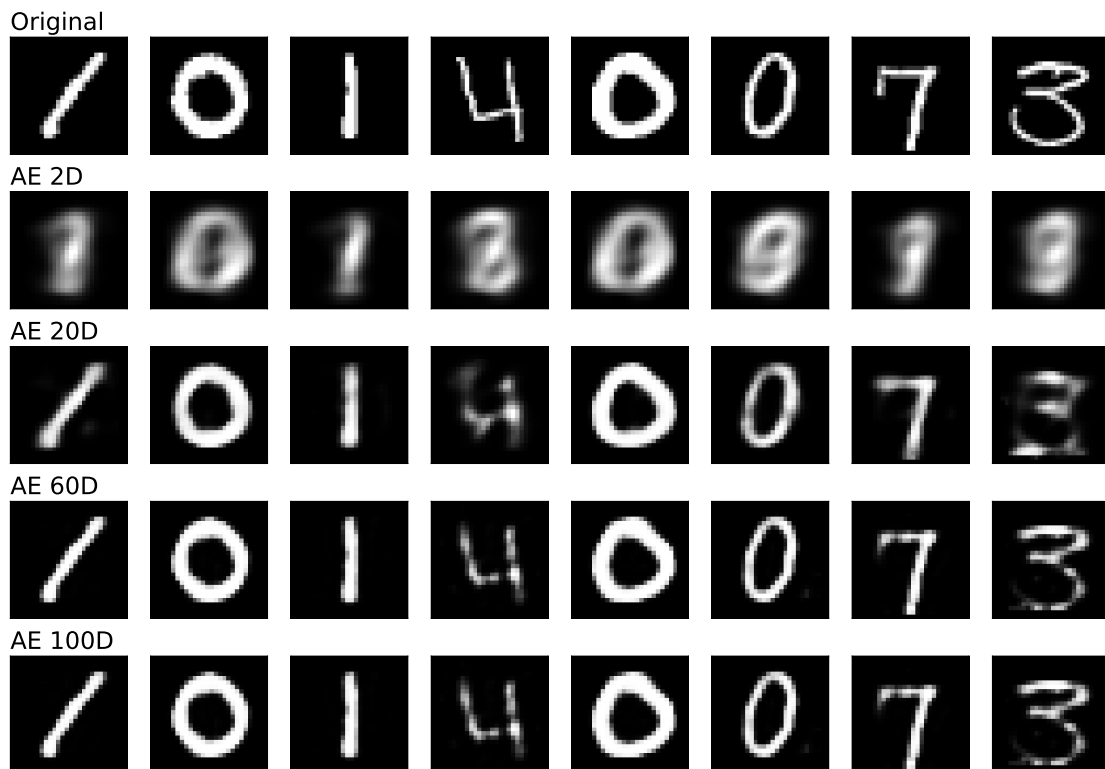


Figure 3.12: Reconstruction of MNIST images using an auto-encoder with one hidden layer and sigmoid activation in the output layer. The first row displays the original images and the following rows the reconstruction using a hidden layer with 2, 20, 60 and 100 neurons, respectively.

metric to evaluate the quality of the model. Therefore, defining meaningful metrics for the tasks we want to perform is crucial to evaluate the algorithm's performance correctly.

An important point that has already been mentioned is that once the training is completed, the performance must always be evaluated on the test set and not on the training or validation sets. This way, we can better understand the generalization error and avoid overestimating performance due to overfitting.

We present some metrics used in this work to evaluate our different algorithms.

3.3.1 Binary classification

Binary classification allows us to divide the data into two different classes: the positive examples (usually written as "1") and the negative examples (usually written as "0"). Therefore, the data classified by the algorithm can be in four situations. If the algorithm classifies them correctly, they are either true positives (TP) or true negatives (TN), depending on their original

		Ground truth	
		Positive	Negative
Prediction	Positive	TP	FP
	Negative	FN	TN

Figure 3.13: Confusion matrix representing the four possible outcomes of a binary classification. The situations where the algorithm predictions are correct are represented in green, whereas the incorrect predictions are shown in red.

class. Similarly, if they are classified incorrectly, they are false positives (FP) or false negatives (FN). We summarize the different situations under the form of a confusion matrix in Fig. 3.13

The algorithm's output can be either binary or a score between zero and one, reflecting the probability of belonging to the positive class. In the latter case, the final classification depends on a threshold score, chosen by the user, that defines the value above which the data is considered to belong to the positive class. If we choose a higher threshold, we generally get fewer false positives, but we also recover fewer true positives. This is reflected in Precision (P) and Recall (R) metrics. Precision, also referred as Purity, measures the proportion of true positive among all objects labeled as positive, while Recall, also called "true positive rate" or "completeness," gives the fraction of true positive that the algorithm identified among all true positive objects in the training set. They are defined as follows:

$$P = \frac{TP}{TP + FP}, \quad (3.21)$$

$$R = \frac{TP}{TP + FN}, \quad (3.22)$$

These two metrics can be combined in one curve, named the Precision-Recall curve (PR-curve), which displays the value of the Precision and Recall for all score thresholds. An example is shown in the upper panel of Fig. 3.14. For a random classifier, the precision is about 0.5 regardless of the score threshold, so such an algorithm produces the red line visible in Fig. 3.14. In contrast, a perfect algorithm produces a PR-curve that touches the upper right corner of the graph. Using this curve, the optimal threshold for the score can be estimated, depending on the limit we set for precision and recall. In the field of lens finding, all positive examples found by the algorithm must be confirmed by visual inspection. Therefore, we generally put more emphasis on precision, since low precision would mean that many false-positive examples would have to be re-inspected, which can be extremely time-consuming. The area under the PR-curve (AUC-PR) gives an indication of performance and allows us to compare different algorithms. AUC-PR ranges from 0.5 to 1, 1 for a perfect classifier and 0.5 for a random classifier.

A similar curve is widely used in binary classification, the Receiver Operating Characteristic (ROC) curve, which plots the true positive rate ($TPR = \frac{TP}{TP+FP}$) versus the false positive rate for all score thresholds ($FPR = \frac{FP}{FP+TN}$). We show an example in the bottom panel of Fig.3.14. Note that recall and true-positive rate are the same. However, since the term false-positive rate is often used in the context of the ROC curve and recall when we talk about the PR-curve, we keep the two denominations. Similar to the PR-curve, the area under the ROC curve (AUC-ROC) can also be used as a performance measure.

The ROC curve and PR-curve are closely related. If a classifier surpasses another in ROC space, it will also be better in PR space (Davis & Goadrich, 2006). However, the PR-curve is sensitive to the imbalance between classes, whereas the ROC curve produces similar results independently of the imbalance ratio. When the imbalance between classes is very high, the ROC curve can give a too optimistic result (Drummond & Holte, 2004). On the other hand, the PR-curve can highlight better the difference between classifiers when the data are highly imbalanced since the precision is affected by the change in class ratio. The PR-curve must, however always be interpreted with the knowledge of the class ratio for the same reason (Brabec et al., 2020). It is also important to highlight that an algorithm that obtains the best performance in terms of AUC-ROC will not necessarily be the best in terms of AUC-PR, especially with highly imbalanced test-sets (Davis & Goadrich, 2006).

Another metric that is of interest when one wants to prioritize precision over recall (or conversely) is the F-beta score. It is defined as follows:

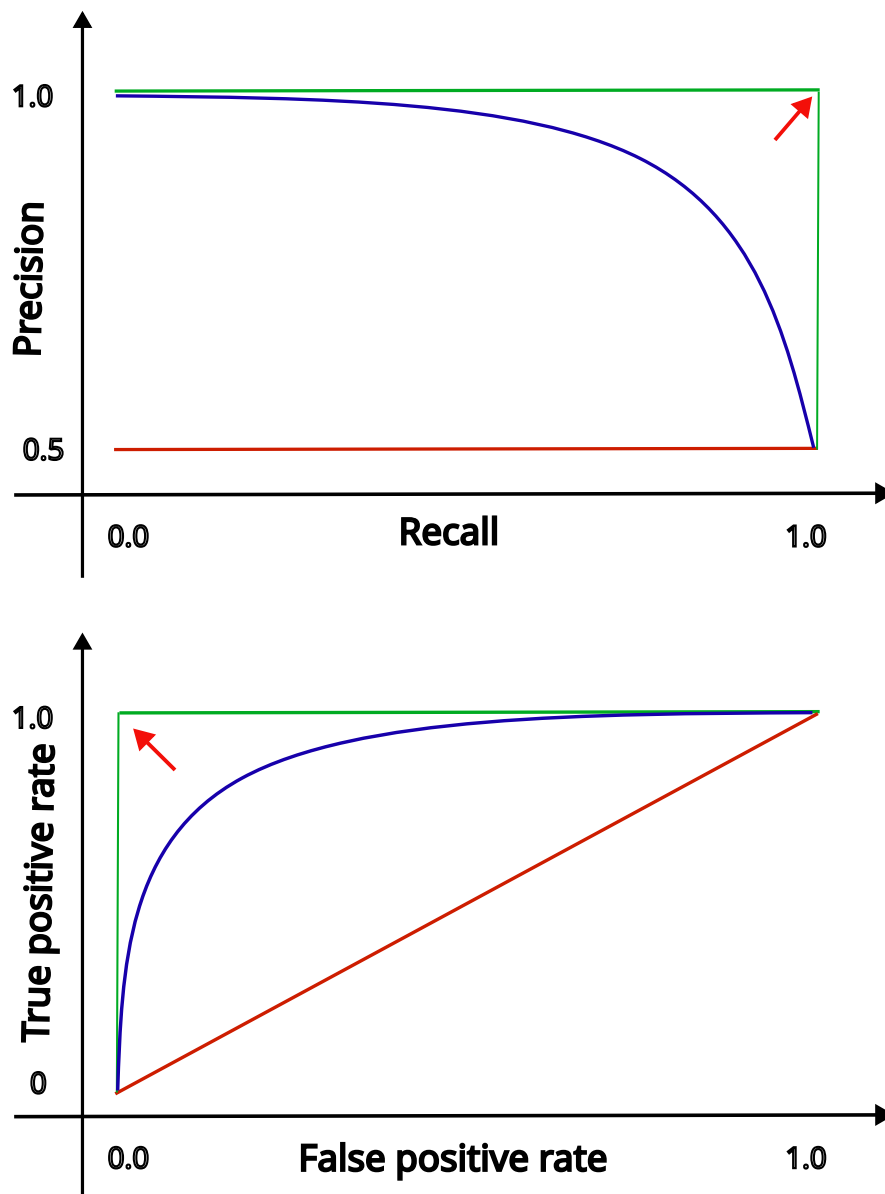


Figure 3.14: Comparison of precision-recall curve and ROC curve *Upper panel:* Illustration of a precision-recall curve. The precision and recall are given for each score threshold. A perfect classifier produces the green curve, while the red curve corresponds to a random classification. The precision-recall curve of an arbitrary classifier (blue curve) lies between these two curves. The closer the curve is to the upper right upper corner (red arrow), the better the classification performance. *Lower panel:* Illustration of a ROC curve. The diagonal (in red) is produced by a random classifier and the green curve by a perfect classifier. The closer the ROC curve of the algorithm (blue curve) is to the top left corner, the better the performance

$$F_\beta = \frac{(1 + \beta^2)}{\beta^2 P + R}, \quad (3.23)$$

where R is the recall and P is the precision. For $\beta < 1$, more emphasis is put on precision. On the contrary, a $\beta > 1$ will give more importance to the recall. For $\beta = 1$, the F-beta score is equivalent to the so-call F-1 score, which returns the harmonic mean of precision and recall. Since the F-beta score depends on the chosen score threshold, we usually indicate the $\max F_\beta$, which is the maximum F-beta score reached on all possible score thresholds.

3.3.2 Image Generation

Evaluating the quality of images generated with an algorithm is not as straightforward as for binary classification or regression tasks. In the context of the generation of astronomical images, the physical soundness of the generated images is the most critical point. The Concentration, Asymmetry and Smoothness (CAS) statistics (Lotz et al., 2004; Conselice, 2003; Bershadsky et al., 2000) are commonly used for morphological classification (e.g. Lotz et al., 2004; Conselice, 2003; Bershadsky et al., 2000) and can be adapted to evaluate sets of simulations of lens. There exist slightly different definitions of the CAS parameters in the literature. In this work, we use the definitions proposed in Lotz et al. (2004) , which we summarize here.

The concentration index (C) measures how "compact" the galaxies are. For example, elliptical galaxies have a higher concentration index than spiral galaxies. It is defined as follows:

$$C = 5 \log \frac{r_{80}}{r_{20}}, \quad (3.24)$$

where r_{80} and r_{20} are the circular radii containing 20% and 80% of the total flux, respectively.

The rotational symmetry of a galaxy's light is characterized by the asymmetry parameter (A). It is calculated in the following way:

$$A = \frac{\sum_{i,j} |I(i,j) - I_{180}(i,j)|}{\sum_{i,j} |I(i,j)|} - B_{180}, \quad (3.25)$$

where I is the original galaxy image, I_{180} the image rotated by 180 degrees, and B_{180} the average asymmetry measured in the background. This last term is introduced to correct the asymmetry caused by the noise. It must be noted that high levels of noise prevent the correct measurement of the asymmetry.

The smoothness parameter (S) quantifies the degree of small-scale structures present in the images. It can be computed as follows:

$$S = \frac{\sum_{i,j} |I(i,j) - I_S(i,j)|}{\sum_{i,j} |I(i,j)|} - B_S, \quad (3.26)$$

where I_S is the image smoothed by a boxcar of width 0.25 times the Petrosian radius of the galaxy, and B_S corresponds to the average smoothness of the background.

4 Finding galaxy-galaxy lenses in imaging data

In the past, images of strong lenses were found essentially by visual inspection by a small group of experts. To reduce the number of candidates to classify, color and magnitude cuts at the catalog level are usually applied. Typical examples of this type of searches are Faure et al. (2008) , Jackson (2008) , Sygnet et al. (2010) , Pawase et al. (2014) . However, future imaging surveys involve such a large amount of data that the visual inspection by experts becomes very cumbersome, even with color and magnitude cuts.

Future and current large-scale imaging surveys are expected to provide large samples of observable strong lenses. To give an order of magnitude, Collett (2015) predicts 400, 12 0000, and 17 0000 observable galaxies lensed by galaxies for DES, LSST, and Euclid, respectively. The difference in the number of expected lenses in the different surveys is mainly due to the difference in spatial resolution, which is key to resolving systems with smaller separation and Einstein radii. This will be particularly the case with Euclid since it is a space-based telescope. Such a large number of lenses forecast to be detected raises the need to design efficient methods to detect strong lenses in imaging data. Visual inspection is, however, still used to validate potential candidates found with other methods.

This chapter reviews the different methods available to search for galaxy-galaxy lenses in imaging surveys. Although this work focuses on CNN-based algorithms, it must be kept in mind that the different methods are not mutually exclusive and can be used simultaneously in future lens searches. Indeed, each method finds different types of candidates, and there is, therefore, a great scientific interest in combining them on the same datasets.

4.1 Citizen science

Citizen science relies on a large number of non-experts volunteers called citizen scientists. They are provided with the essential guidelines to perform the classifications and a visual summary with examples of objects belonging to each category. Because many participants inspect each image, the classification of each object is made more robust by aggregating their

decisions. This approach has been used successfully in different fields of astrophysics, for example, in the morphological classification of galaxy (e.g. Willett et al., 2017) or discovery of transiting exoplanets (e.g. Schwamb et al., 2012).

Citizen science has been applied for the first time to gravitational lens search in the context of the SPACE WARPS project (Marshall et al., 2016; More et al., 2016). This first search involved around 37 000 volunteers who inspected 430 000 images from the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS). They obtained a total of 49 lens candidates. The Space Warps project was then extended with a lens search within Hyper Suprime-Cam (HSC) images which resulted in around 1500 candidates, among which 143 have been classified as "very likely candidates" (Sonnenfeld et al., 2020).

Strong lensing can produce a large variety of lensing configurations, making it challenging to summarize the properties of a good lens candidate in a concise set of textual descriptions or visual examples. Some astrophysical objects, such as ring or spiral galaxies, may fulfill most of the criteria described in the guidelines provided to citizen scientists and are especially difficult to distinguish from lenses for a non-expert eye. These points are critical for applying citizen science to the lens finding search. To overcome these difficulties, the method described in Marshall et al. (2016) proposes a two-step training consisting first of a small tutorial where classification guidelines are given, and various possible lens configurations are shown as examples to the citizen scientist. The second step of training is done during the classification. Real lens images, non-lenses, and lens simulations validated by experts (training examples) are added to the set of objects to classify and displayed randomly during the classification to the citizen scientists. When a training example is displayed and classified, the participants receive immediate feedback, and their performance is stored. The advantage of this "supervised" approach is that the combined weighted scores of all participants obtained on the training examples can be used to weigh the contribution of each particular user. They combine the weighted scores of all participants with a Bayesian pipeline to obtain the probability for an image to be a lens. In addition, at the end of the first classification, some participants are invited to reinspect the candidates in order to lower the false positive rate.

One advantage of citizen science is that humans can easily understand how to extend the guideline to classify unseen lensing configurations. It enables the discovery of more exotic types of lenses than algorithms (More et al., 2016). One good example is the gravitationally lensed hyperluminous infrared galaxy found in the context of the SPACE WARPS project (Geach et al., 2015), which displays unusual bright red arcs. Citizen science requires finding a large number of volunteers to inspect the images. The limited pool of persons interested in participating in astronomical experiments might thus be the critical limitation to the generalization of lens searches (or any other astronomy-related project) with citizen science.

4.2 Automated methods

Citizen science is very efficient to detect strong lenses in images. The recall of known lenses can reach values up to 65 % with non-expert classifier with this method (More et al., 2016). However, it requires a large number of volunteers. It is therefore not possible to use this method for all searches. For this reason, automated methods have to be developed in parallel.

4.2.1 Early methods

The early attempts at automated detection of extended lensing features, which are produced by galaxies acting as sources, consisted of clustering algorithms using pixel intensities coupled with shape measurements in order to detect the presence of elongated and curved objects (e.g. More et al., 2012; Cabanac et al., 2007; Estrada et al., 2007; Seidel & Bartelmann, 2007; Lenzen et al., 2004). In Maturi et al. (2014) the combination of arc detection with color information leads to an increase in the completeness of the sample of candidates and a reduction of false positives compared with simple arc detection in single-band images.

The lensed source features are often hidden behind the light of the lens. In such a case, the detection with arc-detection-based algorithms only is almost impossible. For this reason, methods were developed to remove the lens light and detect possible lensing features in the residuals. A good example is the RINGFINDER tool (Gavazzi et al., 2014), which detects the presence of blue residuals hidden behind early-type galaxies using difference imaging. The method was applied to CFHTLS two-band images (i,g) and returned 2 500 candidates, among which 330 were labeled as possible lenses after a visual inspection. With the same idea of subtracting the foreground galaxy light, Joseph et al. (2014) presents a method based on PCA to subtract the lens light and two algorithms to detect evidence of lensing in the residuals. The main advantage of this approach is that, unlike Gavazzi et al. (2014), it does not rely on models for the lens subtraction and thus is more adaptable to more diverse populations of foreground objects. The idea of PCA lens subtraction was then reused in the lens finding algorithm proposed in Paraficz et al. (2016) and applied to CFHTLS data. They found a sample of 1098 candidates, which was reduced to 109 after visual inspection.

Interestingly, the candidates of More et al. (2012), Gavazzi et al. (2014), Maturi et al. (2014) and Paraficz et al. (2016) do not overlap entirely. This difference can be explained partially by the different preselection functions but also highlights the fact that each type of algorithm is specialized in identifying limited categories of lenses. Indeed, all of these algorithms were rule-based, i.e., the criteria for being a lens must be hard-coded by the user, for example, displaying blue residual or arc-like features. This lack of flexibility suggested the need for a new type of detection method, namely machine learning-based algorithms. Unlike rule-based methods, machine learning algorithms directly infer essential features from the data. They open up the possibility of finding a wider variety of lenses.

Different machine learning methods have been applied to the detection of strong lenses

in images such as support vector machines (SVM) (e.g. Hartley et al., 2017) or Gaussian Mixture Models (GMMs) (e.g. Ostrovski et al., 2017). The recent improvement of graphics processing units (GPU) capacities enabled further astrophysics application of neural networks. In particular, CNNs stand out in the field of strong lens search in imaging surveys. One big advantage of CNN-based detection methods or, more generally, machine learning methods is their flexibility. Since they do not require hard-coded rules, they can easily be adapted to any type of lens, provided we have an adequate training set. We present them in the next section and other neural networks.

4.2.2 Neural networks based methods

We can understand strong gravitational lens search as a binary classification problem, where the two classes are the lenses and the non-lenses. The main issue faced when using supervised machine learning methods to find lenses is the extremely low number of known lenses. Indeed, neural networks need, in general, large training sets, and enough representative examples of each class must be provided to span the entire diversity of possible lens features. For this reason, lens examples must be simulated to constitute labeled training sets. Since this point is crucial, we dedicate Chap. 5 to lens simulations. Recently, unsupervised learning was introduced in the field of lens finding (Cheng et al., 2020). It represents a promising alternative to supervised learning since it does not require labeled data. We summarize here the different applications of the two methods.

Supervised

CNNs are especially suited to find lenses in imaging surveys as they are known to perform well in image classification (e.g. He et al., 2015). One of the first applications of CNNs to lens finding was Petrillo et al. (2017). They applied a CNN architecture to a selection of 21 789 luminous r-band images of red galaxies (LRGs) taken from 255 square degrees of the Kilo-Degree Survey survey (KIDS). To train the network, they used a training set composed of 50 percent of lens simulations and 50 percent of non-lens examples that are either LRGs, false positives obtained from previous searches, or contaminants identified by visual inspection. The network retrieved 761 candidates among which 56 were classified as possible lenses after visual inspection. The visual inspection is used to evaluate the performance of the network, in particular the false positive rate. One important result of Petrillo et al. (2017) is the nature of the false positives, which are essentially ring galaxies, spiral galaxies, or mergers, e.g., types of objects that have features that look like lensing features. Ring galaxies are very difficult to differentiate from lenses since they consist in a luminous core surrounded by a disk that contains young blue stars. Therefore they mimic the shape and color of lensing features. This search was then extended by retraining the network with more complete training sets (R. Li et al., 2021; R. Li et al., 2020; C. E. Petrillo et al., 2019; Petrillo et al., 2018).

Similar efforts have been applied to CFHTLS data by Jacobs et al. (2017) who used a CNN-

based method on three-bands images (i,r,g). They performed two different searches: the first one on a subsample of LRGs selected by color-magnitude cuts at the catalog level similarly as Petrillo et al. (2017) and the second search on all the 171 square degrees of CFHTLS without preselection. The main difference with Petrillo et al. (2017) is that they used a committee of four CNNs, allowing them to obtain a lower false positive rate in the catalog-based search. Unsurprisingly, the search on all CFHTLS led to a larger false positive rate than the catalog-based search and did not enable the discovery of more exotic types of lenses. This highlights the importance of a representative training set since the CNNs recover mainly the lenses similar to those in the training set.

Jacobs et al. (2019a) conducted a search targeting high-redshift strong lenses in DES, using a similar method to Jacobs et al. (2017). This search was then extended to a wider range of redshifts and color-magnitude cut in Jacobs et al. (n.d.) and led to a total of 511 lens candidates. Recently, I contributed to applying a lens finding pipeline using CNN to the first data release of DES (Rojas et al., 2021). Our results are presented in Chap. 8.

Similarly, Cañameras et al. (2020) discovered 330 wide separation galaxy-galaxy lens candidates in the Pan-STARRS 3π survey with a CNN-based approach. In VST Optical Imaging of the CDFS and ES1 fields (VOICE survey), Gentile et al. (2021) trained two CNNs: one on one band data and the second done on three-band data. We conducted the first lens search in the initial 2 500 square degrees of CFIS using a committee of CNN (Savary et al., 2021), our results are summarized in Chap. 8.

Overall, the number of lens candidates was increased by a factor of 1000 thanks to the aforementioned studies, proving that CNNs are a promising tool for future large-scale surveys. In view of these future surveys, their efficiency has been tested on simulated data. For example, this has been done in Lanusse et al. (2018), where they applied an improved architecture to LSST g-band mock images. They achieved completeness of 90 percent for lenses with an Einstein radius larger than 1.4 arcsec and with signal to noise ratio larger than 20 and a rejection rate of 99 percent for the non-lens simulations. They however highlighted the fact that their simulations are not complex enough to evaluate the improvement of this architecture in comparison to more traditional CNNs as in Petrillo et al. (2017). Similarly, CNNs have been tested on Euclid simulations in the context of the gravitational lens finding challenge presented in Sect. 4.3.

Although the number of known lenses is generally thought to be too small to train CNNs, Huang et al. (2020) and Huang et al. (2021) have proven the possibility of using only a limited number of lenses to train CNN in the Dark Energy Spectroscopic Instrument Legacy Imaging Surveys (DECaLS) Data Release 7 and the Dark Energy Spectroscopic Instrument (DESI) Legacy Surveys Data Release 8, respectively. The positive examples of their training set consist only of known lenses taken from the Master Lens Database (Moustakas, 2012). They were able to find 335 and 1210 new lens candidates, respectively. In the context of future large-scale surveys such as Euclid, this opens the possibility to reuse the lens candidates found in the first

iterations of the lens searches to train the future classifiers using only real data or a reduced amount of simulated data. This may limit the biases induced by simulations.

The studies mentioned earlier targeted essentially LRG acting as lenses. However, CNNs are especially suited to search for more exotic deflectors such as edge-on galaxies or quasi-stellar objects acting as lenses because of their flexibility. They, however, require to be retrained with specific training sets.

Unsupervised

The main drawback of CNNs used in a supervised setting is that they struggle to find the more rare type of lenses that are hardly represented in the training set. They also require meticulous manual labeling of all images used to train them. Unsupervised learning algorithms, which do not require labeled data, may represent an interesting alternative. Since the algorithm deduces how to separate the images directly from the data, it is also less prone to suffer from human biases.

The first example of an unsupervised method applied to lens finding is Cheng et al. (2020). They use a combination of a convolutional auto-encoder, which acts as a feature extractor, and a Bayesian Gaussian mixture model, which classifies the data as a function of the features extracted by the auto-encoder. They applied this method to the space-based images provided by the strong gravitational lens finding challenge (Metcalf et al., 2019) and were able to reach an accuracy of 0.77 and an AUC of 0.86 with a probability threshold of 0.5. Although the performance is slightly lower than the supervised CNN's entries of the challenge, this method shows a remarkable ability to pick up different features characteristic of lenses. Moreover, each cluster created by the Bayesian Gaussian mixture model seems to contain a different type of feature, which suggest that in addition to the classification, it is possible to obtain information on the similarity between the candidates.

With the same idea of decreasing the need for labeled data, Stein et al. (2021) applied a self-supervised method to find lenses in the Data Release 9 of the DESI Legacy Survey. Self-supervised learning offers a balanced approach between supervised and unsupervised learning as it uses only a limited number of labeled data (e.g. Caron et al., 2021). They first trained a CNN encoder to learn lower dimensional representations of 43 million images without labels. Using a contrastive loss function, they force the network to encode the similarity between images in the lower dimensional representation. They then used three different methods to find lenses. The first is a similarity search, where they obtained lens candidates similar to the set of known lenses using a similarity measure on the lower dimensional representations of the images. The main advantage of this approach is that it does not require labeled data apart from the set of known lenses. The second and third methods compared a linear classifier applied to the low dimensional features learned by the CNN encoder and the pre-trained CNN encoder fine-tuned specifically for the classification task. Since these two methods are supervised, they require more labeled data. However, since both use the features learned

during the first phase, they require less labeled data than traditional CNN- based methods. Using all these three combined methods, they could identify 1 192 lens candidates on 18 000 DESI legacy survey images.

4.3 Gravitational lens finding challenge 1.0 and 2.0

With the idea of comparing the different lens finding automated methods, the Bologna Lens Factory group launched two strong gravitational lens finding challenges (Metcalf et al., 2019), (Metcalf, in. prep). The Bologna Lens Factory project is led by members of the Strong Lensing Working Group of the Euclid Consortium and aims at gathering a large and diverse set of strong lens simulations that can be used by the scientific community for any kind of application. The first challenge tested the performance of algorithms on both ground-based and space-based simulations, whereas the second challenge was focused on Euclid-like data. We summarize here the results of the first challenge and present our submission to the second challenge.

4.3.1 Challenge 1.0

The first challenge was split into two categories: ground-based with four-band KIDS simulations (u,g,r, i) and space-based with one-band Euclid simulations (VIS). The lens simulations of the ground-based set were obtained with two different methods. Most ground-based simulations (85 percent) are entirely synthetic images where both the source and the lens have analytical profiles. The remaining 15 percent are composite images in which lensing features are added to a real image of LRG of the KIDS survey. The participants in the challenge used various methods such as visual inspection, arc-finder, machine learning methods making inferences from pre-selected features, and CNNs.

The most important result of the challenge is that CNN-based methods outperformed all other entries of the challenges for ground-based and space-based data in terms of AUROC. They also represent the majority of the entries in the top five in terms of TPR0, defined as the highest TPR reached before a non-lens image of the test set is classified as a lens (false positive). Interestingly, the group who participated in the challenge using a visual inspection method ranked sixth in terms of the TPR0 for the space-based and ground-based challenges. It highlights the fact that even classification done by experts can lead to false positives and that automated methods have the potential to outperform experts when the training set has similar properties to the testing set. However, visual inspection was the only one that identified the jackpot-lens, i.e., a lens with two lensed sources, present in the ground-based challenge.

All methods, including human classification, performed worse on composite simulation images than on synthetic simulations. This illustrates the importance of using realistic simulations to evaluate the classifier performance and training machine learning-based algorithms. Indeed, machine learning algorithms can struggle to generalize from simulated data to real data, leading to a drop in performance.

A question raised by this challenge is whether the color information can improve the performance of classifiers. All the methods obtained better AUROC scores in the ground-based part of the challenge. It was hypothesized that the color might help considerably to distinguish lensing features from spiral arms, rings, or other features mimicking lensing. It was tested for CNNs on KIDS data in Petrillo et al. (2018), which trained two similar architectures with three-band images and one-band images. The network trained with color images has more trouble retrieving lenses with redder features and obtains a more significant number of false positives. They suggest that adding mock lenses with more diverse colors can improve the performance of the CNN on color images. However, this highlights the difficulty of using color images since the lens simulations must span all the possible lens configurations (shape of the lens features) and an extensive range of colors for the source. More recently, Gentile et al. (2021) compared the performance of the CNN architecture of Petrillo et al.; Petrillo et al. (2018, 2017) on one-band and three-band data from the VOICE survey. They observed that, on simulated data, the three-band CNN identifies more systems with smaller Einstein radii, whereas the single-band CNN obtains a higher accuracy, especially with larger Einstein radii systems. The one-band CNNs focus on the shape of the lensing features and thus are more efficient to discriminate with false positives such as spiral galaxies. Nevertheless, the multi-band CNNs are able to discover lenses with smaller Einstein radii with the help of the color difference between the lensing features and the deflector. This shows that if multi-band images are available, the performance can be improved by combining the results of CNNs trained with single-band and color images.

4.3.2 Challenge 2.0

Recently, the Bologna Lens Factory (BLF) group launched a second Lens Finding Challenge. The 2.0 version of the competition focuses only on Euclid simulations. The training set consists of 100,000 Euclid-like images. All images are provided in 4 different bands (VIS images of dimension (200,200) and the near-infrared (NIR) Y, J, and H bands of dimension (60,60)). The pixels sizes are 0.1 arcseconds for VIS, and 0.3 arcseconds for NIR bands. Along with the imaging data, a catalog summarizing the properties of the images is provided. We summarize here the entries of the catalog we used for our submission:

- x_{crit}, y_{crit} : coordinates of the center of the critical curve
- z_{source}, z_{lens} : redshifts of the source and the deflector, respectively
- $source_ID$: ID number of the source
- mag_{source}, mag_{lens} : magnitude of the source and the deflector, respectively
- ein_area : Einstein area computed from the largest critical curve
- n_{pix_source} : pixels above 1σ in source images
- n_{source_im} : number of lensed sources

- mag_{eff} : effective magnification of the source

The lenses are defined in the challenge rules by the objects in which $n_{sources} > 0$, $mag_{eff} > 1.6$ and $n_{pix_source} > 20$. The non-lenses are the objects with $n_{sources} = 0$.

The metric used to evaluate the submission is the $\max F_\beta(s)$, which represents the maximum F_β measure for all score thresholds between 0 and 1. The value of β is very small (0.001). Therefore more emphasis is put on the precision than the completeness of the sample of candidates.

The main difficulty in the challenge is that when we apply the definition of lenses provided in the challenge rule, the simulation set contains 49 % of lenses, but many of them do not display visible lens features. In addition, the proportion of lenses in the test set used to evaluate the simulations is not specified and may differ from the proportion in the training set. Another particularity of the challenge is the fact that the non-lens represent only 10% of the set of simulations. Our lab participated in the challenge with six master students, Boris Bergsma, Nicolas Frank, Alexandre Di Piazza, Daniel Forero Sanchez, Andrei Variu, and Yaroslav Ilichenko, supervised by myself. We split the training set in two, keeping 30% of the images in a validation set that was used to monitor the performance of our networks. We adapted and tested three different architectures, a Residual Network (ResNet) inspired by He et al. (2015), the best model of Schaefer et al. (2018) (LASTROmodel), and a variation of Efficientnet architecture B0 (Tan & Le, 2019). The last architecture is a Resnet, in which the depth, width, and resolution are scaled uniformly to limit the usage of computational resources. To refine the classification, we added fully connected and dropout layers at the top of Efficientnet B0. Our modified Efficientnet architecture is presented in Fig. 4.1. The advantage of Efficientnet B0 is that it has been pre-trained on ImageNet data (J. Deng et al., 2009). Hence, we initialized Efficientnet layers of our network with the weights and biases of the pre-trained model. It enabled us to shorten the training by a factor of 4 compared to the two other networks, even though they have a comparable number of parameters.

For all architectures, we implemented a one-band version using only VIS images and a three-band version where the second and third bands correspond to a weighted average of all the NIR bands $0.15 Y + 0.35 J + 0.5 H$ and the H band alone, respectively.

All images were normalized between 0 and 1 before being passed to the CNNs. The CNNs were trained using mini-batch stochastic gradient descent with batches of 128 images flipped randomly along the x - and y -axes. Except for Efficientnet, which was initialized with pre-trained weights from the ImageNet, all models were initialized randomly following the method described in He et al. (2015).

After the training, we evaluated the performance of the CNNs on the validation set. The maximum $F_{0.001}$ value we could achieve with all three architectures was 0.7. We hypothesized that the low performance resulted from the large number of objects labeled as lenses in the training set but with no visible lensing features. Since the metric used to evaluate the challenge,

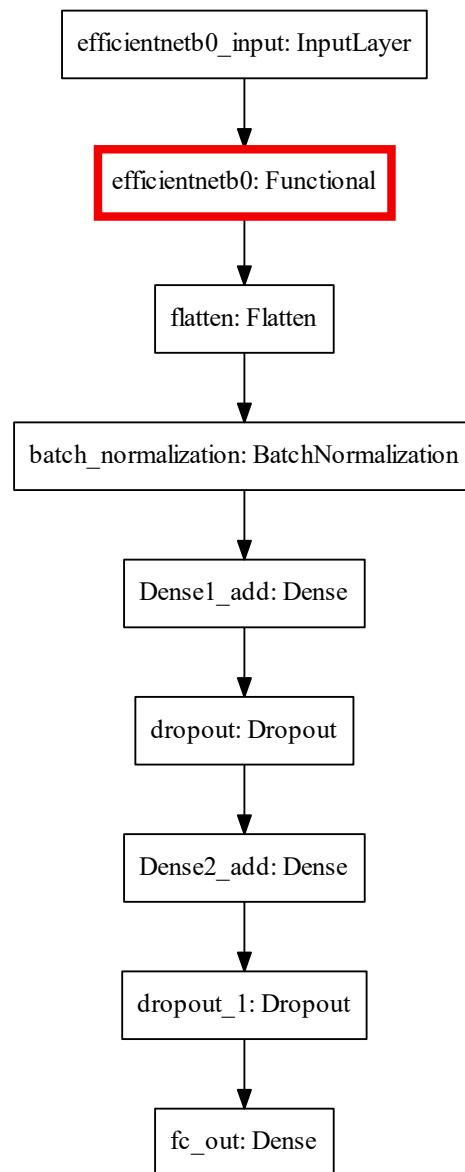


Figure 4.1: Architecture of our modified Efficientnet B0. The red rectangle corresponds to the Efficientnet B0 architecture of Tan and Le (2019) . The layers we added are indicated with black rectangles.

$F_{0.001}$, puts more emphasis on the precision than on the recall, we chose to drastically reduce the set of lenses we use for training by including only the ones with manifest lensing features. We first selected the lenses in which the number of pixels in the lensed source image above 1σ , $n_{\text{pix_source}}$, fulfill the following condition:

- $R_E < 2 \Rightarrow n_{\text{pix_source}} > 90$
- $2 < R_E < 2.5 \Rightarrow n_{\text{pix_source}} > 145$
- $2.5 < R_E < 3 \Rightarrow n_{\text{pix_source}} > 150$
- $3 < R_E < 3.5 \Rightarrow n_{\text{pix_source}} > 300$
- $3.5 < R_E < 4 \Rightarrow n_{\text{pix_source}} > 400$
- $4 < R_E \Rightarrow n_{\text{pix_source}} > 800$

where R_E is the Einstein radius of the lenses.

The reason to use a more restrictive limit for large Einstein radii is that, in this case, the lensed source features are generally more spread, making them more difficult to identify. We also imposed a limitation on the magnitude of the deflector mag_lens :

$$\text{mag_lens} > 20 \tag{4.1}$$

This enables us to remove the images where the lensing features are blended in the light of very bright deflectors. Using these criteria, we selected 12 562 lenses from the 49 218 original lens simulations. Examples of lenses accepted in our training set are shown in Fig. 4.2. As we can see, the lenses are not necessarily at the center of the image. We retrained all our networks with the subset of lenses we selected and the 10 000 original non-lenses images.

With this new training set and our new definition of lenses, we reached $F_{0.001}$ above 0.95 on the validation set with all our models. The results are summarized in the first row of Table 4.1. We found no improvements and, in fact, even a slight decrease in the performance of all our models when using the NIR images. This can be explained by the fact that due to the low resolution, the lensing features in NIR images are very often blended in the noise for NIR images and thus invisible to the networks. We submitted the two best models for the challenge and obtained an $F_{0.001}$ higher than 0.97 on the test set (using the original lens definition). The detailed results are presented in the second row of Table 4.1.

The winner of the challenge, Bom et al. (2022) used a deeper Efficientnet architecture (B2). However, the main difference with our submission lies in the preprocessing method they use to reduce the number of positive examples without visible lensing features, which consist of clipping and contrast enhancement.

Table 4.1: Results of our three architectures after training with the modified set of lenses on VIS band data. The results on the validation set are obtained using our definition of lenses. The test set results were obtained by the organizer on a separate test set that was not available for the participants during the challenge. The Resnet architecture displays no score on the test set since we submitted only our two best models.

Architecture	LASTROmodel	Resnet	Efficientnet B0
$\max F_{0.001}$ (validation set)	0.96	0.95	0.99
$\max F_{0.001}$ (test set)	0.97	-	0.98

The definitive ranking, as well as the detailed results, will be described in Metcalf et al. (in prep.). One interesting observation from the intermediate results is that even if we trained our network with a restricted number of lenses, we obtained a recall value comparable to the other groups who used the original training set. The essential question raised by this point is if a CNN trained on blatant lensing features would be able to generalize and also detect some of the less evident lenses. The detailed results of the challenge will partially answer this interrogation. In the meantime, our results show that too many invisible lensing features prevent the network from learning to identify even the most conspicuous lenses.

4.4 Limitations of CNN-based methods

As we have seen in the previous section, CNNs are a very efficient tool for finding lenses, and numerous new lenses have been found with the advent of this new method. However, using them in larger-scale surveys requires overcoming some challenges. We summarize here the main difficulties encountered during the previous searches and explore solutions.

4.4.1 Low occurrence rate of lenses

Gravitational lensing is a sporadic event, and the set of known lenses is still limited. Therefore, simulations of lenses have to be used in most cases when using a supervised learning approach. Nevertheless, this is not the only issue with the very low occurrence rate of lenses.

Indeed, the classifiers are often trained and tested with sets containing an unrealistic fraction of lenses (typically 50%). However, the probability for an image labeled as a positive example by the algorithm to be a real lens $P(L|C)$ depends on the occurrence rate of lenses $P(L)$ within the dataset investigated. It can be deduced from the Bayes' rule:

$$P(L|C) = \frac{P(L)P(C|L)}{P(L)P(C|L) + (1 - P(L))P(C|NL)}, \quad (4.2)$$

where $P(C|L)$ and $P(C|NL)$ correspond to the precision and the false discovery rate obtained on the testing set, respectively. This effect limits the precision we can reach on real data for a fixed false discovery rate. For a small $P(L)$, $P(L|C)$ will be small even when $P(C|L)$ is

high. Conversely, a fixed precision forces the false discovery rate to be in the same order of magnitude as the occurrence rate to make significant predictions. This mathematical phenomenon is named the "base rate fallacy" and is already known in the field of anomaly detection (e.g. Axelsson, 2000).

The prevalence of lenses is different depending on the survey. Indeed, as discussed in Chap.6, what we define as a lens depends mainly on the presence of visible lensing features. The detectability of lenses depends on parameters related to the telescopes and conditions of observations such as the point spread function, noise level, seeing, and sky-brightness and related to the lensing system such as the Einstein radius (Collett, 2015). The number of lenses visible in photometric surveys has already been estimated for point-source lenses (Oguri & Marshall, 2010) and extended sources lensed by LRG (Collett, 2015) with the help of simulations. However, the occurrence rate of lenses among a general population of galaxies is still unknown.

A solution to limit the importance of the base rate fallacy is to increase the prevalence of lenses in the sample with a preselection of objects. This selection generally takes the form of a color-magnitude-cut, for example, in Jacobs et al. (2017) , Petrillo et al. (2017) , Petrillo et al. (2018) , Jacobs et al. (2019a) , Cañameras et al. (2020) . It must be kept in mind that the presence of lensing features shifts the lenses towards the bluer and brighter regions of the color-magnitude diagram in comparison to similar deflectors without lensing features. The color-magnitude cut must consider this effect, for example, by estimating the extent of the shift with lens simulations as in Jacobs et al. (2019a) . Nevertheless, color-magnitude-cuts may remove some more exotic lenses, such as those with red arcs, and must therefore be as conservative as possible.

4.4.2 High false positive rate

One common observation between all the searches mentioned in Sect. 4.2.2 is the high false positive rate. False positives are usually spiral galaxies, ring galaxies or merging galaxies, and any other object that mimics in one way or another the morphology of lensed system (Petrillo et al., 2017, e.g.). In the case of color images, blue sources can confuse the CNNs (Jacobs et al., 2017, e.g.).

This high false positive rate is partially a consequence of the base rate fallacy mentioned in the previous section. For this reason, a visual inspection of the candidates found with CNNs is, at the moment, inevitable. However, there are ways to reduce the false positive rate and thus diminish the number of images to inspect visually.

One method to reduce the false positive rate is to add examples of common false positives in the training set. It can be done by retraining the network with false positives found in a previous classification (Cañameras et al., 2020; Petrillo et al., 2017, e.g.). However, like lenses, the usual false positives are not common. Therefore simulating the most common contaminants can

be a solution to constitute a sufficiently large training set for neural networks. Nonetheless, false positives are more challenging to simulate analytically than lenses since their luminous profile is particularly complex and can belong to a large variety of astronomical objects. To circumvent this difficulty, we propose a data-driven approach to produce sets of false positive examples.

A first test has been done on color images of spiral galaxies from SDSS using a generative adversarial neural network (GAN) trained with a Wasserstein cost function. This collaborative work is the subject of a paper (Dia et al., 2019, see appendix A) (see Appendix A). In order to improve the reconstruction of small features like spiral arms, we use a progressive training method. It consists of training first the inner layers of the generator with smaller images and then expanding the dimension of the discriminator and the generator simultaneously as the dimension of the images. The galaxies generated by the GAN display complex structures and morphologies. However, some simulations present unrealistic shapes. This aspect must be improved in subsequent versions of the generator, for example, by acting on the latent space of the GAN. To produce various false positives for different surveys, we plan to add the type of galaxy and the survey in the information given to the network during the training. This can be done with a supervised or semi-supervised method using a variation of Conditional GAN architectures (e.g Odena et al., 2016). Training the network with images from different surveys could also decrease the number of training examples needed compared for training a different generator for each survey.

4.4.3 The problem of using simulations in training sets

Because of the small number of known lenses, CNNs are usually trained with simulated lenses. To this end, multiple methods have been developed to simulate lenses (see 5). A decrease in performance is observed when the CNNs trained with lens simulations are applied to real data. This happens partially because of the different prevalences of lenses in the simulated sets and real data but can also be understood as a domain transfer problem, i.e., when the inference set is not drawn from the same distribution as the training set. In such a case, the performance may drop significantly.

One solution proposed by Petrillo et al. (2018) is to build a benchmark set of known lenses from surveys in which many lens searches have been conducted, such as KiDS and DES. This set could be used to compare the performance of the different classifiers on real data. The realization of such a benchmark set will be possible with the future extension of the Master Lens Database (Moustakas, 2012) which is a project that aims to compile all discovered strong gravitational lenses in a unique database. At the time of writing, it contains the lenses discovered up to 2016 representing hundreds of objects. However, its extension (Lemon et al., in prep) will include all the new candidates found in the recent lens searches and allow users to enter new lenses easily. In addition, the database will provide a grade reflecting the quality of the lens candidates, which could give more information when we re-use the candidates to

evaluate CNNs.

Another solution is to use more realistic simulations. To this end, we developed a simulation pipeline presented in Chap. 5. In our simulations, only the lensing effect is simulated. It reduces the risk of creating synthetic features that CNNs can recognize. However, the biases induced by the usage of the simulations must be identified and quantified in the view of future large-scale surveys.

4.5 Summary

This chapter summarized the different methods to find galaxy-galaxy lenses in imaging surveys. We also presented our contribution to the second strong gravitational lens finding challenge. In this work, we essentially focused on CNN-based lens finders. As we have seen, there are some challenges to overcome when using CNNs to find lenses. Galaxy-galaxy lenses are rare events. This has two main consequences. First, the set of known lenses is small. Therefore simulations are needed to train the CNN. Secondly, the low occurrence rate of lenses in real data leads to a significant false positive rate due to the base rate fallacy. Therefore, a visual inspection of the candidates is difficult to avoid. However, solutions exist to lower the false positive rate, such as retraining the CNNs against the false positives found during previous searches. In the future, sets of false positives may also be produced with generative models. The quality of the training examples is crucial for the performance of CNNs. During our participation in the second strong gravitational lens finding challenge, we observed that including a large number of lenses without visible lensing features dramatically lowered the classification's performance.

In the following chapters, we present the parts of our lens-finding pipeline and how we addressed the abovementioned issues.

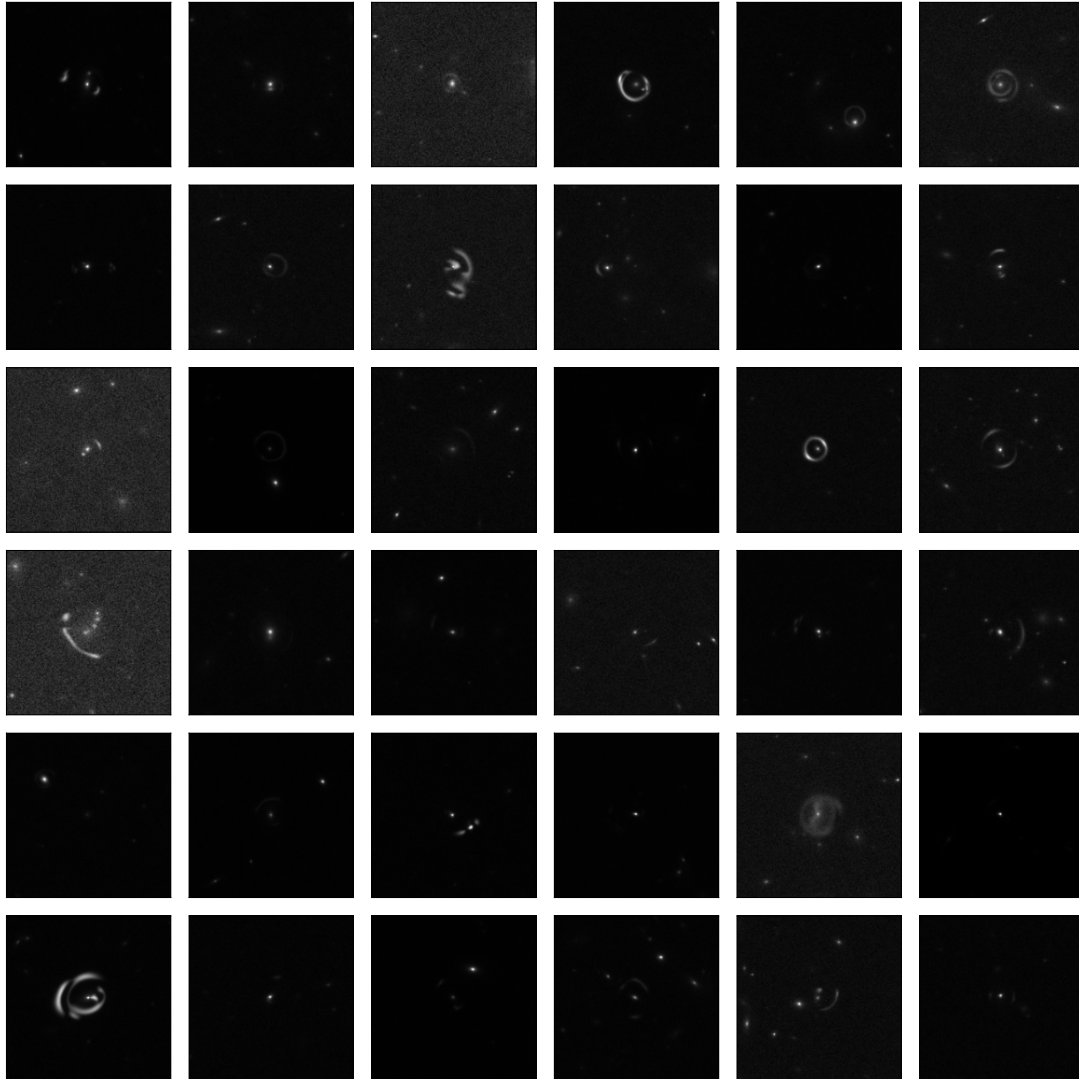


Figure 4.2: Examples of Euclid lens simulations (VIS band) matching the criteria to be included in the positive examples of our training set.

5 Lens simulations

As mentioned earlier, CNNs-based lens finders require a large set of lenses to be trained. Although some recent examples such as Huang et al. (2021) and Stein et al. (2021) have shown the possibility of using a minimal set of known lenses only to train neural networks, simulations of lenses are required for at least the first iterations of most CNN-based lens searches. This will be, in particular true for Euclid. However, we hope to retrain the classifier using only real data or a hybrid set with real lenses and simulations when a sufficient number of candidates and contaminants will be found in the first batch of data of the survey.

One method to produce sets of lenses is to produce entirely synthetic images where all light profiles are analytical models. For example, such an approach is used in Jacobs et al.; Metcalf et al. (2019a, 2019) . This method gives full control of the lensing parameters to the users. Thus, it allows easy production of training sets with specific distributions, i.e., a limited range of Einstein radii and colors. Since, by construction, this type of simulation does not contain instrumental effects or usual "defects" in images such as cosmic rays or companions, a random real background image is often added to the simulations to make them more realistic. However, it is challenging to recreate all of the subtleties of genuine lenses using fully analytical simulations. It is especially true if the foreground object has a more complex light profile, for example, spiral galaxies. An alternative consists of combining real deflector images with simulated sources, as in Petrillo et al. (2017) and Pourrahmani et al. (2018) to have foreground galaxies with a realistic light profile. In such cases, noise and "defects" are included in the images by construction.

CNNs trained with mock lenses may learn to recognize features characteristic of simulated data instead of learning lensing features (Jacobs et al., 2017). A significant drop in performance is observed when applying such CNNs on real data. In some cases, CNNs can even learn features that are invisible to the eye. This has been observed, for example, in Wilde et al. (2022) , where a part of the training set was rotated. The CNN learned the orientation of the images thanks to the asymmetric PSF. The rotated images were, as a consequence, misclassified. For this reason, improving the realism of lens simulations used in training sets is crucial. On the other hand, the computational cost of simulations must stay reasonable since large numbers

of mock lenses must be produced. In the context of future large-scale imaging surveys, we developed a flexible simulation pipeline, which is aimed to be easily adaptable to any survey and type of lens. It is designed to produce both single and multi-band images. With the idea of producing more realistic simulations, we choose a similar approach to Pourrahmani et al.; Petrillo et al. (2018, 2017) by using real foreground images, but go further in the realism of the simulation by using real source images. This method has already been applied in Cañameras et al. (2020). This method allows us to have sources with small structures and realistic magnitudes. With our method, only the lensing effect is simulated. We hope, therefore, to limit the formation of artifacts in the simulations. The pipeline has already been applied to produce r-band lens simulations for CFIS (Savary et al., 2021), and DES (Rojas et al., 2021).

We tested two different lensing software to simulate the lensing effect GLEE software package (Suyu et al., 2012; Suyu & Halkola, 2010) for the production of CFIS lens simulations (Savary et al., 2021) and Lenstronomy¹ (Birrer & Amara, 2018; Birrer et al., 2015) in the context of a DES lens search (Rojas et al., 2021). We used the source galaxies sample of Cañameras et al. (2020) in both cases. The morphology of these galaxies is taken from HST/ACS F814W images and the color is inferred from HSC ultra-deep stacked images. Since HST is a space telescope, we neglect the effect of its PSF on ground-based simulations.

Apart from the foreground galaxy and source images, our simulation pipeline requires the knowledge of the redshift of the source and the foreground galaxy and the velocity dispersion of the deflector as a proxy of the mass.

We describe here the main steps of the simulation pipeline and the difference between the two versions. The pipeline can be decomposed into two parts: The "match" part, where we associate a foreground galaxy with a source (Steps 1 to 3), and the simulation part, where we simulate a lens using the source and the foreground galaxy selected in the first part (Steps 4 to 7). The two parts of the pipeline are summarized in Fig.5.1.

1. In general, the background source images must be deconvolved from the PSF of the telescope. However, as explained above, if we aim to simulate lenses from a ground-based survey and the source images are provided by a space-based telescope, this step can generally be skipped. This step is, however, crucial for Euclid simulations as, in this case, the PSF of the source cannot be neglected.
2. We select a foreground galaxy in the deflector catalog. We assume that the lensing mass model is an SIE (Kormann et al., 1994; Kassiola & Kovner, 1993), which has five free parameters: the Einstein radius, the coordinates of the lens center, the ellipticity (or axis ratio), and the position angle (PA). It is important to highlight that with such a model, the ellipticity and PA are identical for the light and the mass distributions.
3. A random source is extracted from the background galaxy catalog. We ensure that the Einstein radius is sufficiently large to be visible in the lens image. For CFIS simulation

¹<https://github.com/sibirrer/lenstronomy>

we allow the Einstein radius value to lie in the range $0.8'' < R_E < 3.0''$, whereas for DES we chose $1.2'' < \theta_E < 3.0''$. The larger lower limit for DES simulation is due to the lower resolution of DES images. The upper limit depends on the size of the cutouts and ensures that the lensing features lie within the image. If the Einstein radius does not fulfill the conditions, we take another random source from the source catalog. This step is repeated until we can produce a sufficiently large Einstein radius. If, after 100 iterations, no good match is found, we multiply the velocity dispersion by 1.5 and repeat the process. If the procedure fails again we discard the foreground galaxy from our catalog and restart at step one. (See Fig. 5.1)

4. The source is placed randomly in a region of the source plane that can produce a strong lens system. This region is defined differently in the two versions of the pipeline. With *lenstronomy*, the region corresponds to the smallest rectangle encompassing the caustics, whereas with GLEE, the region is defined with a constraint on the total magnification $\mu \geq 2$. In the last case, the possible locations of the sources are therefore limited to the ones producing a total magnification $\mu \geq 2$.
5. Once the source position is determined, the lensing effect is computed using the GLEE software or *Lenstronomy*. This results in a high-resolution image of the lensed source.
6. The PSF of the foreground image is rescaled to the pixel size of the source image. We then convolve the lensed source image with this rescaled PSF. In the case of CFIS and DES simulations, resizing the lensed source image to the pixel size of the foreground image and then convolving directly with the PSF eventually saves computational resources and does not lead to a significant decrease in the quality of the lens simulations. However, to simulate high-resolution lenses, such as for Euclid, it is imperative to rescale the PSF image first if the size and resolution of the source images are greater than the foreground images.
7. The convolved lensed source image is then re-sampled to the pixel size of the foreground image and added to the foreground galaxy image.

5.1 Comparison of the two versions of the simulation tool

The two versions of the pipeline can produce comparable simulations of similar quality. This can be observed, for example, in Fig. 5.2 which depicts four pairs of CFIS simulations produced with the same foreground galaxies and sources using the two pipeline versions. In the pairs, only the position of the source is different. The velocity dispersions and the redshifts of the foreground galaxies and sources are the same. The simulation time is highly dependent on the lens and source image sizes and is similar for the two versions.

With both versions, a significant number of simulations displayed no visible features. In the GLEE version, this problem is more limited since the condition on the magnification helps

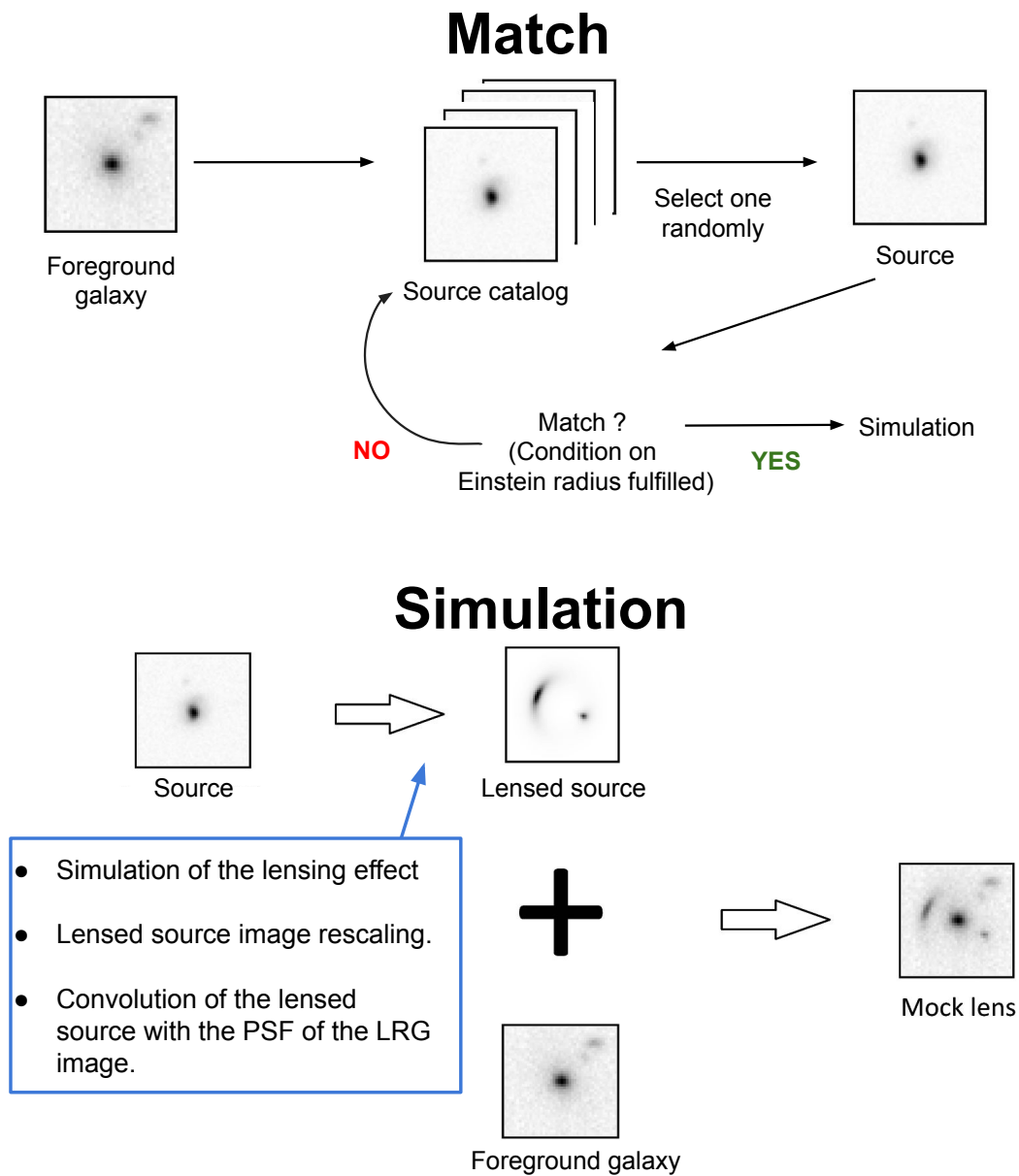


Figure 5.1: Illustration of the two different parts of the simulation pipeline. First, a foreground galaxy and a source are associated (Upper panel). Then, the lens is simulated using the selected foreground galaxy and source.

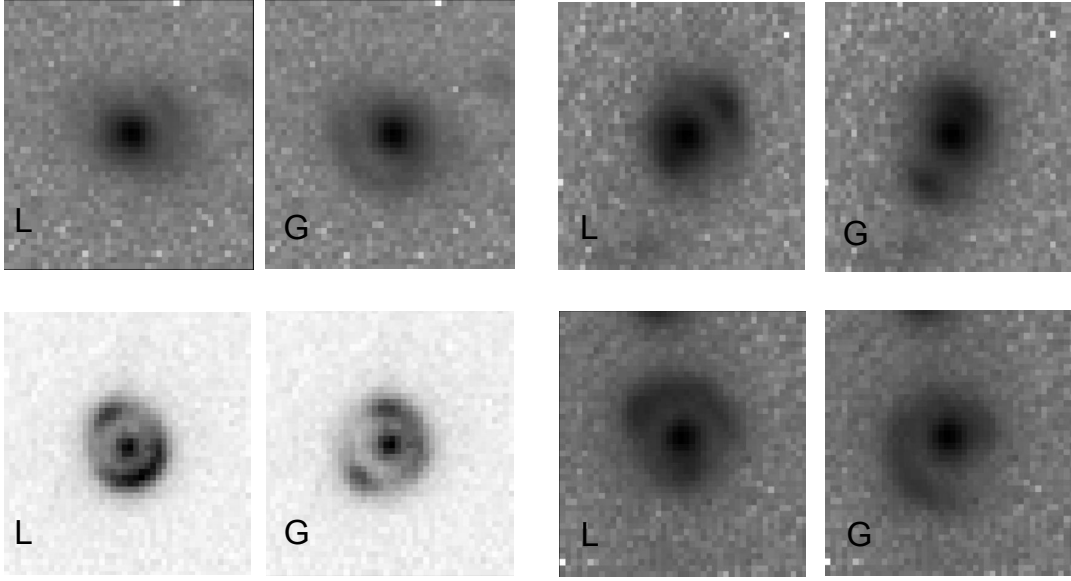


Figure 5.2: Comparison of CFIS lens simulations produced with the two versions of the simulation pipeline. Each pair depicts simulations with similar parameters (identical sources, deflectors, velocity dispersions and redshifts) for the *Lenstronomy* (L) and the *GLEE* (G) version of the pipeline. Only the position of the source is different in each pair.

to ensure the visibility of the features. However, we decided to keep only the simulations for which the sum of the brightness of all pixels of the lensed source is at least 20 times the mean rms value of the sky noise measured in the four corners of the deflector image. In the *Lenstronomy* version, an option allows decreasing the magnitude by a chosen number (boost). This increases the source's brightness and, thus, the visibility of the features. However, with this method, we artificially shift the magnitude distribution of the sources toward brighter ends and thus produce less realistic simulations than with the *GLEE* version. In the *GLEE* version, the condition on the total magnification allows us to increase the probability of producing visible lensing features. Nevertheless, since the condition forces the source to be placed in a higher magnification region, this can reduce the diversity of lensing features.

With our actual method of positioning the source, some lensing configurations occur less frequently. This is, in particular, the case for the cusp configuration with the SIE profile (see Fig. 2.3). This happens because, when we attribute the position in a random uniform way, the probability of falling into the positions producing the cusp position is lower than for fold and cross positions. Since the sources are extended, this effect is less important when producing a training set for galaxy-galaxy lenses. However, this must be taken into account if the pipeline is used to produce lenses with point source sources. For point sources, the function to position the source must be adapted to ensure a sufficient number of cusp configurations.

We presented in this chapter our lens simulation pipeline. It was designed to limit the creation of synthetic features that neural networks may learn. To do so, we use real images for the

deflectors and the sources and simulate only the lensing effect. The simulation tool allows us to produce lenses with different deflectors easily. Recently, the pipeline has been updated to produce lenses with edge-on galaxies to search for such lenses in CFIS (Acevedo, in prep). It is possible to produce simulations for any ground-based imaging survey with this tool. However, for space-based images, an additional step must be added to deconvolve the PSF of the source images.

6 Visual inspection of the lens candidates

As explained in the previous chapters, lens finding algorithms are hampered by a large false positive rate. Even if solutions can lower the false positive rate, because of the low occurrence rate of lensing, the classification of the lens finding algorithms cannot be trusted blindly. For this reason, visual inspection is a crucial step in all lens searches. In addition, the definition of what specific features define a lens can differ between the persons participating in the visual inspection of candidates. In this chapter, we present different tools designed to standardize and ease the visual inspection of sets of candidates. We also propose guidelines that can be used to rate lens candidates from ground-based imaging surveys.

6.1 Visual inspection tool

To provide an efficient method to inspect large samples of lens candidates, I developed a python library containing tools that allow a detailed grading and inspection of images one by one and a rapid selection of potential lenses among images displayed in mosaics. This library is available on Github on <https://github.com/esavary/Visualisation-tool/>. The interest of using this library is to ease the grading of large sets of candidates and standardize the visual inspection between visual inspectors.

6.1.1 Detailed inspection tool

The detailed inspection tool is available in two different versions for 1 band and color images. An example screenshot of the tool is presented in Fig. 6.1.

The detailed inspection tool allows the user to assign a classification (Lens/Maybe Lens/ Non Lens) and sub-classification (Merger / Spiral/ Ring/ Elliptical/ Disk) to each image. If more categories or subcategories are necessary, buttons can easily be added to the tool. The user can add comments about the candidates that are stored with the classifications. This allows, for example, the flagging of images that do not fit into the categories proposed in the buttons or candidates of special interest. Provided that a coordinate list of the candidates is given to

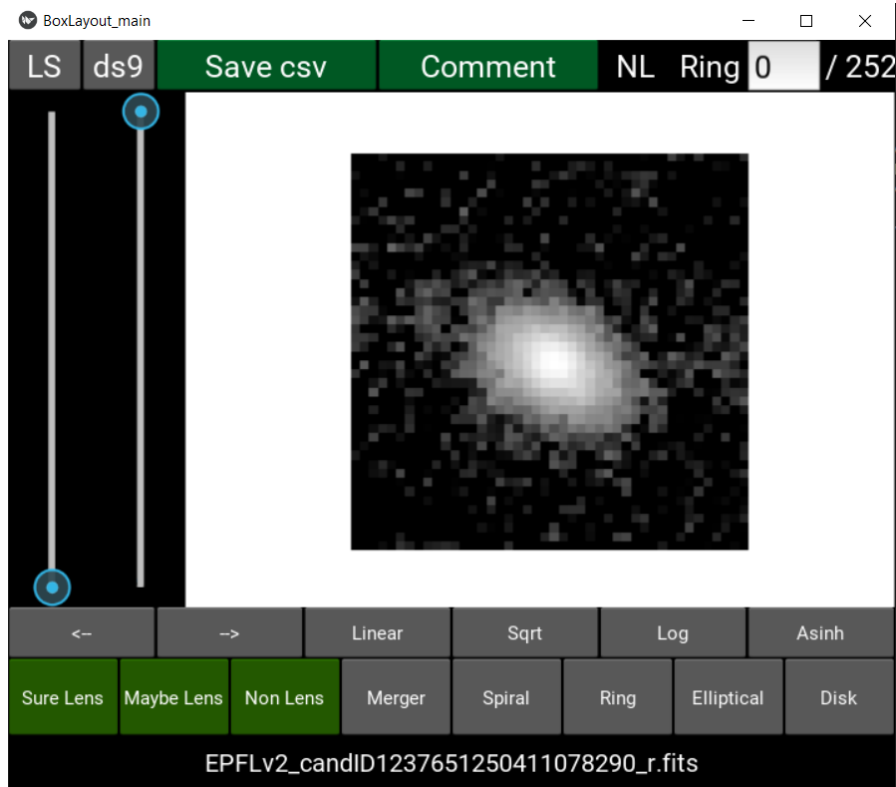


Figure 6.1: Screenshot of the one-band version of the detailed inspection tool. For each image the user can choose a main classification (Lens/Maybe Lens/ Non Lens) and a sub-classification (Merger / Spiral/ Ring/ Elliptical/ Disk).

the code, the user can also visualize Legacy Survey images (LS; Dey et al., 2019) in a separated window when available. This tool also enables the user to open the images in the software SAOImageDS9 (Joye & Mandel, 2003) in case more advanced visual settings are needed.

The user can correct his classification at any time and move forward or backward in the set of images by using the arrows or by specifying the image's position directly to inspect.

Mosaic inspection tool

The mosaic tool allows the user to inspect multiple images simultaneously. It is designed for a quick selection of potential candidates and therefore does not enable detailed labeling of them. It is possible to display either color or one-band images with the mosaic tool. We show a screenshot of the tool in Fig. 6.2.

The images are arranged randomly in frames of 10 by 10 images according to a random seed that can be given as an argument to the code. This allows us to force all visual inspectors to see the images in the same position (By using the same random seed) or in a different order. We also added the possibility to repeat images intentionally from the set if needed for statistics.

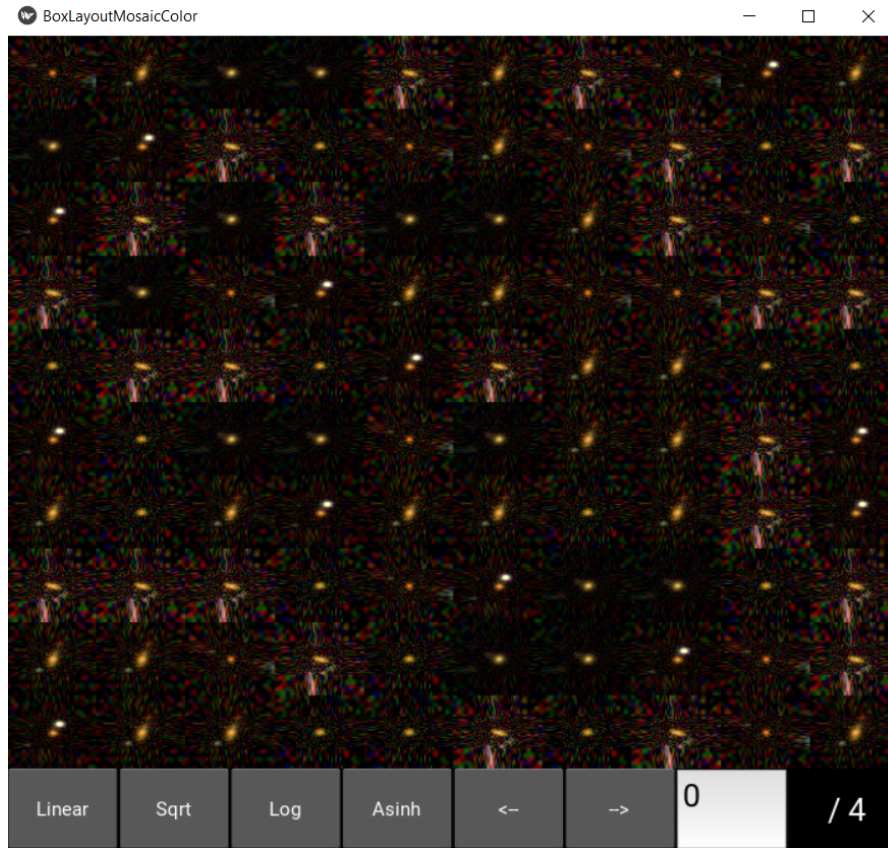


Figure 6.2: Screenshot of the color version of the mosaic tool. The user can flag the potential lenses by clicking on the images

Clicking once on an image changes its classification to "Lens". Clicking a second time on the same image cancels this classification. When an image is selected, it is replaced by an "L". This avoids unnoticed misclick on candidates. Similarly, as in the detailed inspection tool, it is possible to go one frame backward or forward or jump directly to a given page.

In the context of two lens searches in DES (Rojas et al., 2021) and CFIS (Savary et al., 2021), we analyzed the biases induced by the position of the images on the grid of the mosaic tool during the first step of the visual inspection. In the two searches, the candidates were shown at a random position, different for each visual inspector. We display in Fig. 6.3 the normalized mean number of clicks per cell of the mosaic tool computed with the contribution of all visual inspectors participating in each search. As we can see, the cells in the top part of the mosaic tool are slightly more likely to be selected. In the DES search, the bottom part of the Mosaic tool also received more clicks. This confirms the importance of showing the candidates at different positions to the visual inspectors to limit the effect of the location of images during visual inspection.

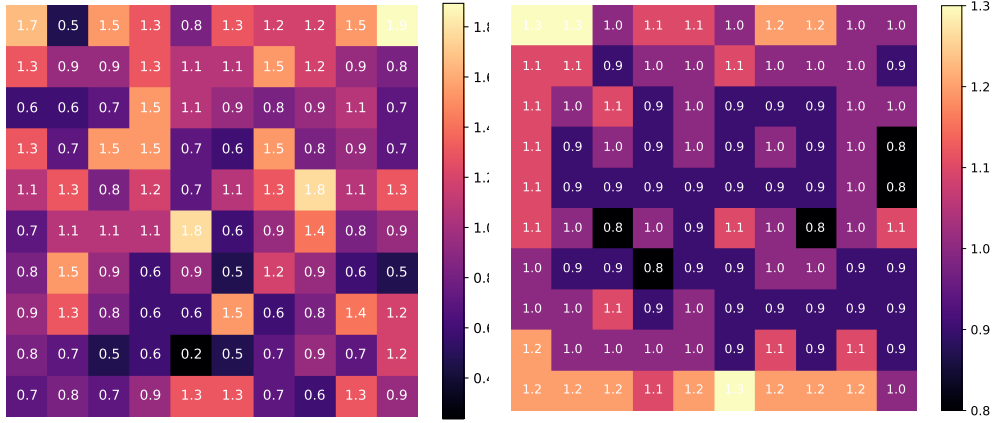


Figure 6.3: Color-map depicting the mean number of clicks in each cell of the Mosaic tool during the visual inspection of the candidates of the lens searches (Savary et al., 2021) (left panel) and (Rojas et al., 2021) (right panel). The values are computed among all visual inspectors participating to the searches and are normalized by the mean number of clicks per cell. Credit of the right panel image: adapted from (Rojas et al., 2021).

6.1.2 Classification guidelines

The notion of what can be a lens may differ significantly among visual inspectors. The grading schemes are also very heterogeneous between the different lens searches. For example Petrillo et al. (2017) , Petrillo et al. (2018) use three categories, "Lens", "Maybe lens" or "Non lens", whereas Jacobs et al. (2017) , Jacobs et al. (2019a) use four numerical categories ranging from 0 to 3. However, what belongs to each category depends on the appreciation of the visual inspector.

In the context of the CFIS and DES lens searches (Rojas et al., 2021; Savary et al., 2021), we established a grading scheme that could easily be adapted for the search of extended lensing features in other ground-based surveys. Our grading scheme is divided into four categories, for which we provide precise guidelines. This is done to standardize the classification between all visual inspectors. We display in Fig. 6.4 illustrative examples for each of the four categories.

The four categories are summarized below:

- **Secure lens (SL):** SL must display clear signatures of strong-lensing as several clear multiple images, signs of counter-image or an Einstein cross or ring. This category contains only the best candidates; no additional observation is needed to confirm their classification.
- **Maybe lens (ML):** ML candidates present a lens-like configuration and a large tangential distortion. However, they are classified as ML because they have no distinct multiple images or all images or arcs are on one side of the galaxy without signs of counter images. Candidates in this category require higher resolution imaging and modeling to confirm

their lens nature. ML category can include spiral or ring galaxies that mimic lensing features.

- Flexion (F) or Single arc: F category includes all galaxies surrounded only by a single distant arc with possible tangential distortion. F category can contain galaxies distorted by weak lensing effect.
- Non-lens (NL): NL candidates are all objects that do not fit into the three previous categories. They can be, for example, rings, spirals, merging or irregular galaxies or companions.

These guidelines will be reused in the lens database currently under development at LAS-TRO/EPFL (Lemon, in prep). The lenses will be provided with observational and model-derived quantities and a grade. Our guidelines serve as the basis for grading extended lens candidates from ground-based surveys.

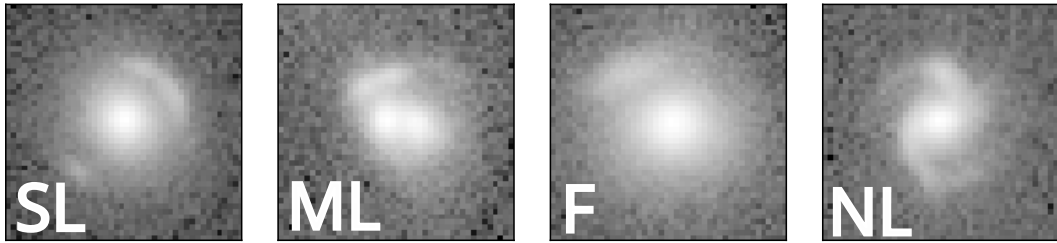


Figure 6.4: Illustrative examples for the four categories of the grading guidelines.

6.2 Method

A detailed inspection of each candidate found by the classifier is laborious and time-consuming for the visual inspectors. For this reason, we designed an efficient method in three steps that aim to reduce the time spent on the inspection of large sets of candidates found by CNNs. The steps are summarized below:

1. Preliminary inspection: We use the Mosaic tool to select all candidates found by the classifier that display possible signs of lensing. This pre-selection aims to eliminate a large part of the false positive, so the selection of possible lenses must be large. Depending on the number of images and time available for the second phase, we can choose to take the union or intersection of the selections of all visual inspectors or only images that obtain a certain number of votes as our final selection.
2. Detailed inspection: We reinspect all the objects selected during the first step with the detailed inspection tool. For all images, we assign a main classification following the grading guidelines and, if the image is a non-lens, a sub-classification to precise its nature.

3. Joint inspection: We investigate each image labelled as ML or SL by at least one visual inspector and decide with the entire group of visual inspectors on one unique grade for each image. As it will be explained later, we introduced this step because of the lack of agreement between the classifications of the visual inspectors. However, with improvements in the guidelines and the method to convey them, this step may become optional in future lens searches.

6.3 Discussion

In this chapter, we presented a method to inspect large set of lens candidates. This method has already been applied to two lens searches Rojas et al. (2021) and Savary et al. (2021). For the visual inspection of the candidates of these two searches, we communicated the grading guidelines in the form of slides displaying a short description of each category along with example images. However, we observed that despite the guidelines, the agreement between the classification of the different visual inspectors was very low.

To illustrate this, we present in Table 6.1 a summary of the classification from all visual inspectors after the detailed inspection step of Savary et al. (2021). In total 4 626 images were graded during this inspection. As we can see, for the same set of images, the proportions of the different categories vary greatly between the different visual inspectors. Category F causes the largest confusion. Some visual inspectors use this category only for galaxies surrounded by very large arcs, whereas others include most galaxies with companions that present a slight deformation. In some cases, objects identified as F by some visual inspectors were classified as SL by others since the boundary between a very large single arc and an incomplete Einstein ring is ambiguous. Even though all visual inspectors identified a similar fraction of SL, they selected different candidates. All visual inspectors classified only 8 from the 4 626 images as SL.

Table 6.1: Percentage of objects classified as SL, ML, F, and NL for each visual inspector after the detailed visual inspection step of Savary et al. (2021)

Visual inspector	SL	ML	F	NL
User 1	0.5	2.1	0.4	97.1
User 2	0.7	9.1	0.5	89.7
User 3	0.7	3.1	1.3	94.8
User 4	0.8	10.2	3.8	85.2
User 5	1.0	16.7	8.8	73.4
User 6	0.8	7.7	4.8	86.7

Although it is impossible to reach a perfect agreement on the candidates, this large discrepancy in the classification of the different visual inspectors shows that there is room for improvement in the phrasing of the guidelines to make them more precise and in the method to present them. This point will be essential for the future lens database (Lemon et al. in prep) since the contributors must provide a grade for their candidates. A reflection has to be done on

methods to homogenize the grading, such as training the visual inspectors on a labeled data-set or including labeled examples within the set of candidates to evaluate the performance of inspectors.

A large-scale study on the efficiency of visual inspection is currently taking place on Zooniverse¹ (Rojas et al. in prep). The experiment consists of classifying 1 500 color images from DES into four categories following simple instructions given on the website. The study involves only professionals working in the field of strong lensing since they are usually the ones that participate in the inspection of candidates but include people with different levels of expertise. At the moment, 69 persons participated in the experiment.

The aim of this work is to understand the limits and biases of visual inspection and what parameters improve or decrease the performance of visual inspectors. It will also estimate the number of candidates missed during the visual inspection step of previous lens searches. This work will provide invaluable help to improve the strategies of visual inspection for large-scale lens searches. In the future, the visualization tools and the guidelines must therefore be updated taking into account the result of this research.

During the visual inspection, the inspectors had access to color images of the objects from the Legacy Survey with a button of the detailed visual inspection tool. In the case of CFIS, this provided additional information on the color difference between the lensing features and the central galaxy. The Legacy Survey images were meant to help grade the candidates, but the main decision was based on the CFIS images. The majority of our candidates displayed blue lensing features and red deflectors. This is visible in Fig.6.5. However, some candidates contain red arcs. In most cases, the Legacy survey images confirmed our decision, and none of the candidates were discarded based on the color image.

¹<https://www.zooniverse.org/projects/krojas26/experts-visual-inspection-experiment/about/research>

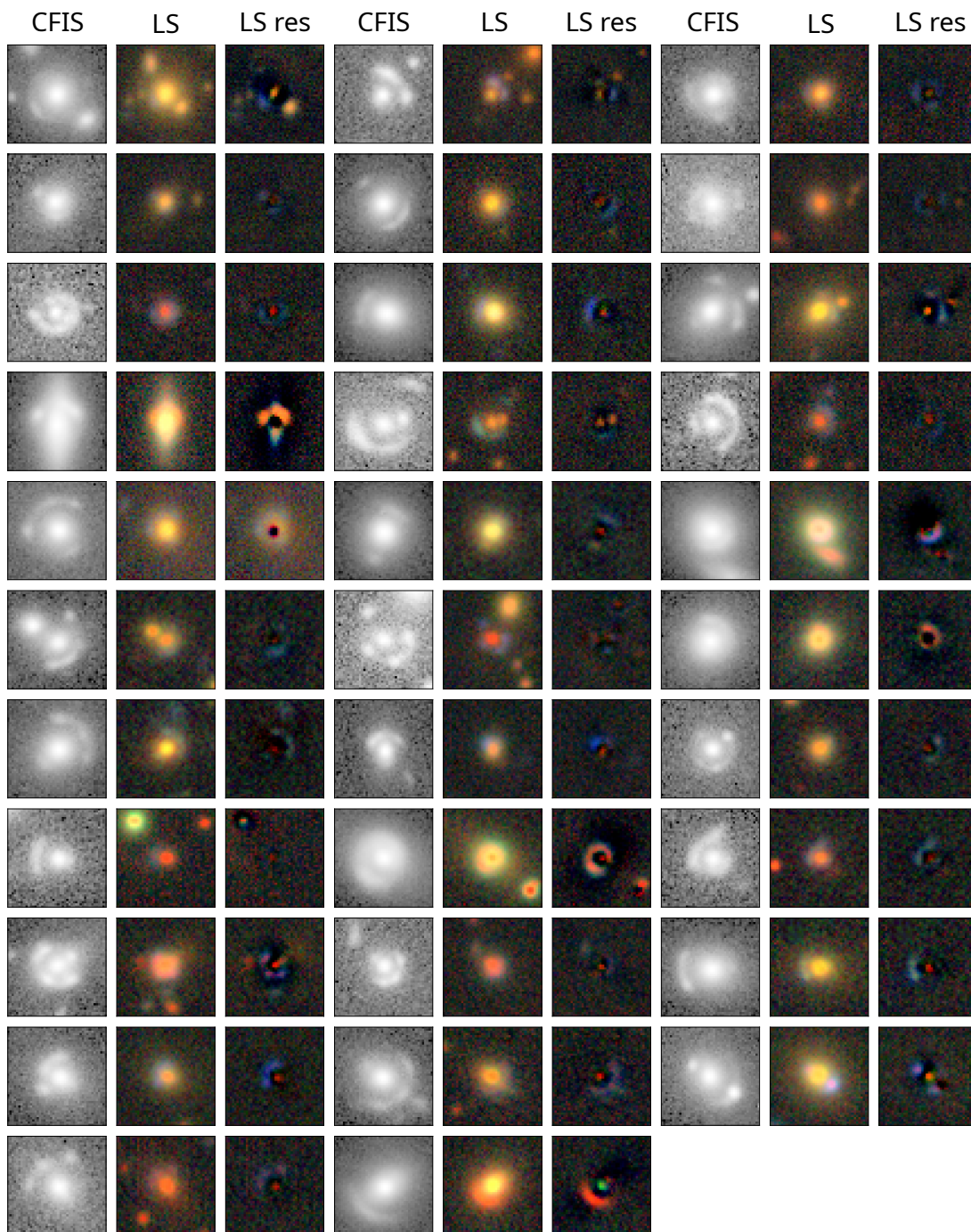


Figure 6.5: Legacy Survey images of the 33 SL candidates of the CFIS lens search. For each candidates, the CFIS image, the Legacy survey image and the model subtracted image are shown.

7 Lens deblending

In most ground-based images of strong lenses, the light of the source is hidden behind the light of the deflector. Deblending the lensing features from the images of lens candidates has multiple interests. First, deblending can reveal the faint structures either in the lens or in the source. Therefore, it can be used either to ease the visual inspection of candidates or as a pre-processing step before classification. Providing deblended images can also reduce the computational cost of modeling the candidates by providing an estimation of the lens model parameters. Finally, deblending enables the source and deflector redshift estimation for color images.

Usually, lens systems are deblended by fitting an analytical light profile on the deflector and subtracting its contribution from the lens image. However, this method can result in large residuals since the deflector's light profile does not perfectly match the analytical profile. Thus, we cannot use this method to highlight the lensing features as the residual may mimic their shape. Moreover, this procedure is hardly applicable to deflectors with complex light profiles since it requires analytical profiles.

Joseph et al. (2014) proposed a machine learning approach that removes the necessity of analytical profiles. The algorithm is trained on large sets of galaxies with morphologies similar to the deflectors. Since the lensing features are rare in the training set, they are not learned by the algorithm. When given an image of a lens, the algorithm returns the image of the deflector without the lensing features. Although very efficient, this method has some limitations. The number of principal components must be adapted for each survey depending on the characteristics of the lens images, such as deflectors' size or noise level. Additionally, the algorithm can interpret the orientation of the galaxies as a principal component. To avoid this, the galaxies must be aligned in the same direction before being passed to the algorithm. Companion galaxies can affect the quality of the reconstruction. It is, therefore, necessary to mask them for the process. The main limitation of the method is the large number of examples needed to learn a particular morphology of deflector. This requirement may be difficult to fulfill for surveys with a small number of objects or with a complex deflector.

The color information provides a valuable help to deblend lensing system images since the color of the source, and the deflectors are usually different. Generally, the lensing features appear bluer than the deflector since the sources are often high redshift star-forming galaxies. In this case, it is possible to use the information of color to deblend the lenses. This property was already used in the first automated lens finding algorithms (Gavazzi et al., 2014). The core principle of this algorithm is to subtract a scaled i-band image from the g-band image to reveal the lensing features. However, this method's application is limited to lenses with early-type galaxy deflectors and, since it is based on color difference, on blue sources.

Recently, Joseph et al. (2016) proposed a method to deblend color images of lenses, Multi-band morpho-Spectral Component Analysis Deblending Tool (MuSCADeT) that uses a combination of color difference and geometrical information. Similarly, as the PCA-based method (Joseph et al., 2014), this technique has the advantage of not requiring the fitting of analytical light distributions with an analytical profile. The idea behind the method is that the source and the deflector signals are sparse in different morpho-spectral representation domains and can thus be separated. This method was used in combination with the *scarlet* software (Melchior et al., 2018) in Rojas et al. (2021) to deblend color images of DES lens candidates. The method is well suited to provide deblended images in the context of a lens finding pipeline as it is entirely automated and does not require analytical light profiles. However, it relies on the assumption that the lensed features are bluer than the deflector and, therefore, will fail to deblend more exotic lenses, i.e., lenses with red arcs or blue deflectors.

The resolution of the NIR bands in Euclid will be much lower than the VIS band. For this reason, the combination of MuSCADeT and *scarlet* may not be optimal in this context since it requires color images. This motivated the development of an alternative automated deblending method based on morphology only that can be applied to single-band data.

Similarly, as the PCA deblending (Joseph et al., 2014), our method is entirely data-driven and does not require assumptions on the light profiles of the deflectors. Encouraged by the good performance of PCA to deblend lenses and the similarities between PCA and auto-encoders, we designed a modified auto-encoder architecture to separate the source and the deflector images.

In this chapter, I present our deblending auto-encoder. I also investigate a more conventional auto-encoder architecture and compared its performance with our model. I also introduce the first application of our method on real data.

7.1 Multi-output auto-encoder

Similar to traditional auto-encoder, our architecture contains an encoder and a decoder. The particularity of our network is that the decoder part is duplicated. This allows us to extract the lensed source and deflector image simultaneously, representing an advantage over the PCA deblending algorithm, which only outputs the deflector. The network also outputs the

reconstructed lens system. To do so, it computes the addition of the reconstructed lensed source and the reconstructed lens through an Add layer. Our algorithm can currently deblend single-band images but will be adapted for color images soon. The detailed architecture is presented in Fig. B.1 of Savary et al. (2021) (see Chap. 8).

We integrated our deblending algorithm in the lens finding pipeline of Savary et al. (2021) . We summarize here the main steps of the training and the main points of the results.

7.1.1 Training

The modified auto-encoder was trained with 10 000 CFIS lens simulations produced following the procedure described in Sect. 5 and 5 000 CFIS images of LRGs. During the training the images are flipped randomly along the x and y axis We use mini-batch stochastic gradient descent with Adam optimization (D. P. Kingma & Ba, 2014). We include a small percentage of non-lenses in training set to teach the algorithm not to deblend the potential false positives. This point is important since the auto-encoder aims to deblend sets of lens candidates before a human visual inspection is performed. The sets can therefore contain a significant amount of false positives. Ideally, we would use the result of the deblending to discard a part of the false positives.

The input of the auto-encoder is the images of the lens systems. The ground truth we use is different from traditional auto-encoders. Instead of using only the input image as labels, we provide three ground truth images: the lens system, the lensed source, and the deflector. For the images of the training set containing an LRG without lensing features, the ground truth images for the lensed source and deflector are an image filled with zero and the LRG image itself, respectively. In doing so, we expect the network to return no lensing feature for a non-lens image.

The three output layers of the network have sigmoid activation functions. Therefore, we normalize the input and ground truth images between 0 and 1 as follows:

$$\frac{X - \min X_{lens}}{\max X_{lens} - \min X_{lens}}, \quad (7.1)$$

where X is the image to normalize and X_{lens} is the image of the lens system. We must highlight that we use the minimum and maximum of the lens system image to normalize all ground truth images, including the lensed source and deflector images. This allows us to conserve the original ratio of intensity between the lensed source and lens images. As the lens image is the sum of the two others, it ensures that its maximal value is larger than or equal to the one from the two images. Consequently, the pixel values for the three images lie between 0 and 1 after the normalization.

To train the network, we designed a custom loss function that takes into account the reconstruction error on the three different ground truth images:

$$L = 0.4 \times L_{\text{deflector}} + 0.4 \times L_{\text{source}} + 0.2 \times L_{\text{lens}}, \quad (7.2)$$

where $L_{\text{deflector}}$, L_{source} and L_{lens} correspond to the binary cross-entropy losses evaluated between the real deflector and the deblended deflector images, the lensed source and the deblended lensed source images, and the lens system and the combination of the deblended deflector and the lensed source image, respectively.

We included the term L_{lens} to improve the lens system's magnitude estimation. However, we attribute a lower weight to this term in the loss function since we want to give more importance to the correct reconstruction of the lensed source and deflector.

7.1.2 Evaluation of the deblended images

Since we aim not only to use the deblending auto-encoder to highlight the lensing features but also to initialize the lens model, we must ensure that the physical information contained in the images, such as the shape of the different components or the small structures, is correctly reconstructed. The reconstruction error metrics that are commonly used in computer vision, such as Mean-squared error or Chi-squared error, provide only little indication on the physical soundness of the reconstructed images. For this reason, we use the CAS parameters described in section 3.3.2 to evaluate the reconstructed images.

In Fig. 7.1 we compare the CAS statistics for a population of 400 simulated lenses and their reconstruction with the auto-encoder. We used the python library `statmorph` to measure the CAS parameters. As we can see, the asymmetry and concentration of the lenses are well reconstructed by the auto-encoder. However, the reconstructed lenses are slightly smoother than the simulated lenses. It indicates that the auto-encoder missed some of the small structures. As we discuss below, this effect is also observed on real data and may partially be due to our choice of training data. However, alternative architectures such as generative adversarial neural networks may improve this issue and must be tested for future versions.

One interest of the deblending is to provide a photometric measurement of the deflector and the source to estimate their redshifts. In simulations, the deblending auto-encoder can provide reliable photometry of the deflector. It tends to reconstruct too bright lensed source features, especially for faint lensed sources. It is visible in Fig. 7.2 which displays the difference in magnitude between the reconstructed lensed source and deflector and their ground truth. It is, however, possible to obtain the photometry of the sources by subtracting the reconstructed deflectors to the lens images with the actual version of the deblending auto-encoder. Including the source photometry in the loss function could theoretically improve this aspect. However, all tests we performed with photometry as an additional term in the loss function led to very unstable training and our model did not converged. More work must be done on the loss function and training set to improve the reconstruction of the small features and the

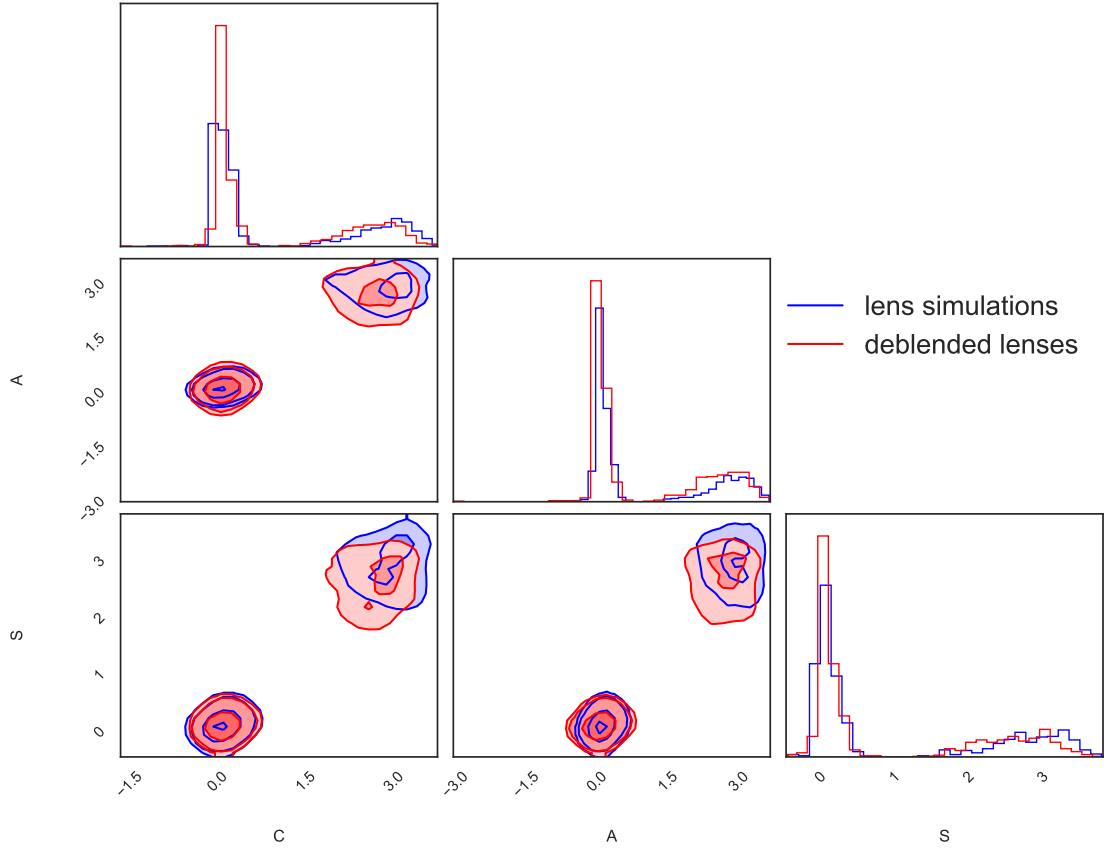


Figure 7.1: Comparison of the CAS statistics for 1000 lens simulations (in grey) and 1000 lenses reconstructed with the multi-output auto-encoder (in blue). C corresponds to the concentration index and measures the compactness of the galaxies. A is the asymmetry parameter, and S is the smoothness parameter which estimates the degree of small-scale structures.

correctness of the photometry of the sources.

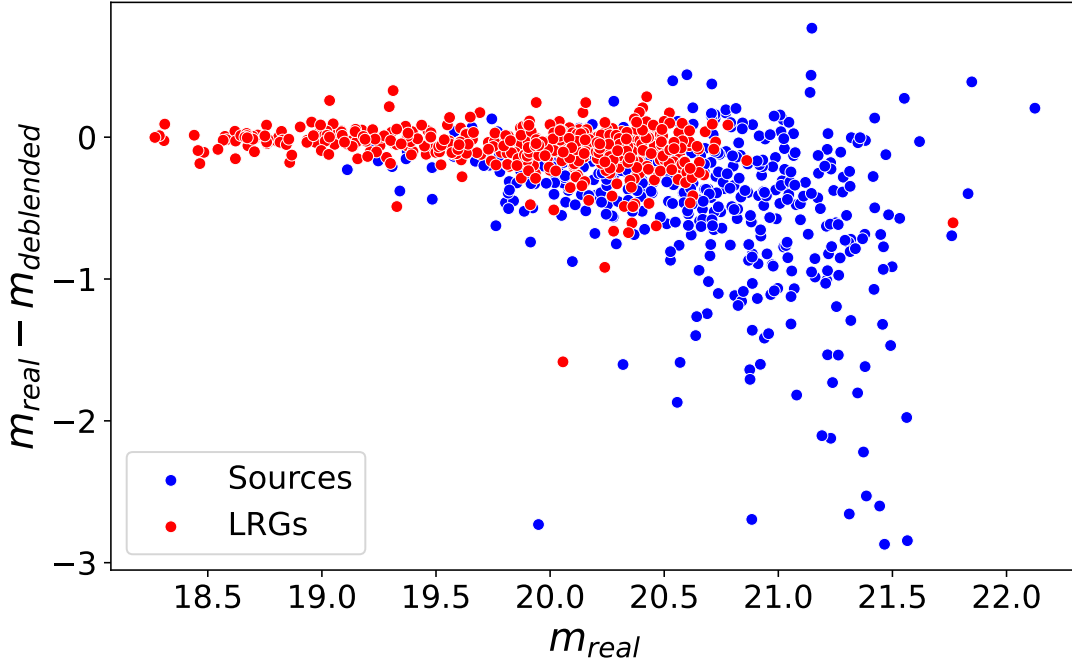


Figure 7.2: Difference in magnitudes between lensed sources and deflectors reconstructed with the multi-output deblending auto-encoder and their ground truth. The magnitude are computed on aperture of 4 arc-seconds.

7.2 Comparison of the two-output auto-encoder with two separated auto-encoders

We study here an alternative for the deblending auto-encoder. It uses two auto-encoders, each specialized in extracting either the lensed source or the deflector in the lens images. We present in Fig. 7.3 the two possibilities. The multi-output model (option A) is our actual deblending auto-encoder, whereas the two separated auto-encoders (option B) are the alternative. The layers inside the encoders, decoders, and bottleneck layers are identical in each case. The detail of these layers is given in Savary et al. (2021) .

The two auto-encoders of option B are trained separately. The ground-truth images given to these models during the training are the lensed sources for the first auto-encoder and the deflector images for the second. The two models were trained with the same training set as Option A using a binary-cross entropy loss.

We show in Fig.7.4 the difference in magnitude between the reconstructed lensed source and deflector and their ground truth obtained with option B. By comparing with Fig. 7.2, which displays the same measurements for the multi-output model, option A, we see that the photometry of the sources is less precise with option B. However, the precision of the

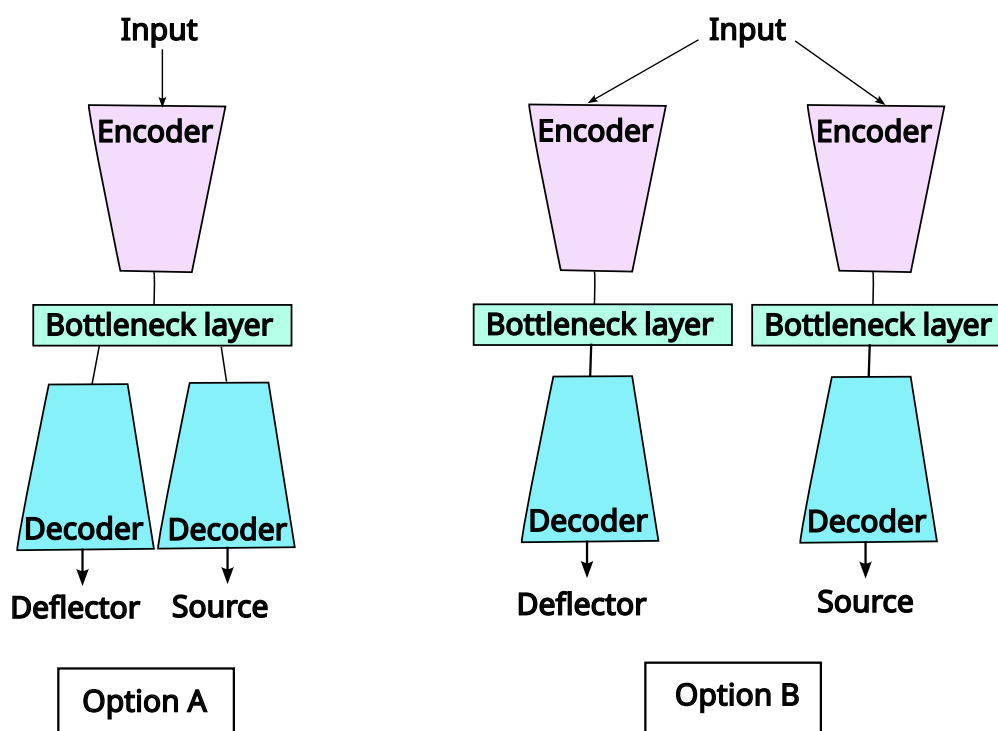


Figure 7.3: Scheme of two possible deblending models. Option A is a multi-output auto-encoder that outputs at the same time the lensed source and the deflector images. Option B consists of two different auto-encoders. One of them extracts the lensed sources and the other the deflector images. The encoder, decoder, and bottleneck layers blocks are identical in each option. The details of the layers inside these blocks can be found in Savary et al. (2021) .

photometry of the deflectors is similar for the two options.

Figure 7.5 displays examples of lens simulations deblended with the multi-output model, option A, and the two separated auto-encoders, option B. We see that the reconstruction of the shape of the lensed source is less blurry with the multi-output model. Additionally, lensing features often appear in the LRG images with option B. Option A and Option B are both trained to reconstruct the companion galaxies in the LRG images. However, we observe that with option B, the companions are systematically mistaken with lensing features and reconstructed in the lensed source images.

Given these results, we think that training the auto-encoder to output the lensed source and the deflector image simultaneously helps it to specialize better on the deblending task. More investigations have to be done to understand the cause of the better performance of the multi-output model.

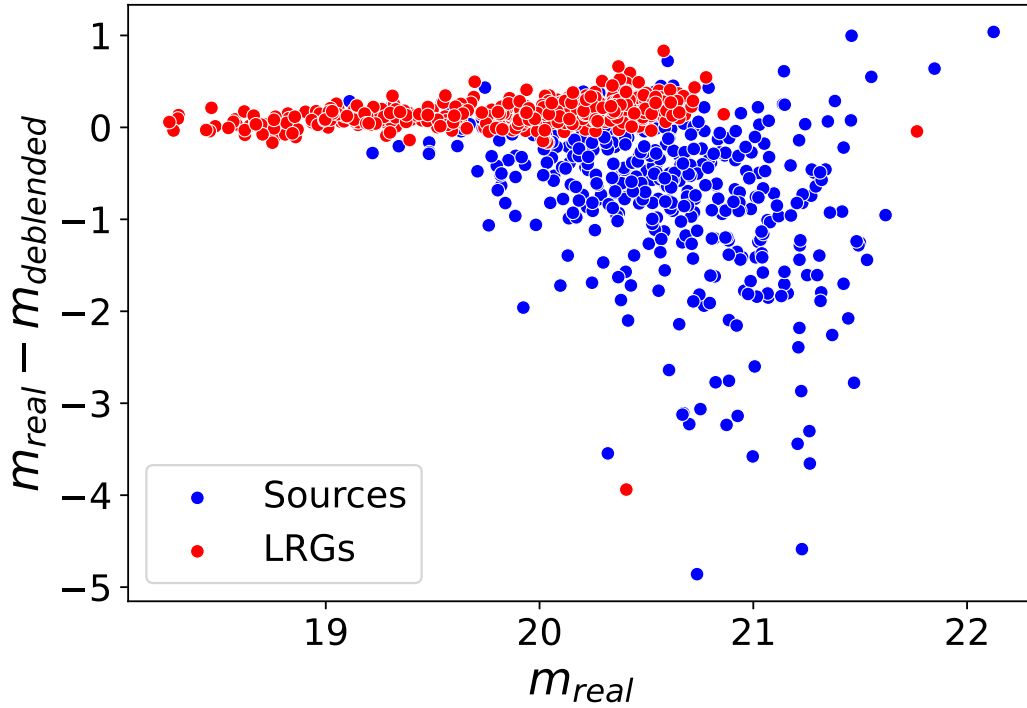


Figure 7.4: Difference in magnitudes between lensed sources and deflectors reconstructed with two different auto-encoders (Option B) and their ground truth. The magnitude is computed on apertures of 4 arc-seconds.

7.3 Application on real data

We applied our multi-output auto-encoder to the 32 SL candidates of Savary et al. (2021) . The deblended images are presented in Fig. 7 of Savary et al. (2021) . We also display in Fig 7.7 the deblending results for a subsample of the ML candidates of Savary et al. (2021) .

The performance of the deblending is generally not affected by the presence of companions in the lens system images, except if the companions are very close to the lensing features (see, e.g., UNIONS J090919+351503 (F) in Fig. 7.7). This is an advantage of the auto-encoders over the other deblending methods. Contrary to what was observed with simulations, in most cases, the auto-encoder does not reconstruct the companions present in the real lens images. This is the case, for example, in UNIONS J121959+340210 (E), UNIONS J165402+444252 (I), UNIONS J080114+455838 (J), UNIONS J135800+492334 (K) or UNIONS J101441+504451 (N) in Fig.7.7. Less frequently, the companion appears in the deflector image as in UNIONS J085335+352650 (H) in Fig.7.7. This is, however, not an issue for photometric measurements as the companions can easily be isolated and masked from the LRG images.

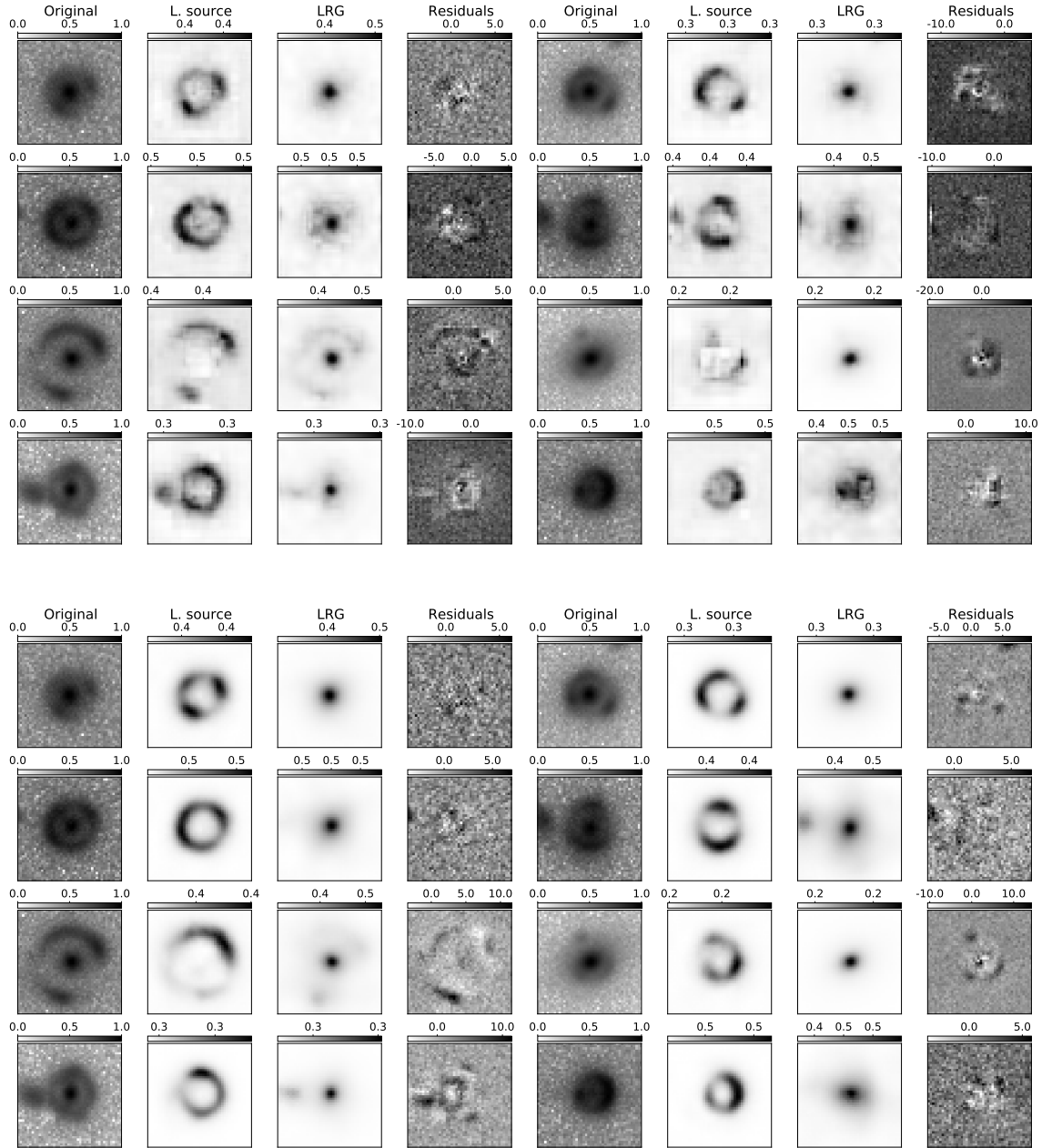


Figure 7.5: Comparison of deblending results on simulated lenses for two different architectures. **Upper panel:** Results of the two separated auto-encoders (option B). **Lower panel:** Results of the multi-output auto-encoder (option A). Two lens systems are shown in each row. For each lens system, we display from left to right: the original image in asinh grayscale; the lensed source and deflector deblended using the auto-encoder; and the standardized residuals.

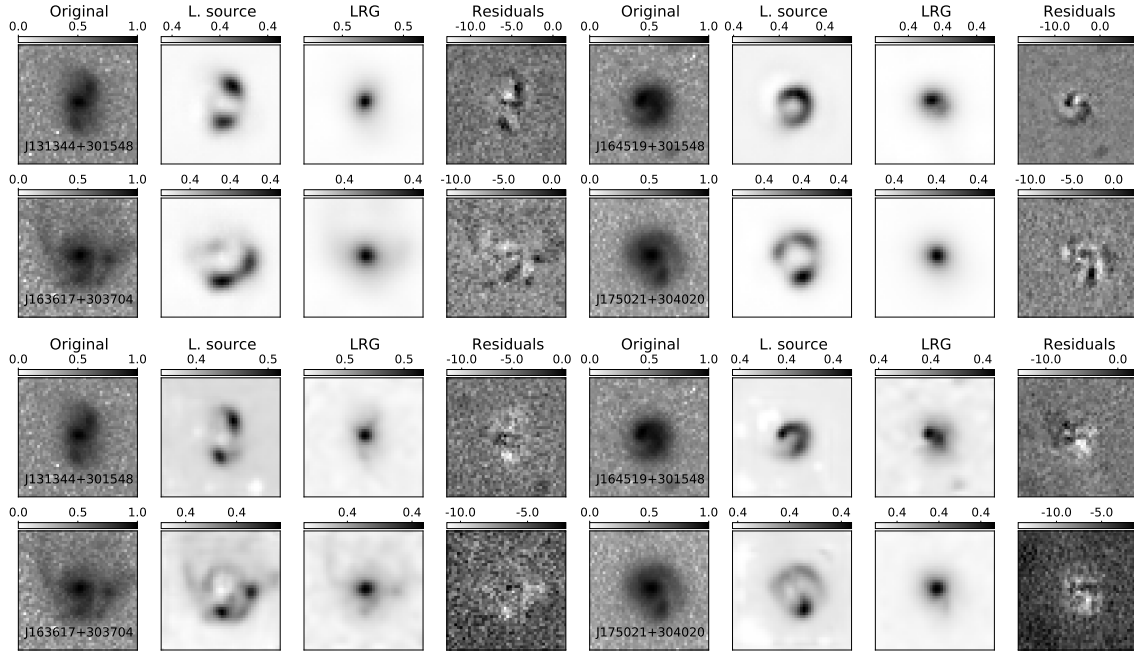


Figure 7.6: Result of the auto-encoder deblending for 4 spirals extracted from the false positive catalog of Savary et al. (2021) . **Upper panel:** Results with the original training set composed with LRG and lens simulations only. **Lower panel:** Results after retraining with 900 additional spiral galaxies in the non-lens examples.

We can observe that the auto-encoder fails to capture some of the high frequencies features. This effect is visible, for example, in UNIONS J182125+510038 (O) Fig.7.7 and to a lesser extent in UNIONS J102147+500808 (L), where the point-like structure is mistaken for arcs by the auto-encoder. We believe this happens partially because the training set contains essentially extended and smooth arc-like features and no point-like substructure. This will be improved with retraining with more detailed simulations when the simulation pipeline is adapted for more diverse lens types. One of the main issues with the implementation of Savary et al. (2021) is that the network tends to find arcs in false-positive images, as shown in the examples presented in the upper panel of Fig.7.6. This is a consequence of our choice of training examples. The network was trained with LRG and lenses only and thus, was not exposed to objects such as spiral, ring, or merger galaxies. We retrained the network with 900 spirals identified during the visual inspection in Savary et al. (2021) . The spirals were added to the non-lens examples of the previous training set. Surprisingly, after retraining, the auto-encoder still considers the spiral arms as lensing features, as shown in the lower panel of Fig.7.6. However, the shape of the spiral arms is restored with more precision, allowing us to distinguish them from genuine lensing features. The next iteration of lens searches within the CFIS survey will provide us with a larger false-positive sample. We expect, therefore, that the subsequent versions of the algorithm will perform better with this issue.

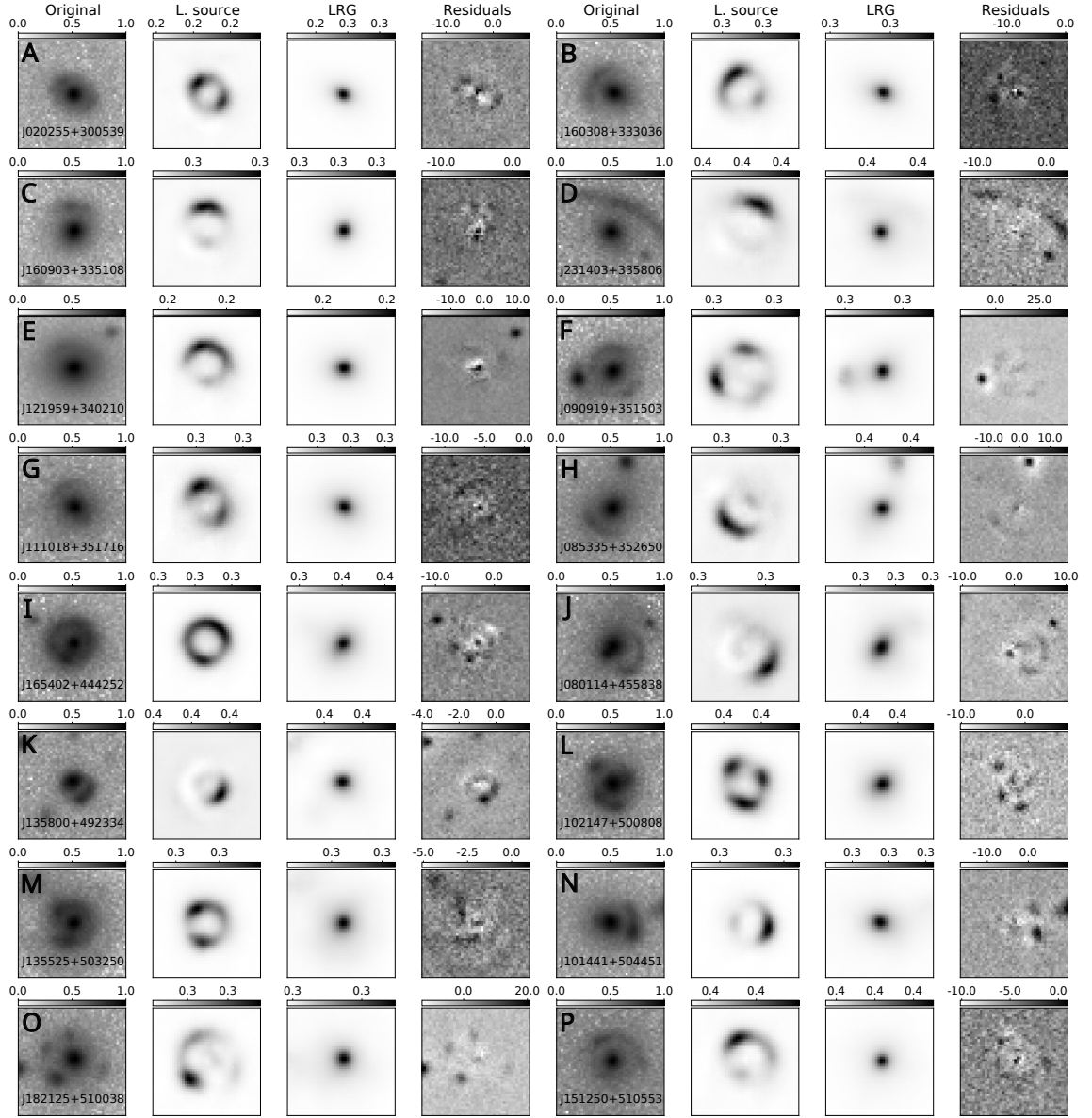


Figure 7.7: Result of the auto-encoder deblending for 16 ML candidates of Savary et al. (2021) .

7.4 Summary

In this chapter, I described a new method to deblend images of gravitational lenses based on a modified auto-encoder architecture. This approach is entirely data-driven since it does not rely on analytic light profiles. The auto-encoder method is easily adaptable to other surveys and different deflector types provided a training set is available. The lens simulations produced to train the CNN-based lens finders can be reused to train the auto-encoder. Therefore, the auto-encoder deblending algorithm can easily be integrated into future lens finding pipelines.

In its actual version, the deblending auto-encoder provides help during the visual inspection by enhancing the visibility of the lensing features. We applied it to the 32 SL candidates in Savary et al. (2021) to establish the presence of lensing features. In the future, the deblended images could be used to initialize lens models. In addition, the deblending may serve as a pre-processing step for classifiers to improve the detection of lenses with small image separation. The next step is to provide a three-band version of the auto-encoder. This will allow us to compare the performance with the standard deblending methods.

8 First application of an end-to-end lens finding pipeline in CFIS and DES

In the light of future wide-field surveys, we built a flexible lens finding pipeline. This pipeline includes modules to detect strong lenses in imaging data, deblend the light from the source and the deflector, and model the lens candidates automatically.

The core of our detection module is a variation of Efficientnet architecture B0 (Tan & Le, 2019), which we used for the lens finding challenge 2.0 (see section 2.0). This model obtained the best performance among all the models we tested in the challenge.

To deblend the lens candidates, we applied two different algorithms. For DES candidates, we used a combination of MuSCADeT and scarlet, which is very appropriate to deblend the color images of LRG lenses with blue features. In the case of CFIS, only r-band images were available for all the candidates. Therefore, as described in Chap 7, we designed an auto-encoder-based method to deblend one-band images. This method is not only useful for CFIS images but also for Euclid since the VIS band will have a higher resolution than the infrared bands.

The automated modeling module was the subject of the master thesis of Mark Maus and will be detailed in an upcoming publication (Maus, in prep.). We summarize here its key features and limitations. The automated modeling algorithm is aimed at modeling automatically large sets of potential lens system images. The modeling is done with the Lenstronomy package. In its current version, the deflectors are modeled with an SIE mass profile with external shear. However, since Lenstronomy is very flexible, one can adapt the mass profile to modelize more complex lenses. The modeling is possible both for one-band or multi-band images. Lens candidate images often contain companion galaxies or artifacts, and the size of the lens features may differ a lot between candidates. With the automated modeling module, they are automatically detected and masked. In addition to providing an estimation of the parameters of the lensing system, the automated modeling enables the distinction between lenses and ring galaxies (Maus, in prep.). The main limitation of the automated modeling occurs when companions reside either close to the deflector or the lensing features. In such a case, the automated masking procedure fails, and the companions must be masked manually. However, future versions of the automated algorithm may take advantage of the deblended images to

improve this issue. This is particularly true with color images, for which the color difference may help to identify the companions. We applied the pipeline to the search of LRG acting as lenses in two different surveys: CFIS and DES.

8.1 Lens finding in CFIS and DES

We describe in this section the details of the two first applications of our lens finding pipeline to CFIS and DES. They are the first applications of an end-to-end lens finding pipeline that enables automated modeling and model-independent deblending of the candidates. The two lens searches allowed us to test different dimensions of the lens finding pipeline. Our galaxy-galaxy lens search is the first applied to CFIS. It will therefore provide new candidates since the excellent-seeing r -band imaging of CFIS allows the unveiling of candidates that have been missed by previous strong lens searches performed from the Southern hemisphere. Indeed, the image quality enables us to detect lenses with smaller image separation. In addition, applying the pipeline to r -band images gives an overview of the potential of our pipeline when applied to the VIS band of Euclid.

Since DES data are public, some galaxy-galaxy lens searches have already been performed in DES (Jacobs et al., 2019a; Jacobs et al., n.d.) and many will follow. Therefore, this survey can serve as a benchmark to compare the performance of our lens finding pipeline with other methods. Since multiple bands are available, we can probe the efficiency of our lens finding pipeline on color data.

8.1.1 Statement of contribution

The following papers, which are accepted for publication in *Astronomy Astrophysics*, present our application of the lens finding pipeline to the initial 2 500 square degrees of CFIS and the first data release of the Dark Energy Survey. The two papers are reproduced with permission from co-authors.

In Savary et al. (2021), the cutouts were prepared by R.Gavazzi, and the lens simulations were produced by myself in collaboration with S.H. Suyu, R.Canameras, S.Schuldt. The adaptation of the CNN was made by myself in collaboration with K.Rojas. The visual inspectors of the candidates were K.Rojas, B.Clément, F.Courbin, J.H.H.Chan, G.Vernardos, and me. The modeling of the candidates was performed by M.Maus. The deblending algorithm was designed by myself. All co-authors contributed to the correction of the paper. In Rojas et al. (2021), the simulations were produced by K.Rojas. We collaborated with K.Rojas on the adaption of the CNN. The candidates were inspected visually by K.Rojas, B.Clément, F.Courbin, J.H.H.Chan, G.Vernardos, C.Lemon, and myself. M.Maus performed the modeling of the candidates and R.Joseph the deblending. All co-authors participated in the corrections of the paper.

Strong lensing in UNIONS: Toward a pipeline from discovery to modeling ★

E. Savary¹, K. Rojas¹, M. Maus¹, B. Clément¹, F. Courbin¹, R. Gavazzi^{2,3}, J. H. H. Chan¹, C. Lemon¹, G. Vernardos¹,
R. Cañameras⁴, S. Schuldt^{4,5}, S. H. Suyu^{4,5,6}, J.-C. Cuillandre^{7,8}, S. Fabbro¹², S. Gwyn¹², M. J. Hudson^{9,10,11},
M. Kilbinger⁷, D. Scott¹³, and C. Stone¹⁴

¹ Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland

² Institut d'Astrophysique de Paris, UMR7095 CNRS & Sorbonne Université, 98bis Bd Arago, 75014 Paris, France

³ Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB30HA, UK

⁴ Max-Planck-Institut für Astrophysik, Karl-Schwarzschild Str. 1, 85748 Garching, Germany

⁵ Technische Universität München, Physik-Department, James-Frank-Straße 1, 85748 Garching, Germany

⁶ Institute of Astronomy and Astrophysics, Academia Sinica, 11F of ASMA, No.1, Section 4, Roosevelt Road, Taipei 10617, Taiwan

⁷ AIM, CEA, CNRS, Université Paris-Saclay, Université de Paris, F-91191 Gif-sur-Yvette, France

⁸ Observatoire de Paris, PSL Research University 61, avenue de l'Observatoire, F-75014 Paris, France

⁹ Waterloo Centre for Astrophysics, University of Waterloo, 200, University Ave W, Waterloo, ON N2L 3G1, Canada,

¹⁰ Department of Physics and Astronomy, University of Waterloo, Waterloo, ON, N2L 3G1, Canada

¹¹ Perimeter Institute for Theoretical Physics, 31 Caroline St N, Waterloo, ON N2L 2Y5, Canada

¹² Canadian Astronomy Data Centre, Herzberg Astronomy and Astrophysics, 5071 West Saanich Rd, Victoria, BC V9E 2E7, Canada

¹³ Department of Physics and Astronomy, University of British Columbia, 6225 Agricultural Road, Vancouver, V6T 1Z1, Canada

¹⁴ Queen's University, Dept. of Physics, Engineering Physics and Astronomy, Kingston, Canada

July 26, 2022

ABSTRACT

We present a search for galaxy-scale strong gravitational lenses in the initial 2500 square degrees of the Canada-France Imaging Survey (CFIS). We designed a convolutional neural network (CNN) committee that we applied to a selection of 2 344 002 exquisite-seeing r -band images of color-selected luminous red galaxies (LRGs). Our classification uses a realistic training set where the lensing galaxies and the lensed sources are both taken from real data, namely the CFIS r -band images themselves and the Hubble Space Telescope (HST). A total of 9460 candidates obtain a score above 0.5 with the CNN committee. After a visual inspection of the candidates, we find a total of 133 lens candidates, of which 104 are completely new. The set of false positives mainly contains ring, spiral, and merger galaxies, and to a lesser extent galaxies with nearby companions. We classify 32 of the lens candidates as secure lenses and 101 as maybe lenses. For the 32 highest quality lenses, we also fit a singular isothermal ellipsoid mass profile with external shear along with an elliptical Sérsic profile for the lens and source light. This automated modeling step provides distributions of properties for both sources and lenses that have Einstein radii in the range $0.5'' < \theta_E < 2.5''$. Finally, we introduce a new lens and/or source single-band deblending algorithm based on auto-encoder representation of our candidates. This is the first time an end-to-end lens-finding and modeling pipeline is assembled together, in view of future lens searches in a single band, as will be possible with Euclid.

Key words. Gravitational lensing: strong – Surveys – Techniques: image processing

1. Introduction

Strong gravitational lensing provides a unique astrophysical tool, via the formation of several distinct images of a high-redshift source. Depending on the source light profile and lensing mass distribution, multiple images of the source can appear as partial arcs or even complete arcs called Einstein rings. Such strongly lensed systems offer a vast range of astrophysical and cosmological applications, from the determination of cosmological parameters (Suyu et al. 2017; Bonvin et al. 2017; Wong et al. 2020) to the study of galaxy mass profiles (Koopmans & Treu 2003; Sonnenfeld et al. 2015; Bellagamba et al. 2017) and halo

substructure (Mao & Schneider 1998; Dalal & Kochanek 2002; Koopmans 2005; Vegetti & Koopmans 2009; Vegetti et al. 2010, 2012; Nierenberg et al. 2013; Vegetti et al. 2014; Hezaveh et al. 2016; Despali et al. 2016; Gilman et al. 2017; Vegetti et al. 2018; Chatterjee & Koopmans 2018; Ritondale et al. 2019). Observations of such lenses provide important calibrations for N-body cosmological simulations (e.g., Peirani et al. 2019; Mukherjee et al. 2021) and allow deeper higher resolution views of faint distant galaxies otherwise too faint to be studied (e.g., Paraficz et al. 2018). However, due to the rarity of lens systems, many of these studies are limited by small sample sizes, prompting targeted lens searches by the community.

Lens searches can be divided into two broad classes: source-selected and lens-selected. The first requires follow-up of a known high-redshift source, in the hope of observing signs

* Tables F.0 and D.1 are only available in electronic form at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via <http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/A+A/>

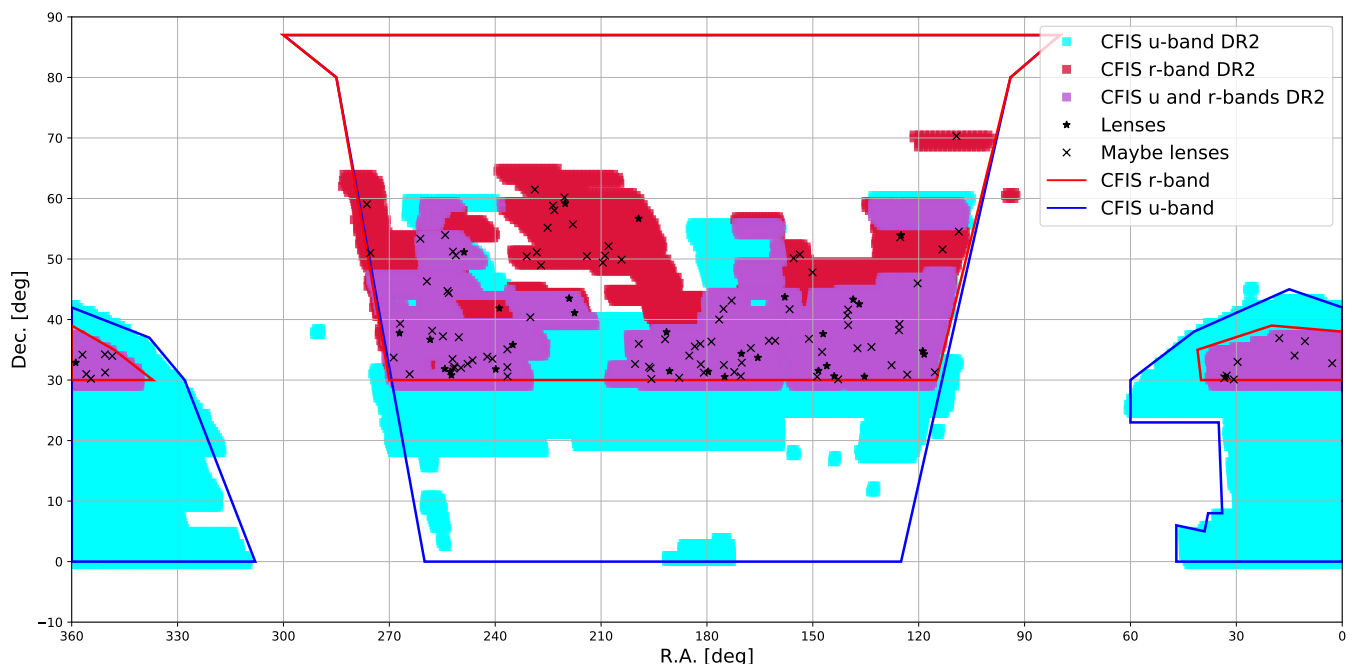


Fig. 1. Illustration of the planned and current CFIS footprint. The contours of the final CFIS-*r* footprint and the CFIS-*u* footprint are shown in red and blue, respectively. Shown are the current areas covered in the *r* band for DR2 (in red), in the *u* band (in blue), and where *r*-band and *u*-band data are available simultaneously (in violet). In this work the *u* band is used, when available, for visual inspection purpose only. Also indicated is the position of the 32 highest quality candidates (stars) and the 101 maybe lenses (crosses) obtained after the joint visual inspection of the CNN-committee candidates. Of these candidates 104 are new.

of strong lensing. Examples include early lens searches, such as the Cosmic Lens All-Sky Survey (CLASS, Myers et al. 2003; Browne et al. 2003) and the SDSS Quasar Lens Search (SQLS, Oguri et al. 2006). In contrast, lens-selected searches look for signs of a lensed high-redshift source in imaging or spectroscopy in known samples of massive galaxies. Some of the most well-studied lens systems come from the Sloan Lens ACS survey (SLACS) (Bolton et al. 2006, 2008), the BOSS Emission Line Lens Survey (BELLS, Brownstein et al. 2012), and the SLACS for The Masses Survey (S4TM, Shu et al. 2015), which pre-selected high velocity-dispersion galaxies from SDSS, with signs of emission lines of a higher redshift source blended in the spectra. High-resolution HST imaging subsequently confirmed many of these systems as lenses. Lens searches are not limited to the visible domain. Similar efforts are done at longer wavelengths, particularly in the submillimeter regimes, even if the search methods are then very different and mostly done by taking advantage of the magnification bias at the catalog level. Typical examples are searches in the Herschel Astrophysical Terahertz Large Area Survey (Bussmann et al. 2013; Wardlow et al. 2013; Nayyeri et al. 2016; Negrello et al. 2017), the South Pole Telescope Survey Data (SPT) (Vieira et al. 2010, 2013), and the Planck all-sky survey (Cañameras et al. 2015). With numerous ongoing and upcoming wide-field imaging surveys, such as *Euclid* (e.g., Laureijs et al. 2011; Amiaux et al. 2012), *Roman* (Spergel et al. 2015), and the Rubin Observatory Legacy Survey of Space and Time (LSST) (e.g., Ivezić et al. 2019), lens samples can be built directly from imaging data. Previous imaging-only samples have come from visual inspection of HST images (Faure et al. 2008; Pawase et al. 2014) or ground-based imaging with the help of citizen science (e.g., Sonnenfeld et al. 2020). With increasing depth and survey areas, visual searches alone become unsustainable and require automated techniques to condense the sample size. Early work included ring-finding algorithms, as was

done with the SL2S sample (Cabanac et al. 2007; Gavazzi et al. 2012) or model-aided search in HST (Marshall et al. 2009) and HSC (Chan et al. 2015). More recently, machine-learning methods have been applied to ground-based surveys, including the Kilo Degree Survey (KiDS, e.g., Petrillo et al. 2017; Kuijken et al. 2019; Petrillo et al. 2019); the Dark Energy Survey (DES, e.g., The Dark Energy Survey Collaboration 2005; Jacobs et al. 2019b,a), and the Hyper Suprime-Cam SSP Survey (HSC, e.g., Aihara et al. 2018; Sonnenfeld et al. 2018; Wong et al. 2018; Sonnenfeld et al. 2019; Chan et al. 2020; Sonnenfeld et al. 2020; Cañameras et al. 2020; Jaelani et al. 2020, 2021).

The application of machine-learning techniques to lens searches encompasses a wide range of methods, from support vector machines to deep neural networks. However, in recent years convolutional neural networks (CNNs) have emerged out thanks to their well-tested reliability for image classification (He et al. 2015). In particular, the top five algorithms of the first strong gravitational lens finding challenge (Metcalf et al. 2019) were mainly CNN-based. In this challenge CNNs were able to recover 50% of the lenses. However, false positives, such as ring galaxies, spirals, mergers, or galaxies with companions, were a severe problem for CNNs trained with overly simplistic simulations. Using a training set that is as realistic and exhaustive as possible is thus crucial. The current number of known lenses in the CFIS footprint is very small, which complicates the composition of a training set for machine-learning algorithms. For this reason, we still use simulations in this paper. However, recent lens searches conducted in the Dark Energy Spectroscopic Instrument Legacy Imaging Surveys' Data Release 8 have proven the possibility to train neural networks with a small number of lenses (Huang et al. 2020, 2021). We therefore hope to reuse our lens candidates to train the next versions of classifiers with real lenses only or a mix of simulations and real lenses. A typical approach to producing training sets for CNNs is to use entirely

synthetic images, as in [Jacobs et al. \(2019b\)](#), where the lens light, the lens mass, and the light profile of the source are analytical. Random images are then taken from real data and added to the simulated “clean” lenses to introduce instrumental effects and more realistic features, such as companions around the central galaxy. The main advantage of this approach is the ability to control the distribution of the lensing parameters. However, it may be difficult to reproduce all the complexity of real lenses with this method.

One approach to mitigating this consists in creating training sets that combine images of real foreground galaxies with simulated sources, as in [Petrillo et al. \(2017\)](#) and [Pourrahmani et al. \(2018\)](#). In this case the training set contains deflectors with more realistic light profiles. In the present work we go one step further toward more realistic simulations by also using a real image for the background source, as in [Cañameras et al. \(2020\)](#). The main difference with [Petrillo et al. \(2017\)](#) and [Pourrahmani et al. \(2018\)](#) is that in our case only the lensing effect is simulated; the shape and the light profile of both the deflector and the source are taken from real data. In general, any classification performed by the CNNs is imperfect and candidates must be confirmed by follow-up observations. For efficient use of telescope time, the candidates are first visually inspected in order to remove the most obvious false positives. The timescale for this visual inspection must remain reasonable, in particular for future large-scale surveys. Therefore, it is crucial to keep a very low false positive rate. A common way to reduce the occurrence of false positives is to increase the proportion of the most common misleading objects in the negative training set (e.g., [Cañameras et al. 2020](#)). However, these must be taken as often as possible from real data. Simulating negative examples is not reliable enough, and using modified images from other surveys involves re-sampling and PSF mismatch, not to mention K-corrections and evolution effects due to redshift mismatch between surveys. We thus believe it is important to provide a catalog of false positives taken from our specific data. These false positives can be used to improve the training sets of future searches in CFIS.

In this paper we use CNNs to look for lensed galaxies in 2 500 deg² of the excellent-seeing *r*-band imaging of CFIS. This complements past and ongoing searches mostly carried out in the south. We first describe the data in Sect. 2, along with our machine-learning method and simulation pipeline in Sect. 3, and then carry out a visual inspection of the machine-classified objects in Sect. 4 to remove false positives. In Sect. 5 we present a single-band method based on auto-encoders that separates the lens and source light and enhances the contrast of each component. We additionally carry out automated mass modeling of the best candidates in Sect. 6, and derive basic properties of the lens and source populations. Finally, in addition to our best candidates, we provide a catalog of false positives, which will be useful for future lens searches based on neural networks. To our knowledge, it is the first time a full lens-finding and modeling pipeline has been presented for single-band data. Although it specifically targets galaxies lensed by luminous red galaxies (LRG), we expect to make it more general when looking for lenses in the full CFIS footprint and for all types of lens galaxies.

2. Data

The Canada France Imaging Survey (CFIS) is an ongoing legacy survey using the Canada-France-Hawaii Telescope (CFHT), a 3.6 m telescope at the summit of Mauna Kea in Hawaii. CFIS is a component of the multi-band Ultraviolet Near Infrared Optical Northern Survey (UNIONS). This effort will provide the nec-

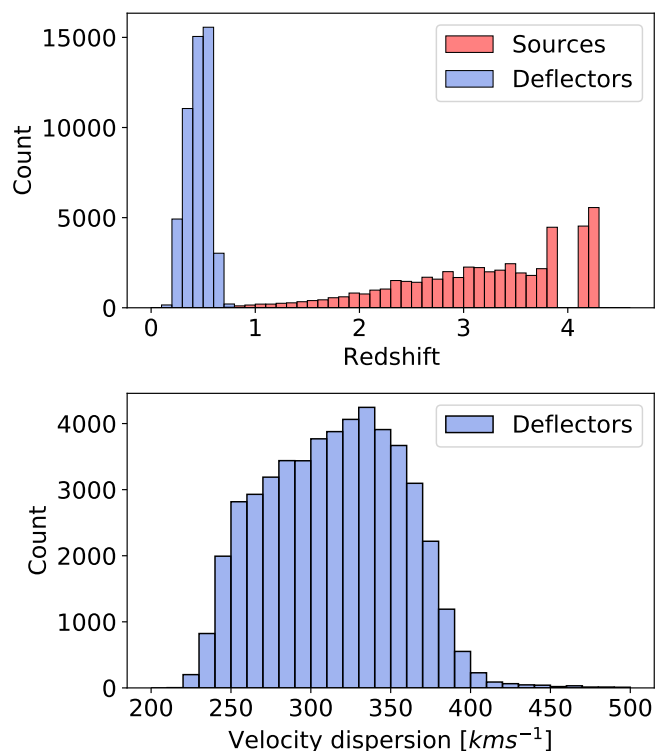


Fig. 2. Statistics of the deflectors and sources used in the lens simulations. *Upper panel:* Redshift distribution of the sources and deflectors used for the lens simulations. *Lower panel:* Measured velocity dispersion distribution for the deflectors.

essary ground-based optical counterparts of the forthcoming *Euclid* space mission, along with stand-alone immediate scientific applications ([Ibata et al. 2017](#); [Fantin et al. 2019](#), [Guinot et al. in prep.](#)). When completed, the survey will have imaged 8 000 deg² of the northern sky in the *u* band (CFIS-*u*) and 4 800 deg² in the *r* band (CFIS-*r*). The imaging data used in this work are from the CFIS Data Release 2 (DR2; see Fig. 1), covering around 2 500 deg². CFIS-*r* has exquisite image quality with a median seeing of 0.6″, down to a depth of 24.1 (point source, 10 σ rms). CFIS-*u* has a median seeing of 0.8″ to a depth of 23.6 (point source, 10 σ rms). Here we use the CFIS-*r* footprint to search for lenses with CNNs based purely on high-resolution morphological information. Since not all *r*-band images have *u*-band counterparts, the CNNs use only *r*-band information. However, CFIS-*u* is used, when available, to refine the sample of candidates found through the CNNs by visual inspection. For each target, we have also produced models of the point spread function (PSF) and its spatial variations across co-added images which were reduced, processed, and calibrated at the Canadian Astronomical Data Centre using an improved version of the MegaPipe pipeline ([Gwyn 2008](#)). For each lens candidate we exploited the model PSF obtained with PSFEx ([Bertin 2011](#)) to produce an image of the local PSF oversampled by a factor of 2. Weight images along with other data quality diagnostics are also produced for each candidate in each of the available bands. Details of the spatial variations of the PSF may not always be well accounted for in such a model performed on stacked data, involving largely dithered exposures. This is, however, not a major issue for the strong lens modeling applications in this work.

Our goal is to provide an automated pipeline to find, deblend, and model high-quality galaxy-scale lenses. Our sample is by construction lens-selected, meaning that we look for lensed sys-

tems among a large sample of pre-selected LRG. These objects are bright and massive, and are therefore expected to have the largest possible lensing cross section. (Turner et al. 1984).

2.1. Data selection for the lens search

A reliable color selection of LRGs is not possible with CFIS data alone. Even with color information, it is necessary to account for the fact that LRGs acting as lenses have colors biased toward the blue with respect to LRGs without lensing features: lensing LRGs are blended with the lensed image of a background galaxy (which is often blue). Fortunately, the CFIS footprint is entirely included in the first part of Pan-STARRS1 (PS1), hence we used PS1 to carry out our color selection, thus accounting for the blue bias. This has already been implemented by Cañameras et al. (2020) and consists of a color cut in the PS1 3π catalog, broadly matching the aperture magnitudes and colors of 90 000 Pan-STARRS simulations of lensing systems. This photometric selection is very large, since it was designed to include 96% of the mock lens, and thus hopefully all LRG lens galaxies. In return, however, it may contain a large number of interlopers such as spirals and rings due to the Pan-STARRS data quality. After a cross-match with this catalog, we obtained 2 344 002 images to carry out our lens search.

2.2. Data selection for the simulated training set

Our simulation set is constructed from real data (i.e., a deflector from CFIS imaging data and a background source from HST images), as described in Sect. 3.2. The image stamp size in this work is $8.17''$ per side corresponding to 44 pixels.

The selection of deflectors is taken directly from Cañameras et al. (2020). This corresponds to a subsample of the LRG spectroscopic sample (Eisenstein et al. 2001), which uses color-magnitude cuts to select intrinsically red and luminous galaxies. They have SDSS spectra, and thus also velocity dispersion (σ_*) and redshift (z) estimates. In Fig. 2 we summarize the spectral properties of our selection, which spans the ranges $200 < \sigma_* < 500 \text{ km s}^{-1}$ and $0.1 < z < 0.7$. After a cross-match with the whole CFIS- r catalog from DR2, we obtained 624 170 LRG images, which form the basis of our training set.

The background galaxies were taken from the sample of Cañameras et al. (2020). We used galaxy morphologies from HST/ACS F814W images and converted to r band using HSC ultra-deep stacked images. The original stamps have a size of $10''$ per side and the same pixel size as the HST/ACS F814W image (i.e., $0.03''$). Since the PSF of the HST images is much sharper than the CFIS PSF, we neglected its effect during the simulation process and we did not attempt to deconvolve the HST images from their PSF. All sources are included in the COSMOS2015 photometric catalog (Laigle et al. 2016) and in the Galaxy Zoo catalog (Willett et al. 2017). The redshift information for our sources, when available, was obtained from public spectroscopic catalogs (Lilly et al. 2007; Comparat et al. 2015; Silverman et al. 2015; Le Fèvre et al. 2015; Tasca et al. 2017; Hasinger et al. 2018). When no spectroscopic redshift was available, the best photometric redshift estimate from Laigle et al. (2016) was used. Then with all this information combined we obtained high-resolution r -band images of unlensed sources with known redshift. These selected foreground LRGs and background HST sources provided the basis for building our training set for the CNN search, as described in Sect. 3.2

3. Method

One way to address the problem of lens detection is to consider it as a binary image classification task where the positive class members are the lenses and the negative class members are the galaxies without lensing features. CNNs are especially suited to this task as they are able to detect local correlations of two-dimensional features in images (Lecun et al. 2015). The convolutional layers of CNNs can be understood as a set of kernels that act as specific feature detectors.

3.1. Classifier

In this work we use a recent class of CNNs called EfficientNets (Tan & Le 2019). They outperform the most common CNN architectures on the classification of images from different standard data sets while using a smaller number of parameters (see Fig. 1 of Tan & Le 2019). This is achieved by scaling uniformly the depth, width, and resolution as a function of the available computing resources.

In our case we did not scale the models ourselves, but used the models already implemented in the Keras application programming interface (API) (Chollet et al. 2015). It includes eight versions of EfficientNet, named B0 to B7, depending on the number of free parameters involved. The dimensions of our images, 44 pixels per side, are much smaller than the dimensions of the images of the standard machine-learning data sets used in Tan & Le (2019). Therefore, we used the B0 architecture, which contains the smallest number of parameters of the eight models. The B0 architecture, from the Keras API is pre-trained on ImageNet data (Deng et al. 2009). We took advantage of this pre-training by reusing the parameters of the trained model instead of initializing them randomly, which allowed us to speed up the training. However, since the dimensions of our images are different, we adapted the size of the first layer and the last fully connected layers and randomly initialized the parameters of these layers. The classification was performed using the so-called ensemble-averaging method. This consists of separately training models with the same architecture but different subsets of the training set and combining their results in order to reduce the variance of the predictions. In the following we call the set of models a “committee” and each individual model a “member” of the committee. Here, we use three separate instances of EfficientNet B0 as our committee members. We also tested versions of the committee with more instances. This did not lead to a significant improvement in the quality of the classification.

3.2. Design of the training set

Since we performed a binary classification we needed to build a training set containing images with either positive or negative labels. The negative examples were drawn randomly from the lens search sample described in Sect. 2.1. The negative examples may contain a few real lenses, but we expected that this would have only a marginal effect on the performance of the network since the prevalence of gravitational lenses is very low. Creating a set of positives examples required more preparation. Not enough lenses have been confirmed in the CFIS footprint to build a training set that spans the full diversity of lens systems. We therefore generated a set of simulated lenses using the pipeline described in Schuldt et al. (2020). We present below the most important steps of this process.

We first selected an LRG image from the deflector catalog constructed in Sect. 2.2 and assigned a mass profile to the se-

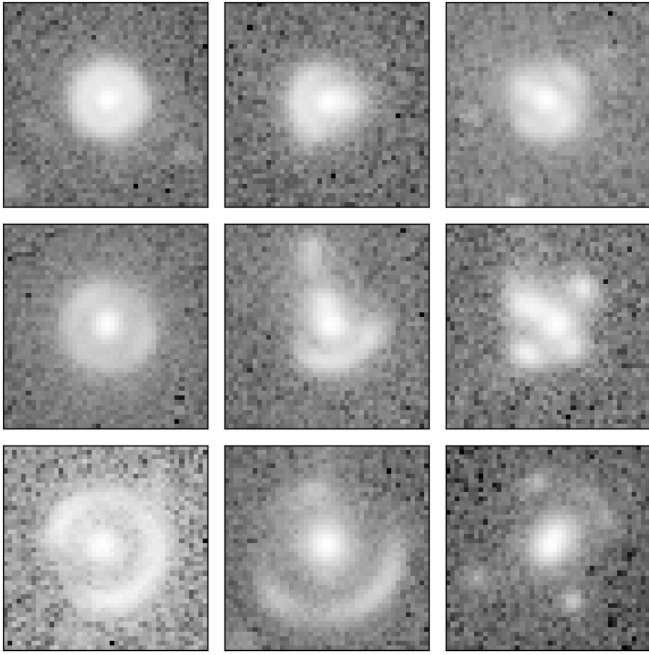


Fig. 3. Examples of simulated galaxy-scale systems of the CFIS survey. Each stamp is the sum of a real CFIS r -band image of a galaxy, to which a lensed HST galaxy convolved with the CFIS PSF is added. Each stamp is $8.17''$ per side, and the pixel size is $0.18''$.

lected galaxy assuming a simple parametric model, the singular isothermal ellipsoid (SIE) (Kassiola & Kovner 1993; Kormann et al. 1994). This mass model has five free parameters: the Einstein radius, the coordinates of the lens center, the ellipticity (or axis ratio), and the position angle (PA). The lens center coordinates were fixed to the center of the deflector image, whereas the values of the axis ratio and the position angle were derived from the second moment of the lens light profile. In this model we note that the ellipticity and PA are assumed to be the same for the light and mass distribution.

A source was then randomly selected from the source catalog. Knowing both the deflector redshift and velocity dispersion, we computed the Einstein radius and checked, given the redshift of the source, that it fell in the range $0.8'' < R_E < 3.0''$. The lower limit was chosen to prevent lensing features becoming blended with the deflector light. If the Einstein radius was outside the given range, we randomly selected another source from the catalog. We repeated this until a matching source was found, otherwise after 100 iterations we increased the velocity dispersion (and hence the Einstein radius) of the deflector by 50% and repeated the process. The goal was to obtain a sufficient number of simulations. If no match was found after increasing the velocity dispersion, we discarded this deflector from our catalog. Since this boost involved only a few objects with small velocity dispersion at the lower end of the distribution shown in Fig. 2, we did not expect this to introduce a morphological bias in our training set.

In the next step the position of the source was chosen randomly in the source plane. However, we imposed a total magnification constraint, $\mu \geq 2$. This limit corresponds to the minimum magnification threshold to produce multiple images. Choosing a higher limit gives more striking lensing features, but also artificially increases the proportion of Einstein rings among the simulations. This may bias the classifier toward this class of objects

or even lead to more false positives, such as ring galaxies. Only the positions resulting in a magnification $\mu \geq 2$ were considered, hence constraining the source to be within or close to the caustics and resulting in multiple images. Once the source position was chosen, we computed a high-resolution image of the lensed source using the GLEE software package (Suyu & Halkola 2010; Suyu et al. 2012).

As a final step, the CFIS PSF was re-sampled to the HST pixel size and the image of the lensed source was convolved with this re-sampled PSF. The result of the convolution was then down-sampled back to the pixel size of CFIS and added to the deflector image. Our simulations are therefore a hybrid between simulations and real data (i.e., built from the CFIS data themselves for the lens and from deep HST images for the source).

For some of the simulations produced with this method the lensing features are too faint or too heavily blended with the deflector light. Including images with indistinguishable lensing features may increase the false positive rate. Therefore, we used only the simulations for which the sum of the brightness of all pixels of the lensed source was at least 20 times the mean rms value of the sky noise measured in the four corners of the deflector image. We then proceeded to a rough visual inspection of all simulations above this threshold to remove images with lensing features blended in the deflector light. This resulted in 10 600 accepted lens simulations, of which we show a few examples in Fig. 3. This number is relatively small. However, since the precision and recall on the validation set are close to perfect, we do not expect that increasing the size of the training set would have a significant impact on the performance measured on validation data.

3.3. Pre-processing and training

Before being passed to the CNNs, all images were normalized so that the full dynamical range lies between 0 and 1. We also applied a logarithmic stretch in order to enhance the contrast of the lensing features. After this pre-processing, the data were separated into three different sets: (1) the training set (80 % of the training data); (2) the test set (10 %), and (3) the validation set (10%). The validation set is used both to monitor the training process and to define the conditions to end the training, whereas the test set is used only at the end of the training to evaluate the performance of the committee.

Our images were only available in the r band, but the EfficientNet architectures from the Keras API are built to handle three-band images. Hence, we transformed our single-band images into a three-band data cube by duplicating three times the images before passing them to the network. In doing so, we were able to use the pre-trained version of the network, allowing us to shorten the training time. We trained the three members of the committee independently (each member being an instance of the EfficientNet B0 architecture), each with a different subset of the training set. The subsets were constructed using a different fraction of lenses; the fraction was drawn randomly in the range 0.2 to 0.5 to mitigate the tendency of the network to learn the fraction of positive examples seen in the training set. The instance trained with the lower fraction will be less optimistic and find fewer lenses, but will reach a lower false positive rate. However, the fraction of lenses is set to 0.5 in the validation and test sets.

To train the CNNs we performed a mini-batch stochastic gradient descent using binary cross-entropy as a loss function and an Adam optimization (Kingma & Ba 2014). The batches contained 128 images picked from the data set and flipped randomly along the x - and y -axes using the Data Augmentation method

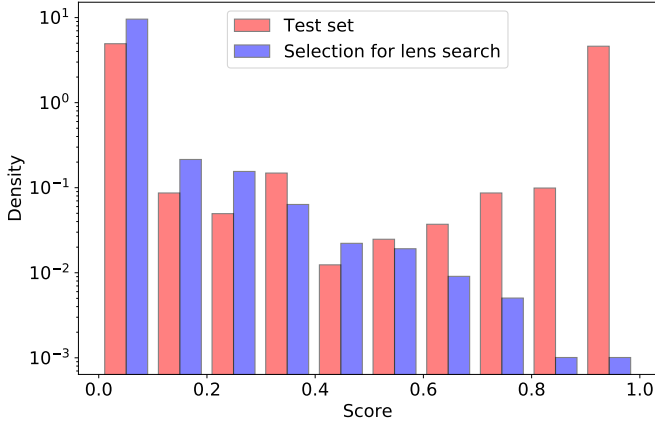


Fig. 4. Probability density distribution of the score assigned by the committee for the test set (red), containing 50% of lens simulations, and 50% of galaxies without lensing features taken from the lens search sample. In blue are shown the scores predicted by the network on the data set containing real images described in Sect. 2.1. In the lens search sample the proportion of lenses is by definition unknown.

from Keras. Overfitting is one common pitfall encountered during the training of machine-learning algorithms. It occurs when the algorithm learns the specifics of the training data, and thus is not able to generalize on new data. When overfitting occurs, the classification error on the training data becomes very small, while the error on previously unseen data starts to grow. Since the batches contain different images each time, the data augmentation procedure allows us to artificially increase the size of the training set and limit the risk of overfitting. Since our network was previously trained on ImageNet data, we only needed to fine-tune the parameters. Therefore we started directly with a relatively low learning rate of 10^{-4} .

The maximum number of epochs, which is the number of times the machine-learning algorithm is allowed to see the entire training data set, was fixed at 200. However, in order to optimize the training time, we used the early stopping procedure from the *Keras* API, which interrupts the training before reaching the maximum number of epochs if the validation loss is no longer improving. More precisely, the training stops if the validation loss reaches a plateau, or if it increases during ten consecutive epochs, or if neither of the other two conditions is met after 200 epochs. Using early stopping with a validation set allows us to interrupt the training if the classifiers start to overfit the training data. At the end of the training, the weights and biases are restored to the epoch achieving the smallest validation loss. The three members of the committee are combined such that the final output corresponds to the mean of all the outputs assigned by the three independent networks.

3.4. Candidate detection and performance of the classifier

Convolutional neural networks are generally not invariant under rotation and the final output of the committee, hereafter referred to as the score, can change significantly for the same image if this image is rotated in different ways. In some extreme cases, an image with a high score can even fall under the selection threshold after a rotation due to statistical fluctuations. In order to mitigate this effect, we rotated and flipped all images in seven different ways: three rotations of 90, 180, and 270 degrees, and flips along the x -axis of all rotations including the flip of the original unrotated image. We then considered the mean of the scores

given by the committee in all directions as the final score. The final scores of the committee range from 0 to 1. If the classifier were ideal, we would expect the scores of galaxies without lensing features to be 0 and the scores of the lenses to be 1. Figure 4 shows the distribution of scores assigned by the committee on all images from the lens search sample and for the test set. The test set distribution is not bi-modal, indicating that our classifier is not perfect. In the case of the lens search sample the distribution has only one peak centered on zero, and decreases exponentially afterward. This can be explained partially with the very low prevalence of lenses in the lens search sample and by the fact that the committee may less easily identify lenses than simulations, as explained below. The performance of the network is evaluated using two metrics on the test set: Precision (P) and recall (R). The precision or purity indicates the fraction of true lenses among all images labeled as lenses, whereas the recall, also called “true positive rate” or “completeness,” gives the fraction of true lenses recovered by the committee among all the true lenses of the training set. They are defined as

$$P = \frac{TP}{TP + FP}, \quad (1)$$

$$R = \frac{TP}{TP + FN}, \quad (2)$$

where TP , FP , and FN are the number of true positives, false positives, and false negatives, respectively.

The number of true positives, true negatives, false positives, and false negatives from Eqs. 1 and 2 depend on the score we chose as a cutoff threshold (i.e., the score above which the images are considered to be lens candidates). We show in Fig. 5 the precision and recall values for all cutoff thresholds between 0 and 1. Choosing a high cutoff threshold increases the precision as the number of false positives decreases, but lowers the recall since fewer true positives are included. In Fig. 5 we observe that the precision and recall on the test set stay fairly high, independently of the cutoff threshold. Since a large part of the contaminants in CNN-based lens searches are spiral galaxies, we estimated the proportion of spiral false positives for each cutoff threshold. Therefore, we evaluated 8 200 CFIS images of spiral galaxies with our trained CNNs. All of these spirals were taken from the Galaxy Zoo catalog (Willett et al. 2017). The fraction of spirals mislabeled as lenses as a function of the cutoff threshold is presented in Fig. 6 and shows that the contamination rate falls below 0.001 for any CNN score above 0.5. For scores higher than 0.8 there are no spiral false positives.

Taking into account both the precision and recall curves, we chose a cutoff threshold of 0.5 for our lens search in the real CFIS data. This results in a precision of 1 and a recall of 0.96 on the test set. However, these results must be interpreted with caution, since the committee may have learned to recognize the simulations, and the performance may decrease on real data. In addition, it should be kept in mind that the real occurrence rate of strong lensing events is very low. The probability for an image classified as a candidate to be a real lens ($P(L|C)$) can then be deduced from the Bayes rule to take into account the occurrence rate ($P(L)$) of lenses in the data set as

$$P(L|C) = \frac{P(L)P(C|L)}{P(L)P(C|L) + (1 - P(L))P(C|NL)}, \quad (3)$$

where $P(C|L)$ and $P(C|NL)$ are the precision and false discovery rate obtained on the testing set, respectively. If $P(L)$ is very low,

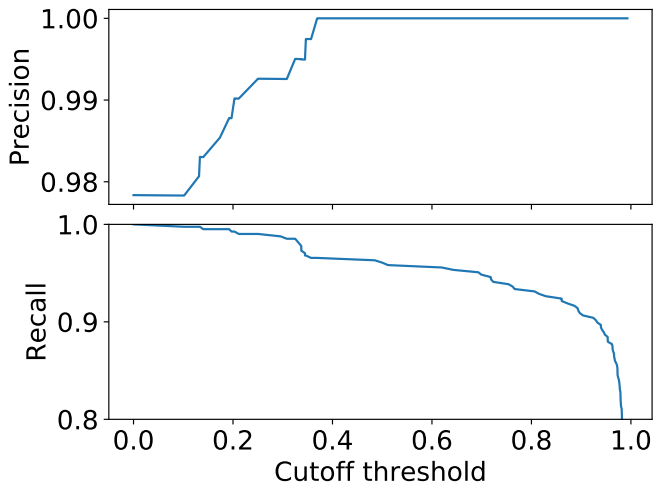


Fig. 5. Precision and recall as functions of the cutoff threshold in CNN scores applied to a test set composed of 1 060 galaxies without lensing features taken from our selection for the lens search and 1 060 lenses taken from our simulation set.

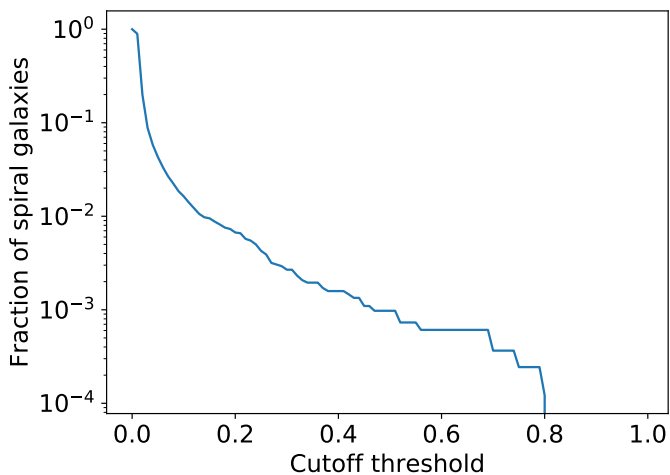


Fig. 6. Fraction of CFIS *r*-band images of 8 200 Galaxy Zoo spiral (Willett et al. 2017) galaxies mislabeled as lenses as a function of the cutoff threshold on the CNN-committee score.

$P(L|C)$ will stay low even when the precision is very high. This effect, called the base rate fallacy, is known to limit the performance of intrusion detection algorithms (e.g., Axelsson 2000) and may be non-negligible for lens detection.

4. Visual inspection of the candidates

Of the 2 344 000 images selected for our lens search, 9 460 obtained a CNN score higher or equal to 0.5. All these candidates were visually inspected independently by the six authors of this paper (E.S., B.C., K.R., F.C., J.C., and G.V.). We separated the candidates following the score attributed by the committee into five CNN-score bins of 0.1 in size and inspected them separately. This separation may introduce some biases in the visual inspection, meaning that the users may give more optimistic grades for objects in the bins corresponding to the highest score. In order to have a consistent visual inspection between the users and the bins, we defined common guidelines and designed a quick

and flexible python tool¹ to display the images in two different ways: a mosaic tool to display a large number of images simultaneously and a single-object display tool to review each object in detail. Both tools can handle single-band and color images.

For each score bin, we first performed a crude pre-selection of potential lenses using the mosaic tool. This tool displays a grid of 10×10 cutouts so that users can quickly flag any of the systems. The images are arranged randomly in the mosaic for each user, hence minimizing any biases depending on the order and positions in which the images are shown, for example due to varying attention levels of users through the inspection process. We also flagged misclassified ring galaxies in this pre-selection in order to improve our set of negative examples for future lens searches.

At the end of this pre-selection, we obtained a list of possible lens candidates and ring galaxies. This sample unavoidably contained other objects, such as spirals and interacting galaxies that resemble lenses, which were inspected in more detail in a second step (see below). In other words, this first step rejects any object that can be immediately demoted to a false positive and keeps the rest. We then considered the union of all objects selected by at least one of the six users (4 626 images).

For all images selected during the mosaic inspection, we proceeded with a more detailed inspection using our single-object tool. This tool displays one single stamp at a time, but offers a dynamic contrast control, more classification options, and a direct link to the Legacy Survey (LS; Dey et al. 2019) cutouts, when available. The LS data are shallower than the CFIS data, but have color information. We note that the LS color images are not displayed systematically during the classification process to avoid the users' decisions being driven by the color information. They are rather used to support the *r*-band CFIS classification when a candidate requires further data to make a decision. In this step we classify the images into four categories: secure lenses (SLs), maybe lenses (MLs), single arcs (SAs), and non-lenses (NLs).

Our single-object classification follows the same guidelines as in Rojas et al. (2021), where the SL category includes images displaying obvious signatures of strong-lensing features, such as multiple images or arcs with counter images. The second category, ML, corresponds to images that exhibit structures compatible with lensing, but that would require further investigation with lens models or follow-up with higher resolution data, higher signal-to-noise data, or spectroscopy. When single arcs with small distortions and no counter-images are seen, we use the category SA. Naturally, the SA category contains fortuitous object alignments or galaxies with a curved shape that may not be due to lensing. All the images that do not belong to the first three categories are labeled NLs. The NL objects can be subclassified into three subcategories: rings, spirals, and mergers. However, these subcategories are used by the users only if they are very confident about their classifications. This allows us to obtain a catalog of the most common false positives as a by-product of this search. These samples are valuable for future searches with CNNs being trained against false positives that can be reliably identified. At this stage, however, the size of this false positive sample remains too small to retrain the CNNs used in this work.

We obtained a total of 1 423 images selected as MLs or SLs by at least one user (out of the six users). However, the agreement between the potential lens candidates selected by the dif-

¹ <https://github.com/esavary/Visualisation-tool>

ferent users remained low after the two-step visual inspection. More details about the classification are given in Appendix A. In order to obtain an agreement, we reinspected all images selected by at least one user in a joint visual inspection session. In doing so, we considered all 1 423 objects that were identified as SL or ML by at least one person and chose one unique grade for each object after a discussion among all the users. In this last part, we also showed the image of the candidates in the u band when available, and used only the categories SL, ML, and NL to further clean the sample. It should be noted that this process is very selective as we require the agreement of all users to grade an image as SL or ML.

After the joint inspection we obtained 32 objects classified as SLs and 101 MLs, which are shown in Fig. 7 and Fig. C.1, respectively. These represent 1.4 % of all the candidates selected by the committee of CNNs. After a cross-match with Vizier (Ochsenbein et al. 2000), Simbad (Wenger et al. 2000), the Master Lens database, and the catalogs of various lens-finding papers with candidates or confirmed lenses included in the CFIS footprint (e.g., Cañameras et al. 2020; Chan et al. 2020; Jaelani et al. 2020; Huang et al. 2021; Talbot et al. 2021), we obtained 15 new SL and 89 new ML candidates. The results of the cross-match and the coordinates of the candidates are presented in Table F.0. With our visual inspection, we also obtained three catalogs with 238 mergers, 361 ring galaxies, and 950 spiral galaxies identified by at least one user, which can be used to expand our negative sample for future searches. Examples of each category are shown in Fig. E.1. Because we consider the unions of all votes to include a candidate in our false positive list, there is a small overlap between the three catalogs. Among all images labeled as ring or spiral by any user, 11 were finally classified as ML or SL after the joint inspection step. We therefore removed them from the final spiral and ring catalog. The relatively small number of ML and L candidates in our final catalog in comparison with the number of rings, spirals, and mergers can partially be explained by the visual inspection method: we require a unanimous decision of all users during the joint inspection to include objects in the ML or L, whereas only one vote is sufficient for an image to be included in the false positive catalogs.

5. Lens–source deblending with auto-encoders

Independently of the lens search itself, it is desirable to provide reliable deblending of the lens and source light of the candidates without relying on a lens model. First, deblending reduces the dynamical range of the data and allows faint structures to be seen more clearly either in the lens or in the source. Second, it allows the remeasurement of clean photometry of the lens and source for future photometric redshifts estimates when color information become available (e.g., from public release of other surveys). Finally, it can be used to initialize lens model parameters when implementing composite profiles with both stellar and dark mass.

Rojas et al. (2021) propose a method for deblending lens candidates based on the scarlet² (Melchior et al. 2018) and MuSCADeT (Joseph et al. 2016) algorithms. However, this method is not directly applicable to our case since it requires color images. Therefore, we present here a fully data-driven alternative approach based on a class of neural networks called auto-encoders, with the goal of deblending our 32 SL candidates. In general, neural networks find a mapping, $Y = f(X)$, between the inputs X and the labels Y . In the case of auto-encoders the labels

are the inputs themselves. In other words, the mapping made by the auto-encoder rather writes as $X = f(X)$. Auto-encoders can be decomposed into two symmetrical parts: the encoder and the decoder. If the dimensions of the layers decrease from the two ends of the auto-encoder to the central layer, the network is able to learn a simplified representation of the original input. Thus, auto-encoders may be used for data compression, feature learning, dimensionality reduction, and denoising. The architectures derived from auto-encoders, like variational auto-encoders, can also be used as generative models, to then generate realistic images of galaxies (Lanusse et al. 2020).

The scheme of the auto-encoder we used for the deblending is presented in Fig. B.1. The input of the network is the image of the lensed system. Unlike traditional auto-encoders, the decoder is split into two parallel parts. The first part extracts the lensed source, whereas the second extracts the deflector image. The dimension of the inner dense layers correspond to the flattened dimension of the last convolutional layer of the encoder and the first of the decoder part. In the end, we obtained three different outputs: the lensed source, the deflector and the lens system reconstructed by the auto-encoder. The reconstructed lens system was obtained by summing the lensed source and deflector images derived with the two different parts of the decoder.

5.1. Training process

We trained the auto-encoder using 10 000 simulations of lenses taken from the sample described in Sect. 3.2 and 5 000 images of LRGs from the spectroscopic LRG selection detailed in Sect. 2. Before the training all images were normalized between 0 and 1. We also set aside 20% of them to constitute the validation set, with the rest as the training set. For each image, we used the following for the ground truth: the image of the lens system itself, and the lensed source and deflector images obtained in the final step of the simulation pipeline before the two images were combined. In the case of LRG-only images, we define the lensed source image as an array of zeros and the deflector image as the image of the LRG. The loss function takes into account the three different outputs and gives more weight to the part containing the lensed source and deflector terms in order to put emphasis on the accurate deblending of the images. It is defined as

$$L = 0.4 \times L_{\text{deflector}} + 0.4 \times L_{\text{source}} + 0.2 \times L_{\text{combined}}, \quad (4)$$

where $L_{\text{deflector}}$, L_{source} , and L_{combined} are respectively the binary cross-entropy losses computed between the true deflector image and the deblended deflector, between the lensed source image and the deblended lensed source, and between the lens system image and the combination of the deblended deflector and the lensed source. We also tested a combination of mean squared error losses instead of binary cross-entropy losses. However, the version trained with mean squared error failed to correctly restore the shape of the lensed source. In most of the case the lensing features appeared incomplete or distorted, or were absent.

We set the maximum number of epochs to 200 and use early stopping to avoid overfitting.

5.2. Secure lens deblending

We present in Fig. 7 the result of the auto-encoder deblending for the 32 SL systems. The auto-encoder correctly captures the general shape of the lensed source features and the deflector. The main advantage of this method is that it does not rely on any assumptions about the light profile of the lens and the source,

² <https://github.com/pmelchior/scarlet>

and is therefore also able to deblend complex lenses. This is true in particular for UNIONS J155923+314712 in which the deflector is an edge-on spiral. However, as is seen in the residual column of Fig. 7, it tends not to correctly deblend some of the high-frequency features from the original images. The auto-encoder clearly distinguishes companion galaxies from lensing features since in general, if companions are reconstructed by the auto-encoder, they appear in the deblended LRG image but not in the deblended lens source image (see Fig. 7). One exception is UNIONS J113952+303204, where one of the companions is very close to the arc of the lensed source and is then mistaken for a lensing feature.

In its current implementation, our method provides reliable photometry of the deflectors, according to tests performed on simulations, but performs less well on the photometry of the lensed source because auto-encoders do not capture all high-frequency features, especially the fainter ones. This results in a loss of flux in some of the deblended lensed sources, and indeed the residuals displayed in Fig. 7 show signal at the locations of small features of the lensed sources. However, we can still obtain reliable photometry of the lensed sources by subtracting the deblended deflectors from the original images and then by carrying out the photometric measurements on the subtracted images. Future work with auto-encoders will focus on better representation of the high-frequency signals contained in the data.

In conclusion, inspecting the auto-encoders results allows us to confirm the presence of promising potential lensing features in our SL sample. The deblended images support our classification for all our SL candidates. In some cases they enhance the visibility of features that are hardly visible in the original low-contrast images, and thus can provide significant help during the visual inspection step of future lens searches.

6. Automated modeling of best lens candidates

Future lens surveys will discover tens of thousands or hundreds of thousands of galaxy-scale strong lenses. From these lenses it will be necessary to define subsamples of objects suited to specific science goals. For example, not all lenses are useful for constraining dark matter substructures and not all lenses are useful for studying galaxy evolution, constraining the initial mass function, or inferring the lens mass-to-light ratio. It is therefore crucial to obtain a basic characterization of the lens and sources properties of the candidates already at the level of the discovery catalog. In the absence of redshift information this boils down to the Einstein radius of the system, its external shear, and the light properties of the lens and of the source. In this section we use our 32 best objects in the SL category as a test bench for a simple automated lens characterization pipeline.

6.1. Modeling pipeline

We model the lens mass, lens light, and source light profiles using simple analytical profiles. We use a SIE profile to model the lens mass distribution to which we add external shear (SIE+ γ_{ext} model). The light distributions of both the lens and the source are represented as single elliptical Sersic profiles. As we show below, these simple models are sufficient to fit most of our lens candidates.

The pipeline is based on the *Lenstronomy*³ Python package (Birrer et al. 2015; Birrer & Amara 2018) and has two main steps, a pre-sampling optimization, followed by a full Markov

chain Monte Carlo (MCMC) sampling. The pre-sampling step uses the particle swarm optimization (PSO) method (Kennedy & Eberhart 1995), which ensures that we initialize the MCMC with model parameter values close to the maxima of the posterior probability distribution. We then perform the MCMC sampling using the *emcee*⁴ package, which is a Python implementation of the affine-invariant Markov chain Monte Carlo ensemble sampler (Goodman & Weare 2010; Foreman-Mackey et al. 2013). While lens modeling is a fairly easy task for isolated lenses, it is complicated by the presence of intervening objects unrelated to the lens, which introduce spurious light contamination. Such objects should be masked to avoid being mistakenly identified as lensed images of the source. This masking procedure is addressed in different ways by different authors. Shajib et al. (2020) modeled 23 lenses from the SLACS sample (Auger et al. 2009), and specifically chose systems that do not contain any contaminating sources of light. Nightingale et al. (2018) did not restrict their sample, but adopted a circular mask with a fixed radius of 3.9'', which selects only the regions of the data dominated by the lensed source light.

In our case the masking algorithm is designed to adapt to systems with very different angular sizes. Figure 8 depicts the steps of the following algorithm:

1. A Laplacian of Gaussian (LoG) filter is applied to the image to highlight areas of the image with a strong flux gradient.
2. All pixels whose flux is below a threshold of $6\sigma_{\text{sky}}$, where σ_{sky} is the rms background noise, are set to zero.
3. All of the nonzero pixels in the filtered image are located, and the locations of the peaks are identified. Peaks are defined as local maxima in the image detected using the `peak_local_max` function of the *skimage* Python package. We require that detected maxima be separated by more than one pixel from each other in order to be considered a peak.
4. The detected objects near the center of the image are assumed to be the lensing galaxy and lensed images and/or arcs from the source light. These objects are used to estimate the angular size of the lens–source system. This is done by sorting the list of detected peaks by their distance from the image center: the first peak corresponds to the lens galaxy–LRG, and the second detected object is one of the images of the lensed source. In order to not mask part of the light from the lensed source image, the lens size is estimated to be eight pixels larger than the distance from the center to the second detected object. All of the brightest pixels farther from the center are treated as contaminant light to be masked. The mask itself is created by using the subset of all nonzero valued pixels from step 3 whose location is farther from the center than the lens system size (red cross-hatched areas outside the black circle in Fig. 8).
5. The final mask is a Boolean array of the same shape as the original image, containing zeros for all pixels that are to be ignored in the modeling and a value of one elsewhere. At each pixel marked with a red plus sign (+) in Fig. 8, the surrounding pixels within a circular area with a radius of two pixels are set to zero. The final mask is shown in the last frame of Fig. 8. For the majority of the 32 SLs we modeled, a visual inspection of the robustness of the automated masked convinced us that it always masked out all the spurious neighboring light that would have been deleted by hand on an object-by-object basis. Only very minor corrections would have been applied (see discussion in Sect. 6.2).

³ <https://github.com/sibirrer/lenstronomy>

⁴ <https://github.com/dfm/emcee>

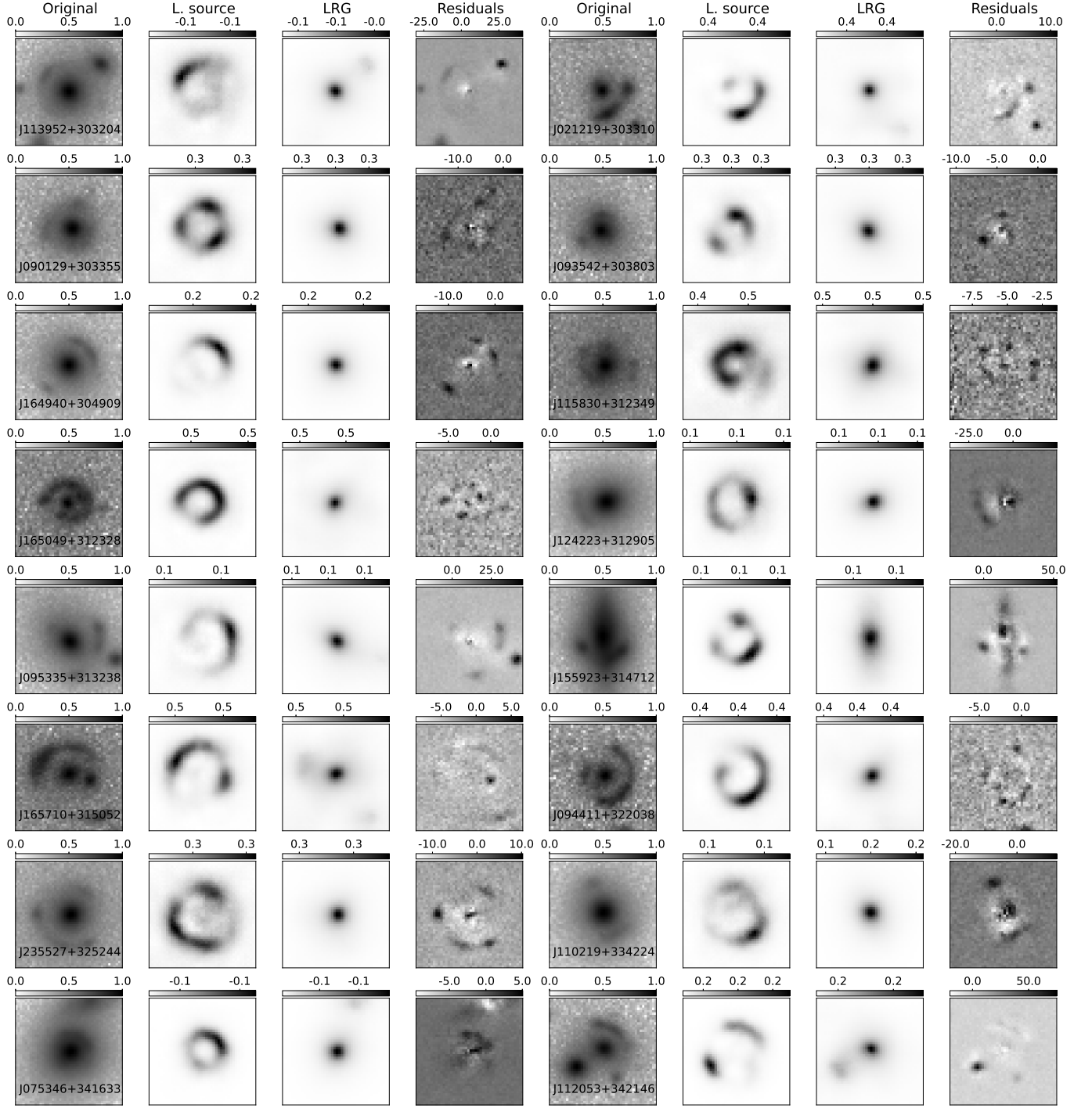


Fig. 7. Results of the auto-encoder debending for our 32 SL candidates. Two objects are displayed in each row. For each object, we show the original image displayed using a asinh grayscale, and the lensed source and lens light debended using the auto-encoder and the scaled residuals.

Before applying `lenstronomy` to our 32 SLs, we adopt realistic priors on the different parameters. First, for the mass and light profiles of the lens, we constrain the axis ratios between the semi-minor and semi-major axes, $q = b/a$, using Gaussian priors centered on a value of $q = 0.8$, with a standard deviation of

$\sigma = 0.1$. This choice was motivated by the results presented in [Kelvin et al. \(2012\)](#), where over 100 000 galaxies of the Galaxy And Mass Assembly (GAMA) survey were modeled, finding distributions in eccentricities peaking at $(1 - q) \simeq 0.2$. In addition, we expect some similarity between the ellipticity of the deflector

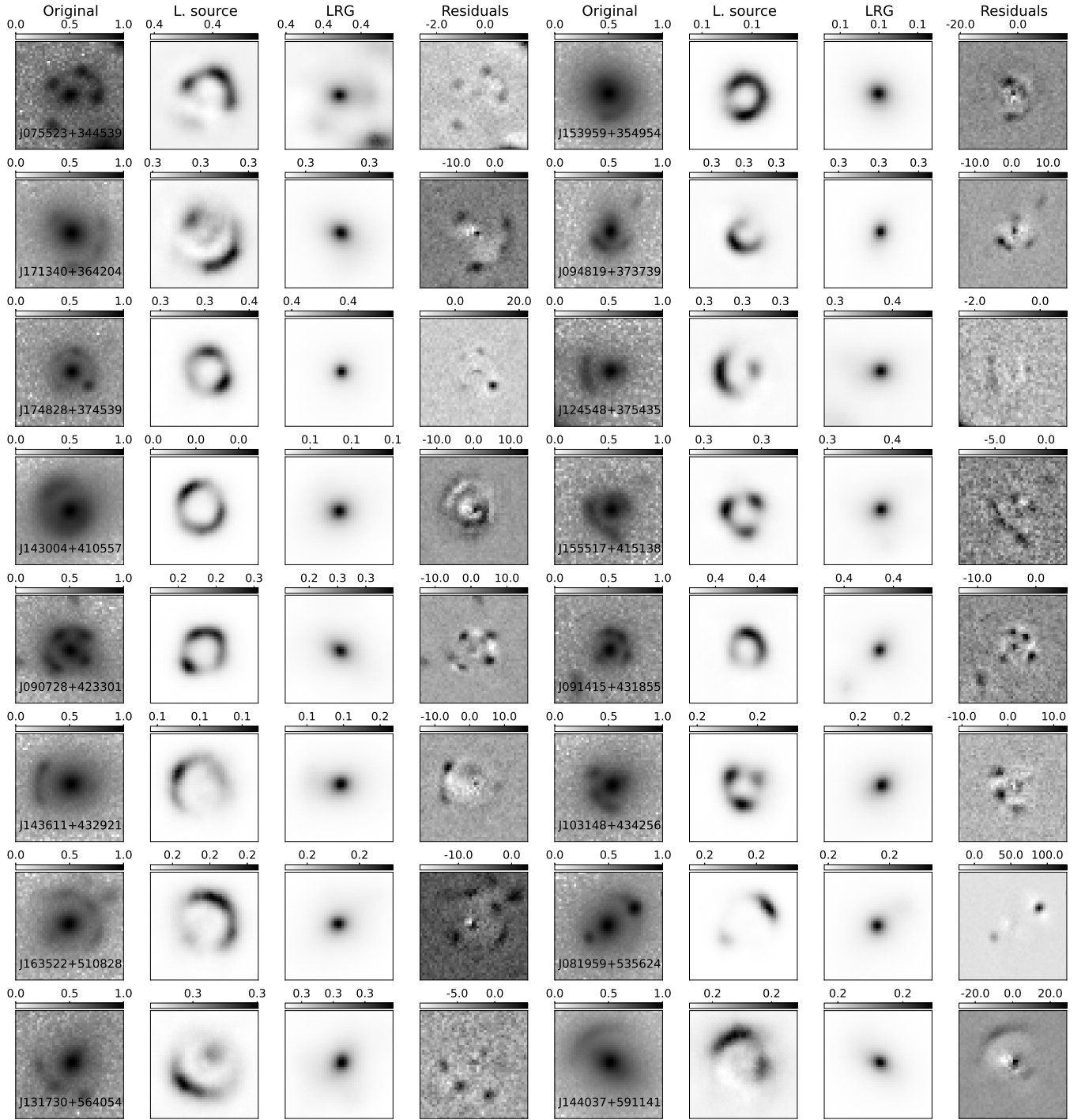


Fig. 7. Cont.

mass and deflector light profiles; however, small deviations are allowed. We therefore apply Gaussian priors, with $\sigma = 0.01$, on both of the ellipticity parameters of the lens mass (e_1^m, e_2^m) that are centered on the corresponding values of the light profile (e_1^l, e_2^l).

In addition to the priors on the ellipticity parameters, we also constrain the effective radius R_{eff} and Sersic index n_s of the

source light through the use of prior probability distributions obtained from a catalog of 56 062 galaxies from the COSMOS survey that were modeled using a single Sersic profile to serve as a training set for the GALSIM⁵ galaxy image simulation software (Rowe et al. 2015). We show these prior distributions in R_{eff} and

⁵ <https://github.com/GalSim-developers/GalSim>

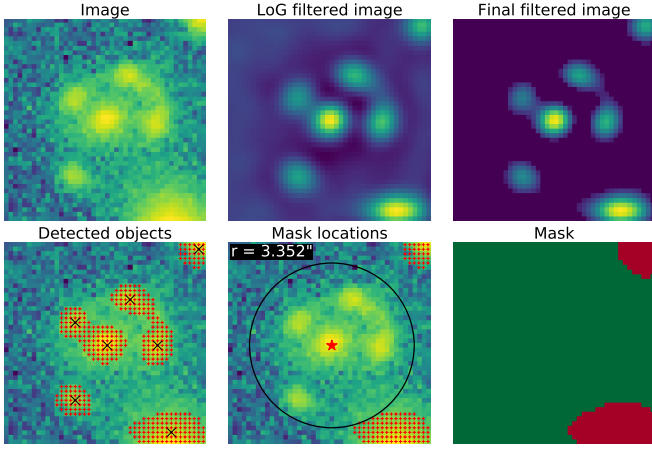


Fig. 8. Illustration of our masking procedure. The upper row shows the original CFIS image in the r band, followed by the image after applying the LoG filter in the top middle panel. This image is then thresholded and all pixels below $6\sigma_{\text{sky}}$ are set to zero, as shown in the top right panel. The bottom row shows the detected peaks, whose centroids are marked with black crosses. The bottom middle panel displays the estimated lens size as a black circle and the red areas indicate what we consider as contaminants. These are masked as the red area on the bottom right image.

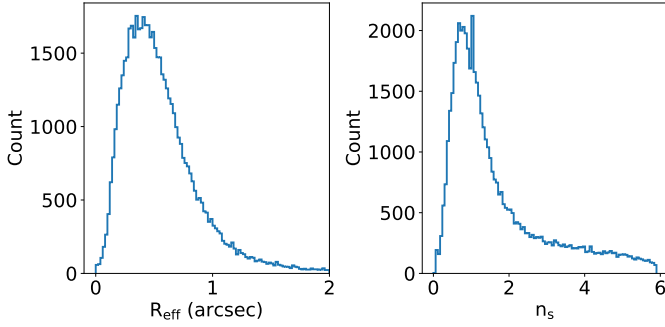


Fig. 9. Priors used for the effective radius R_{eff} and Sersic index n_s of the source light. The priors are derived from 56 062 galaxies from the COSMOS survey.

n_s in Fig. 9. For simplicity, we do not assume any covariance in the prior distribution of these two parameters nor do we assume any covariance with the deflector's flux or magnitude.

6.2. Modeling results

We apply our *lenstronomy*-based pipeline with the priors described above to the best 32 lenses found with the CNN search. Figure 10 shows histograms for the model parameters describing the deflector mass, the source light, and the deflector light profile. The Einstein radii of the lenses are in the range $1.2'' < R_E < 2.5''$ and the external shear strengths are all 0.3 or less. The CNN is biased to find lenses with Einstein radius matching the range of the training set, but the Einstein radius range $1.2'' < R_E < 2.5''$ in the SL sample also highlights the fact that the visual inspection predominantly selects obvious wide-separation lenses with de-blended counter images. Seven lenses have models with a shear compatible with zero, but since no account for the strong correlation of external shear and internal ellipticity is made, this must be interpreted carefully. For the lens light the effective radius and Sersic index distributions peak at $2.5''$ and 5.0 , respectively,

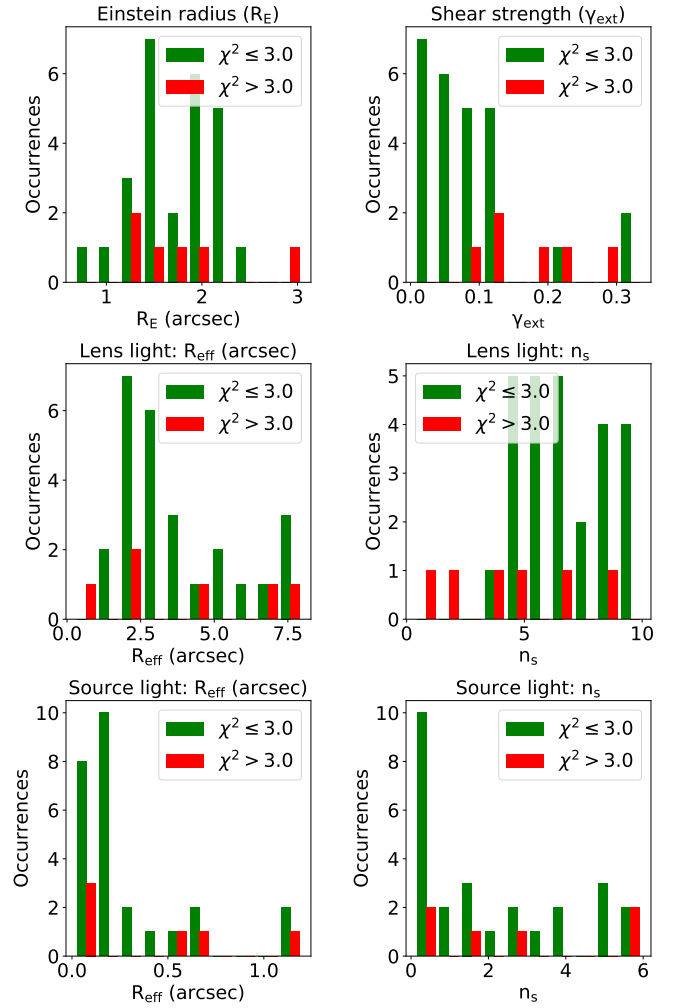


Fig. 10. Parameter distributions from our lens modeling results using the SIE+ γ_{ext} model. The top row gives the Einstein radii and external shear values of the mass model. The middle and bottom rows show the effective radius and Sersic index, for the lens and source light, respectively. The models with an acceptable reduced χ^2 (≤ 3.0) are shown in green and those with a reduced $\chi^2 > 3.0$ are shown in red (see text). After a manual fix of the masking procedure, the bad fits improved to $\chi^2 < 3.0$.

which is not surprising as our lenses are selected among a sample of LRGs. The source galaxies are generally much smaller than the deflector LRGs, with the distribution in effective radius peaking at $R_{\text{eff}} \sim 0.2''$ and with Sersic index peaking at $n_s \sim 1.0$, also unsurprising given that source galaxies are often low mass and/or star-forming disks. In Figs. D.1–D.7 we show mosaics of the modeling results for the 32 lenses. Table D.1 lists the best-fit parameters obtained from these fits. In spite of the simplicity of our models the residuals are acceptable for most systems (i.e., with a mean reduced χ^2 close to 1.0) (see Table D.1). However, setting a limit of 3 on the reduced χ^2 allows us to spot outliers. These objects are indicated with a red rectangle in Figs. D.1–D.7 and show that the bad fits are due to complex lens light profiles beyond Sersic, inaccurate masking, or objects in the source plane being modeled as if they were in the lens plane or vice versa. A notable example is UNIONS J155923+314712 (first row of Fig. D.3) which has a bright elongated deflector that is not well described by a single Sersic profile. This specific example also shows the limitation of our pre-selection of LRG deflectors as

obviously this system has a lens with an edge-on disk and a bulge component. In all cases with $\chi^2 > 3.0$, we were able to immediately identify the problem and correct it in a simple way, bringing the new χ^2 value close to 1.0 (objects indicated with an asterisk in Table D.1).

Four additional objects have $\chi^2 > 3$ (UNIONS J113952+303204, UNIONS J165710+315052, UNION J075346+341633, UNIONS J112053+342146). For three of these the mask produced by our automated procedure is simply not large enough to cover all of the light contaminants. In the case of UNIONS J165710+315052, part of the lensed source light in the bottom right corner of the image is mistaken for a companion, while the contaminant directly to the right of the deflector is treated as a lensed source image. This is a very specific configuration that would be extremely difficult to accommodate in an automated masking procedure for thousands of objects, as will be the case with future wide-field surveys. Nonetheless, since we only have 32 objects, we can afford to create new customized masks, and we show in Figs. D.1–D.4 the results with new masks directly beneath the results with the automated masks. In each subpanel the two modeling results are enclosed in the red dashed rectangles, and the manual masks significantly improve the residuals.

We have a total of six modeling failures. Of these six, one fails due to the deflector light being more complex than for the rest of the sample, while three are due to imperfect masking that can certainly be improved in future versions of the pipeline. One of the failures, UNIONS J081959+535624, is in fact known as a lensed quasar (Inada et al. 2009) and therefore fails because our modeling procedure does not allow for point sources (see Fig. D.8). This point will be addressed in future work. For the 25 successful automated models, the average modeling time is 2 hours for a $44 \text{ pixel} \times 44 \text{ pixel}$ stamp, which is well achievable in the context of future surveys like *Euclid*, which is expected to find 20–30 SLs per day.

7. Discussion and conclusions

In this paper we presented the design of an automated pipeline to find galaxy-scale strong lenses using convolutional neural networks and applied it to the CFIS wide-field optical imaging survey being carried out with the 3.6 m CFHT in Hawaii. We used only the deep and sharp r -band images for which the median seeing is $0.6''$ down to a 10σ depth of $r = 24.6$. We used $2\,500 \text{ deg}^2$ of CFIS in the present work since the survey is still ongoing; it is expected to reach a total area of $5\,000 \text{ deg}^2$ of the northern sky, when completed.

In training our CNNs, we used data-driven simulations where the light distribution of the lens plane is taken directly from the data. This naturally includes the PSF, companions, and noise properties of the actual data and any companion galaxies. The lens mass, modeled as a SIE profile, was derived from the measured velocity dispersion of our sample of LRGs, which we further adapted to ensure that the lensing features were systematically visible in the training set. Our background sources are from the HST images of the COSMOS survey, and are convolved by the local CFIS PSF after lensing.

Starting from a sample of CFIS r -band images of LRGs selected both from spectroscopy and color-cuts, we used a committee of three neural networks, leading to a CNN-based sample of 9 460 objects passing a CNN score of 0.5, averaged over the three members of our committee. The adopted threshold of 0.5 on the CNN identification is based on experiments with our val-

idation set and on the CNN score obtained for real CFIS images of spiral galaxies identified in GalaxyZoo.

Even though the precision of the CNN committee is extremely high using a score threshold of 0.5, the large sample of two million galaxies to classify implies a large number of false positives among the 9 460 objects passing our threshold due to the low prevalence of lenses in real data, as explained in Sect. 3.4. Fortunately, visually inspecting 9 460 objects by eye is still doable and this task was done by six independent authors of this paper. Even though strict and homogeneous rules were set for the visual classification, we note that the six human classifiers still had very different opinions on what a lens is and what it is not, meaning that any automated CNN may still require a time consuming human check, for example with citizen-science projects for future wide-field space surveys like *Euclid*, *Rubin-LSST*, or *Roman*. More effort should be devoted to visual inspection methods to reach better consensus between classifiers.

Following the visual inspection, we found 32 objects with striking lensing features and 101 objects that show strong signs of lensing but that need further data to confirm (i.e., higher resolution, and deeper imaging and/or spectroscopy). This represents around 0.05 candidates per square degree, which is much lower than the 1.95 and 11.95 lenses per square degree estimated in Collett (2015) for LSST and *Euclid*, respectively. This number is however comparable with that found by Cañameras et al. (2020), from which we take our lens search sample, who obtain 0.0117 lenses per square degree. The slightly larger number of lenses per square degree in our case can be explained by the higher resolution of CFIS r -band images and the different lower limit on the Einstein radii of the simulations from the training set.

A by-product of our simulations set is that we were able to train auto-encoders to learn the lens light and lensed-source light separately and we then deblended the lens plane from the source plane for all 133 objects. We see this process, or future evolutions of it, as a way to infer photometric redshifts for the lens and source if many bands are available, which will be the case in future wide-field imaging surveys. With the CFIS r -band data alone, we still find our application of auto-encoders useful to evaluate the quality of the lens candidates, especially for the smallest Einstein radii and/or for objects with strong contrast between the lens and source light.

Finally we developed a simple lens modeling pipeline based on the *lenstronomy* software in which we adopt an SIE mass profile with external shear. We also developed an automated masking procedure to enforce only relevant objects to be modeled and avoid objects unrelated to either the lensed source or the lensing galaxy. The optimization process using particle swarm optimization followed with Markov chain Monte Carlo sampling takes on the order of two hours per object for our $44 \text{ pixel} \times 44 \text{ pixel}$ stamps. With the present ground-based data, even with deep high-resolution ground-based imaging, we find that the SIE plus shear model is sufficient to fit most of the data. The main sources of failure are the following: (1) the lens light was too complex to be described by a single Sersic profile (1 object out of 32); (2) the masking procedure failed to capture the entire extent of the objects to be masked (2 objects out of 32); and (3) the misidentification of source images that should not have been masked (1 object out of 32). We consider only the last case to be a real limitation to a fully automated procedure.

We also produced a catalog of contaminants that mimic lensing geometry (i.e., 238 mergers, 369 ring galaxies, and 961 spiral galaxies). All these contaminants are provided in electronic form as they can be useful for future lens searches in order to train CNNs against false positives.

We demonstrated the possibility to build an automated pipeline to find, deblend, and model lenses in future large-scale surveys, even though there are still challenges to overcome. In particular, human intervention was required at two steps of the pipeline: in the verification of the lens candidates and, to a lesser extent, in the modeling. The importance of the visual inspection step may be decreased in future versions of the pipeline by retraining the CNNs with our catalog of false positives and by combining the information obtained with the deblending and modeling with the classification score provided by the committee. Since the CFIS imaging data are among the best available so far in terms of depth and seeing, our results, although subject to improvements, can be seen as an illustration of what can be achieved with in a single Rubin-LSST band built by stacking some of the best seeing epochs after about a month of data acquisition. With a spatial resolution three times better than the best CFIS images, *Euclid* (and then *Roman*) will give us access to a larger number of small separation systems, especially with Einstein radii smaller than $1''$. These data will also help us to decide on our less secure candidates, since space-based imaging will provide high signal-to-noise ratios for the lensing features of the lenses presented here, allowing us to test more complex mass models, and thus probe astrophysical questions like galaxy evolution and the structure of their dark matter halos.

Acknowledgements. This work is supported by the Swiss National Science Foundation (SNSF) and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (COSMICLENS: grant agreement No 787886). This work is based on data obtained as part of the Canada-France Imaging Survey, a CFHT large program of the National Research Council of Canada and the French Centre National de la Recherche Scientifique. Based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA Saclay, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l'Univers (INSU) of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work was supported in part by the Canadian Advanced Network for Astronomical Research (CANFAR) and Compute Canada facilities. GV has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 897124. RC, SS, and SHS thank the Max Planck Society for support through the Max Planck Research Group for SHS. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (LENSNOVA: grant agreement No 771776). RG thanks IoA and the Churchill College in Cambridge for their hospitality and acknowledges local support from the French government. This research has made use of the VizieR catalogue access tool, CDS, Strasbourg, France (DOI : 10.26093/cds/vizier). The original description of the VizieR service was published in [Ochsenbein et al. \(2000\)](#). This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France [Wenger et al. \(2000\)](#).

References

- Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, *PASJ*, 70, 54
- Amiaux, J., Scaramella, R., Mellier, Y., et al. 2012, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 8442, *Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave*, ed. M. C. Clampin, G. G. Fazio, H. A. MacEwen, & J. Oschmann, Jacobus M., 84420Z
- Auger, M. W., Treu, T., Bolton, A. S., et al. 2009, *ApJ*, 705, 1099
- Axelsson, S. 2000, *ACM Trans. Inf. Syst. Secur.*, 3, 186–205
- Bellagamba, F., Tessore, N., & Metcalf, R. B. 2017, *MNRAS*, 464, 4823
- Bertin, E. 2011, in *Astronomical Society of the Pacific Conference Series*, Vol. 442, *Astronomical Data Analysis Software and Systems XX*, ed. I. N. Evans, A. Accomazzi, D. J. Mink, & A. H. Rots, 435
- Birrer, S. & Amara, A. 2018, *Physics of the Dark Universe*, 22, 189
- Birrer, S., Amara, A., & Refregier, A. 2015, *ApJ*, 813, 102
- Bolton, A. S., Burles, S., Koopmans, L. V. E., et al. 2008, *ApJ*, 682, 964
- Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., & Moustakas, L. A. 2006, *ApJ*, 638, 703
- Bonvin, V., Courbin, F., Suyu, S. H., et al. 2017, *MNRAS*, 465, 4914
- Browne, I. W. A., Wilkinson, P. N., Jackson, N. J. F., et al. 2003, *MNRAS*, 341, 13
- Brownstein, J. R., Bolton, A. S., Schlegel, D. J., et al. 2012, *ApJ*, 744, 41
- Bussmann, R. S., Pérez-Fournon, I., Amber, S., et al. 2013, *ApJ*, 779, 25
- Cañameras, R., Nesvadba, N. P. H., Guery, D., et al. 2015, *A&A*, 581, A105
- Cañameras, R., Schuldt, S., Suyu, S. H., et al. 2020, *A&A*, 644, A163
- Cabanac, R. A., Alard, C., Dantel-Fort, M., et al. 2007, *A&A*, 461, 813
- Cao, X., Li, R., Shu, Y., et al. 2020, *MNRAS*, 499, 3610
- Chan, J. H. H., Suyu, S. H., Chiueh, T., et al. 2015, *ApJ*, 807, 138
- Chan, J. H. H., Suyu, S. H., Sonnenfeld, A., et al. 2020, *A&A*, 636, A87
- Chatterjee, S. & Koopmans, L. V. E. 2018, *MNRAS*, 474, 1762
- Chollet, F. et al. 2015, *Keras*, <https://keras.io>
- Collett, T. E. 2015, *ApJ*, 811, 20
- Comparat, J., Richard, J., Kneib, J.-P., et al. 2015, *A&A*, 575, A40
- Dalal, N. & Kochanek, C. S. 2002, *ApJ*, 572, 25
- Deng, J., Dong, W., Socher, R., et al. 2009, in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255
- Despali, G., Giocoli, C., Angulo, R. E., et al. 2016, *MNRAS*, 456, 2486
- Dey, A., Schlegel, D. J., Lang, D., et al. 2019, *AJ*, 157, 168
- Eisenstein, D. J., Annis, J., Gunn, J. E., et al. 2001, *AJ*, 122, 2267
- Fantini, N. J., Côté, P., McConnachie, A. W., et al. 2019, *ApJ*, 887, 148
- Faure, C., Kneib, J.-P., Covone, G., et al. 2008, *ApJ Sup.*, 176, 19
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *Publications of the Astronomical Society of the Pacific*, 125, 306
- Gavazzi, R., Treu, T., Marshall, P. J., Brault, F., & Ruff, A. 2012, *ApJ*, 761, 170
- Gilman, D., Agnello, A., Treu, T., Keeton, C. R., & Nierenberg, A. M. 2017, *MNRAS*, 467, 3970
- Goodman, J. & Weare, J. 2010, *Communications in Applied Mathematics and Computational Science*, 5, 65
- Gwyn, S. D. J. 2008, *PASP*, 120, 212
- Hasinger, G., Capak, P., Salvato, M., et al. 2018, *ApJ*, 858, 77
- He, K., Zhang, X., Ren, S., & Sun, J. 2015, *arXiv e-prints*, arXiv:1502.01852
- Hezaveh, Y. D., Dalal, N., Marrone, D. P., et al. 2016, *ApJ*, 823, 37
- Huang, X., Storfer, C., Gu, A., et al. 2021, *ApJ*, 909, 27
- Huang, X., Storfer, C., Ravi, V., et al. 2020, *ApJ*, 894, 78
- Ibata, R. A., McConnachie, A., Cuillandre, J.-C., et al. 2017, *ApJ*, 848, 128
- Inada, N., Oguri, M., Shin, M.-S., et al. 2009, *AJ*, 137, 4118
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, 873, 111
- Jacobs, C., Collett, T., Glazebrook, K., et al. 2019a, *ApJ Sup.*, 243, 17
- Jacobs, C., Collett, T., Glazebrook, K., et al. 2019b, *MNRAS*, 484, 5330
- Jaelani, A. T., More, A., Sonnenfeld, A., et al. 2020, *MNRAS*, 494, 3156
- Jaelani, A. T., Rusu, C. E., Kayo, I., et al. 2021, *MNRAS*, 502, 1487
- Joseph, R., Courbin, F., & Starck, J. L. 2016, *A&A*, 589, A2
- Kassiola, A. & Kovner, I. 1993, *ApJ*, 417, 450
- Kelvin, L. S., Driver, S. P., Robotham, A. S. G., et al. 2012, *Monthly Notices of the Royal Astronomical Society*, 421, 1007
- Kennedy, J. & Eberhart, R. 1995, in *Proceedings of ICNN'95 - International Conference on Neural Networks*, Vol. 4, 1942–1948 vol.4
- Kingma, D. P. & Ba, J. 2014, *arXiv e-prints*, arXiv:1412.6980
- Koopmans, L. V. E. 2005, *MNRAS*, 363, 1136
- Koopmans, L. V. E. & Treu, T. 2003, *ApJ*, 583, 606
- Kormann, R., Schneider, P., & Bartelmann, M. 1994, *A&A*, 284, 285
- Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, *A&A*, 625, A2
- Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJ Sup.*, 224, 24
- Lanusse, F., Mandelbaum, R., Ravanbakhsh, S., et al. 2020, *arXiv e-prints*, arXiv:2008.03833
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, *arXiv e-prints*, arXiv:1110.3193
- Le Fèvre, O., Tasca, L. A. M., Cassata, P., et al. 2015, *A&A*, 576, A79
- Lecun, Y., Bengio, Y., & Hinton, G. 2015, *Nature*, 521, 436
- Lilly, S. J., Le Fèvre, O., Renzini, A., et al. 2007, *ApJ Sup.*, 172, 70
- Mao, S. & Schneider, P. 1998, *MNRAS*, 295, 587
- Marshall, P. J., Hogg, D. W., Moustakas, L. A., et al. 2009, *ApJ*, 694, 924
- Melchior, P., Moolekamp, F., Jerdee, M., et al. 2018, *Astronomy and Computing*, 24, 129
- Metcalf, R. B., Meneghetti, M., Avestruz, C., et al. 2019, *A&A*, 625, A119
- Mukherjee, S., Koopmans, L. V. E., Metcalf, R. B., et al. 2021, *MNRAS*, 504, 3455
- Myers, S. T., Jackson, N. J., Browne, I. W. A., et al. 2003, *MNRAS*, 341, 1
- Nayyeri, H., Keele, M., Cooray, A., et al. 2016, *ApJ*, 823, 17
- Negrello, M., Amber, S., Amvrosiadis, A., et al. 2017, *MNRAS*, 465, 3558
- Nierenberg, A. M., Oldenburg, D., & Treu, T. 2013, *MNRAS*, 436, 2120
- Nightingale, J. W., Dye, S., & Massey, R. J. 2018, *MNRAS*, 478, 4738
- Ochsenbein, F., Bauer, P., & Marcout, J. 2000, *A&AS*, 143, 23
- Oguri, M., Inada, N., Pindor, B., et al. 2006, *AJ*, 132, 999
- Paraficz, D., Courbin, F., Tramacere, A., et al. 2016, *VizieR Online Data Catalog*, J/A+A/592/A75
- Paraficz, D., Rybak, M., McKean, J. P., et al. 2018, *A&A*, 613, A34
- Pawase, R. S., Courbin, F., Faure, C., Kokotanekova, R., & Meylan, G. 2014, *MNRAS*, 439, 3392

- Peirani, S., Sonnenfeld, A., Gavazzi, R., et al. 2019, MNRAS, 483, 4615
- Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2019, MNRAS, 482, 807
- Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2017, MNRAS, 472, 1129
- Pourrahmani, M., Nayyeri, H., & Cooray, A. 2018, ApJ, 856, 68
- Ritondale, E., Vegetti, S., Despali, G., et al. 2019, MNRAS, 485, 2179
- Rojas, K., Savary, E., Clément, B., et al. 2021, Strong lens systems search in the Dark Energy Survey using Convolutional Neural Networks
- Rowe, B. T. P., Jarvis, M., Mandelbaum, R., et al. 2015, Astronomy and Computing, 10, 121
- Schuldt, S., Suyu, S. H., Meinhardt, T., et al. 2020, HOLISMOKES – IV. Efficient mass modeling of strong lenses through deep learning
- Shajib, A. J., Treu, T., Birrer, S., & Sonnenfeld, A. 2020, arXiv e-prints, arXiv:2008.11724
- Shu, Y., Bolton, A. S., Brownstein, J. R., et al. 2015, ApJ, 803, 71
- Shu, Y., Bolton, A. S., Mao, S., et al. 2016, ApJ, 833, 264
- Shu, Y., Brownstein, J. R., Bolton, A. S., et al. 2017, ApJ, 851, 48
- Silverman, J. D., Kashino, D., Sanders, D., et al. 2015, ApJ Sup., 220, 12
- Sonnenfeld, A., Chan, J. H. H., Shu, Y., et al. 2018, PASJ, 70, S29
- Sonnenfeld, A., Jaelani, A. T., Chan, J., et al. 2019, A&A, 630, A71
- Sonnenfeld, A., Treu, T., Marshall, P. J., et al. 2015, ApJ, 800, 94
- Sonnenfeld, A., Verma, A., More, A., et al. 2020, A&A, 642, A148
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv e-prints, arXiv:1503.03757
- Suyu, S. H., Bonvin, V., Courbin, F., et al. 2017, MNRAS, 468, 2590
- Suyu, S. H. & Halkola, A. 2010, A&A, 524, A94
- Suyu, S. H., Hensel, S. W., McKean, J. P., et al. 2012, ApJ, 750, 10
- Talbot, M. S., Brownstein, J. R., Dawson, K. S., Kneib, J.-P., & Bautista, J. 2021, MNRAS, 502, 4617
- Tan, M. & Le, Q. V. 2019, arXiv e-prints, arXiv:1905.11946
- Tasca, L. A. M., Le Fèvre, O., Ribeiro, B., et al. 2017, A&A, 600, A110
- The Dark Energy Survey Collaboration. 2005, arXiv e-prints, astro
- Turner, E. L., Ostriker, J. P., & Gott, J. R., I. 1984, ApJ, 284, 1
- Vegetti, S., Despali, G., Lovell, M. R., & Enzi, W. 2018, MNRAS, 481, 3661
- Vegetti, S. & Koopmans, L. V. E. 2009, MNRAS, 392, 945
- Vegetti, S., Koopmans, L. V. E., Auger, M. W., Treu, T., & Bolton, A. S. 2014, MNRAS, 442, 2017
- Vegetti, S., Koopmans, L. V. E., Bolton, A., Treu, T., & Gavazzi, R. 2010, MNRAS, 408, 1969
- Vegetti, S., Lagattuta, D. J., McKean, J. P., et al. 2012, Nature, 481, 341
- Vieira, J. D., Crawford, T. M., Switzer, E. R., et al. 2010, ApJ, 719, 763
- Vieira, J. D., Marrone, D. P., Chapman, S. C., et al. 2013, Nature, 495, 344
- Wardlow, J. L., Cooray, A., De Bernardis, F., et al. 2013, ApJ, 762, 59
- Wenger, M., Ochsenein, F., Egret, D., et al. 2000, A&AS, 143, 9
- Willett, K. W., Galloway, M. A., Bamford, S. P., et al. 2017, MNRAS, 464, 4176
- Wong, K. C., Sonnenfeld, A., Chan, J. H. H., et al. 2018, ApJ, 867, 107
- Wong, K. C., Suyu, S. H., Chen, G. C. F., et al. 2020, MNRAS, 498, 1420

Appendix A: Visual inspection results

We present in Table A.1 the results of the single-object inspection for each user. The agreement between the human classifiers is low, as can be observed in Fig. A.1, which displays the number of images labeled as either ML or SL by each user and the numbers shared between each pair of users. The discrepancy between users is less pronounced for the NL images. The category SA causes the most confusion since no image is labeled as SA by all users, and the overlap between the users selecting the greatest and lowest number of SA (User 1 and User 5) is only five objects. As a result, despite the classification guidelines, the number of ML or SL candidates varies greatly between users (see Fig. A.1). Even so, the agreement between users is the greatest for the SL category, hence showing a good consensus for the objects with the most striking lensing features.

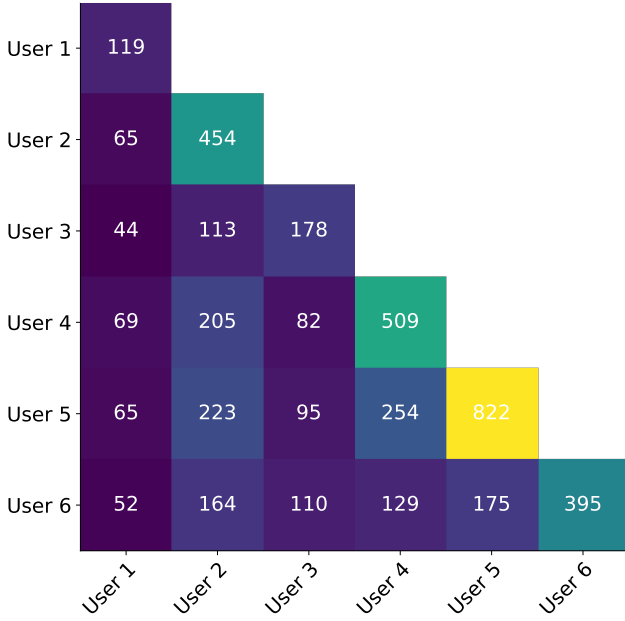


Fig. A.1. Illustration of the overlap between the visual classification of different users. The diagonal terms show the number of SL and ML labeled by each user during the first steps of the inspection, whereas the nondiagonal terms correspond to the number of ML and SL objects that are shared between the users in the corresponding rows and columns.

Appendix B: Auto-encoder architecture

We present in Fig. B.1 the architecture of the deblending auto-encoder. In this scheme all layers are represented by a box and the connections between the different layers are shown with arrows. The decoder is separated in two independent and symmetrical sections that specialize in extracting the lensed source features and the deflector images. For both the encoder and the decoder we use a combination of dense and convolutional layers. The decoder part has three output layers. The last two convolutional layers output the deblended lensed-source and deflector images and the Add layer returns the sum of the two deblended images. We applied “Relu” activations to all neurons of the network except in the last three layers for which we used sigmoid activations in order to keep the range of the output between -1 and 1.

Appendix C: Maybe lens candidates

We show in Fig. C.1 the 101 ML candidates obtained at the end of the visual inspection process. It includes all candidates displaying convincing probable lensing features but that require follow-up observations to be confirmed.

Appendix D: Modeling results

The best-fit parameters for the 32 SL obtained after the modeling procedure are listed in Table D.1. We show in Figs. D.1–D.7 a mosaic with the modeling results for the 32 SL candidates.

Appendix E: Examples of contaminants

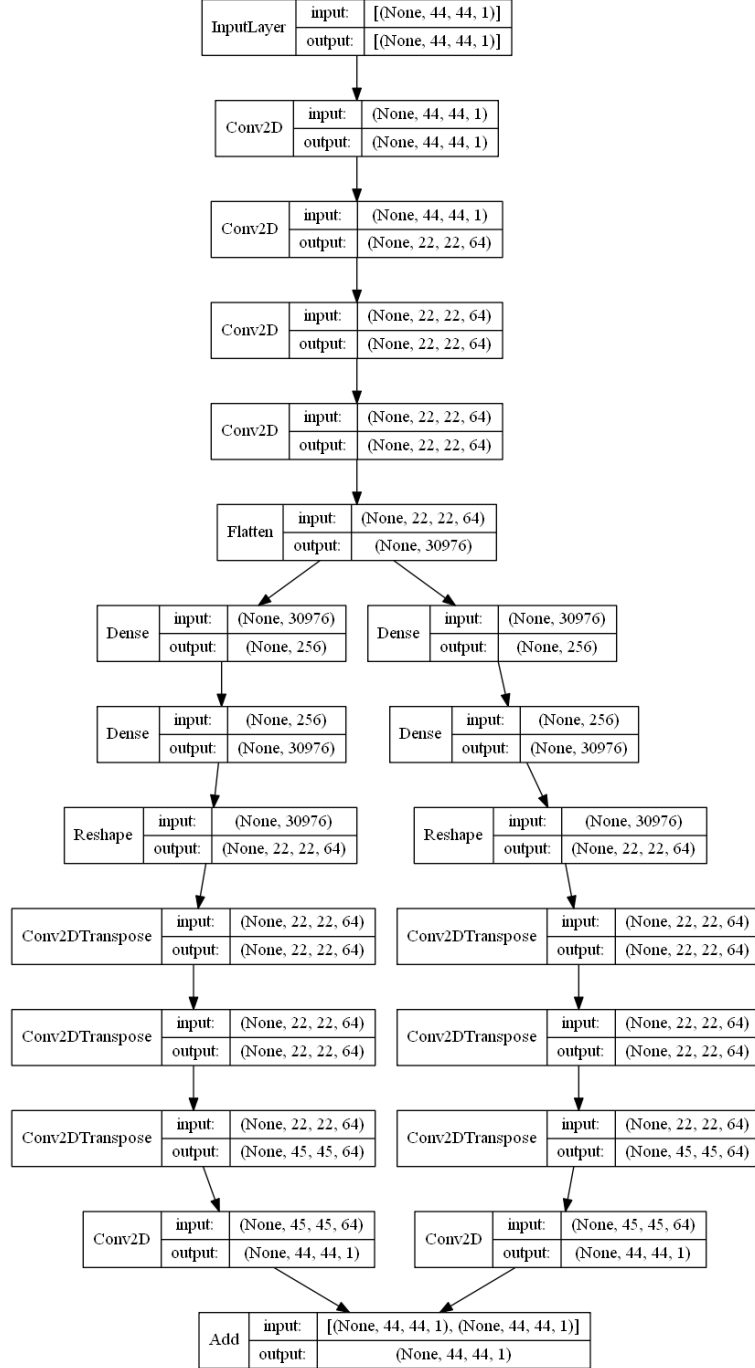
We present in Fig. E.1 examples of images taken randomly from the 238 mergers and 361 ring galaxies, and the 950 identified after the visual inspection.

Appendix F: Lens candidates

We present in Table F.0 all the candidates obtained after the last stage of the visual inspection. For each one the score given by the CNN committee is indicated.

Table A.1. User results of the detailed visual inspection.

Classification	User 1	User 2	User 3	User 4	User 5	User 6
NL	4492	4151	4389	3941	3398	4012
SA	17	23	61	178	408	221
ML	96	423	144	474	775	357
SL	23	31	34	35	47	38

**Fig. B.1.** Architecture of the debinding auto-encoder. Each box represents a layer and the arrows show the connections between the different layers of the networks. The names indicated in the rectangles correspond to the different layer subclasses of the Keras API used in the model. The dimensions of the input and output of each layer are indicated in brackets. For convolutional layers the last dimension corresponds to the number of filters

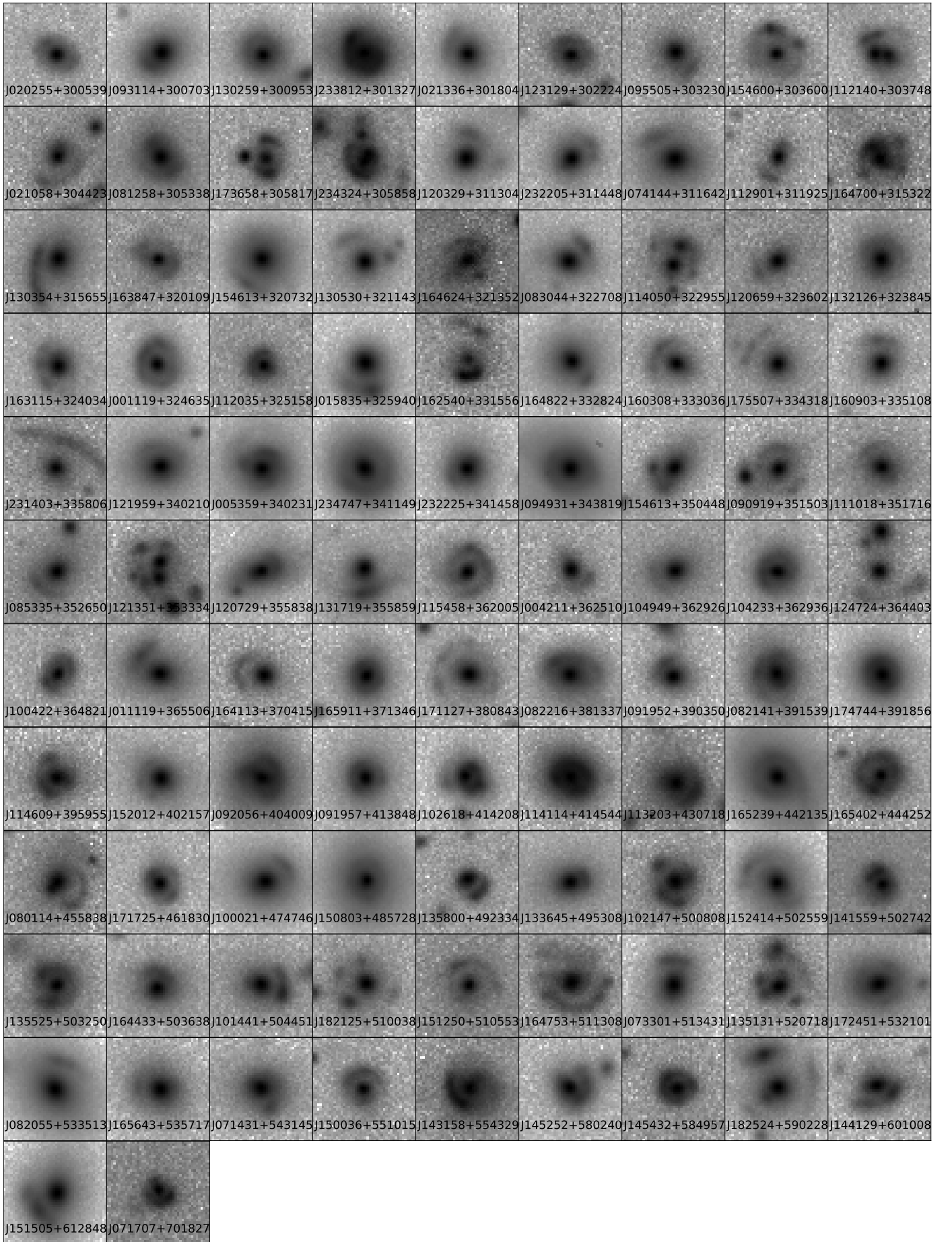


Fig. C.1. Images classified as maybe lenses (MLs) after the final visual inspection step.

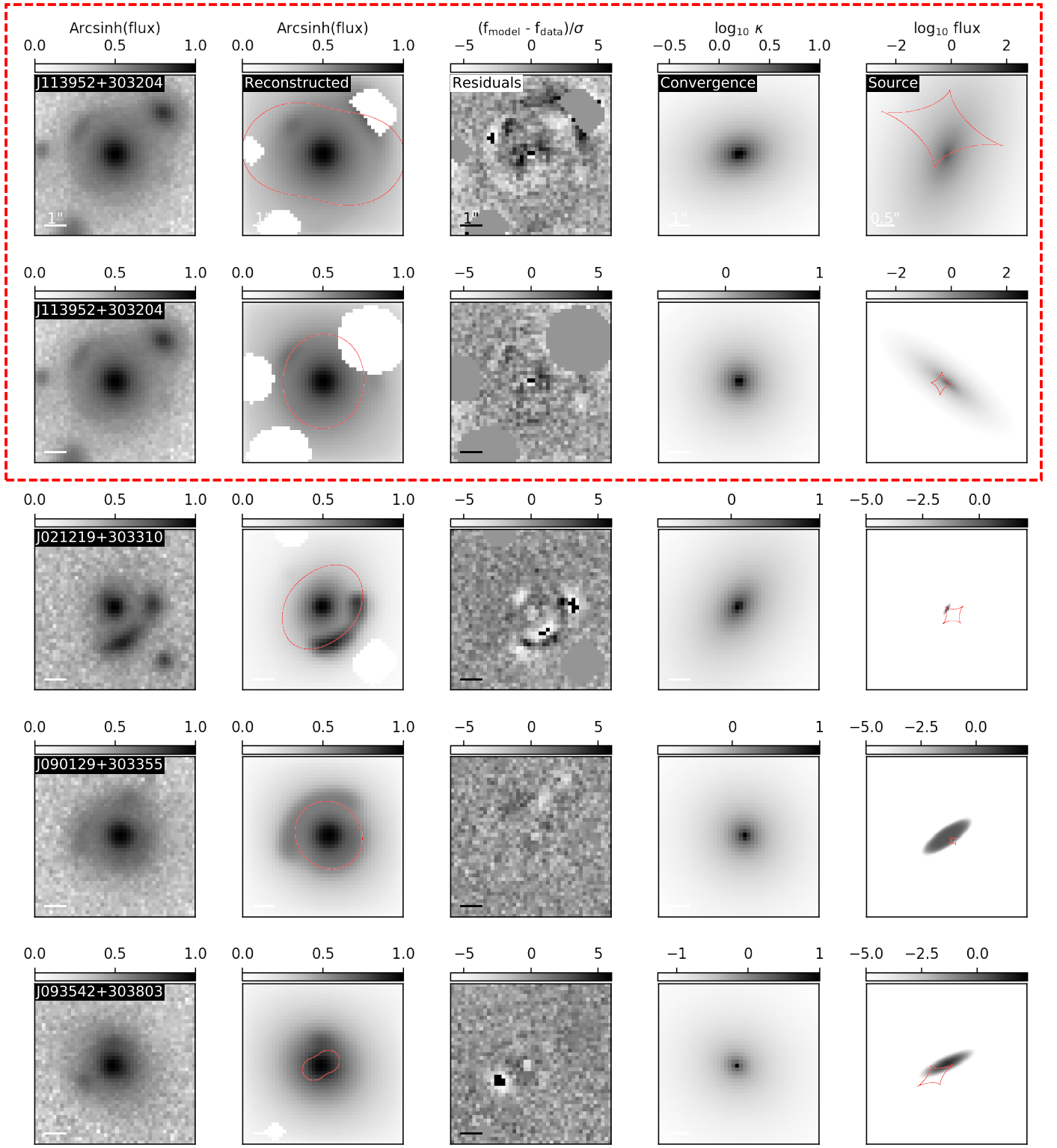


Fig. D.1. Modeling results for the first four of the 32 SL lens candidates. Shown inside the red dashed box are two modeling results for the same image, but with different masks. The top row corresponds to the results from the automated masking procedure, and in the bottom row to the results after applying a custom mask. *1st column:* CFIS *r*-band image. *2nd column:* Image reconstruction using best-fit model parameters. The white regions are masked pixels corresponding to locations of neighboring objects in the observed image. In red we show the critical lines of the lens model. *3rd column:* Normalized residual map of the image reconstruction. *4th column:* Lens mass model convergence map. *5th column:* Reconstructed source light profile (unlensed). In red are shown the caustic lines of the lens model.

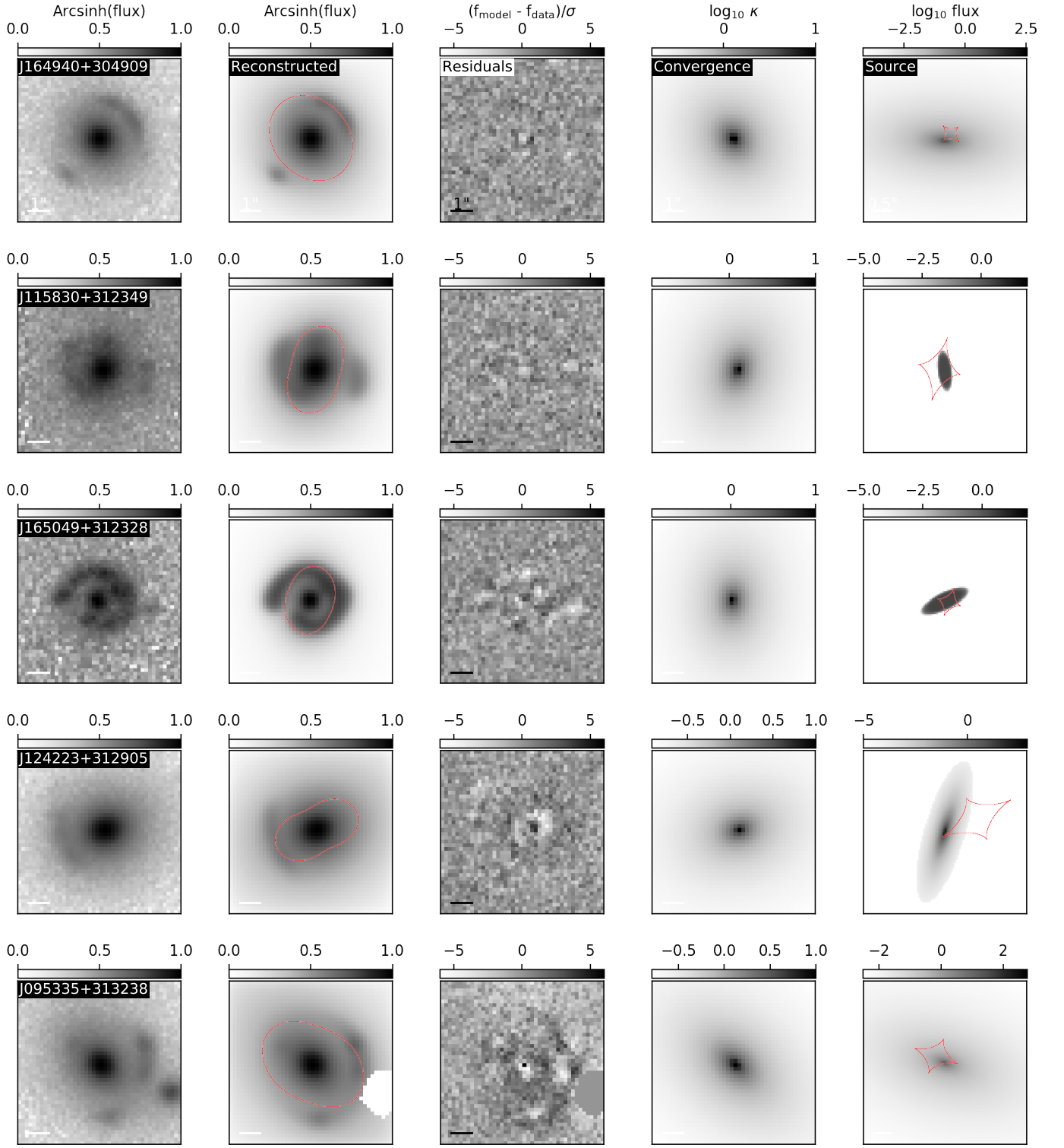


Fig. D.2. Continued from Fig. D.1

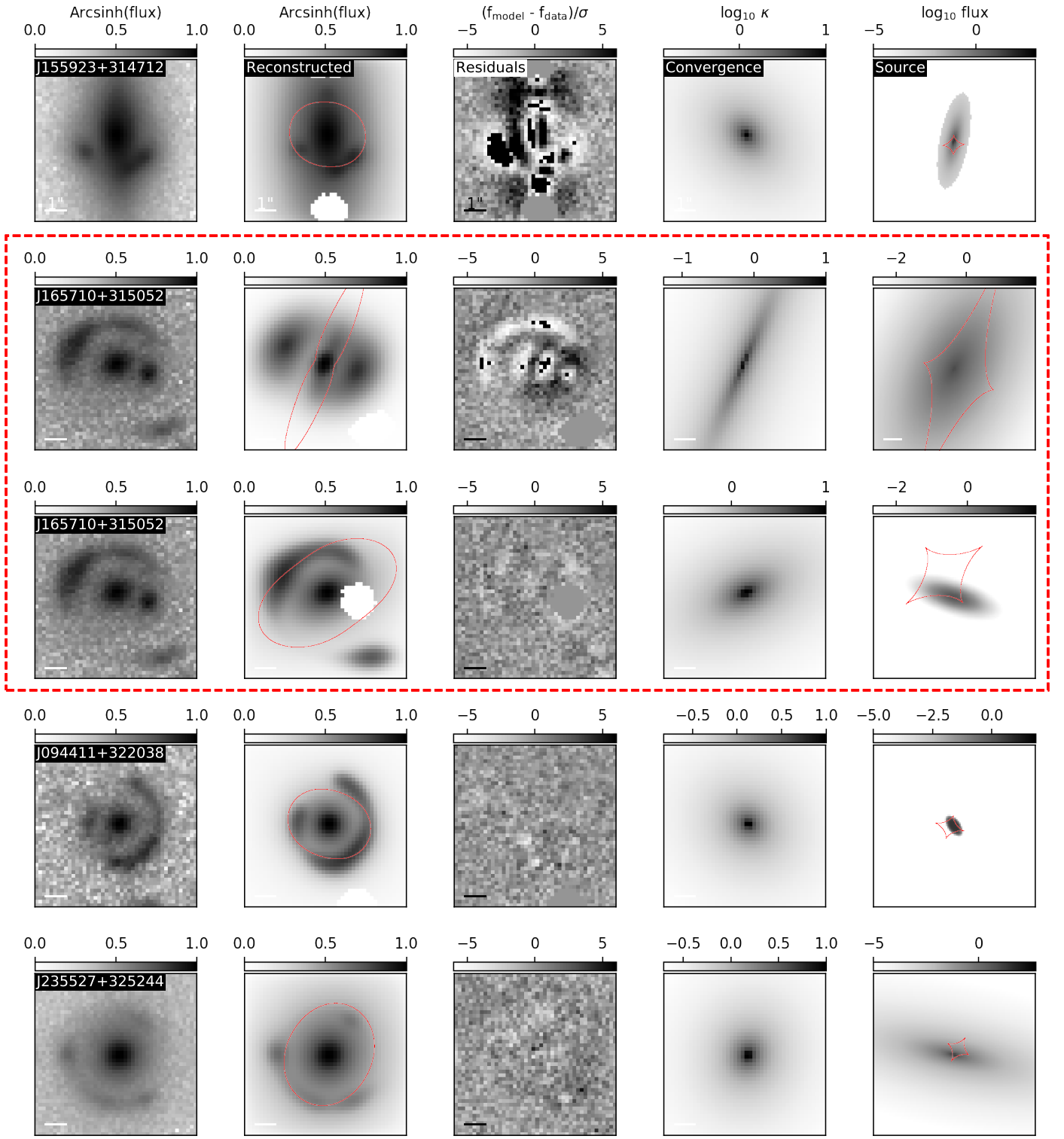


Fig. D.3. Continued from Fig. D.2

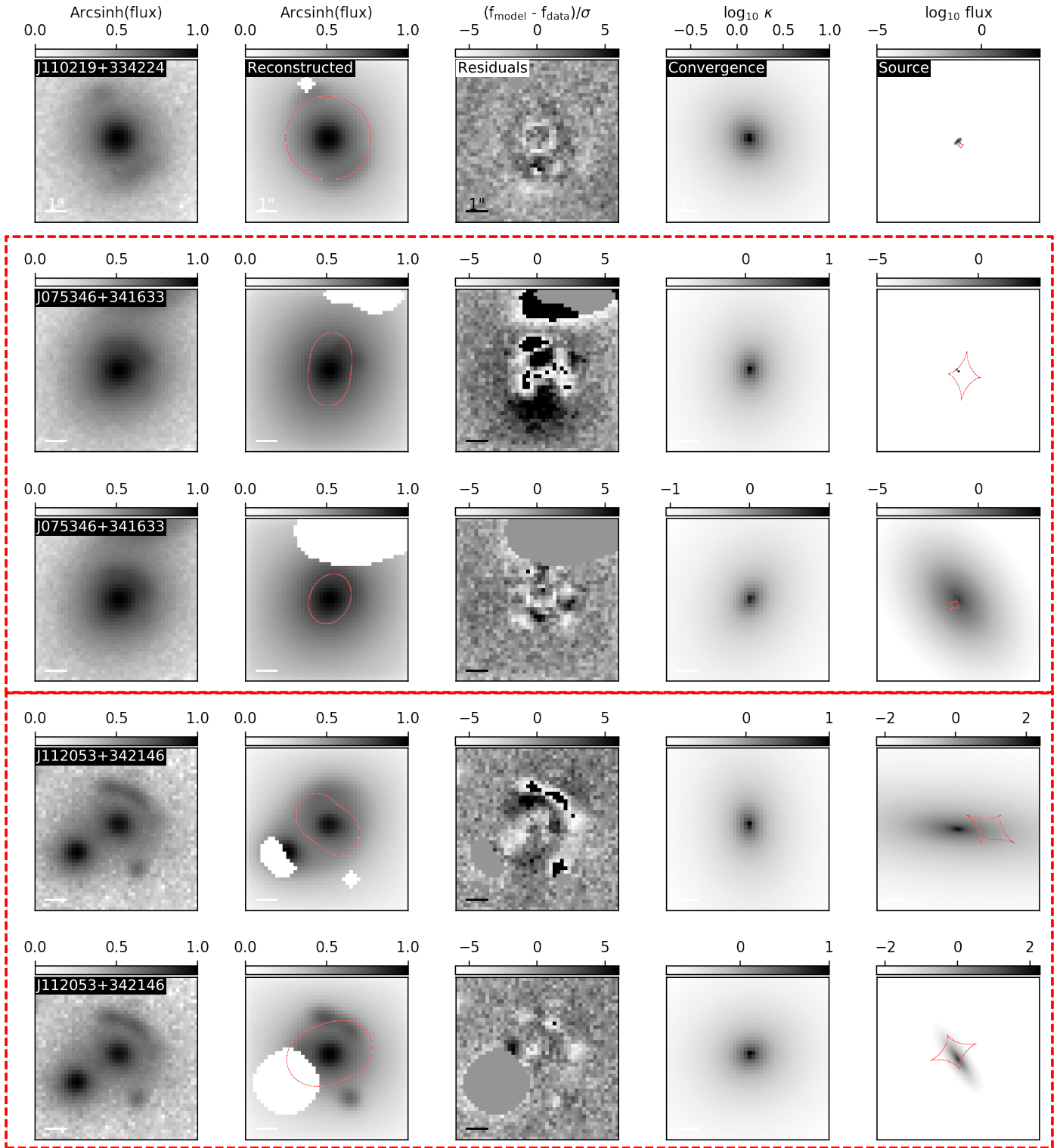


Fig. D.4. Continued from Fig. D.3

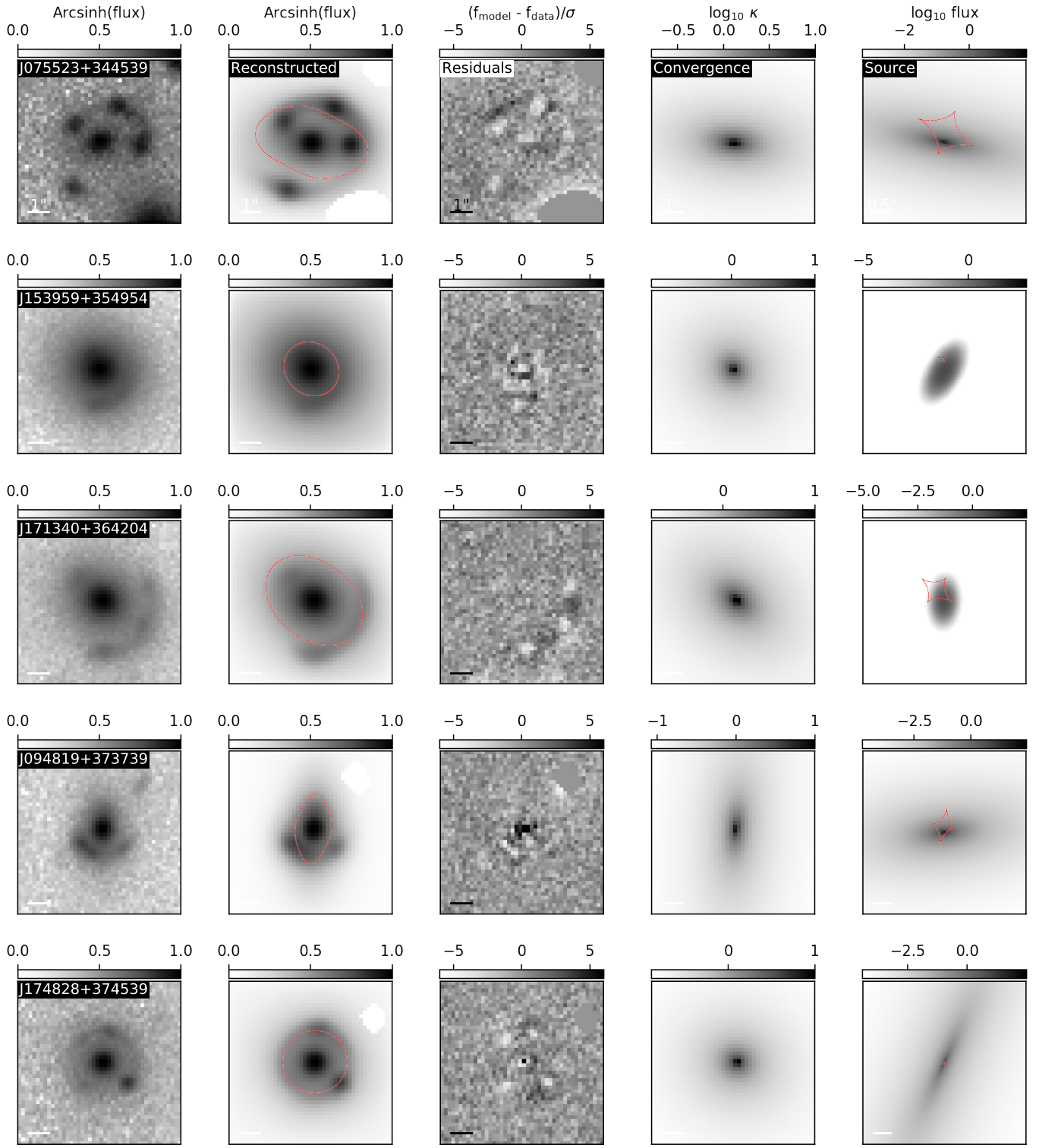


Fig. D.5. Continued from Fig. D.4

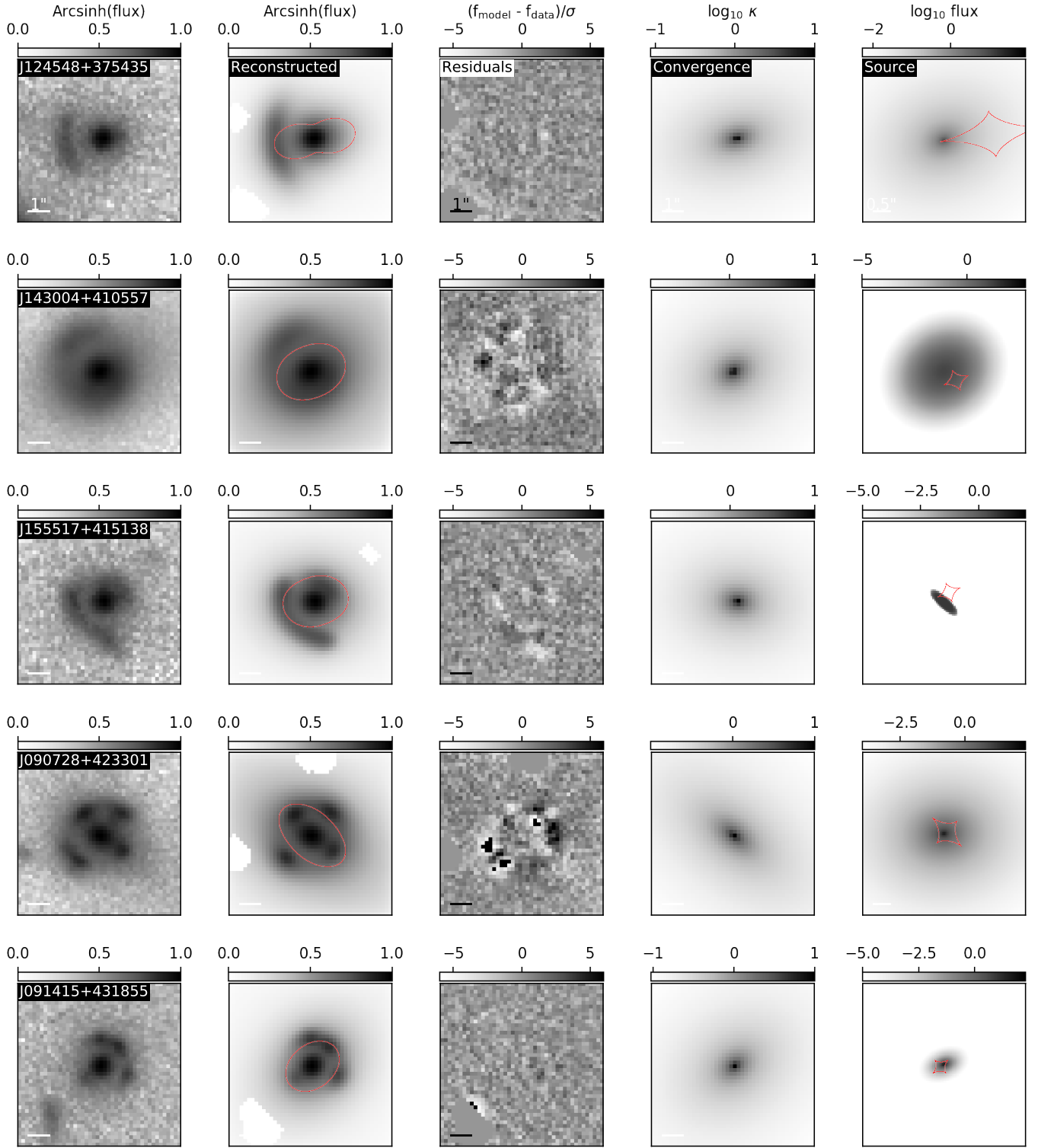


Fig. D.6. Continued from Fig. D.5

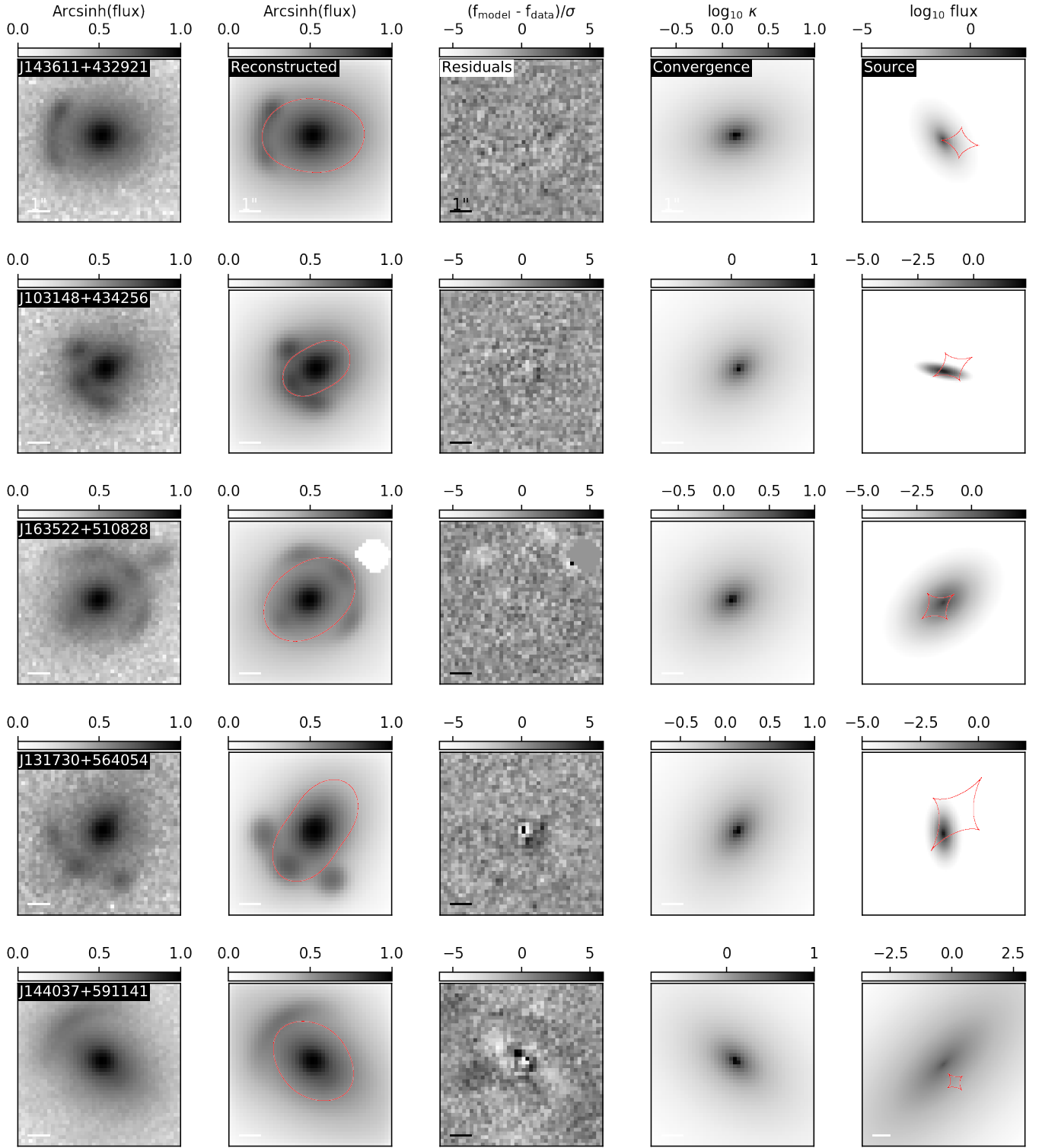


Fig. D.7. Continued from Fig. D.6

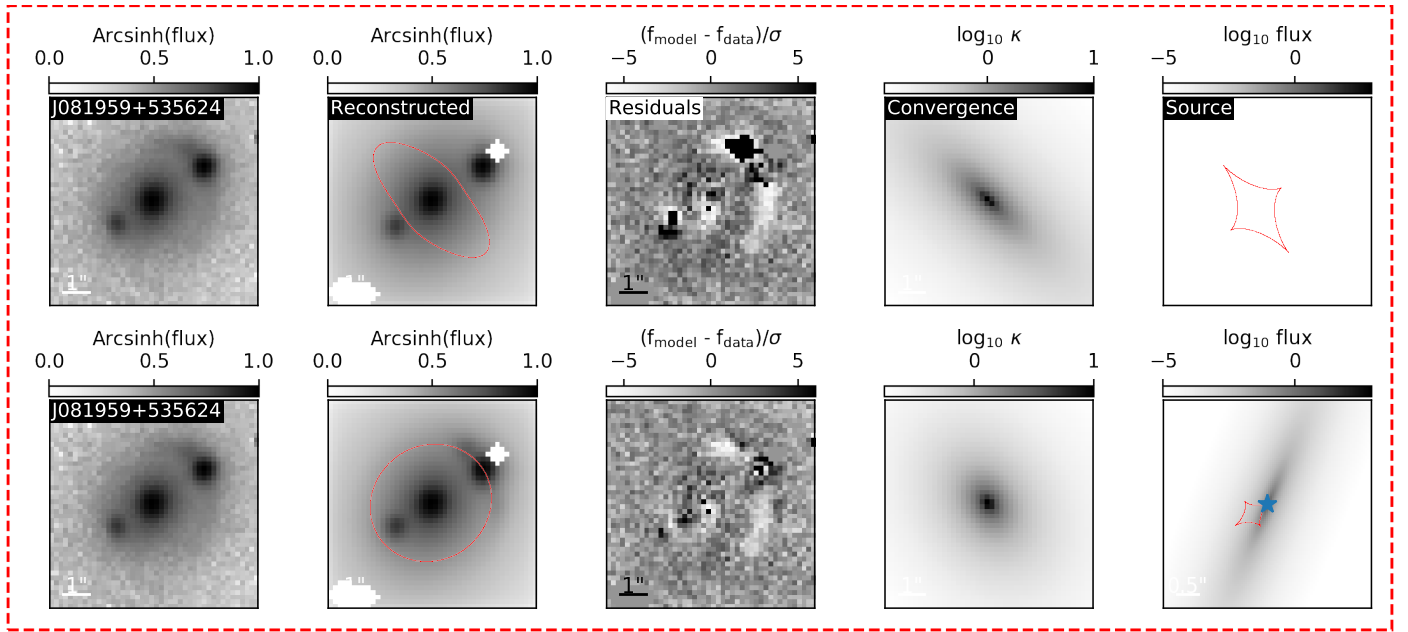


Fig. D.8. Continued from Fig. D.7

Table D.1. Lens mass best-fit parameters from lens modeling.

Candidate	Reduced χ^2	R_E (arcsec)	q_m	ϕ_m (deg)	γ_{ext}	ϕ_{ext} (deg)
UNIONS J113952+303204	3.36	$3.048^{+0.380}_{-0.049}$	$0.737^{+0.191}_{-0.041}$	-15^{+29}_{-15}	$0.215^{+0.248}_{-0.015}$	-65^{+63}_{-4}
UNIONS J113952+303204*	2.19	$2.204^{+0.013}_{-0.013}$	$0.980^{+0.021}_{-0.021}$	-76^{+61}_{-59}	$0.072^{+0.011}_{-0.010}$	3^{+8}_{-8}
UNIONS J021219+303310	1.88	$2.044^{+0.003}_{-0.012}$	$0.670^{+0.006}_{-0.008}$	-56^{+1}_{-1}	$0.068^{+0.002}_{-0.003}$	-77^{+2}_{-3}
UNIONS J090129+303355	1.04	$1.694^{+0.010}_{-0.009}$	$0.957^{+0.022}_{-0.023}$	49^{+31}_{-31}	$0.027^{+0.008}_{-0.008}$	-38^{+16}_{-15}
UNIONS J093542+303803	1.35	$0.670^{+0.029}_{-0.033}$	$0.890^{+0.023}_{-0.022}$	28^{+13}_{-12}	$0.337^{+0.023}_{-0.024}$	57^{+5}_{-5}
UNIONS J164940+304909	0.92	$2.061^{+0.007}_{-0.007}$	$0.855^{+0.013}_{-0.012}$	55^{+7}_{-5}	$0.031^{+0.004}_{-0.005}$	-61^{+8}_{-8}
UNIONS J115830+312349	0.98	$1.633^{+0.021}_{-0.019}$	$0.732^{+0.023}_{-0.021}$	-75^{+5}_{-5}	$0.138^{+0.010}_{-0.011}$	13^{+4}_{-4}
UNIONS J165049+312328	0.98	$1.404^{+0.012}_{-0.012}$	$0.716^{+0.028}_{-0.028}$	-87^{+7}_{-7}	$0.073^{+0.011}_{-0.011}$	45^{+8}_{-9}
UNIONS J124223+312905	1.19	$1.511^{+0.011}_{-0.009}$	$0.785^{+0.015}_{-0.016}$	-18^{+5}_{-5}	$0.217^{+0.007}_{-0.008}$	59^{+2}_{-2}
UNIONS J095335+313238	2.09	$2.241^{+0.004}_{-0.009}$	$0.672^{+0.007}_{-0.008}$	42^{+2}_{-2}	$0.094^{+0.003}_{-0.005}$	79^{+1}_{-2}
UNIONS J155923+314712	15.45	$1.760^{+0.004}_{-0.004}$	$0.788^{+0.004}_{-0.020}$	51^{+2}_{-2}	$0.134^{+0.002}_{-0.003}$	72^{+1}_{-1}
UNIONS J165710+315052	3.21	$1.562^{+0.021}_{-0.030}$	$0.144^{+0.010}_{-0.011}$	-68^{+1}_{-1}	$0.175^{+0.011}_{-0.007}$	16^{+3}_{-2}
UNIONS J165710+315052*	1.12	$2.810^{+0.008}_{-0.009}$	$0.593^{+0.013}_{-0.012}$	-26^{+2}_{-2}	$0.080^{+0.006}_{-0.006}$	31^{+4}_{-4}
UNIONS J094411+322038	1.00	$1.871^{+0.007}_{-0.007}$	$0.888^{+0.022}_{-0.023}$	43^{+14}_{-14}	$0.089^{+0.007}_{-0.008}$	-84^{+4}_{-4}
UNIONS J235527+325244	1.36	$2.381^{+0.002}_{-0.003}$	$0.873^{+0.012}_{-0.011}$	-75^{+7}_{-6}	$0.046^{+0.005}_{-0.003}$	43^{+4}_{-4}
UNIONS J110219+334224	1.39	$2.105^{+0.007}_{-0.007}$	$0.900^{+0.015}_{-0.016}$	60^{+9}_{-10}	$0.052^{+0.006}_{-0.006}$	68^{+6}_{-6}
UNIONS J075346+341633	17.30	$1.372^{+0.004}_{-0.004}$	$0.760^{+0.035}_{-0.032}$	-84^{+18}_{-7}	$0.129^{+0.011}_{-0.006}$	7^{+20}_{-6}
UNIONS J075346+341633*	2.06	$1.133^{+0.006}_{-0.006}$	$0.766^{+0.005}_{-0.008}$	-66^{+1}_{-2}	$0.012^{+0.002}_{-0.001}$	-87^{+15}_{-4}
UNIONS J112053+342146	4.37	$1.352^{+0.009}_{-0.008}$	$0.695^{+0.016}_{-0.014}$	85^{+4}_{-4}	$0.294^{+0.005}_{-0.005}$	-68^{+1}_{-1}
UNIONS J112053+342146*	1.97	$1.776^{+0.018}_{-0.062}$	$0.902^{+0.041}_{-0.045}$	-17^{+19}_{-21}	$0.142^{+0.007}_{-0.004}$	69^{+3}_{-1}
UNIONS J075523+344539	1.48	$2.134^{+0.006}_{-0.008}$	$0.544^{+0.011}_{-0.008}$	6^{+1}_{-2}	$0.124^{+0.004}_{-0.004}$	-28^{+2}_{-2}
UNIONS J153959+354954	1.30	$1.338^{+0.008}_{-0.008}$	$0.892^{+0.018}_{-0.018}$	46^{+9}_{-9}	$0.006^{+0.006}_{-0.006}$	80^{+69}_{-69}
UNIONS J171340+364204	1.15	$2.251^{+0.009}_{-0.009}$	$0.732^{+0.015}_{-0.015}$	41^{+3}_{-3}	$0.015^{+0.005}_{-0.005}$	-78^{+23}_{-23}
UNIONS J094819+373739	1.25	$1.234^{+0.005}_{-0.004}$	$0.400^{+0.012}_{-0.012}$	-85^{+1}_{-1}	$0.130^{+0.006}_{-0.007}$	87^{+2}_{-2}
UNIONS J174828+374539	1.09	$1.567^{+0.029}_{-0.030}$	$0.939^{+0.050}_{-0.063}$	24^{+54}_{-65}	$0.026^{+0.004}_{-0.009}$	48^{+7}_{-9}
UNIONS J124548+375435	0.91	$1.136^{+0.063}_{-0.065}$	$0.653^{+0.025}_{-0.024}$	-14^{+4}_{-4}	$0.310^{+0.038}_{-0.039}$	83^{+4}_{-4}
UNIONS J143004+410557	1.96	$1.508^{+0.004}_{-0.004}$	$0.808^{+0.009}_{-0.009}$	-29^{+3}_{-3}	$0.055^{+0.004}_{-0.004}$	78^{+4}_{-4}
UNIONS J155517+415138	1.13	$1.402^{+0.010}_{-0.011}$	$0.787^{+0.026}_{-0.026}$	5^{+9}_{-8}	$0.089^{+0.012}_{-0.010}$	48^{+6}_{-9}
UNIONS J090728+423301	2.52	$1.511^{+0.003}_{-0.003}$	$0.579^{+0.011}_{-0.010}$	39^{+1}_{-1}	$0.031^{+0.003}_{-0.003}$	3^{+6}_{-6}
UNIONS J091415+431855	1.02	$1.249^{+0.005}_{-0.005}$	$0.766^{+0.018}_{-0.018}$	-37^{+3}_{-3}	$0.020^{+0.006}_{-0.005}$	33^{+14}_{-14}
UNIONS J143611+432921	0.97	$2.146^{+0.022}_{-0.007}$	$0.710^{+0.014}_{-0.018}$	-13^{+3}_{-3}	$0.080^{+0.007}_{-0.005}$	-52^{+7}_{-5}
UNIONS J103148+434256	1.04	$1.392^{+0.005}_{-0.005}$	$0.756^{+0.015}_{-0.015}$	-38^{+4}_{-4}	$0.119^{+0.005}_{-0.005}$	71^{+2}_{-2}
UNIONS J163522+510828	1.17	$2.092^{+0.007}_{-0.008}$	$0.760^{+0.015}_{-0.015}$	-36^{+3}_{-3}	$0.040^{+0.006}_{-0.005}$	49^{+8}_{-8}
UNIONS J131730+564054	1.10	$2.020^{+0.028}_{-0.039}$	$0.684^{+0.023}_{-0.026}$	-56^{+6}_{-7}	$0.138^{+0.020}_{-0.024}$	35^{+7}_{-14}
UNIONS J144037+591141	1.61	$1.979^{+0.010}_{-0.011}$	$0.668^{+0.013}_{-0.013}$	38^{+2}_{-2}	$0.073^{+0.007}_{-0.007}$	25^{+5}_{-5}
UNIONS J081959+535624	5.88	$1.929^{+0.014}_{-0.018}$	$0.393^{+0.023}_{-0.021}$	43^{+2}_{-2}	$0.093^{+0.007}_{-0.011}$	15^{+4}_{-9}
UNIONS J081959+535624*	2.69	$2.291^{+0.004}_{-0.005}$	$0.744^{+0.014}_{-0.014}$	58^{+3}_{-2}	$0.185^{+0.003}_{-0.007}$	57^{+1}_{-1}

Notes. R is the Einstein radius, q_m is the axis ratio (semi-minor/semi-major axis), and ϕ_m is the position angle of the lens mass SIE model. The parameters γ_{ext} and ϕ_{ext} correspond to the strength and angle of the external shear, respectively. We indicate with an asterisk the parameters for our best fits (i.e., for the manually customized mask) when applicable.

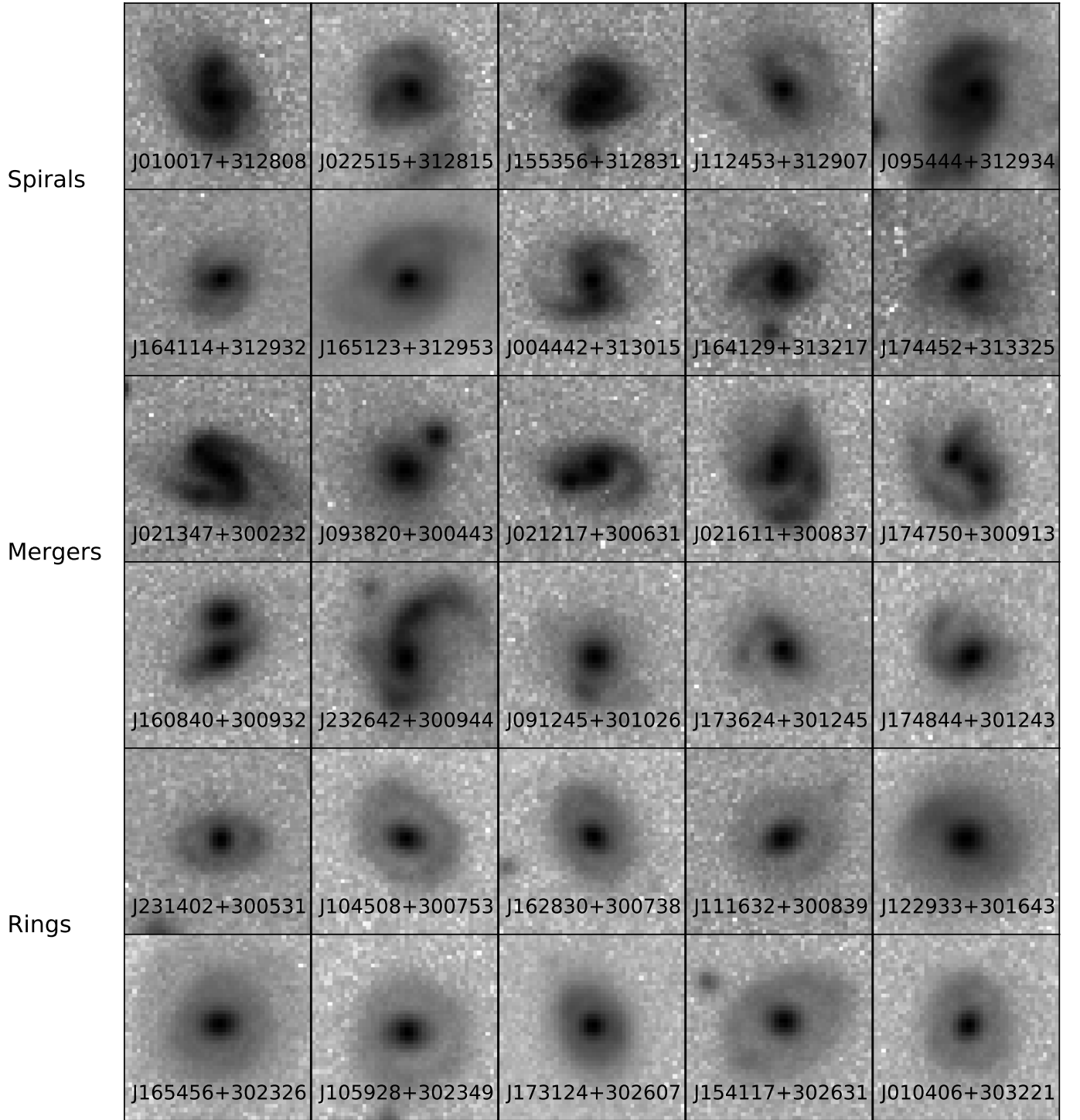


Fig. E.1. Examples of images classified by at least one person as spirals (first two rows), mergers (two rows in the middle), or ring galaxies (last two rows) during the visual inspection. Each of them were identified by at least one user.

Table F.0. Lens candidates obtained after the final stage of visual inspection.

Name	RA	DEC	CNN score	References
UNIONS J165710+315052	254.294217	31.847955	0.99	[8], [11]
UNIONS J115830+312349	179.62672	31.397025	0.79	None
UNIONS J131730+564054	199.378529	56.68182	0.97	None
UNIONS J143611+432921	219.046299	43.489219	0.98	[10]
UNIONS J090728+423301	136.86744	42.550401	0.99	[8],[12], [11]
UNIONS J113952+303204	174.968967	30.534686	0.99	None
UNIONS J103148+434256	157.952047	43.715754	0.99	[8], [11]
UNIONS J174828+374539	267.120502	37.760876	0.99	None
UNIONS J163522+510828	248.845539	51.141378	0.99	None
UNIONS J143004+410557	217.517011	41.099238	0.98	[1]
UNIONS J171340+364204	258.419705	36.701363	0.97	None
UNIONS J021219+303310	33.082568	30.552995	0.99	[11]
UNIONS J235527+325244	358.865775	32.879142	0.99	[8]
UNIONS J153959+354954	234.999752	35.831739	0.98	None
UNIONS J124548+375435	191.451962	37.909935	0.99	None
UNIONS J095335+313238	148.399699	31.544166	0.99	None
UNIONS J094411+322038	146.049312	32.344091	0.99	[8]
UNIONS J165049+312328	252.706648	31.391165	0.99	[8]
UNIONS J081959+535624	124.999119	53.940081	0.7	[2]
UNIONS J090129+303355	135.371629	30.565434	0.87	None
UNIONS J155923+314712	239.847219	31.786779	0.87	[8]
UNIONS J155517+415138	238.823966	41.860723	0.99	[6],[9],[8],[12]
UNIONS J093542+303803	143.925535	30.63432	0.82	None
UNIONS J124223+312905	190.598116	31.484828	0.99	[11]
UNIONS J164940+304909	252.418703	30.819194	0.99	[8]
UNIONS J112053+342146	170.22424	34.362826	0.99	None
UNIONS J075346+341633	118.442572	34.276024	0.98	[5]
UNIONS J075523+344539	118.847955	34.761005	0.99	[4],[7]
UNIONS J094819+373739	147.081482	37.627583	0.99	None
UNIONS J144037+591141	220.155419	59.194928	0.94	None
UNIONS J110219+334224	165.579915	33.706789	0.96	None
UNIONS J091415+431855	138.564324	43.31547	0.99	[12]
UNIONS J112901+311925	172.254715	31.323845	0.95	None
UNIONS J143158+554329	217.995754	55.724873	0.64	[3]
UNIONS J164753+511308	251.974251	51.219046	0.84	None
UNIONS J104949+362926	162.455302	36.490762	0.85	None
UNIONS J150036+551015	225.150662	55.170865	0.99	None
UNIONS J115458+362005	178.741706	36.334761	0.99	None
UNIONS J164822+332824	252.092057	33.473465	0.95	None
UNIONS J082141+391539	125.422624	39.260887	0.6	None
UNIONS J094931+343819	147.380711	34.638801	0.88	None
UNIONS J113203+430718	173.014645	43.121823	0.92	None
UNIONS J152414+502559	231.059906	50.433218	0.72	None
UNIONS J141559+502742	213.997269	50.461768	0.82	None
UNIONS J092056+404009	140.233516	40.669275	0.81	None
UNIONS J154613+320732	236.55542	32.12574	0.76	None
UNIONS J163115+324034	247.813118	32.676273	0.84	None
UNIONS J080114+455838	120.309377	45.9774	0.96	None
UNIONS J071431+543145	108.63086	54.529382	0.96	None

Table F.0. continued.

Name	RA	DEC	CNN score	References
UNIONS J232205+311448	350.521187	31.246802	0.97	None
UNIONS J091952+390350	139.970218	39.064101	0.83	None
UNIONS J001119+324635	2.83263	32.77663	0.99	None
UNIONS J114114+414544	175.309537	41.762407	0.9	None
UNIONS J021058+304423	32.744232	30.739869	0.98	[11]
UNIONS J005359+340231	13.498377	34.042151	0.75	None
UNIONS J085335+352650	133.397432	35.447405	0.99	None
UNIONS J004211+362510	10.549333	36.419549	0.63	None
UNIONS J121351+353334	183.466463	35.559484	0.84	None
UNIONS J172451+532101	261.216655	53.35028	0.83	None
UNIONS J164624+321352	251.600986	32.231342	0.75	None
UNIONS J074144+311642	115.434351	31.27839	0.79	None
UNIONS J165239+442135	253.165234	44.359909	0.92	None
UNIONS J015835+325940	29.649733	32.994532	0.95	None
UNIONS J124724+364403	191.85275	36.734418	0.51	[8],[11]
UNIONS J233812+301327	354.552069	30.224249	0.73	None
UNIONS J131719+355859	199.329276	35.983279	0.94	None
UNIONS J132126+323845	200.360122	32.645929	0.98	None
UNIONS J171127+380843	257.863509	38.145351	0.86	None
UNIONS J135131+520718	207.883099	52.121882	0.99	[12]
UNIONS J083044+322708	127.685572	32.452379	0.96	None
UNIONS J121959+340210	184.997596	34.036332	0.95	None
UNIONS J101441+504451	153.67308	50.747535	0.98	None
UNIONS J112035+325158	170.146207	32.866338	0.73	None
UNIONS J020255+300539	30.732029	30.094366	0.53	None
UNIONS J151250+510553	228.208395	51.098312	0.995	None
UNIONS J154613+350448	236.555281	35.080241	0.95	[11]
UNIONS J104233+362936	160.640653	36.49354	0.98	None
UNIONS J163847+320109	249.696824	32.019203	0.84	None
UNIONS J164433+503638	251.140747	50.610568	0.97	None
UNIONS J154600+303600	236.503001	30.600112	0.86	None
UNIONS J133645+495308	204.191213	49.885661	0.82	None
UNIONS J152012+402157	230.051775	40.365852	0.93	None
UNIONS J145432+584957	223.634534	58.832696	0.99	[12]
UNIONS J091957+413848	139.990868	41.646885	0.99	None
UNIONS J130530+321143	196.376255	32.195307	0.82	None
UNIONS J073301+513431	113.256743	51.575494	0.98	[11]
UNIONS J164700+315322	251.753434	31.889623	0.59	None
UNIONS J175507+334318	268.781414	33.72189	0.56	None
UNIONS J232225+341458	350.605096	34.249556	0.94	None
UNIONS J112140+303748	170.420714	30.630111	0.95	None
UNIONS J111018+351716	167.576918	35.287779	0.88	None
UNIONS J145252+580240	223.220491	58.04453	0.99	[12]
UNIONS J102147+500808	155.447437	50.135595	0.99	None
UNIONS J130354+315655	195.97702	31.9487	0.69	None
UNIONS J021336+301804	33.403042	30.301241	0.9	None
UNIONS J182524+590228	276.351457	59.041283	0.77	None
UNIONS J173658+305817	264.245714	30.971567	0.95	None
UNIONS J082216+381337	125.570203	38.227187	0.69	None
UNIONS J165643+535717	254.180951	53.954788	0.55	None
UNIONS J182125+510038	275.357473	51.010559	0.89	None
UNIONS J234747+341149	356.947866	34.196982	0.79	None
UNIONS J150803+485728	227.013274	48.957797	0.55	None
UNIONS J135800+492334	209.504125	49.392834	0.99	None
UNIONS J090919+351503	137.330317	35.25094	0.99	[8]
UNIONS J130259+300953	195.745854	30.164815	0.93	None
UNIONS J120729+355838	181.87263	35.977381	0.83	None
UNIONS J095505+303230	148.771722	30.541826	0.87	[12]
UNIONS J071707+701827	109.279891	70.307512	0.78	None
UNIONS J082055+533513	125.232609	53.587001	0.84	None
UNIONS J120329+311304	180.874396	31.218011	0.92	None

Table F.0. continued.

Name	RA	DEC	CNN score	References
UNIONS J093114+300703	142.812224	30.117749	0.96	None
UNIONS J160308+333036	240.786202	33.510113	0.99	[8]
UNIONS J231403+335806	348.51442	33.968598	0.5	None
UNIONS J171725+461830	259.357806	46.308496	0.99	None
UNIONS J165911+371346	254.797098	37.22968	0.96	None
UNIONS J120659+323602	181.749009	32.600781	0.88	None
UNIONS J174744+391856	266.934787	39.315775	0.97	None
UNIONS J135525+503250	208.857143	50.547389	0.98	None
UNIONS J102618+414208	156.575959	41.702331	0.88	None
UNIONS J164113+370415	250.304317	37.070838	0.99	None
UNIONS J114609+395955	176.537956	39.998664	0.99	None
UNIONS J151505+612848	228.771508	61.480081	0.97	None
UNIONS J114050+322955	175.209408	32.498829	0.96	None
UNIONS J100021+474746	150.090655	47.796124	0.99	None
UNIONS J123129+302224	187.874467	30.37355	0.95	None
UNIONS J162540+331556	246.4202	33.265777	0.81	[11]
UNIONS J160903+335108	242.265886	33.852487	0.86	None
UNIONS J011119+365506	17.832415	36.918446	0.97	None
UNIONS J144129+601008	220.374691	60.169088	0.88	None
UNIONS J100422+364821	151.093545	36.805996	0.73	None
UNIONS J234324+305858	355.852619	30.982916	0.9	None
UNIONS J165402+444252	253.51002	44.71453	0.98	None
UNIONS J081258+305338	123.242546	30.894062	0.95	None

Notes. The first section corresponds to the secure lens candidates (SL) and the bottom section to the maybe lens candidates (ML)

References. [1] Bolton et al. (2008), [2] Inada et al. (2009), [3] Paraficz et al. (2016), [4] Shu et al. (2016), [5] Shu et al. (2017), [6] Sonnenfeld et al. (2018), [7] Cao et al. (2020), [8] Cañameras et al. (2020), [9] Chan et al. (2020), [10] Jaelani et al. (2020), [11] Huang et al. (2021), [12] Talbot et al. (2021)

Strong lens systems search in the Dark Energy Survey using Convolutional Neural Networks

K. Rojas¹, E. Savary¹, B. Clément¹, M. Maus¹, F. Courbin¹, C. Lemon¹, J. H. H. Chan¹, G. Vernardos¹, R. Joseph², R. Cañameras³, and A. Galan¹

¹ Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland

² Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA

³ Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany

August 7, 2022

ABSTRACT

We present a search for strong lens galaxy-scale systems in the first data release of the Dark Energy Survey (DES), from a color-selected parent sample of 18 745 029 Luminous Red Galaxies (LRGs). A Convolutional Neural Network (CNN) is used to grade this LRG sample with values between 0 (non-lens) and 1 (lens). Our training set of mock lenses is data-driven, i.e. using lensed sources taken from *HST*-COSMOS images and lensing galaxies from DES images of our LRG sample. A total of 76 582 cutouts obtained a score above 0.9. These were visually inspected and classified into two catalogs. The first one contains 405 lens candidates, where 90 present clear lensing features and counterparts, while the others 315 require more evidence, such as higher-resolution imaging or spectra, to be conclusive. A total of 186 candidates are newly identified by our search, of which 16 are among the 90 best ones. The second catalog includes 539 ring galaxy candidates. This catalog will be a useful false positive sample for training future CNNs. For the 90 best lens candidates we carry out color-based deblending of the lens and source light without fitting any analytical profile to the data. This method is shown to be very efficient in the deblending, even for very compact objects and for objects with complex morphology. Finally, from the 90 best lens candidates we select 52 systems having one single deflector, to test an automated modeling pipeline which successfully modeled 79% of the sample within an acceptable amount of computing time.

Key words. Gravitational lensing: strong – Surveys – Techniques: image processing

1. Introduction

Gravitational lensing is the phenomenon by which light rays are deflected by a gravitational field. In *strong* lens systems it is possible to observe multiple images, arcs or rings of a distant source around a foreground galaxy, group or cluster. In such cases, the systems can serve as important tools to study diverse and fundamental questions about the Universe. Some examples are the study of luminous and dark matter components of the deflector (Kochanek & Dalal 2001; Oguri et al. 2002; Davis et al. 2003; Jiménez-Vicente et al. 2015), measuring the Hubble constant H_0 using time delays (Falco et al. 1997; Vuissoz et al. 2007; Bonvin et al. 2017; Wong et al. 2020; Millon et al. 2020), and constraining the dark energy equation of state (Biesiada et al. 2010; Collett & Auger 2014; Cao et al. 2012, 2015). However, most of these applications are limited by the paucity of known systems; only a few hundred such systems are confirmed. Therefore, an effort on the discovery and confirmation of more lenses is required.

Since the serendipitous discovery of the first lensed quasar (Walsh et al. 1979), discovery methods have developed depending on novel datasets and techniques. Recent searches include: algorithms based on identifying lens features like Arc-finder (Alard 2006), Ringfinder that searches for blue features blended with red light (Gavazzi et al. 2014), Principal Component Analysis (PCA) of galaxies to search for lensed features in the residual images using machine learning (Joseph et al. 2014; Paraficz et al. 2016), and CHITAH which evaluates point source config-

urations as possible lensed images using lens modeling (Chan et al. 2015). In recent years the growing amount of available data has motivated the use of more automated techniques like Artificial Neural Networks (ANNs) (Rosenblatt 1957) and in particular Convolutional Neural Networks (CNNs; LeCun et al. 1989). This technique is based on supervised machine learning algorithms capable of solving complex problems such as pattern recognition or image classification when a proper training set is provided.

The biggest challenge of using CNNs for lens finding is creating a robust training set that contains diverse lens systems for positive examples and non-lens galaxies, including some that can be mistaken as lenses like spirals, rings, and mergers, as negative examples. We currently lack sufficient numbers of known examples of both lens systems and common false positives. The only solution is to then simulate them as realistically as possible. This has already been addressed in several ways: fully simulating images using analytical profiles for both the lens and source (Jacobs et al. 2019a), using an analytical profile for the source but a real image of the lens (Petrillo et al. 2019b), and using real data for both the deflector and background galaxy (Cañameras et al. 2020). The fully-analytical approach has the advantage of having full control over all parameters to create a sample as varied as possible, but lacks the ability to mimic features of real images like artifacts, noise, and companion galaxies.

The main differences between previous searches depend on mock simulation methods, use of single or multi-band data, and the architecture design. Previous searches include: the Canada-

France-Hawaii Telescope Legacy Survey (CFHTLS) (Jacobs et al. 2017), the Kilo Degree Survey (KiDS) (Petrillo et al. 2017, 2019a,b; He et al. 2020), the Dark Energy Survey (DES) Year 3 (Jacobs et al. 2019a,b), the Dark Energy Spectroscopic Instrument (DESI) Legacy Imaging Surveys (Huang et al. 2020, 2021), the Pan-STARRS 3 π survey (Cañameras et al. 2020), the VST Optical Imaging of the CDFS and ES1 fields (VOICE survey) (Gentile et al. 2021). Overall, these studies have shown that CNNs are a promising tool, listing thousands of new lens candidates. However, all of them rely on large amounts of human visual inspection afterwards to compile the final candidate list. Improving the training process with realistic lenses and diverse types of galaxies is therefore of key importance when the next generation of surveys like Euclid Space Telescope (Laureijs et al. 2011) and the Rubin Observatory Legacy Survey of Space and Time (LSST, LSST Science Collaboration et al. 2009; Ivezić et al. 2019) start producing data. Current lens finding efforts, including some performed in simulated data (Lanusse et al. 2018; Avestruz et al. 2019), serve as important preparation, as it is expected that over 100 000 new strong lensing systems will be discovered (Collett 2015), and a visual inspection process is not affordable.

In this work our main aim is to find new strong lensing systems. As a secondary goal we want to characterize a subset of our false positives. We perform our search in the footprint of the Dark Energy Survey (DES) (The Dark Energy Survey Collaboration 2005). The data description is presented in Section 2. Details of the simulation procedure of our training sets are in Sect. 3. Sect. 4 presents details of both the CNN training and validation. We grade the parent sample using the CNN model, and we select the ones with higher score to perform a dedicated visual inspection to identify the best lens candidates and subclassify false positives. In Sect. 5 we present our visual classification procedure for the best graded candidates from the CNN, presenting both the best lens candidates and a compilation of ring galaxy candidates that will help to improve future lens finding searches. Finally in Sect. 6 we show the results of an automatic modeling tool on a sample of our best candidates.

2. Data selection

We use data from DES, which used the Dark Energy Camera (DECam, Honscheid & DePoy 2008; Flaugher et al. 2015) on the Blanco 4-m telescope at Cerro Tololo Inter-American Observatory (CTIO), Chile. DECam is a 570 Megapixel camera with a field of view of 2.2 square degrees and a pixel size of 0.27". Observations were performed in the optical *grizY* bands. The first DES Data release (DES-DR1, Abbott et al. 2018) contains images taken over the first three years of operation, covering an area of 5186 deg². The images from DR1 have been coadded, and each filter has been re-scaled to have a fixed zero point of 30 mag.

We use the NOAO Data Lab (Fitzpatrick et al. 2016) service to build our sample from the `des_dr1.galaxies` catalog and selected a sample of Luminous Red Galaxies (LRGs) in order to maximize the lensing cross-section (e.g. Turner et al. 1984). To do so, we applied the following cuts in color and magnitude:

$$\begin{aligned} 1.8 < g - i < 5, \\ 0.6 < g - r < 3, \\ 18 < r < 22.5, \\ g > 20, \\ i > 18.2, \end{aligned} \quad (1)$$

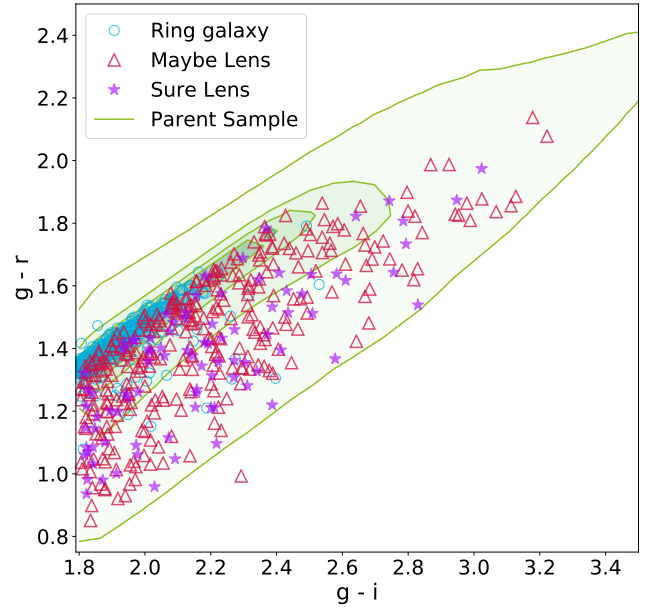


Fig. 1. Color-color diagram of our parent sample of ~19 million galaxies. The plot displays where most galaxies in our selection are located (see Eq. 1). The green shaded area shows the density of galaxies from the parent LRG sample. The green solid lines shows the 1, 2, 3, and 6 σ contours. The lens candidates and contaminants from this work are shown in overlay: Sure Lens (purple stars), Maybe Lens (open red triangles), and ring galaxies (open cyan circles). See Sect. 5.3 for details.

where the magnitude used is the `mag_auto` column reported in the DES data release. Our color selection is summarized in Fig. 1 and is similar to the one adopted by Jacobs et al. (2019b). However, we slightly widened the $g - i$ range and adopted a brighter magnitude limit in the r -band. This selection allows us to better account for the contamination of the lens light by the bluer color of any putative lensed source, but also increases the probability that other types of galaxies, e.g. mergers, spirals, or ring galaxies, can be selected in the sample. The result is a sample of 18 745 029 galaxies located in 10 388 coadded tiles from DES DR1, which we refer to as the parent sample in this work. We downloaded all the g , r , and i bands of these tiles, and generated cutouts around the galaxies of 50×50 pixels, corresponding to $\sim 13 \times 13''$. When the Point Spread Function (PSF) is required for the simulation process and modeling, PSFex (Bertin 2011) is run on the relevant coadd tile, extracting a model of the PSF from the FITS image, allowing us to retrieve a PSF at any position on the tile.

3. Simulated galaxy-scale lenses

Since our search for galaxy-scale strong lenses will use a CNN, we require a training set that mimics as closely as possible both the lenses we want to find and also the common non-lensed systems. When simulating lenses, we adopt an approach where the training set is data-driven, in the sense that both the images of the lenses and the background sources are obtained from real data. The outline of our procedure is as follows. First, we select a sample of high-redshift background galaxies with available high-resolution imaging and accurately measured colors (Sect. 3.1).

Then, we select a sample of LRGs that will act as deflectors and match them to the background sources to create pairs of lens-source suitable for simulations (Sect. 3.2). Finally, we lens the source light using the lens equation – whose parameters (deflection angles) are defined by the lens – to produce a simulated image of a lens system that combines images from both samples (Sect. 3.3).

3.1. Background galaxies selection

We want realistic source galaxies, imaged at *HST*-resolution, with color information and high S/N. Such a catalog of background sources has already been compiled by Cañameras et al. (2020). The galaxies in this catalog were selected from the Galaxy Zoo catalog (Willett et al. 2017), and are in the COSMOS2015 photometric catalog (Laigle et al. 2016). All the objects categorized as galaxies in these catalogs were picked, with no previous selection in color or magnitude, as the depth is limited by the Galaxy Zoo selection, i.e. down to F814W ~ 23.5 . Stars and artifacts were manually removed from the sample, but also extended galaxies and galaxies with nearby companions were removed, leaving a final sample of 52 696 objects. Spectroscopic redshifts were obtained from several follow-up surveys (Lilly et al. 2007; Comparat et al. 2015; Silverman et al. 2015; Le Fèvre et al. 2015; Tasca et al. 2017; Hasinger et al. 2018), and photometric redshifts from Laigle et al. (2016) for the ones lacking spectra.

To create high-resolution *gri*-images of our sources, we created cutouts combining the morphological information from *HST*/ACS F814W high-resolution images (Leauthaud et al. 2007; Scoville et al. 2007; Koekemoer et al. 2007) and the color information from Hyper Suprime Cam (HSC) ultra-deep stack images (Aihara et al. 2018). The detailed procedure to combine the information from these two surveys is described in Cañameras et al. (2020) who followed the steps described in Griffith et al. (2012). In summary, *HST*/ACS F814W images were aligned and rescaled as if they were observed in the HSC *i*-band. These HSC images were then resampled to the resolution of the *HST*/ACS F814W images and were multiplied by an illumination map obtained by dividing the *HST*/ACS F814W image by the HSC *i*-band image. Each galaxy stamp has a size of $10 \times 10''$ and a pixel size of $0.03''$, i.e. *HST* resolution and PSF but with the HSC observed colors. The morphology of the source is the same in each band. Since the *HST* PSF is much sharper than that of the ground-based DES images we do not deconvolve our stamps from the *HST* PSF, which would introduce noise and possible artefacts.

3.2. Lens-source association

Ideally, we would want spectroscopic redshift and velocity dispersions for each member of our LRG sample but the vast majority are lacking this information. To cope with this limitation, we perform a prediction of those parameters using a simple K-nearest-neighbors (KNN) algorithm, assuming that other galaxies with similar *gri* magnitudes will also have similar redshifts and velocity dispersions. For a reference data set where colors, redshifts and velocity dispersions are available, the KNN algorithm provides a match between the *gri* magnitudes and the redshift and velocity dispersion of the galaxies. It then predicts the redshift and velocity dispersions for new data based on the K-objects with the most similar colors. We trained the algorithm with 1 400 000 SDSS galaxies that match the color-magnitude

cuts of the parent sample and have redshift and velocity dispersion measurements available. We tested the model on another set of 99 382 spectroscopically-confirmed SDSS galaxies, obtaining the predictions for the parameters from the ten nearest neighbors in the *gri* color space of the training set. We found that the *rms* scatter in the predictions was $\sigma_z = 0.06$ for the redshift, and $\sigma_{vel} = 69 \text{ km s}^{-1}$ for the velocity dispersion. Finally, we used this model to predict the most likely redshift and velocity dispersion for each of our galaxies in the parent sample. The distributions of the predicted redshifts and velocity dispersions are shown in Fig. 2.

We now pair LRGs with source galaxies, requiring that our simulations have a uniform distribution in Einstein radii spanning $1.2'' < \theta_E < 3.0''$. We choose a conservative lower limit on θ_E because we noticed that, given the average seeing in the *gri* bands of $1.12''$, $0.97''$, and $0.88''$ respectively (Abbott et al. 2018), simulations with $\theta_E < 1.2$ create lensing features that are too close or blended with the lens galaxy, and can easily be mistaken as non-lenses, e.g. galaxies with extended disks. To evaluate θ_E we use the redshift of the source galaxy, the redshift of the lens, and the velocity dispersion of the lens. To match a source to each LRG, we first take a random lens galaxy from the parent sample and compute the Einstein radii for all source galaxies. We then form lens-source pairs that produce Einstein radii falling only within our desired bounds and filling bins that produce an uniform distribution. In case no lens-source pairs satisfy the θ_E conditions we artificially increase the velocity dispersion of the lens galaxy up to 1.5 times its original value. If still no pair satisfies our criteria we discard the LRG. Note that, as illustrated in Fig. 2, this results in a high bias of the velocity dispersion distribution of the lenses. Although this procedure tends to produce lenses with dark matter halos larger than the predicted from the actual galaxy velocity dispersion, this ensures that the lensing features are clearly noticeable to the CNN.

Finally, we enforce the final Einstein radii distribution to be uniform. In other words, our training set is not representative of the true distribution of Einstein radii on the sky, but gives equal probability to all possible values, allowing for more discriminative power in our trained CNN.

3.3. Lensing simulation

We now combine these components to create realistic images of lenses. We adopt the Singular Isothermal Ellipsoid (SIE) as our lensing mass model, which is defined by the Einstein radius (θ_E), the Position Angle (PA) and axis ratio converted into a complex ellipticity (e_1, e_2), and the central position (x_1, x_2). As mentioned before, the Einstein radius follows a uniform distribution, while the other parameters are acquired individually according to the light distribution of each lens galaxy. Our simulation toolbox uses the Python package *Lenstronomy*¹ (Birrer et al. 2015; Birrer & Amara 2018). The first step is to determine a simple but realistic representation for the mass of the lensing object. The Einstein radius is calculated using the lens and source redshifts as well as the lens velocity dispersion of the lens derived in Sec. 3.2. The ellipticity parameters and mass centroid are estimated by fitting an elliptical Sérsic profile to the DES *r*-band image of the LRG. We optimized the fitting procedure using 50 iterations of Particle Swarm Optimization (PSO; Kennedy & Eberhart 1995), with 50 particles. This simple model provides us with parameters for a mass distribution that broadly follows the light distribution of the brightest object in the image. Note

¹ <https://github.com/sibirrer/lenstronomy>

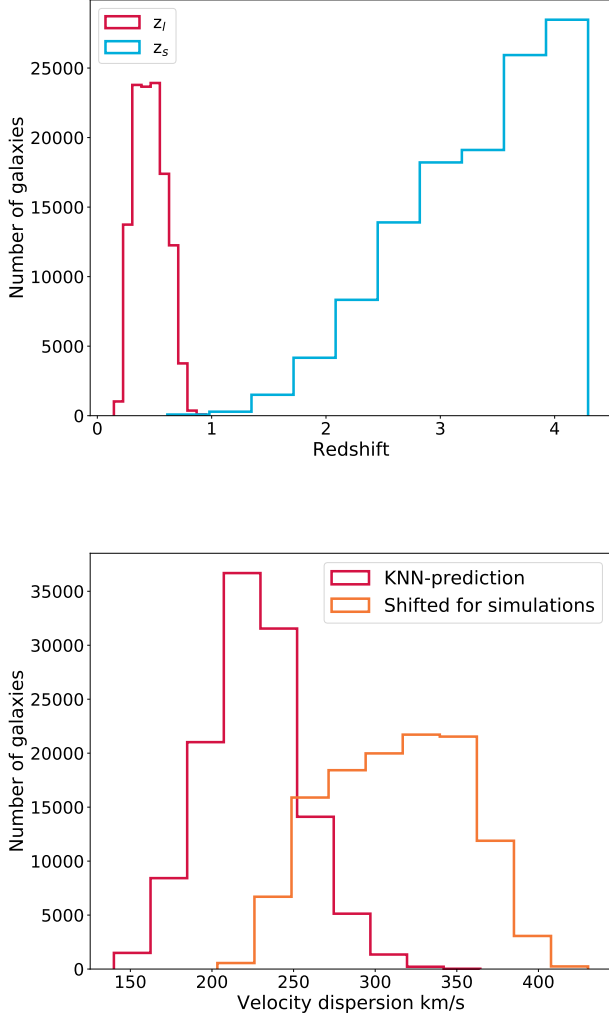


Fig. 2. *Top panel:* redshift distributions of the lenses (red) and sources (blue) in the simulated training set. *Bottom panel:* lens velocity dispersion distribution, predicted from the K-nearest neighbor algorithm (red) and shifted to match the criteria of pairing lenses and sources (orange), as described in Sec. 3.2. Note that the actual lens velocity dispersions used in the simulations are shifted to higher values, so that lensing features can be seen even at the DES resolution.

that limiting to 50 iterations results in ellipticity parameters that are not perfect and that naturally mimic the effect of deviations of the dark matter profile with respect to the light, without introducing extra complexity in the simulation pipeline. Even though it could produce a few lenses with exotic properties, e.g. very elliptical mass profiles or unusually large dark matter halos, it was found to be adequate for our goal of building realistic simulations in the vast majority of the cases.

The second step is to deflect the light rays from the source according to the lensing mass model. To ensure we can distinguish the final lensed source features against the lens galaxy light we first increase the original source brightness by one magnitude. To decide where in the source plane our background galaxy is located, we select a random position inside a square that encloses the caustic curves (curves that mark the location of the maximum magnification and delimit the region inside which a source will be multiply-imaged). Then, we perform a ray-tracing sim-

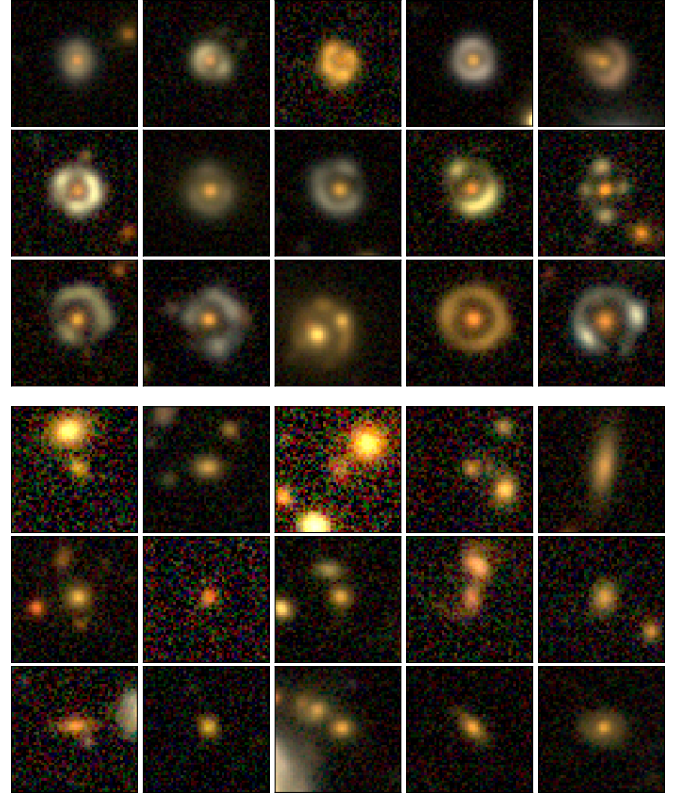


Fig. 3. *Top:* examples of simulated lenses based on real DES images. Stamps are ordered by increasing Einstein radii (top-left to bottom-right). The top row corresponds to $\theta_E=1.2-1.8''$, the middle row to $\theta_E=1.8-2.4''$, and the bottom row to $\theta_E=2.4-3.0''$. *Bottom:* examples of LRGs used either as non-lenses during the training of the CNNs (see Sec. 4.1) or as objects onto which we inject a lensed source to build simulated lens systems, as shown in the top panel. All the cutouts are 50 pixels on-a-side, corresponding to $13''$.

ulation to map the source image onto the image plane and we further convolve the resulting lensed source with the relevant stamp PSF. To convert this image into the DES characteristic pixel resolution we down-sample the pixels from $0.03''$ (HST) to $0.27''$ (DES), and re-scale the flux to match the DES zero points in each filter. As a last step, we add the convolved, resized, and flux-normalized image of the lensed source to the original image of the LRG lens. The latter has, by construction, the right DES PSF and noise properties. Thus, our simulations preserve the characteristics of the original image, such as background noise, seeing, the presence of artifacts, and neighbouring galaxies or stars in the field of view.

To build the multi-band *gri* simulations we use the same mass model for all bands, with its parameters derived only from the *r*-band, and lens the source image in each band according to this model. We then add the lensed source in each band to the corresponding image of the lens taken from the DES images in the *g*, *r* and *i* bands. Our final set of simulated galaxy-scale lenses consists of 100 000 systems with a uniformly distributed Einstein radius in the range $1.2'' < \theta_E < 3.0''$. Examples of these stamps are shown in Fig. 3 as well non-lens objects.

4. Lens Finding using CNN

Artificial Neural Networks (ANNs; Rosenblatt 1957) consist of an interconnected group of nodes which are typically organized

into the so-called input, hidden, and output layers. In particular, CNNs (LeCun et al. 1989) – which are especially good at solving image classification problems (He et al. 2015) – have hidden layers that are of key importance, as they highlight the patterns in the data using a series of convolutional, pooling, normalization and fully connected layers. The level of abstraction in the pattern features increases with the depth of the convolutional layers, helping classification of objects into different classes. Here, we train a CNN to recognize strong lens systems against isolated red galaxies.

4.1. CNN training

The training set, consisting of 50×50 pixel cutouts in each of the *gri*-bands, is composed of two equal subsets: the first being 100 000 simulated lens cutouts from Sect. 3, and the other containing LRGs that were not used in the simulation process. We labeled our data using 1 for lenses and 0 for non-lenses. We kept 20% of each sample as a validation set. Before training the CNN we pre-processed our data by normalizing each image brightness to range between 0 and 1. We also augmented our sample by flipping each image horizontally and vertically. Data augmentation increases the probability that the network correctly classifies different orientations of the same image, but it does not transform the CNN into a rotationally invariant one. To achieve this, a different architecture must be used that is not explored in this work. The training process is performed using the Keras Deep Learning framework (Chollet et al. 2015).

Our CNN uses a model from the EfficientNet family (Tan & Le 2020), which has been designed to achieve better performance than other CNNs. The network of this model uses a compound coefficient to scale the depth, width, and resolution, which are key parameters for obtaining better accuracy and efficiency. EfficientNet implementation in Keras counts with 8 different variants B0-B7, whose depth, width, and resolution parameters have been carefully selected and tested to produce good results. The complexity and requirements of the models grow as we move from B0 to B7. As running a more complex model also implies the use of more computational resources, we decided to use an EfficientNet-B0, whose architecture is described in (Tan & Le 2020), and is sufficient for our classification task and the characteristics of our data. After this CNN model we add a sequence of fully connected hidden layers. The network has a total of 4 182 205 trainable parameters.

During training, the neural network learned how to grade images of galaxies and distinguish between lenses and non-lenses. At each iteration the network analyzed subsets of 32 images. When all iterations are completed through the entire training set, it is counted as one epoch. Within each epoch the accuracy and loss of the model is monitored using the validation set. We minimize a binary cross-entropy loss function using a stochastic gradient descent optimizer (Adam) with a learning rate of 0.0001, and stop the training if either the loss value does not improve by more than 0.0001 during 10 epochs, or if 100 epochs are reached.

4.2. Evaluation of the CNN performance

The network provides a score, S_{CNN} , between 0 and 1, for each processed image. This means that those images classified as lenses obtain $S_{\text{CNN}} \sim 1$ while non-lenses obtain $S_{\text{CNN}} \sim 0$. The training process was performed in a single GPU Nvidia GTX 1080 Ti in about 8 hours. It converged, within our criteria above, after 57 epochs and achieved a 99.9 (99.8) percent accuracy in

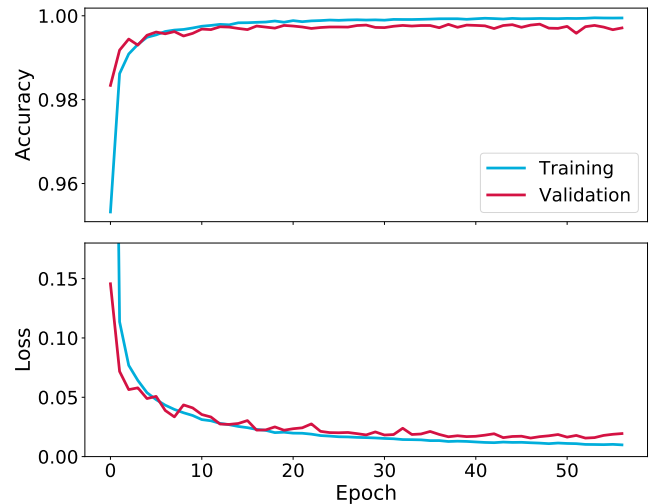


Fig. 4. Learning curves for the accuracy (top) and loss (bottom) for the training (blue) and validation (red) sets, as a function of epoch. In the case of loss we crop the y-axis, but the initial training set loss is 0.76.

the training (validation) sets and a loss of 0.01 (0.02). This near-perfect accuracy achieved in the training set might be understood as overfitting, thus, to evaluate this possibility we compared the loss and accuracy learning curves for the training and validation sets (Fig. 4). For both loss and accuracy we see that after 10 epochs the training set reaches a stable point with minimal changes, while the validation set follows the same trend with a small gap showing less accuracy and more loss than the training set, as is expected. The lack of overfitting signs, i.e. training loss continues decreasing and/or the validation loss starts increasing again after several epochs, leads us to the conclusion that our model is able to learn and generalize this classification problem.

In order to evaluate the performance of the CNN we built two test sets. The first one contains 40 000 cutouts with the same characteristics as the training set, i.e. half simulated lenses and half LRGs. The second one has 636 cutouts where half are known lenses or lens candidates (visually selected to have noticeable lensing features) and half are LRGs not seen by the CNN during training. The known lenses are taken from the Master Lens Database² and the candidates from Jacobs et al. (2019b,a). The purpose of this second test set is to have a more realistic idea of the performance of the CNN in grading real strong lens systems instead of simulations. The distribution of S_{CNN} for both test sets (Fig. 5) shows that objects labelled as lenses are concentrated around $S_{\text{CNN}} > 0.9$ and non-lenses around $S_{\text{CNN}} < 0.1$, as expected.

To evaluate the number of lenses correctly identified, we use a Receiver Operating Characteristic (ROC) curve, (Fig. 6) which shows the True Positive Rate (TPR) against the False Positive Rate (FPR), and are naturally both functions of the decision threshold applied to the score. It illustrates the performance of a binary classifier in discriminating between the two classes as the decision threshold is varied. The first test set shows a very good performance reaching an accuracy of 99.7% and a loss of 0.02. From the ROC curve we see that choosing $S_{\text{CNN}} = 0.5(0.9)$ gives a TPR = 99.8% (99.4%) and a FPR = 0.21% (0.12%). On the other hand, in the second test set the performance of the net-

² <http://admin.masterlens.org/index.php>

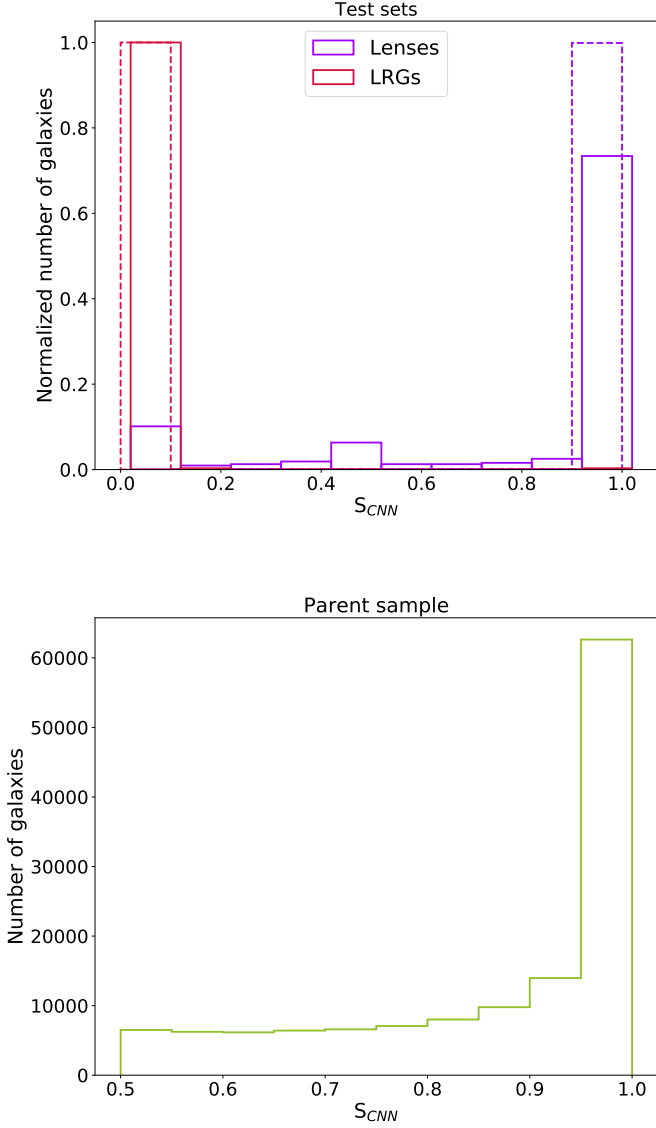


Fig. 5. CNN score (S_{CNN}) distribution for the different datasets. The top panel shows the S_{CNN} distribution of both test sets: lens simulations (dashed line) and real confirmed/candidate lenses (solid line). Both datasets contain images labeled as lenses (purple lines) and real LRGs labeled as non-lenses (red lines). The two test-sets are normalized to their corresponding maximum value in the distribution. We shifted the x-axis of the second test set distribution by 0.02 for clarity. The bottom panel shows the S_{CNN} distribution for the objects in the Parent sample with scores above 0.5. We crop the x-axis for visualization as 99% of the sample is below 0.5.

work decreases obtaining an accuracy of 89.6%, and a loss of 0.44, with $TPR = 76.1\%$ (65.7%) and $FPR = 0.31\%$ ($>0.01\%$) for $S_{CNN} = 0.5(0.9)$. Thus, whereas the accuracy in the second test set is still high and the network did not grade any LRG above 0.9, the loss and TPR are significantly worse than for the dataset with similar characteristics to the training set. We think that this decrease in the performance of the CNN is because it was trained to recognize lens simulations which, despite being created in a fully data-driven way, lack the diversity and uniqueness of some strong lens systems, e.g., multiple deflectors, distortions produced by substructures or external sources, etc. For example, most of the false negatives in this second test set are

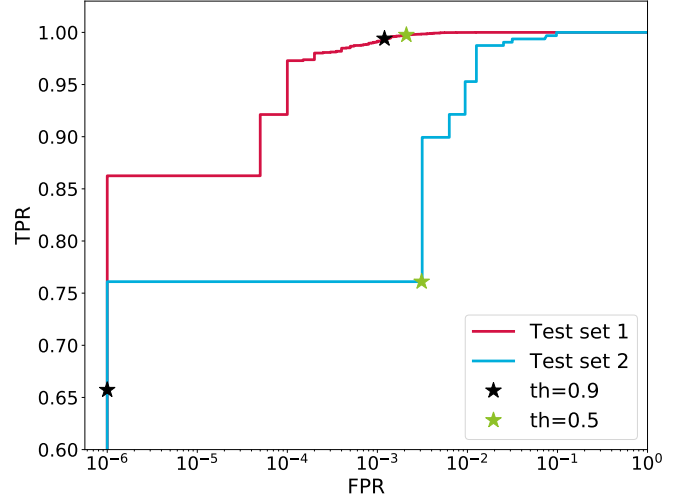


Fig. 6. Receiver Operating Characteristic (ROC) curve for the test set containing simulations (test set 1 in red) and confirmed/candidates lenses (test set 2 in blue); both datasets contain real LRGs as non-lens examples. The FPR is plotted on a logarithmic scale to aid visualization. The TPR and FPR for $S_{CNN} = 0.5$ (green) and $S_{CNN} = 0.9$ (black) are also shown for each set.

compact lens systems or have lensing features that are too faint to be properly recognized. Nevertheless, we found that our model is able to generalize and accomplish the goal of successfully classifying a high percentage of strong lens systems, although we are aware that in a realistic scenario we misclassify more objects compared to the simulations, as Fig. 5 and Fig. 6 show.

When we applied the CNN to our parent sample, we found that 98.6% of cutouts obtained $S_{CNN} \leq 0.1$, 133 322 obtained $S_{CNN} \geq 0.5$ (Fig. 5, bottom panel), and 76 582 cutouts obtained $S_{CNN} \geq 0.9$. The choice of $S_{CNN} \geq 0.9$ is driven both by the ROC shown in Fig. 6, and also the resulting number of candidates being reasonable for human inspection.

5. Visual inspection

The 76 582 cutouts scored above $S_{CNN} = 0.9$ by the CNN were visually inspected by 7 authors of this work (K.R., E.S., B.C., F.C., C.L., J.C., and G.V.).

5.1. Visualization tools and guidelines

We created two visualization tools³: one to quickly select lens candidates from many objects displayed simultaneously in a mosaic configuration, and one to visually inspect each individual object in more detail and classify them into specific categories.

The mosaic tool simultaneously displays 100 color cutouts, each of which the user can mark for selection. The user can choose a random seed for displaying the images in a random fashion on the grid, to avoid all users seeing each object at the same location in the grid. This has the objective of preventing any possible bias from the position of the object on the mosaic coupled with the different level of concentration when looking at many mosaics in a row. This turned out to be very efficient, as illustrated by the "heat-map" of user grid selections displayed

³ <https://github.com/esavary/Visualisation-tool>

in Fig. 7, that are fairly flat, with a small bias towards selecting more objects from the top, bottom and left row for this particular example. With the mosaic tool, we classify into only two categories, i.e. objects that we select as displaying potential strong-lensing features, and the rest that we discard from any following step of the visual inspection.

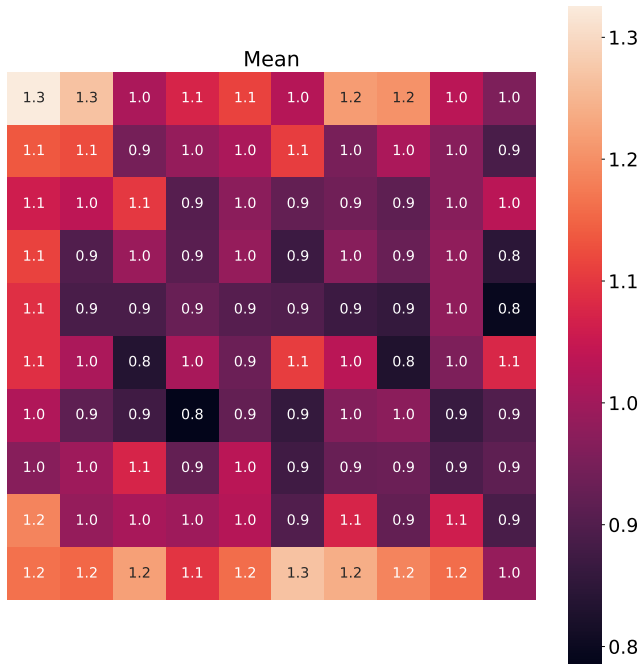


Fig. 7. Heat-map for the normalized mean number of times that each cell in the Mosaic was clicked on, among all seven visual inspectors during Phase 1 of the visual inspection. The values in each cell were obtained by calculating the mean of the total number of clicks per-cell among the 7 users, we normalized these values by a factor of 24.78 that represents the mean number of clicks in a cell for this specific classification.

The second visualization tool allows us to inspect one by one all lens candidates selected with the mosaic tool. In doing so, we display the *gri* color stamps allowing the user to change the display scale and color-map. With this tool, we classify each object into one of four categories: 1- Sure Lens, 2- Maybe Lens, 3- Single Arc, and 4- Non Lens. In addition, we define 5 sub-categories for objects classified as Non Lens: 1- ring, 2- spiral, 3- elliptical, 4- disk and 5- merger.

In order to achieve a more consistent classification among users, we all agreed to follow the same guidelines for the four main categories. “Sure Lens” is selected when the cutout shows a clear strong lensing configuration without the help of a higher resolution image. This means that several clear multiple images can be identified or that there are signs of a counter-image. “Maybe Lens” is chosen if the object shows a promising lensing-like configuration but a clear identification of multiple images is not possible visually. This category also includes cases where several objects or a single arc-like object lie on one side of the central galaxy but no clear counter-image can be distinguished on the other side. In this case, high-resolution imaging or spectroscopy will be required to decide whether it is a false positive or a genuine lens. When there is a single image object or a single arc far away from the central galaxy with signs of tangential distortion, the cutout goes to the category “Single Arc”. Finally, everything else that does not fit these categories is classified as “Non Lens”.

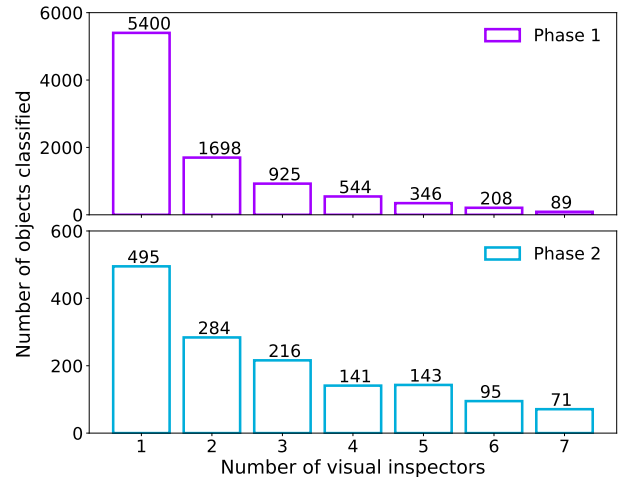


Fig. 8. Number of objects classified by a certain number of visual inspectors in phase 1 (top panel; to select lenses and ring galaxies), and phase 2 (bottom panel; to select only the ring galaxies from phase 1). Objects selected by 7 visual inspectors represent a 100% in agreement among users, while the sum along the different bins give us the union, i.e. the number of objects selected by at least one user. The exact number in each bin is shown at the top of each bar.

5.2. Visual selection procedure

We used both tools in 4 different phases to ensure that we have a clean sample of not only potential lens candidates but also a sub-sample of contaminants, such as ring galaxies, which are a source of confusion for CNNs and a matter of debate among visual inspectors. The steps carried out are described below:

1. Lens and ring galaxy selection. Using the mosaic tool we selected from the 76 582 cutouts all objects that presented signs of lensing features or looked like ring galaxies in one category and we discarded the rest. An average of 2 478 cutouts were selected per visual inspector, the normalized mean distribution of clicks per cell are shown in Fig. 7. A total of 9 210 objects was selected by at least one user, while 89 of them were selected by all the users unanimously (see Figure 8, top panel).
2. Ring galaxies selection. We used the mosaic tool to select only ring galaxies from the 9 210 objects. A mean of 230 cutouts were selected per visual inspector, but only 71 were classified by all seven, while a total of 1 445 were selected by at least one (see Figure 8, bottom panel).
3. Lens systems classification. We visually inspected all the 9 210 objects selected in phase 1 using the one by one visualization tool, looking specifically for lens systems. Here, we showed again the classified ring galaxies from the previous step as a consistency check (users should re-classify them as rings, or at least not classify them as lenses). We classified each object into: Sure lens, Maybe lens, Single Arc and Non-lens. Optionally, if a Non-lens was clearly identified by the user as a spiral, merger or ring galaxy, the object was sub-classified into the corresponding category. From this visual inspection we obtained a total of 275 Sure lens, 2 666 maybe lens, 2 602 Single Arc and 9 125 non lenses, classified at least by one visual inspector. In a unanimous agreement among the 7 visual inspectors we counted only 6 sure lenses, 1 maybe lens, 1 Single Arc and 4 716 non lenses. On the

other hand, K.R., E.S., B.C., F.C., and J.C. sub-classified 359 ring galaxies, 22 mergers and 49 spirals with an agreement of 50% among the visual inspectors. In Table 1 we summarize the individual classification by category and subcategory of each user and in Fig. 9 we show examples of objects classified in each category and sub-category, except "Sure Lens".

4. Final group visual inspection. We implemented this last step due to the lack of agreement among users for the main classification categories (Sure lens and Maybe lens), as the Fig. 10 shows. In this figure we can see for example that User 2 classified 1396 objects in both categories, being the one that classified more cutouts as candidates, but the overlap with other users is not higher than 321 (with User 4 and User 5). This step was performed by K.R., E.S., B.C., F.C., J.C. and G.V. all together. Using the one by one visualization tool we revised 2 690 objects selected as Sure or/and Maybe lens by at least one visual inspector. The aim was to obtain a final selection of potential candidates that can be suitable for follow-up high-resolution imaging and spectroscopic confirmation and to avoid spending telescope time on false positives. We classified them into the two main categories as follows: 81 sure lenses and 296 maybe lenses. This represents 0.5% of the sample with $S_{\text{CNN}} \geq 0.9$ and 0.002% of the initial LRG selection sample.

An extra visual inspection step was performed for the cutouts classified by the CNN with scores between $0.8 < S_{\text{CNN}} < 0.9$ (hereafter referred to as bin80) by K.R., E.S. and B.C. with the purpose of quantifying how many objects we could have missed by selecting only those with $S_{\text{CNN}} > 0.9$. Similarly to the previous analysis we first used the mosaic visualization tool to inspect the 17 779 cutouts, selecting 190 potential lens candidates, which were then inspected one by one. A set of 24 objects were classified as lenses by at least one visual inspector and 115 as maybe lenses. Finally, K.R., E.S., B.C., F.C., J.C. and G.V. conducted a group visual inspection of the 190 firstly selected candidates to compile a final sample with 9 sure lenses and 19 maybe lenses that were then added to our candidate list. In total, only 0.2% of the data visually inspected in the bin80 was considered as a lens candidate, while for all the objects with $S_{\text{CNN}} \geq 0.9$ we selected 0.5% of the cutouts in the categories Sure or Maybe Lens. Furthermore, taking into account that the amount of cutouts classified in the bin80 is about 4 times smaller than those with $S_{\text{CNN}} \geq 0.9$, we conclude that the number of expected candidates with $0.8 < S_{\text{CNN}} < 0.9$ was very low, showing that we reached a point of diminishing returns which would make the additional human visual inspection of images classified with $S_{\text{CNN}} < 0.8$ ineffective.

5.3. Final Catalogs

As a final product we present two main catalogs: one containing lens candidates and one containing ring galaxy candidates. We assigned a Visual Inspection Score (VIS) to each candidate, computed using the percentage of visual inspectors that classified it into a certain category. In the case of lens candidates, we used the percentage of users that classified a system as either a "Sure lens" and as "Maybe lens". We summed these percentages to obtain a "strong lensing percentage". Then, we considered this percentage as the final visual inspection score for lens systems (VIS_L). In the case of ring galaxies we had candidates from step 2 (using the mosaic tool) and/or step 3 (the one by one method). We average the percentage of users who classified each object as a ring in each step, and present this as the final visual inspection

score for Rings (VIS_R). If the candidate was selected only by one of the tools, the final score obtained is the one corresponding to that classification (i.e. not the average).

The ultimate catalog of lens systems can be split in two categories, "Sure lens" (SL) with 90 systems (Fig. 15 to 20) with prominent lensing features and counterpart images, and "Maybe Lens" (ML) with 315 systems that show promising lensing features but for which more evidence, such as higher resolution imaging and spectra, is needed (see Fig. 9 for examples). From the figures we can conclude that a large portion of our candidates are group- and cluster-scale lenses. This is mostly because to be able to identify them in ground-based data, most of them should have an Einstein radius above $1''$.

The CNN and visual inspection scores of both the SL and ML candidates is shown in Figure 11. Here, we clearly see that most "Sure Lens" systems are clustered towards the upper right corner, indicating that in general they obtained a high score from both methods, while very few of them had either CNN scores below 0.95 or visual inspection scores below 0.5. On the other hand, a large majority of "Maybe Lens" objects did not receive a high visual inspection score, including two that originally were rejected by visual inspectors, but upgraded after the group visual inspection. Several of the "Maybe Lens" objects still got very high scores from the CNN, indicating that the visual inspection step is needed to refine the final catalog.

In order to identify lens candidates that were not previously published, we cross-matched our final catalog with available astronomical databases such as Vizier (Ochsenbein et al. 2000), Simbad (Wenger et al. 2000), the Master Lens database, and other lens-finding papers including Wong et al. (2018); Jacobs et al. (2019b,a); Petrillo et al. (2019b); Cañameras et al. (2020); Jaelani et al. (2020); Huang et al. (2020, 2021). As a result, we found that our catalog contains 219 previously identified candidates (74 SL, 145 ML), including at least 5 spectroscopically-confirmed systems, and 186 new candidates (16 SL, 170 ML). The detailed information for these systems can be found in Table 2, available at the CDS.

Our second catalog is composed of ring galaxy candidates classified during two different steps of the visual inspection process. We identified 1 445 ring galaxy candidates during the second visual inspection step using the mosaic tool, while 985 galaxies were classified into this category by at least one visual inspector using the one by one tool in the third step. A cross-match between these two selections gave an intersection of 854 galaxies for a total of 1 576 ring galaxies selected by at least one user using either of the two methods. The final catalog was built by selecting the objects picked by at least 50% of the visual inspectors using the mosaic or the one by one tool, resulting in 539 ring galaxy candidates. In Fig. 9 we present the 6 top-graded candidates and in Table 3, available at the CDS, we detail the information for the full sample.

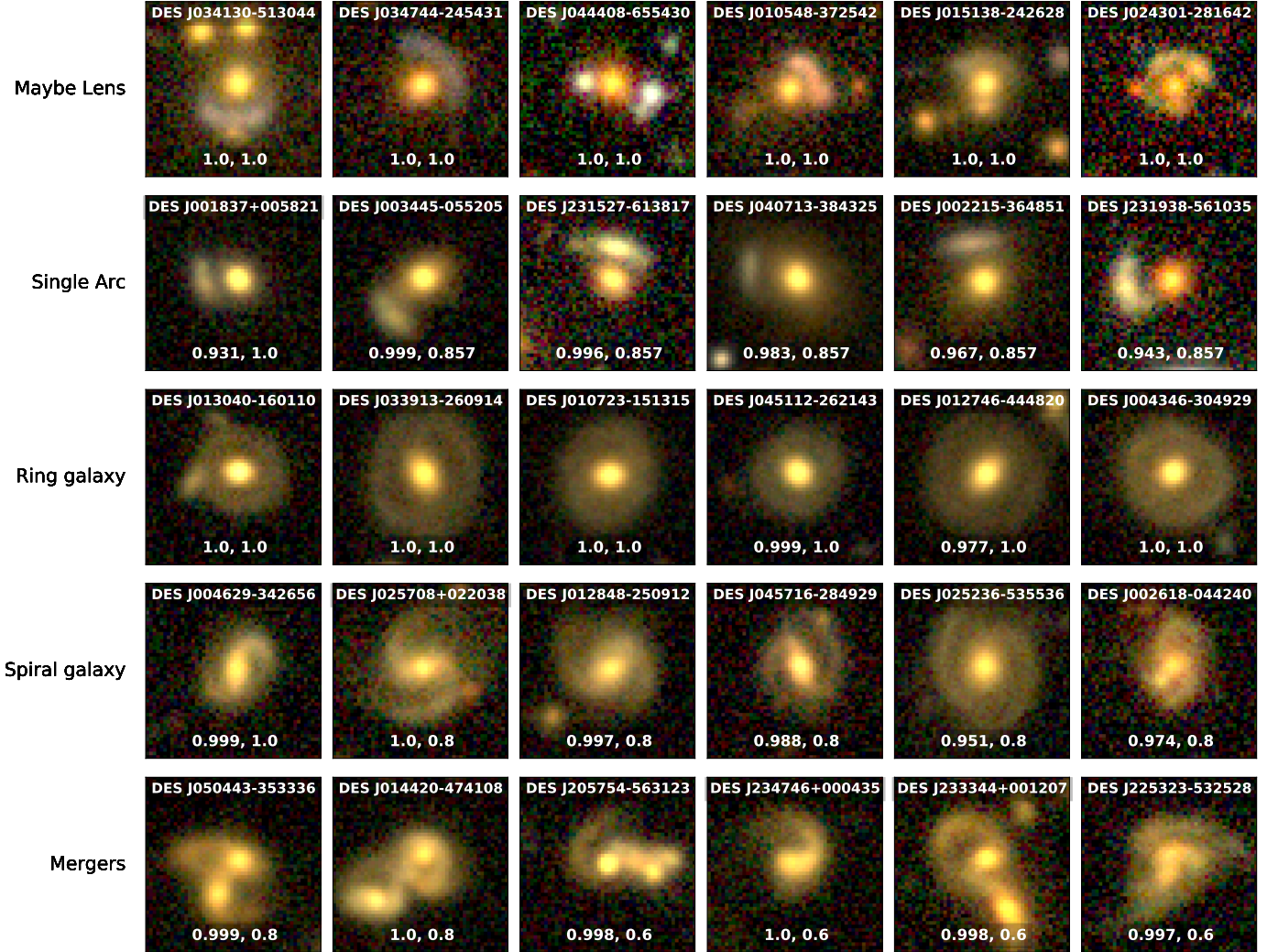
Finally a catalog using the classification "Single Arc" will be created after a more detailed analysis of these objects, but this is beyond the scope of the current work. We expect that this catalog could serve as a probe for works like Birrer (2021).

5.4. Lens-source decomposition

In the ground-based observations of most strong lensing systems, the light from the source and lens galaxies are blended. In order to better visualize our 90 lens candidates in the Sure Lens catalog, we designed a prototypical automated procedure for deblending the light from lens and source galaxies based primarily on their colors.

Table 1. Classification and sub-classification details per visual inspector during Phase 3

Classification	User 1	User 2	User 3	User 4	User 5	User 6	User 7
Sure Lens	116	41	79	146	120	19	90
Maybe Lens	612	1355	492	691	849	203	141
Single Arc	654	1421	540	300	812	473	26
Non Lens	7828	6393	8099	8073	7429	8515	8953
Sub-classification							
Spiral	33	100	651	59	-	35	-
Ring	713	111	393	364	-	563	-
Merger	112	9	246	70	-	15	-

**Fig. 9.** Example of objects classified as "Maybe Lens" or "Single Arc", and sub-classified as "Ring galaxy", "Spiral", or "Merger", "Sure Lens" category is shown in Fig. 15. At the top of each image is the name of each system, while at the bottom we show the CNN score (S_{CNN}) and visual inspection score (VIS) obtained in the corresponding category.

Due to the complexity of the light profile of the lensed sources we choose to represent them in a non-parametric way using undecimated isotropic wavelets (starlets, Starck et al. 2007), as implemented in the Multi-band morpho-Spectral Component Analysis Deblending Tool **MuSCADET** (Joseph et al. 2016). Starlets are a family of functions that allow free-form modelling of images at various spatial scales and present advantages for modelling smooth galaxy profiles as discussed in the **MuSCADET** paper. Estimating both color and morphology of sources requires a large number of parameters, larger than the number of pixels

in the starlet-decomposed image, making it a degenerate problem. To overcome this we use a combination of the **Oscarlet** (Melchior et al. 2018) and **MuSCADET** algorithms. In both methods, multi-band images are modeled as sums of factorised components, where each object i in an image has a 2-dimensional surface brightness S_i with as many pixels as there are in the image bands, and a spectrum A_i with as many entries as there are bands (see details in Melchior et al. 2018; Joseph et al. 2016), such that:

Table 2. Excerpt of the Maybe lens catalog available online.

Candidate	RA	Dec	S_{CNN}	VIS_L^a	References
DES J034130–513044	55.378331	-51.512411	1.00	1.00	[7] [10]
DES J034744–245431	56.935562	-24.908741	1.00	1.00	[9] [10]
DES J044408–655430	71.034707	-65.908598	1.00	1.00	This work
DES J010548–372542	16.450174	-37.428457	1.00	1.00	[10]
DES J015138–242628	27.909990	-24.441314	1.00	1.00	[18]
DES J024301–281642	40.754315	-28.278515	1.00	1.00	This work
DES J025052–552411	42.717809	-55.403251	1.00	1.00	[10]
DES J014358–470037	25.995764	-47.010469	1.00	1.00	This work
DES J001718+015818	4.325557	1.971828	1.00	1.00	[10]
DES J040349–352601	60.955780	-35.433763	0.99	1.00	This work
DES J225146–441220	342.943254	-44.205688	0.99	1.00	[7]
DES J002056–594016	5.236669	-59.671225	0.99	1.00	This work
DES J011758–052717	19.494766	-5.454924	0.98	1.00	[10] [13]
DES J015904–345009	29.767747	-34.835994	0.98	1.00	This work
DES J015009–030438	27.537943	-3.077297	0.98	1.00	[9] [10]

Notes. ^(a) Visual Inspection Score for Strong lens systems

References. [1] Cabanac et al. (2007), [2] Limousin et al. (2009), [3] More et al. (2012), [4] Maturi et al. (2014), [5] More et al. (2016), [6] Paraficz et al. (2016), [7] Diehl et al. (2017), [8] Wong et al. (2018), [9] Jacobs et al. (2019b), [10] Jacobs et al. (2019a), [11] Petrillo et al. (2019b), [12] Cañameras et al. (2020), [13] Huang et al. (2020), [14] Jaelani et al. (2020), [15] Lemon et al. (2020), [16] Li et al. (2020), [17] Nord et al. (2020), [18] Huang et al. (2021)

Table 3. Excerpt of the ring galaxy candidates catalog available online.

Candidate	RA	Dec	S_{CNN}	VIS_R^a
DES J013040–160110	22.666806	-16.019599	1.00	1.00
DES J012733–151618	21.888203	-15.271692	1.00	1.00
DES J010723–151315	16.847733	-15.221047	1.00	1.00
DES J004346–304929	10.942121	-30.824795	1.00	1.00
DES J033913–260914	54.805084	-26.154158	1.00	1.00
DES J045112–262143	72.804088	-26.362060	1.00	1.00
DES J012542–231630	21.427496	-23.275137	1.00	1.00
DES J003809–224742	9.537798	-22.795153	1.00	1.00
DES J012843–350926	22.183211	-35.157252	0.99	1.00
DES J041502–404547	63.762073	-40.763330	0.99	1.00
DES J010902–450634	17.258607	-45.109657	0.98	1.00
DES J012746–444820	21.942424	-44.805651	0.98	1.00
DES J004837–330630	12.156848	-33.108576	0.98	1.00
DES J024746–243851	41.941910	-24.647757	0.97	1.00
DES J021101–315721	32.757447	-31.956016	0.97	1.00

Notes. ^(a) Visual Inspection Score for Ring galaxy candidates.

$$Y = \sum_{i < o} A_i S_i + N, \quad (2)$$

where Y is a multi-band cube of images, o is the number of objects in the scene and N is the noise map.

The strategy implemented in MuSCaDeT only allows for crude estimates of source colors, based on principal component analysis of pixel fluxes. Instead, the *scarlet* software is able to estimate the colors of each source in the field provided that the morphology is constrained to be a monotonic profile. Monotonicity of galaxy profiles from the center out does not suit the description of complex lensed sources, hence the need for MuSCaDeT to model strongly lensed galaxies in a non-parametric way coupled with sparse regularization. In short, MuSCaDeT is used to model the 2-dimensional profile of galaxy images, including a complex lensed source, while *scarlet* recovers the colour of the objects.

Scarlet requires detection of the brightest pixel of each source to model, which is a challenging and ill-defined problem in the case of strongly lensed galaxies, where lensed features are often multi-modal and strongly blended with the deflector's light. In order to circumvent this issue and make sure we capture (lensed) sources with a bluer spectrum than the central LRGs, we allow *scarlet* to model one source with Starlets, initialised with a "blue" spectrum. This allows *scarlet* to capture blue features with complex morphologies that might not have been detected due to blending, while limiting degeneracies with other sources. The blue, normalised spectra used for initialisation are empirically set to [0.4, 0.4, 0.2], where each of the three values reflect the relative contributions to g , r and i bands respectively. Other non-lensed sources are detected using the *Qsep* package (Bertin & Arnouts 1996). We run *sep* on a filtered version of the images. The filtering is done by computing the starlet decomposition of an image and setting to zero the coefficients that

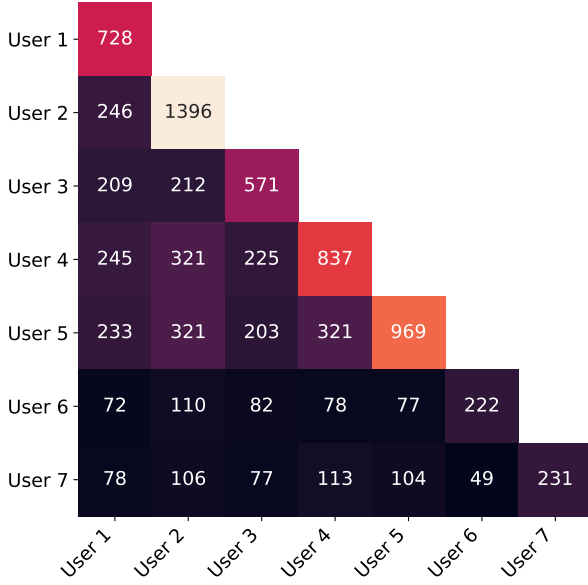


Fig. 10. Correlation among different classifiers for the categories "Sure Lens" and "Maybe Lens" after Phase 3. The values in the diagonal represent the total number of objects classified into both categories for each user, while in the adjacent rows and columns the number of objects that both users classified into the same category.

contribute to low frequencies before reconstructing the image. This amounts to a high pass filtering that favours peak detection. The position of the brightest pixels of objects detected by *sep* are fed as entries to *scarlet*. For each object detected with *sep*, *scarlet* estimates a spectrum (flux in each band). From these spectra obtained with *scarlet*, we select the bluest and reddest spectra by finding those that maximize the scalar product between the normalized spectra $[0.667, 0.333, 0]$ (for blue) and $[0, 0.333, 0.667]$ (for red). This ensures that two components with different colors are extracted, with the expectation that the red component features the morphology of the LRG and its neighbours, while the blue component extracts the morphology of the lensed star-forming background galaxies.

The summary of the procedure for deblending strong gravitational lens candidates is as follows:

- Detection of sources in the image using the source extraction package, *sep* (Bertin & Arnouts 1996) on a *starlet*-filtered version of the image where only the first two levels of the *starlet* decomposition are used. We set detection threshold to 1 noise standard deviation of the noise upon running *sep*. This may seem aggressive and potentially leading to a shredding of the objects upon deblending, but the smoothness of the images imposed by the PSF prevents such and effect. Furthermore, shredding is not an issue as we only intend to capture the spectra of the reddest and bluest objects.
- Initialization of *scarlet* sources: one extended source per detected object plus one *starlet* component with blue spectra. *Scarlet* uses a target PSF of Gaussian profile with a standard deviation of 0.5 pixel. This is the target resolution to which all bands are uniformly deconvolved to.
- Run *Scarlet* for up to 200 iterations and extract spectra for each source in the field of view by simply measuring the flux in each source.

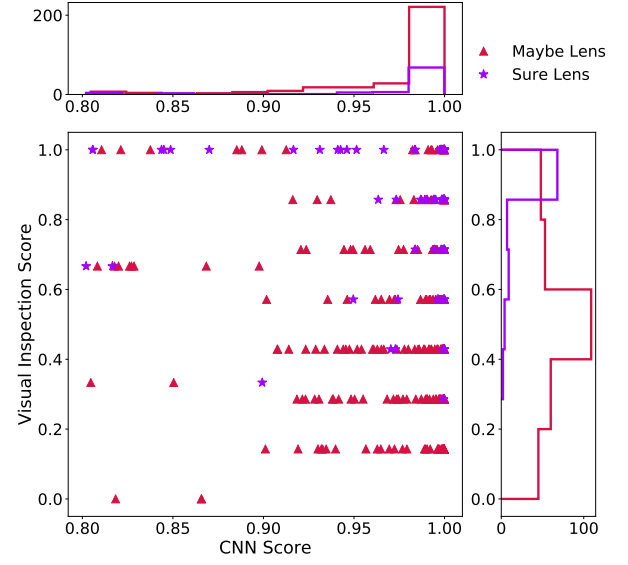


Fig. 11. CNN score (S_{CNN}) against the visual inspection score (VIS_L), with their respective distributions for the final catalog of lens candidates containing 405 systems. The distribution of the category "Sure Lens" is in purple, and "Maybe Lens" is in red, while in the scatter plot "Sure Lens" systems are represented by stars and "Maybe Lens" systems by triangles.

- Identify the bluest and reddest sources through scalar product with predefined red and blue spectra.
- Run *MuSCADET* with the red and blue spectra for 200 iterations. The threshold for *starlet* reconstruction is set to 5 sigma of the standard deviation of the noise. This means that the model reconstructs features that are 5-sigma above noise levels.
- Extract red and blue components by computing the difference between the multi-band images and the model for each *MuSCADET* component.

The details of the processing can be found in the notebook that was used to generate these images: [Lens-Deblend](#).

The results for our best lens candidates are shown in Figs. 15–20 and display for each system the red residuals, i.e. the data from which the blue model has been subtracted, and the blue residuals, i.e. the data from which the red model has been subtracted. In the following we refer to these as R_r and R_b respectively, defined as:

$$R_{b,j \in \{g,r,i\}} = Y_j - A_{r,j} S_{r,j}, \quad (3)$$

$$R_{r,j \in \{g,r,i\}} = Y_j - A_{b,j} S_{b,j}, \quad (4)$$

Where $A_{r,j} S_{r,j}$ and $A_{b,j} S_{b,j}$ are the models for the red and blue components in each band j .

The results in Figs. 15–20 show that the lens and source light can be deblended efficiently without fitting any analytical profile. The effectiveness of the method to deblend the profiles comes mostly from the spectral decomposition of the objects and on their representation on an array of pixels to which we apply sparse regularization with wavelets (*starlets*). This procedure is well suited to automated use in a pipeline but assumes that lensed sources are significantly bluer than the lens light. This is the case for most of our lenses as by construction our lens

finding method is based on a preselection of objects that favors such a configuration. Still, we do have objects where the lens-source color contrast deviates significantly from our assumption. In this case the deblending works less efficiently and we see leakage of flux between the lens and source. Another case of leakage, leading to sub-optimal deblending can be observed in systems where the image contains sources with colors different from those of the lenses or sources. In this case, since the whole image is modeled as 2 fields of light, the spectra of the color components tends to offset towards an average spectrum that better matches all the colors in the patch. This can be observed in systems DES J013522-423223, DES J024911+004848 and DES J010826-262019, where the blue components contain light from the lens galaxies and contain objects with colors different from that of the main deflector. These shortcomings are motivation enough for further refinement of our deblending, in particular with focus on using *scarlet* to better model individual, non-lensed sources, which is beyond the scope of this paper. Finally it is important to emphasize that our light deblending confirms our visual grading and does not discard any of our best candidates.

6. Model

We developed an automated modeling pipeline in order to further explore the highest rated lens candidates obtained from the visual inspection. Our candidate sample is very heterogeneous, containing galaxy, group and cluster scale systems. Thus, in order to perform this automatic modeling we split the sample and selected through visual inspection only the images in which there appeared to be a single lens galaxy as a deflector. The 52 images selected for modeling are labeled with an "M" in the mosaics of Figs. 15 - 20.

This pipeline allows us to efficiently model large samples of lens candidates acquired in current and future lens finding efforts, and to explore the model parameter distributions in search of meaningful trends.

6.1. Automated Modeling Pipeline

We modeled the images using single elliptical Sérsic profiles for the light distributions of both the deflector and source. For the mass distribution of the deflector, we use a Singular Isothermal Ellipsoid profile (SIE) along with an additional external shear component (γ_{ext}). The simplicity of these profiles allows us to model many lens candidates efficiently, while still fitting most images well enough for us to observe large-scale trends in the properties of the sample. The pipeline supports multi-band fitting, so we fit the DES lens candidates using images in the g , r , and i bands. We used a separate elliptical Sérsic profile for each of the three photometric bands when fitting the deflector and source light components, although we fixed the center positions between bands and we added priors to bound the semi-minor and semi-major axes. The deflector mass profile is shared across all bands.

The modeling pipeline was entirely written in Python and makes use of the *Lenstronomy* lens modeling package (Birrer et al. 2015; Birrer & Amara 2018). For parameter optimization we used the Particle Swarm Optimization (PSO) (Kennedy & Eberhart 1995), and to estimate the variances in the sample we used a Markov Chain Monte Carlo (MCMC) sampler. For each image, the pipeline first performed a chain of pre-sampling PSOs before running the sampling with the MCMC. The MCMC is performed using an affine-invariant MCMC ensemble sampler

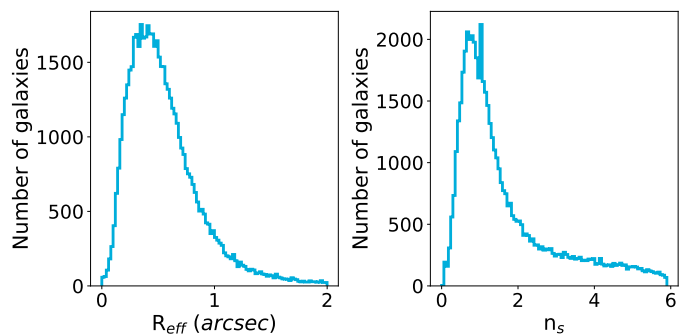


Fig. 12. Prior distributions for the effective radius (R_{eff}) and the Sérsic index (n_s) used for constraining source light parameter values.

(Goodman & Weare 2010; Foreman-Mackey et al. 2013), which is implemented using the *emcee*⁴ Python package.

In order to obtain realistic results on the parameters, we introduced priors that punitively discourage extremes in some of the model parameter values. While we can not assume anything about the position angles of the Sérsic or SIE profiles, we use Gaussian priors on the ratio between the semi-minor and semi-major axes, q . The Gaussian prior was centered on a value of $\bar{q} = 0.8$, in accordance with the distributions of 138 269 galaxies from the Galaxy And Mass Assembly (GAMA) database that were modeled in Kelvin et al. (2012). This prior is applied first to the r -band, and then to allow only small variations between bands in the light components of the model, the Gaussian prior of the other two bands is centered on the result obtained in the r -band. We also used a similar Gaussian prior method to constrain the deflector mass eccentricity and position angle to values close to those of the deflector light. Lastly, we also applied a prior distribution for the effective (half-light) radius, R_{eff} , and Sérsic index n_s parameters of the source light. The source priors we used came from the Sérsic parameter distributions of 56 062 galaxies from the COSMOS survey. This data was used as a training set in the development of the *GalSim*⁵ software. (Rowe et al. 2015). We show these distributions in Fig. 12.

When modeling the lens candidates, it is common for image cutouts to contain neighboring objects in the field of view that are unrelated to the lens system. Light contamination from these "satellites" can be mistaken as originating from the lensed source if not masked properly. In the literature, this problem is handled differently by various authors. Shajib et al. (2020) excluded systems contaminating satellites in their sample, and modeled only isolated lenses from the SLACS survey (Auger et al. 2009). Nightingale et al. (2018) did not pre-select isolated lenses, but instead masked all pixels outside of a circular region with a fixed radius of $3.9''$.

In our case, we designed the pipeline to be flexible in handling a large variety of lens system configurations and sizes. The steps of our masking procedure are illustrated in Fig. 13, and begins with applying filters in order to identify the brightest regions in the image as well as their centroid locations. We first applied a Laplacian of Gaussian (LoG) filter to detect areas with rapid changes in flux. Next we took all remaining pixels with flux less than a threshold of six times the rms background, and set them to zero. This results in a final filtered image with only the areas of the image containing the most light. We find the centroid locations of these areas by finding the local maxima, or peaks, in

⁴ <https://github.com/dfm/emcee>

⁵ <https://github.com/GalSim-developers/GalSim>

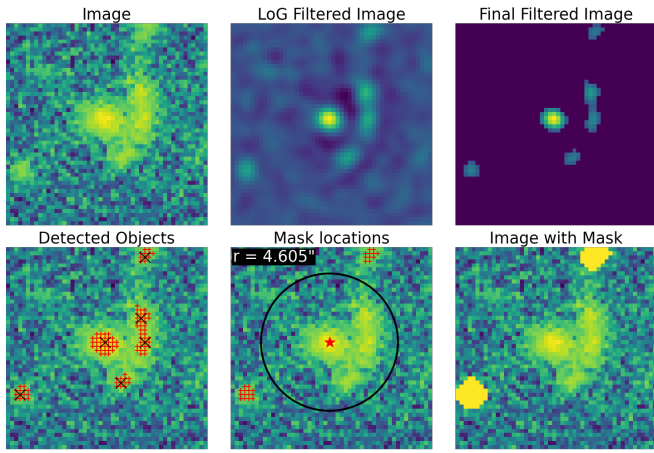


Fig. 13. Illustration of the automated masking procedure using an example DES image. In the upper row we show (from left to right) the original image, the image after applying an LoG filter, and the result of setting pixels with flux below the threshold to zero. In the bottom row we show in the leftmost frame the original image annotated with the remaining pixels from the filtering step (red '+' marks), along with the detected peaks (black 'x' marks). In the middle we show the estimated size of the lens system with a black circle, as well as the detected bright pixels that are considered contaminant light. These pixels are then used for the mask, and we show the areas covered in the rightmost panel by setting the corresponding pixels to a large constant value.

the final filtered image. For our masking algorithm we made use of both these peak locations as well as the pixel values – these are labeled with black and red markings in the bottom left panel of Fig. 13, respectively. The peak locations are used first for determining the center of the lens system, i.e. the position of the deflector galaxy, and assume that this is the peak detected object nearest to the image center. Because the deflector is assumed to be an LRG, we use the reddest available band (*i*-band) for this step. Next, we take the detected peaks in the bluest band (i.e. the *g*-band) to estimate the lens system size. This is because the source galaxies in lens systems are usually younger, more active galaxies, meaning that the lensed source light will be more prominent in the bluest image band. We assume that the second closest detected peak to the center is the first of the lensed images of the source. We also assume that the furthest lensed image from the deflector is not more than $1.5''$ further out than the nearest one. Therefore, our estimated lens system radius is the distance from the deflector to the closest lensed source object plus $1.5''$. We show this estimation as a black circle enclosing the lens system in the bottom middle panel of Fig. 13.

Using the estimated size of the lens system from the *g*-band image and the location of the deflector obtained from the *i*-band, we create a circular mask for each band that is centered on the deflector location and only covers detected bright pixels outside of the circular region with our determined size. The mask itself is a boolean array with the same shape as the original data, and has the value of zero at any pixel that is to be ignored in *Lenstronomy* computations and ones everywhere else. In the bottom right panel we illustrate the coverage of the mask by setting all of the "ignored" pixels of the original image to a large constant value.

On average our pipeline took 4.3 hours to model a *gri* DES system. This includes reading data, the masking process, and performing the modeling sequence to find the best parameters that describe the lens candidate.

6.2. Modeling Results

Using our automated pipeline, we modeled 52 of the systems in the "Sure Lens" catalog that appeared to have only a single galaxy as a deflector. We show in Figs. 21–33 a sequence of images to visualize the modeling results in the *r*-band, including the corresponding image, a reconstructed image, normalized residuals, convergence map, and the reconstructed source light. In Tab. ?? we present the best model parameters obtained for each system, and we show the obtained distributions for the Einstein Radii, the external shear, the effective radius and Sérsic index for the lens and source light, in the histograms in Fig. 14.

We obtained acceptable fits for 41 systems, which represent 79% of the sample, and we observed 11 failures in the fitting, which we define as fits with mean reduced χ^2 per pixel above $\chi^2 = 1.5$. In the lens mass components, we observed Einstein Radii, R_E , distributed between $\sim 1''$ and $\sim 3.5''$. For the external shear strengths we observed $\gamma_{\text{ext}} \lesssim 0.47$ for all lenses except for one in which the fit failed. The distribution of the values have a peak at 0.14, these are typical shear values for strong lens systems (Keeton et al. 1997). For the effective radii, R_{eff} and Sérsic indices, n_s , of the lens light profiles in the *r*-band, we observed peaks at $R_{\text{eff}} \sim 2''$ and $n_s \sim 5$, respectively. Because the CNN searches selected lens systems from a catalogue of LRGs, we expect to obtain deflector light parameters that are typical for LRGs, and that is indeed what we recover. For the parameter distributions of the source light, we observe the effective radii and Sérsic indices peaking at $R_{\text{eff}} \sim 0.2''$ and $n_s \sim 1$, respectively. This is also the expected behavior for smaller, low-mass galaxies that are usually dominating the lensed galaxy source population.

When modeling these lenses, the primary source of failures lies in the masking procedure. For example, the estimated size for the lens system is either slightly too small or too large, resulting in parts of the lensed source light being masked, or neighboring contaminants not being masked and instead treated as lens features. This happened for four systems with failed models (DES J060653–585843, DES J015216–583842, DES J032216–523440, and DES J051047–263222) and for two considered as having acceptable fits (DES J034713–453506, DES J040822–532714). Since it is common for images to contain companion objects very close to the lens systems, there is a small margin for error in determining the lens system size. For the system, DES J012042–514353, the contaminant is actually residing among the lensed images of the source, a situation which can not simply be handled with a more precise measurement of the lens size. A method would be needed for better untangling the contaminant light from the lens features. Finally, there are two lens systems (DES J010553–053419, DES J041809–545735) in which the contaminant light distributions were spread out enough that the mask failed to adequately cover all this light. In general we need to improve our masking procedure to avoid these problems during an automatic fitting of the lens. In the meantime, for all of these systems for which the masking algorithm did not perform well, we recreated masks by hand and performed the modeling a second time. These results are shown in the rows directly below the original results for the specific system, and both sets of results are enclosed in a red dashed box in Figs. 21–33. Each time, we see a significant improvement in residuals after using the better mask. On the other hand, the rest of the models that are considered as failures (DES J010659–443201, DES J021159–595624, DES J024803–061606, DES J202855–523118) do not show an obvious reason for it, but are likely due to the compactness of

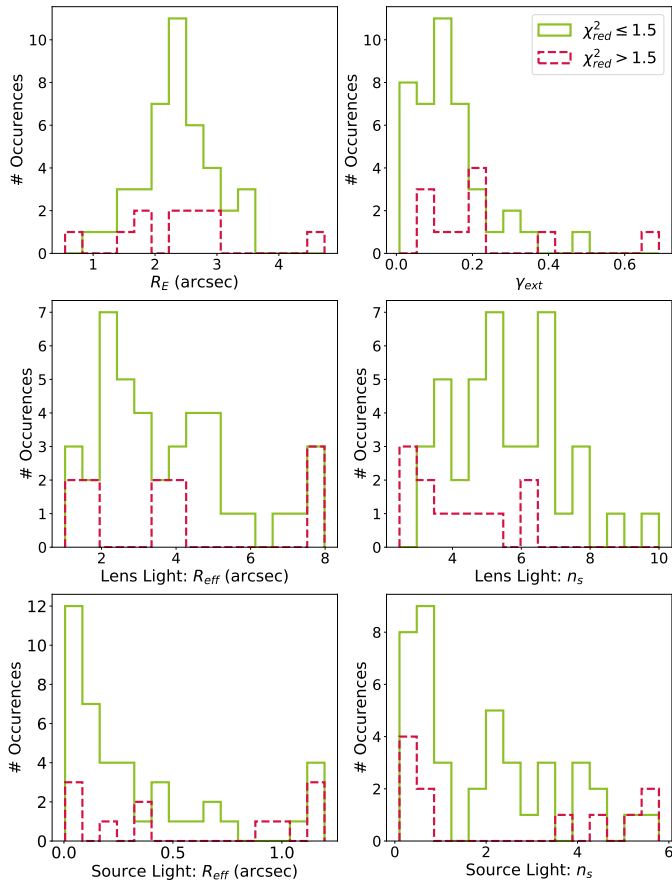


Fig. 14. Model best-fit parameter distributions for the lens mass, lens light, and source light profiles. Results for which *reduced* $\chi^2 \leq 1.5$ are shown in green and those with *reduced* $\chi^2 > 1.5$ are displayed in red. Top left: Einstein radii in arcseconds. Top right: External shear strength, γ_{ext} . Middle left: Sérsic half-light radii of the lens light. Middle right: Sérsic indices of the lens light. Bottom left: Sérsic half-light radii of the source light. Bottom right: Sérsic indices of the source light.

the system, faint lens features, or the complexity in the shape of the source. For these cases we need further investigation to find a general solution to improve their models in the automatic pipeline.

7. Conclusions

We used DES-DR1 to search for galaxy-scale strong lensing systems using a Convolutional Neural Network that carries out a binary classification of optical images in the g , r and i bands. In doing so, we targeted massive galaxies, i.e. LRGs, which were selected using a wide color-magnitude cut accounting for realistic color contamination by the putative background star-forming blue galaxies.

The design of our training set was data-driven in the sense that real DES images of LRGs were used to mock the light distribution of the lens plane. Real images of galaxies from the COSMOS HST were used to mock lensed sources. This helps ensure diversity in colors and morphologies for the sources and lenses, but also preserves the sky background characteristics, galaxies and/or stars acting as companions, and any artifacts in the images.

We used these data-driven simulations as positive examples to train a Convolutional Neural Network, while we used a portion of the LRG sample as negative examples, despite some pre-

vious searches that included as negative examples other types of galaxies. We concluded after analyzing the results from the visual inspection that the lack of representation of other types of galaxies was not important as they were not the most relevant source of false positives. LRGs with bluish satellites near the line-of-sight are our most important contaminants. The CNN was trained and validated using a total of 200 000 images, half of them being mocked lens systems labeled as 1, and the other half being LRGs labeled as 0. Evaluating our model on a test set built from images with the same characteristics of the training set gave us an accuracy of 99.7%. On the other hand a small test set built with 300 lens candidates and the same proportion of LRGs, gave us a more realistic evaluation reaching an accuracy of 89.6%.

Applied to the 18 745 029 LRGs drawn from our color-magnitude selection, we obtained 76 582 images with CNN scores above or equal to 0.9, that several authors visually inspected. To do so, we created guidelines to separate them into different categories: “Sure Lens”, “Maybe Lens”, “Single Arc”, “Non Lens”, and subcategories: “Ring galaxy”, “Spiral galaxy”, “Merger” for objects falling in the “Non Lens” category. To perform the classification we used a mosaic visualization tool displaying 100 images at once, as well as a one-by-one visualization tool that displayed the color composite image and each band for one object at a time. We classified 0.5% of the 76 582 images as lens candidates, 81 falling in the “Sure Lens” category and 296 in the “Maybe Lens” category. Additionally we inspected the 17 779 cutouts with a CNN score in the range $0.8 < S_{\text{CNN}} < 0.9$, with only 0.2% of the images classified as lens candidates, i.e. 9 “Sure Lens” objects and 19 “Maybe Lens”. The visual inspection of these low-score lenses allowed us to conclude that the reward for inspecting images with scores below 0.9 was very poor compared with the amount of work. We therefore did not consider systems with even lower scores at all.

From our visual inspection we created two main catalogs: a lens candidates catalog and a ring galaxy candidates catalog, the latter being our main source of contaminants. The first catalog contains a total of 405 lens candidate systems: 90 “Sure Lens” and 315 “Maybe Lens”. Out of these, 186 were totally new systems and 219 were identified (but not necessarily confirmed) in previous searches. We debled the lens and source light for our 90 “Sure Lens” systems using the MuSCADET software, which does not involve any profile fitting, but uses the color contrast between the lens and source together with sparse regularization. This was successful in debleding most of the cases, where there were clear differences in the colors of the lens and source. The second catalog contains 539 ring galaxy candidates. We expect to use this ring catalog in the future to improve the training of machine learning algorithms in the recognition between lenses and ring galaxies. Still, 539 objects is not much to train CNNs and further work, e.g. with Generative Adversarial Networks, is likely to be needed.

Finally we selected from the “Sure Lens” category the 52 systems that apparently had one well-defined galaxy as a deflector to test an automated modeling pipeline. The relatively simple SIE + γ_{ext} and elliptical Sérsic profiles used in the modeling appear to be sufficient in describing these lens systems, and additional complexity is not necessary for the purposes of this automated modeling pipeline, at least with the image quality of DES-DR1. We successfully modeled 41 of these systems, while the other 11 failed mainly due to problems in the masking algorithm, especially in the estimation of the lens system size. To address these failures we plan improvements in a future version of the pipeline including the use of the decomposed images from

MuSCADeT to initialize the code and find the correct position and size of the system.

The outcomes of this lens finding work in DES-DR1 include a catalog with 405 meticulously-selected lens candidates that can serve as a start for spectroscopic confirmation. In our selection, we did our best to privilege quality of the candidates over their quantity. The methods and tools studied, developed and presented here have room for improvement, but serve as a preview of what can soon be achieved for the future generation of surveys, i.e. LSST, Roman telescope and the Euclid mission.

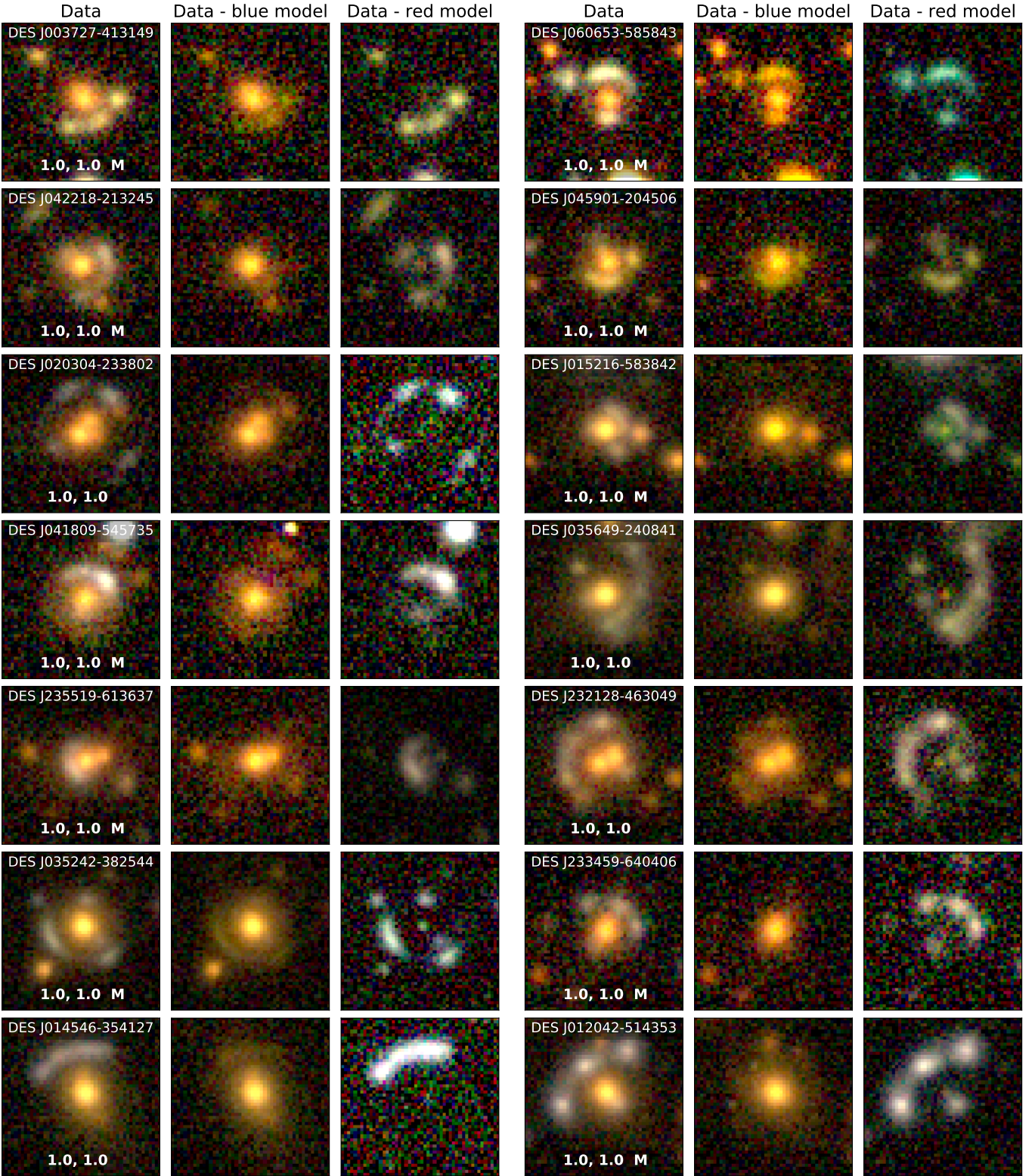


Fig. 15. Images for the 90 lens candidates in the category Sure lens and their corresponding decomposition performed with MuSCADeT. In the first and fourth columns we have the gri-composite image of the system, the name is on the top, while the CNN score and the visual inspection score (VIS_L) are displayed at the bottom of each image. Additionally we marked with a "M" those that we modeled in sect. 6. Columns 2 and 5 show the subtraction of the blue model from the respective data. Columns 3 and 6 show the subtraction of the red model from the respective data, isolating the lensing features.

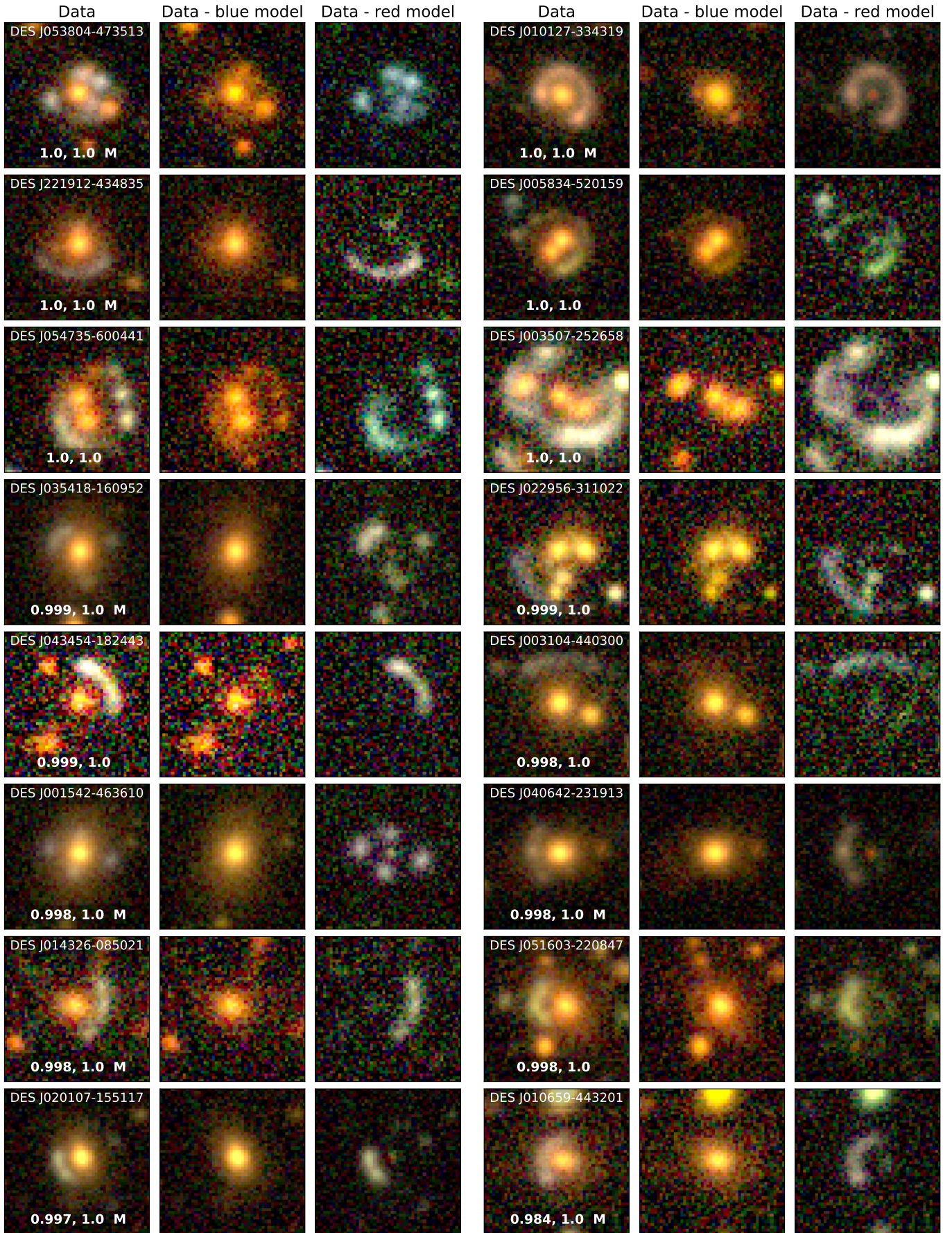


Fig. 16. Continued from Fig. 15

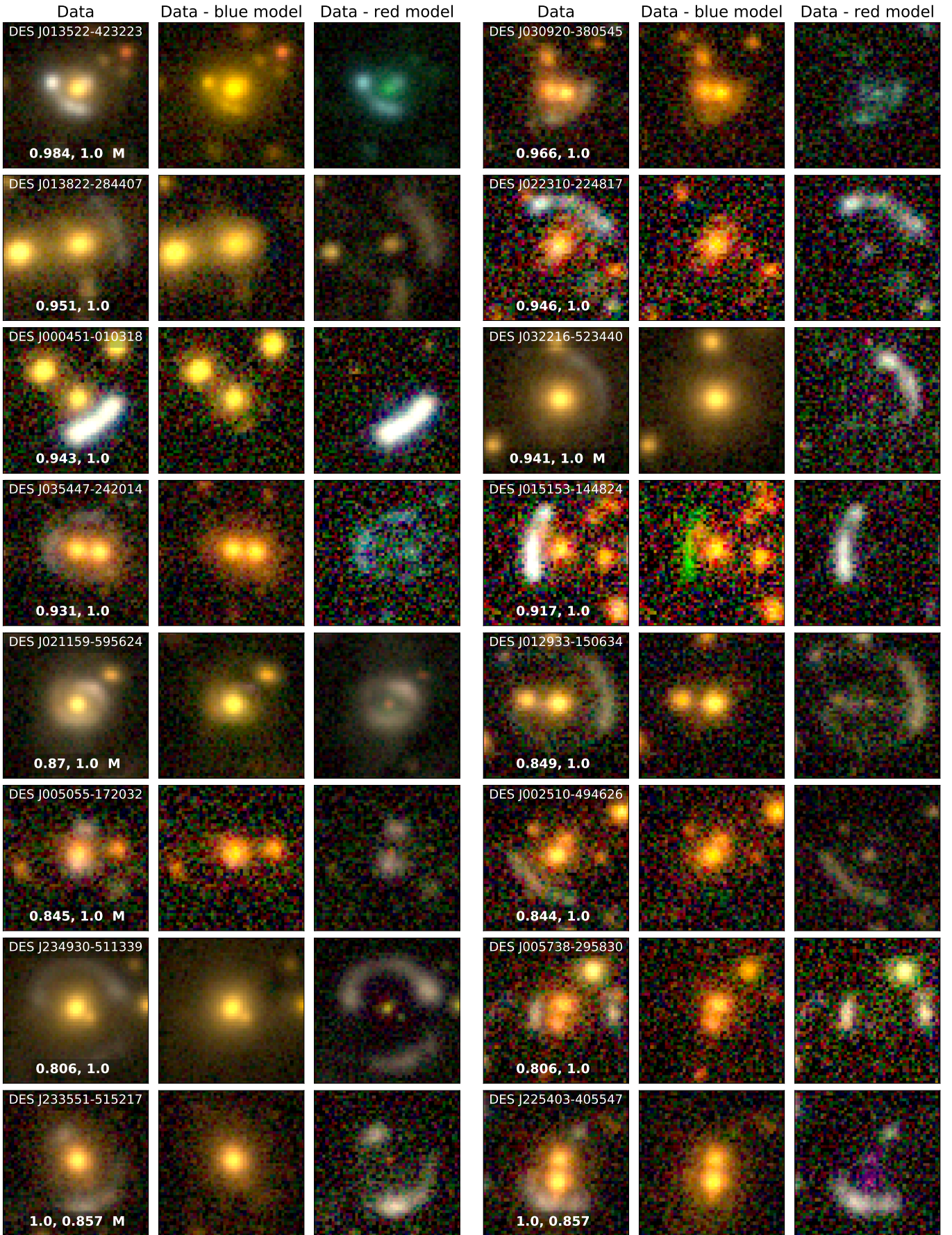


Fig. 17. Continued from Fig. 16

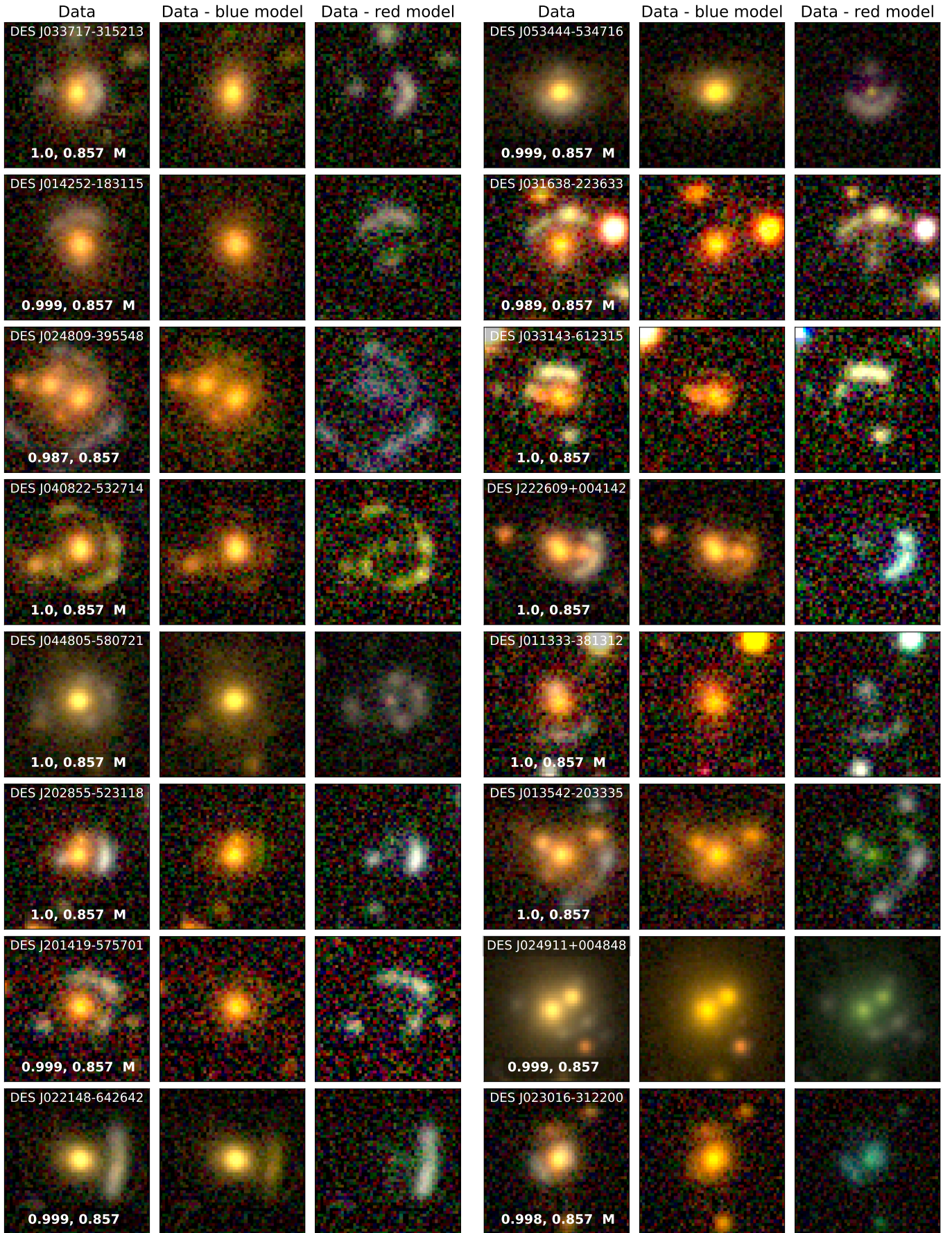


Fig. 18. Continued from Fig. 17

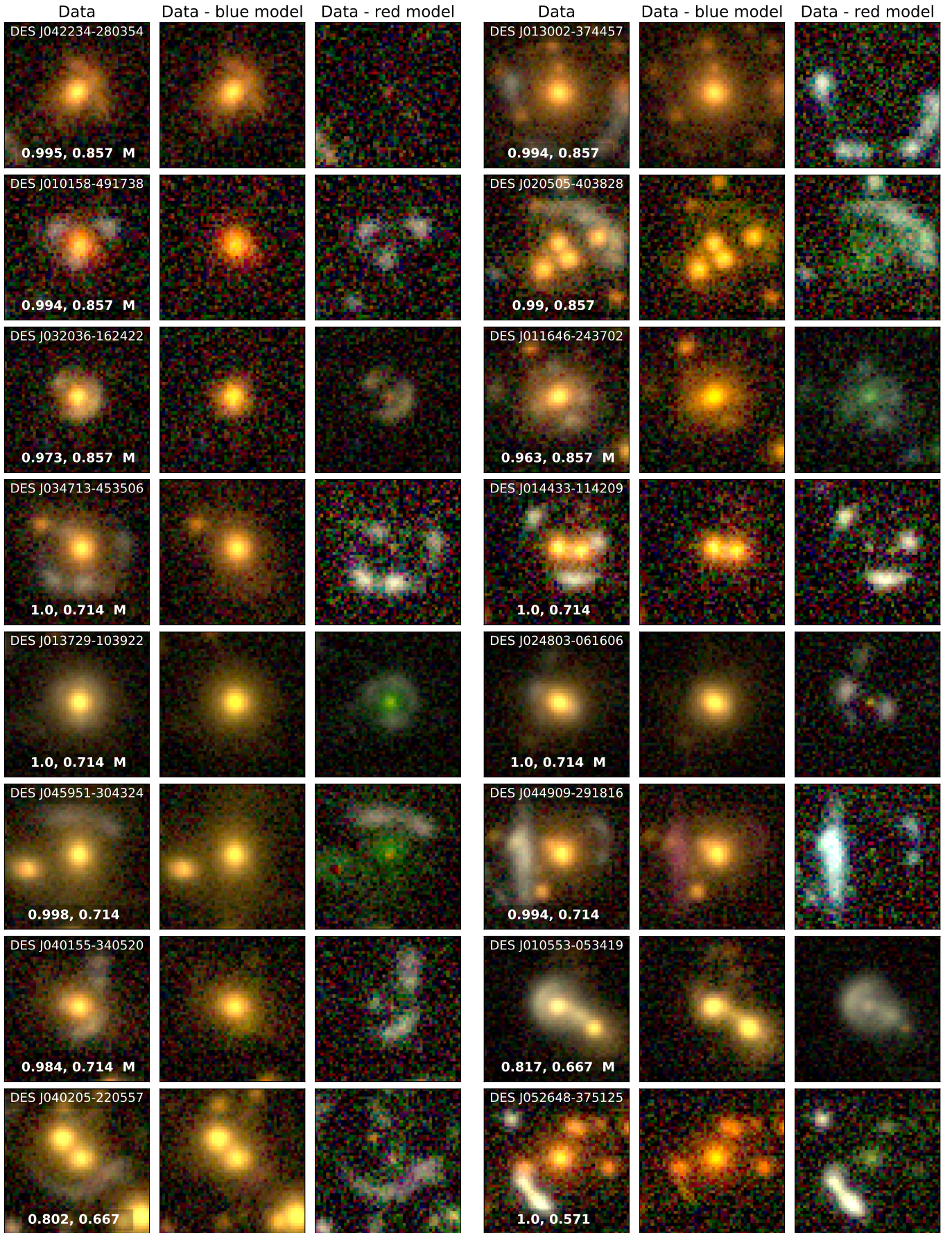


Fig. 19. Continued from Fig. 18

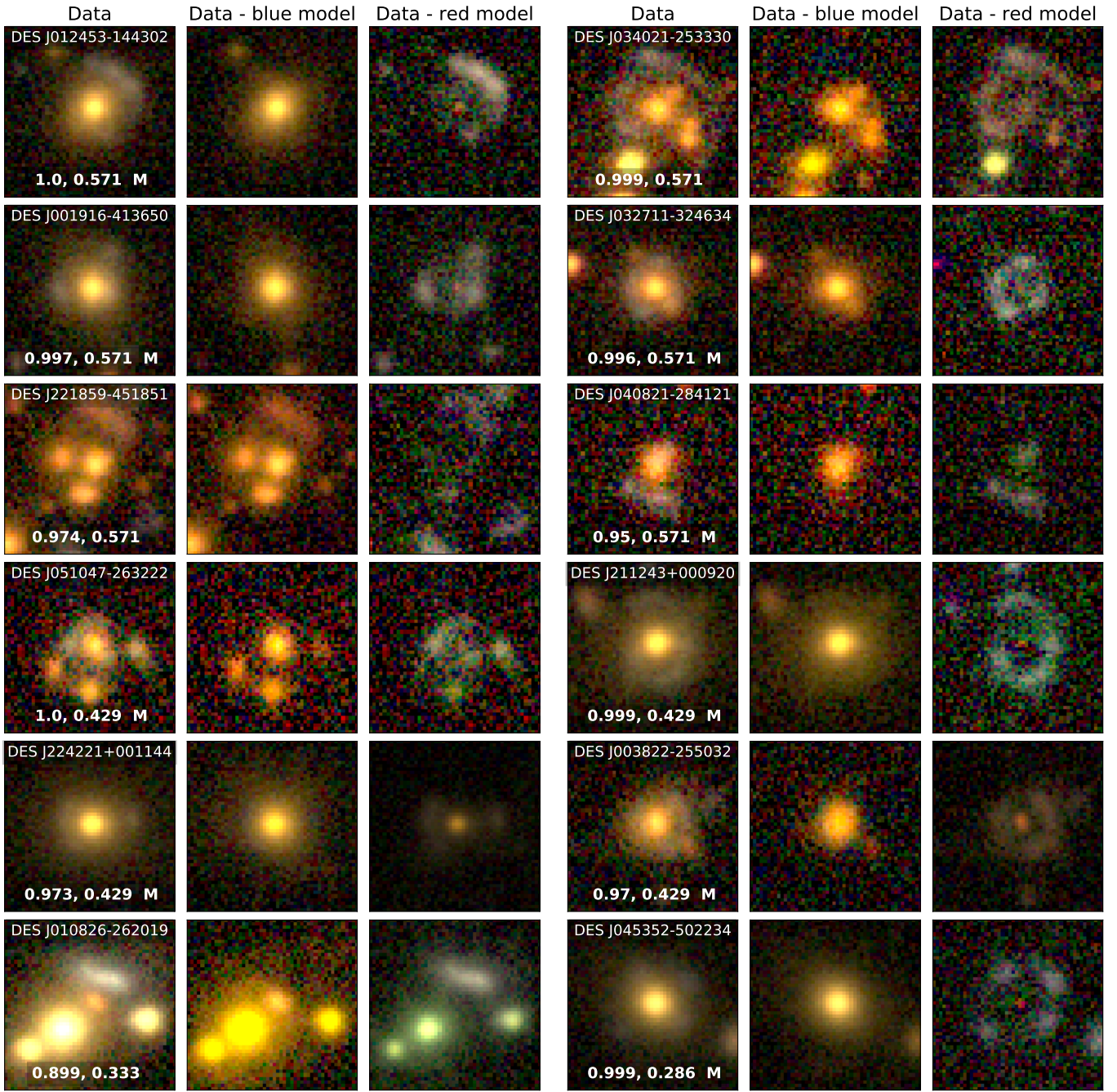


Fig. 20. Continued from Fig. 19

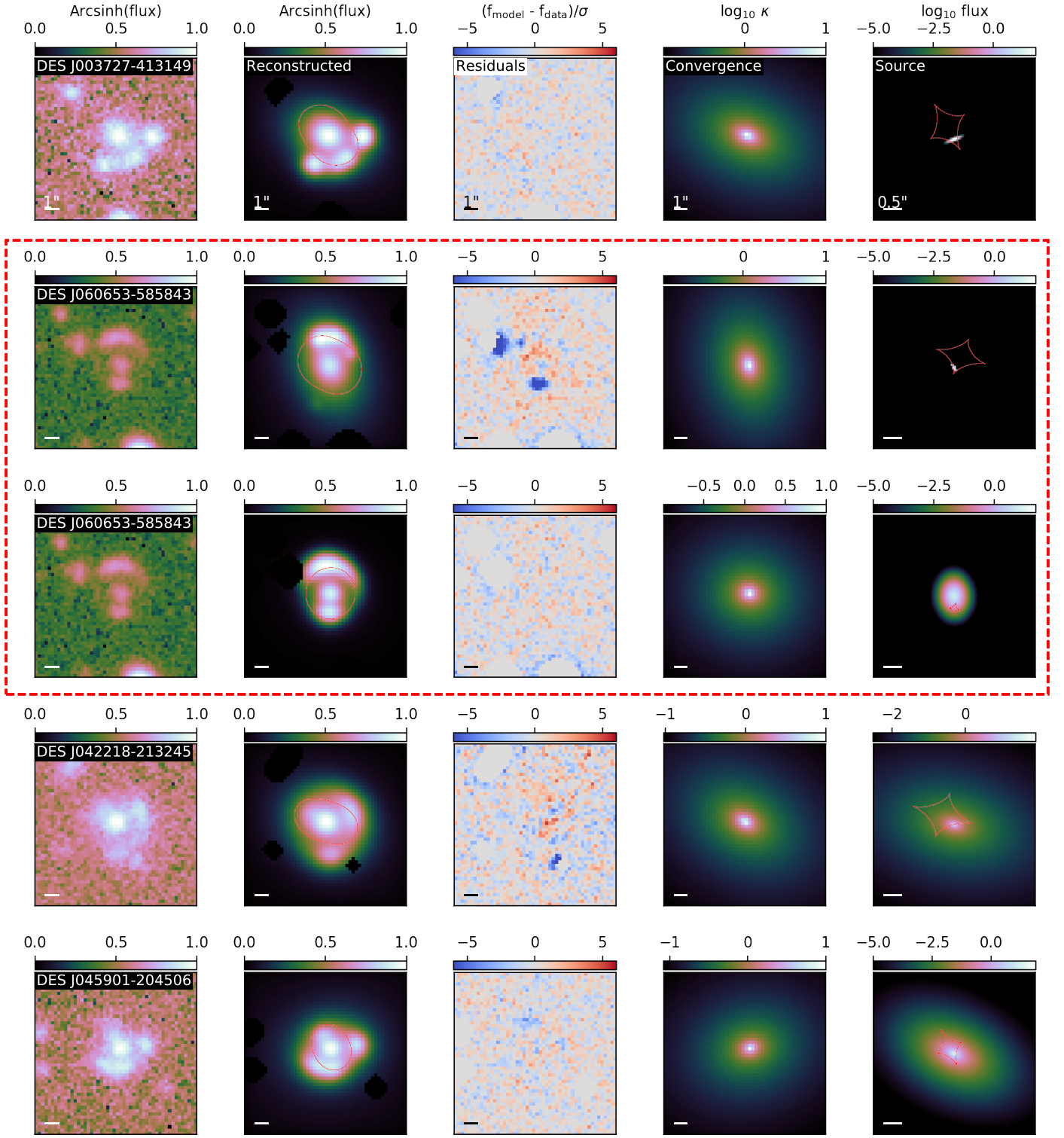


Fig. 21. Modeling results for the 52 lens candidates that appear to have only a single lens galaxy acting as a deflector. *1st column:* Observed DES image of the lens system in the *r*-band. *2nd column:* Reconstructed image using best-fit model parameters. The black regions are "masked" pixels that are ignored in the modeling as they contain light from contaminant objects in the image. The red curves are the critical lines of the lens model. *3rd column:* Normalized residual map showing the difference between the best-fit model and the original data. *4th column:* Convergence map of the lens model. *5th column:* Reconstructed source light profile (un-lensed). The caustic curves are shown in yellow.

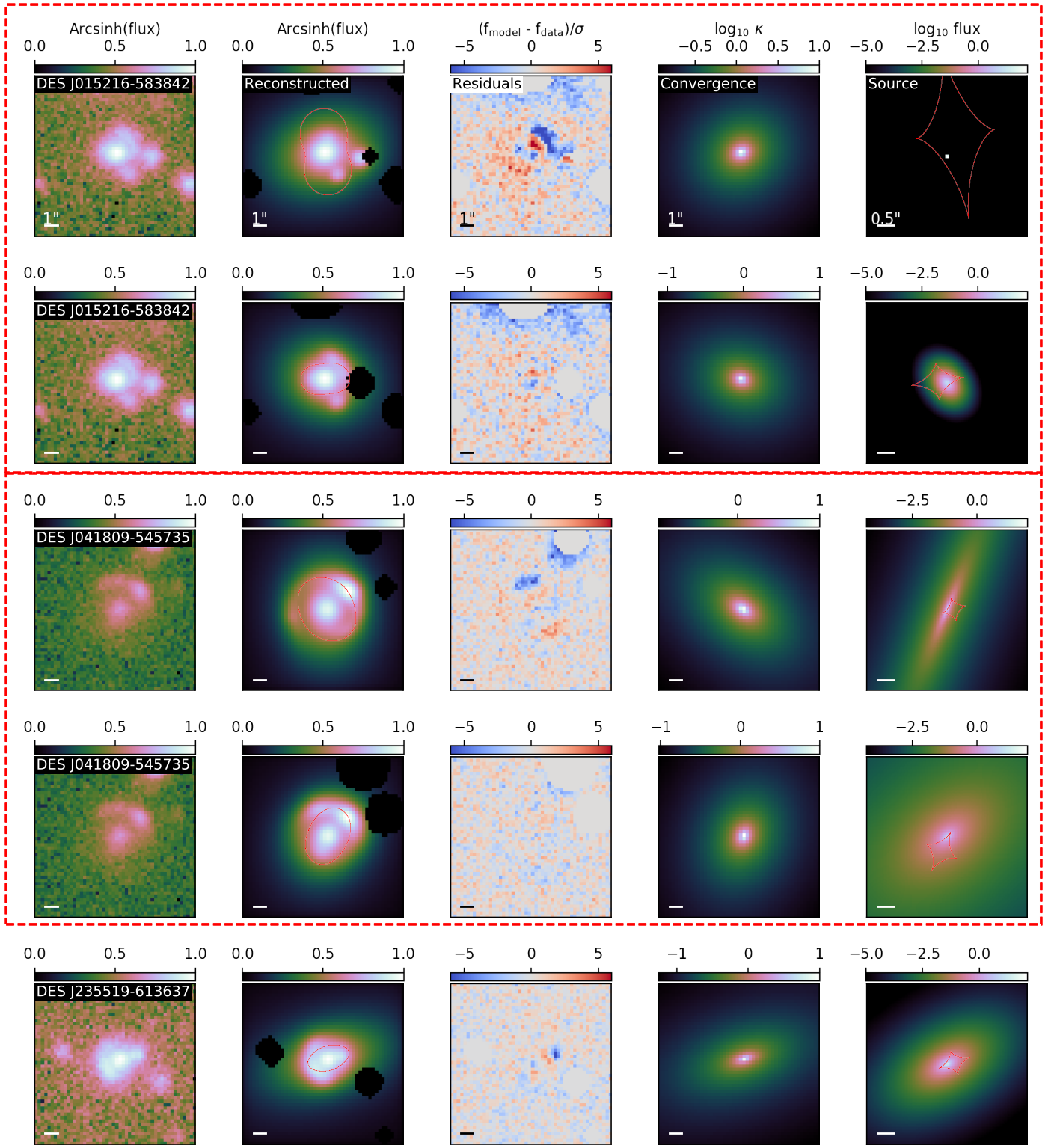


Fig. 22. continued from Figure 21

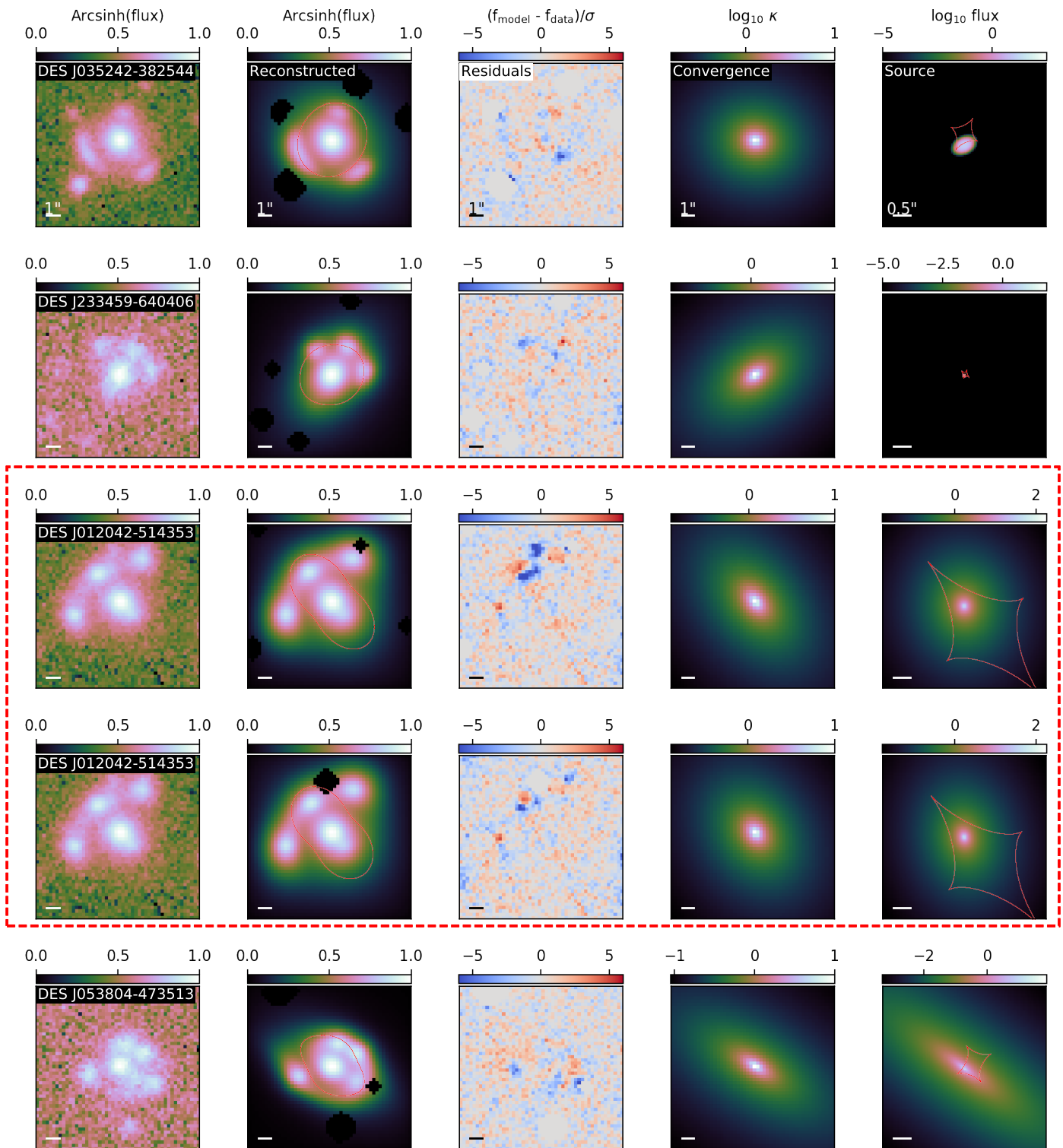


Fig. 23. continued from Figure 22

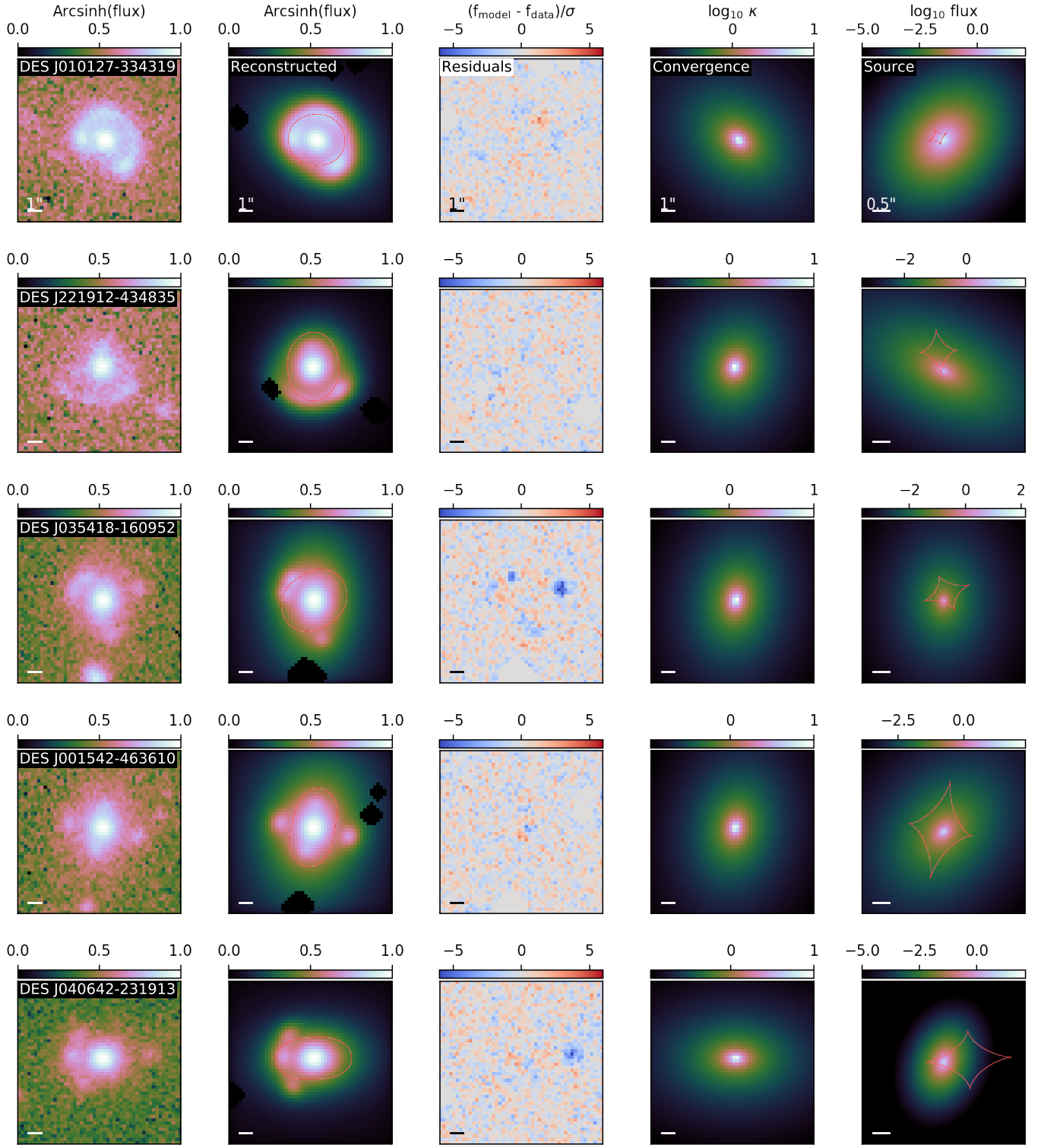


Fig. 24. continued from Figure 23

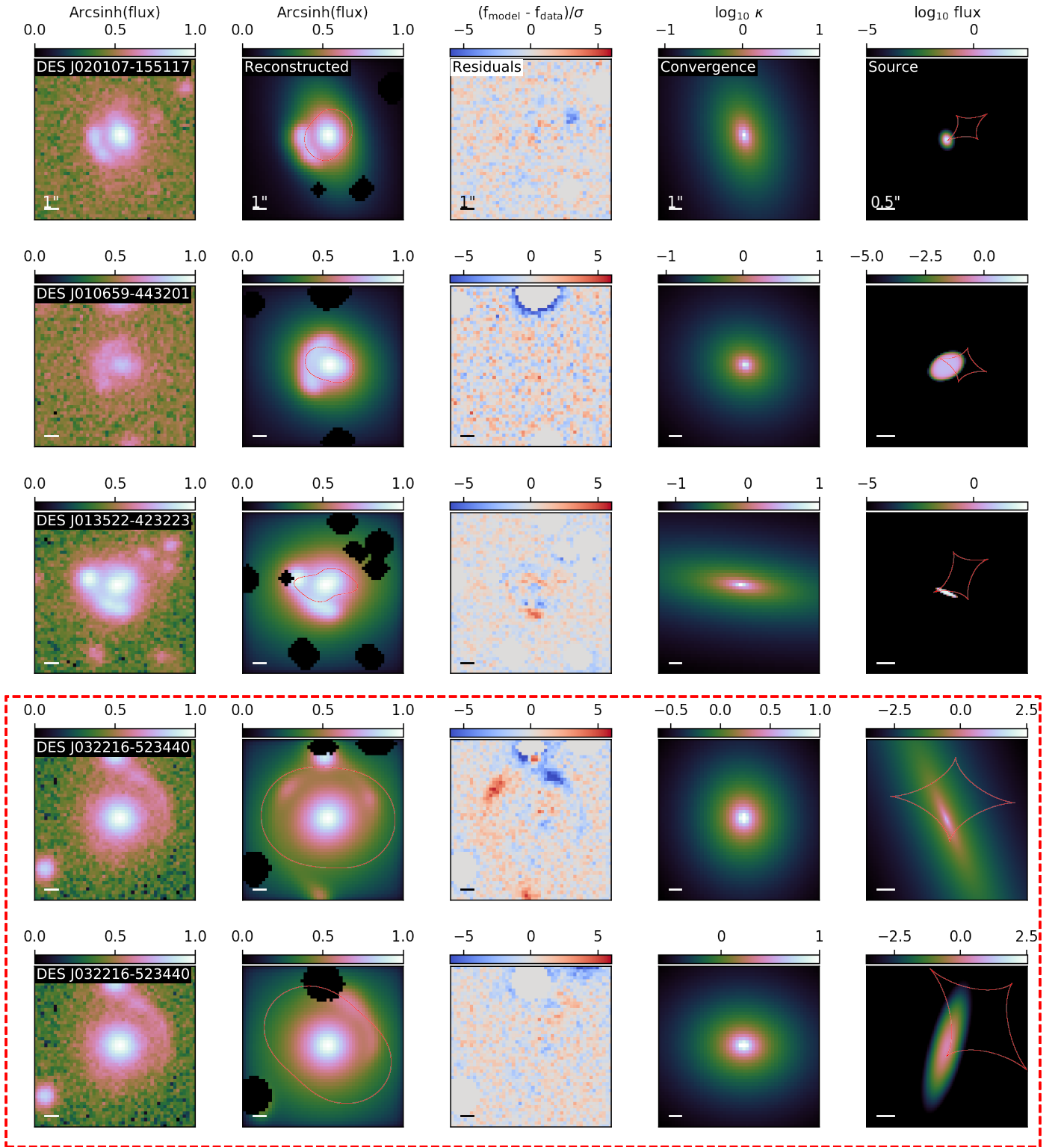


Fig. 25. continued from Figure 24

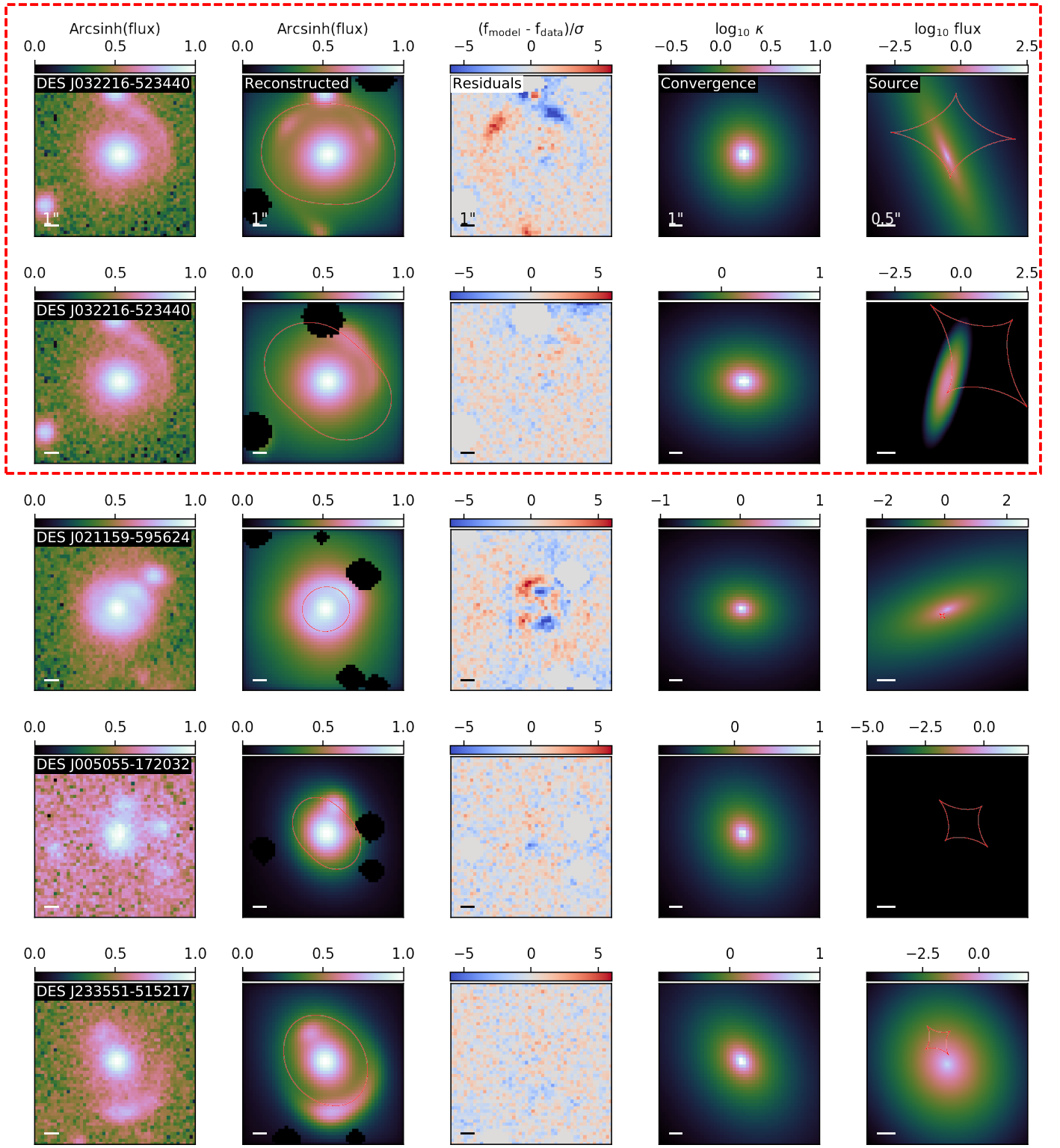


Fig. 26. continued from Figure 25

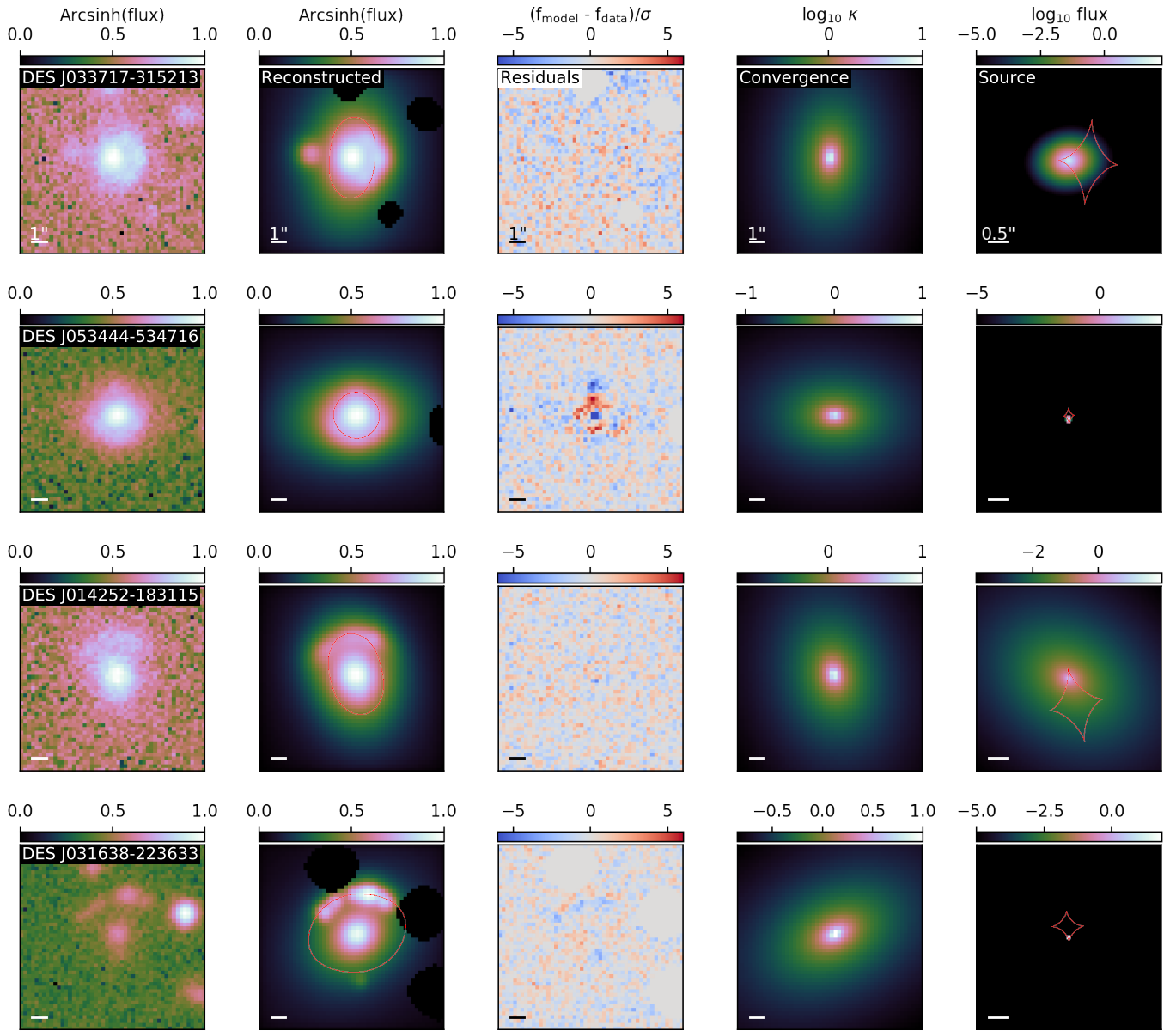


Fig. 27. continued from Figure 26

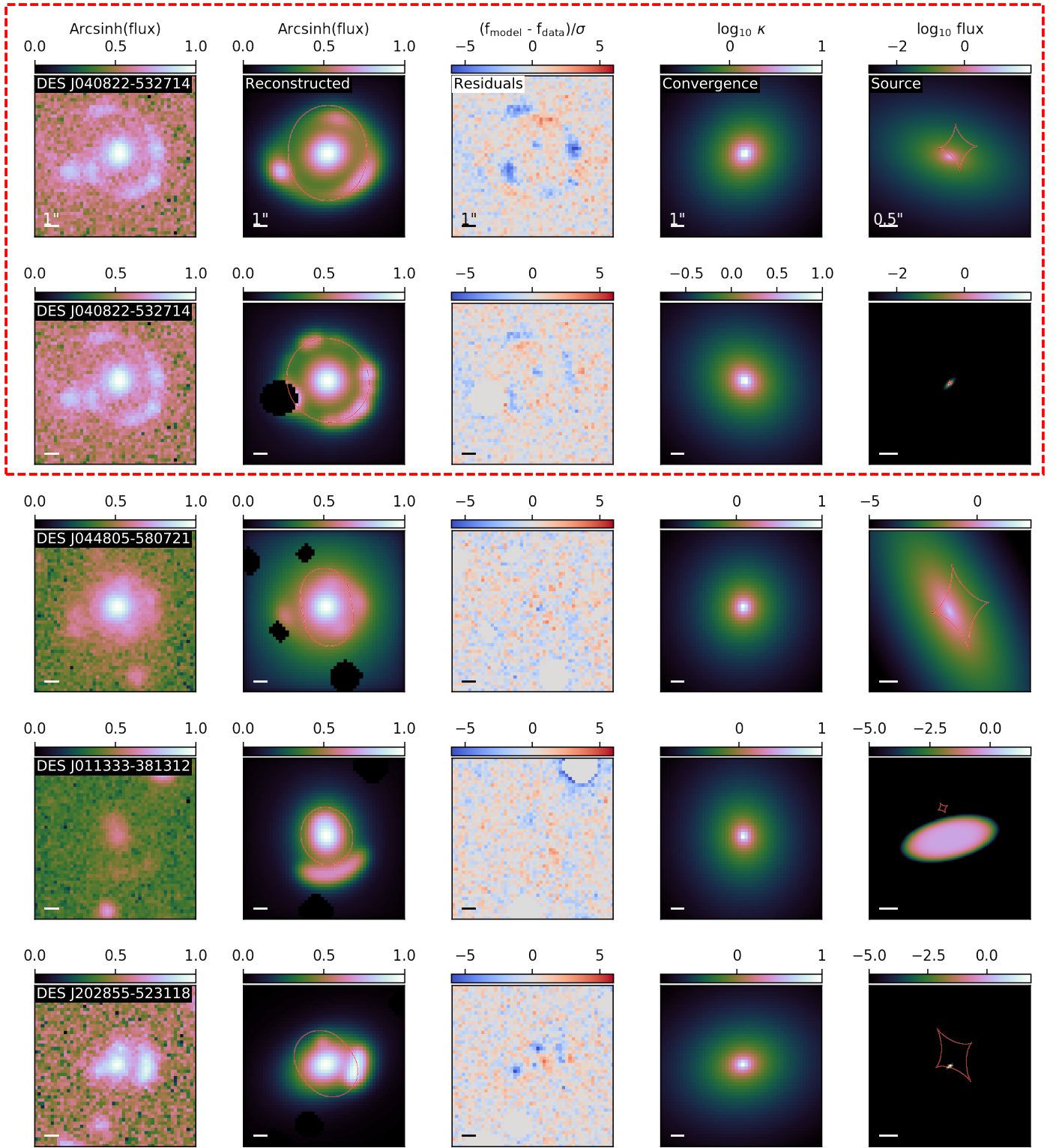


Fig. 28. continued from Figure 27

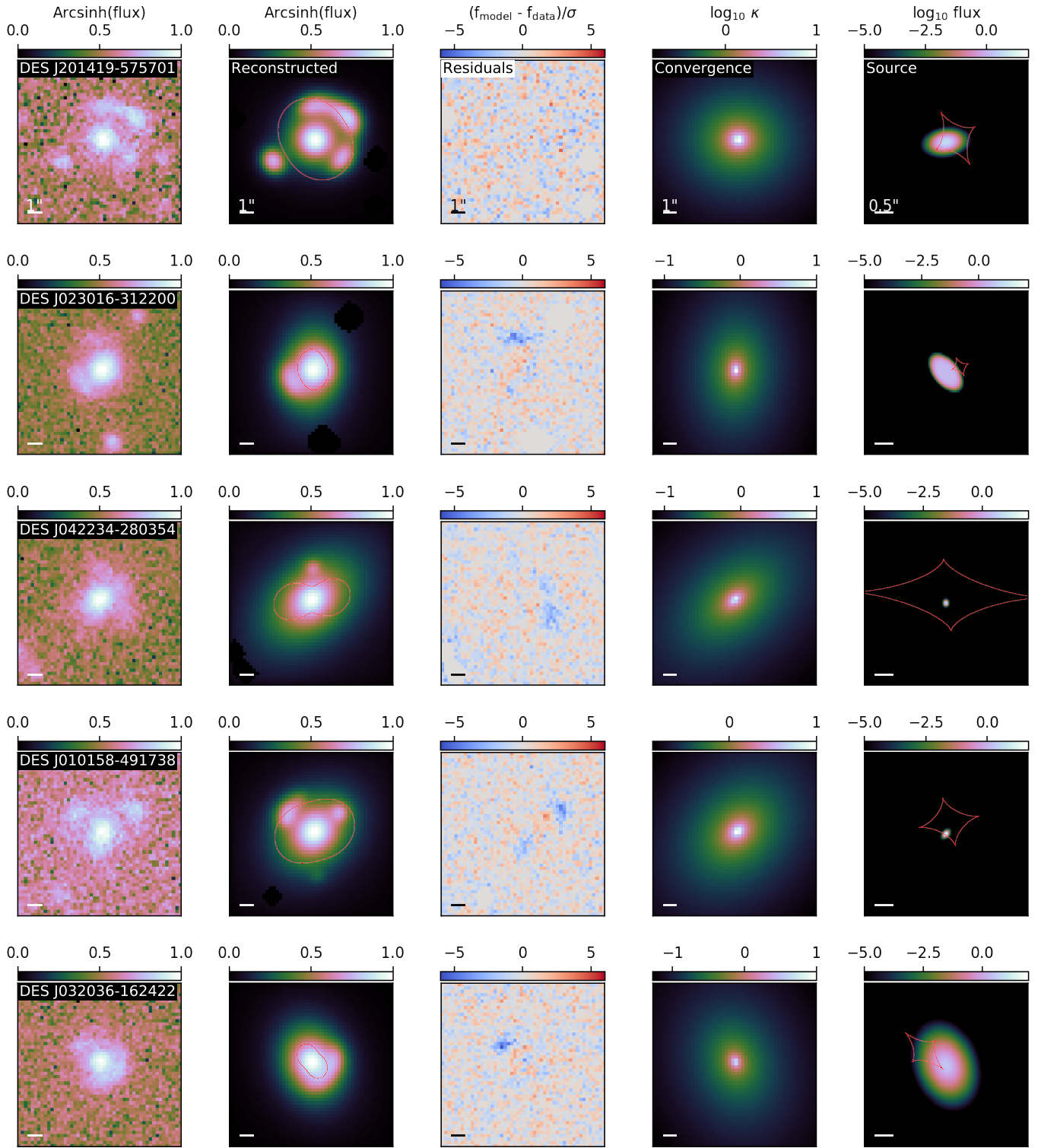


Fig. 29. continued from Figure 28

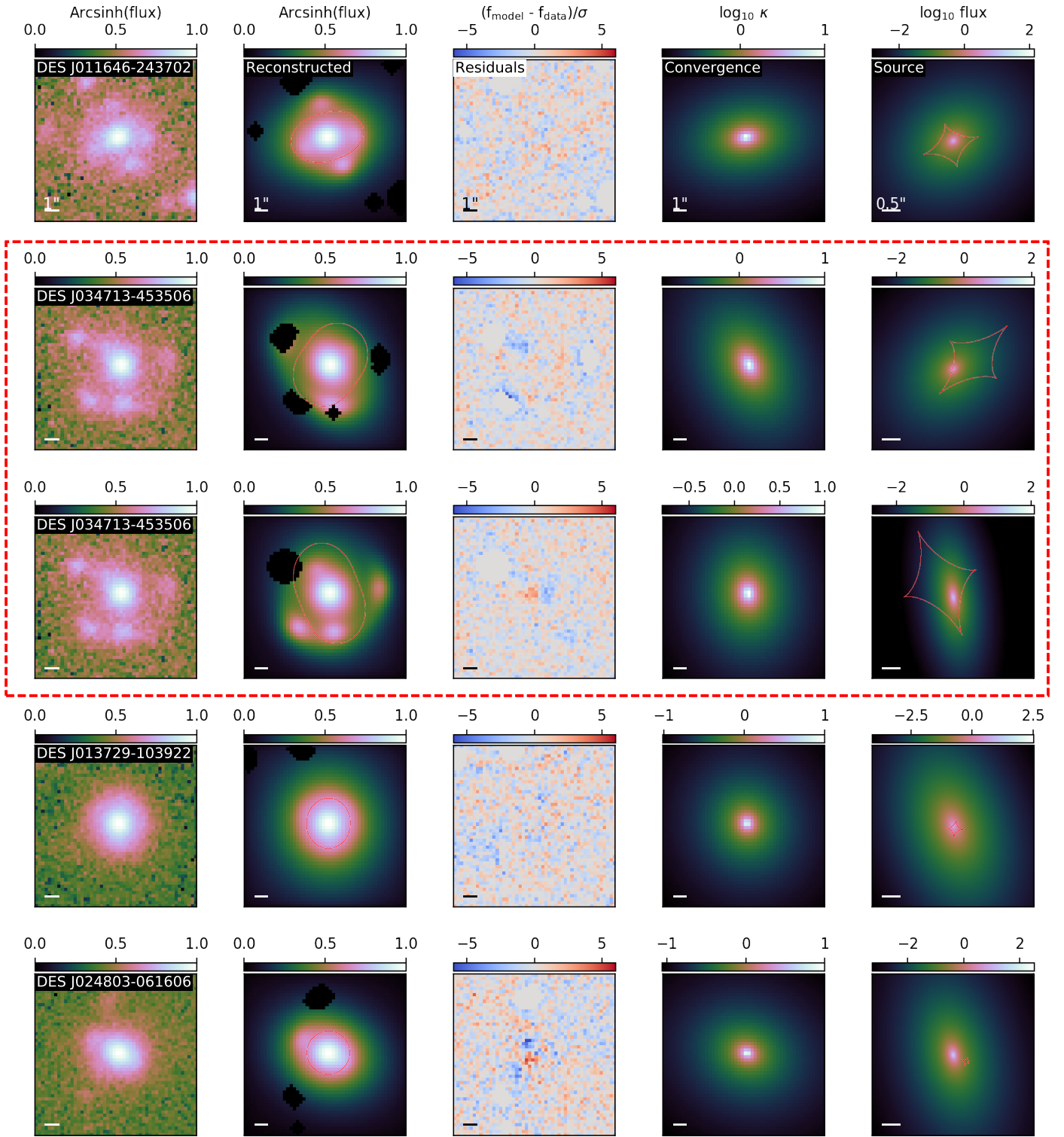


Fig. 30. continued from Figure 29

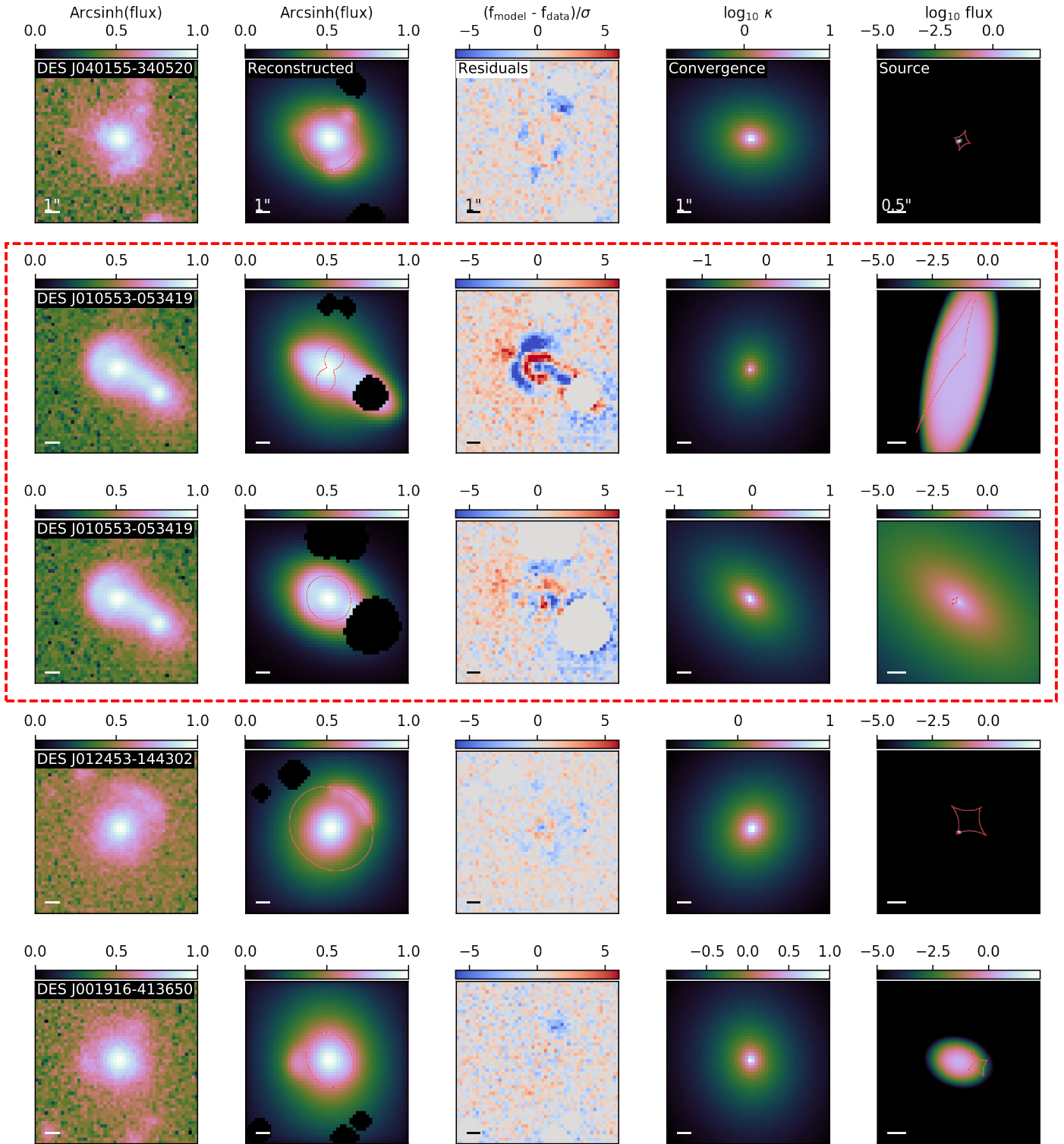


Fig. 31. continued from Figure 30

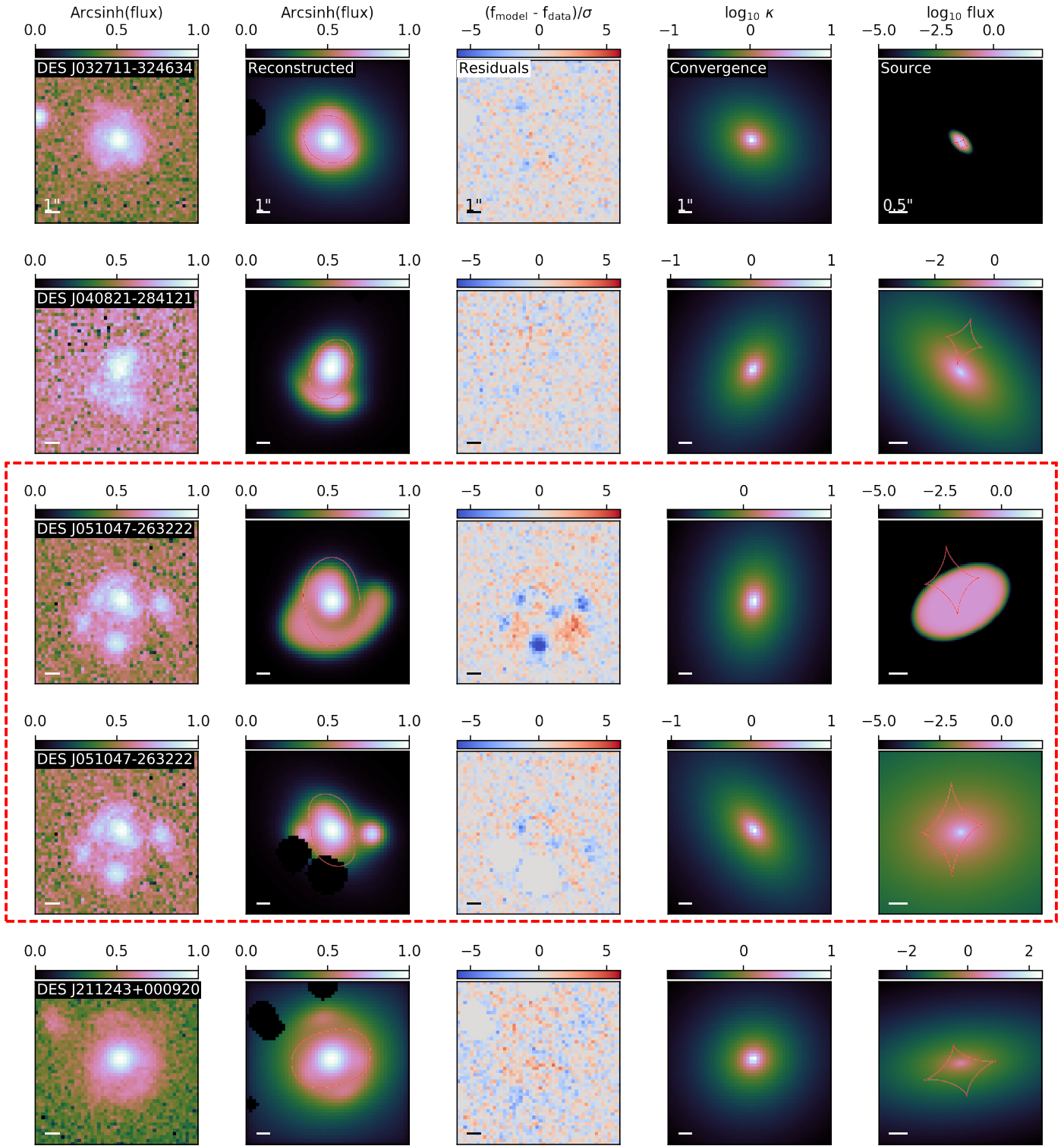


Fig. 32. continued from Figure 31

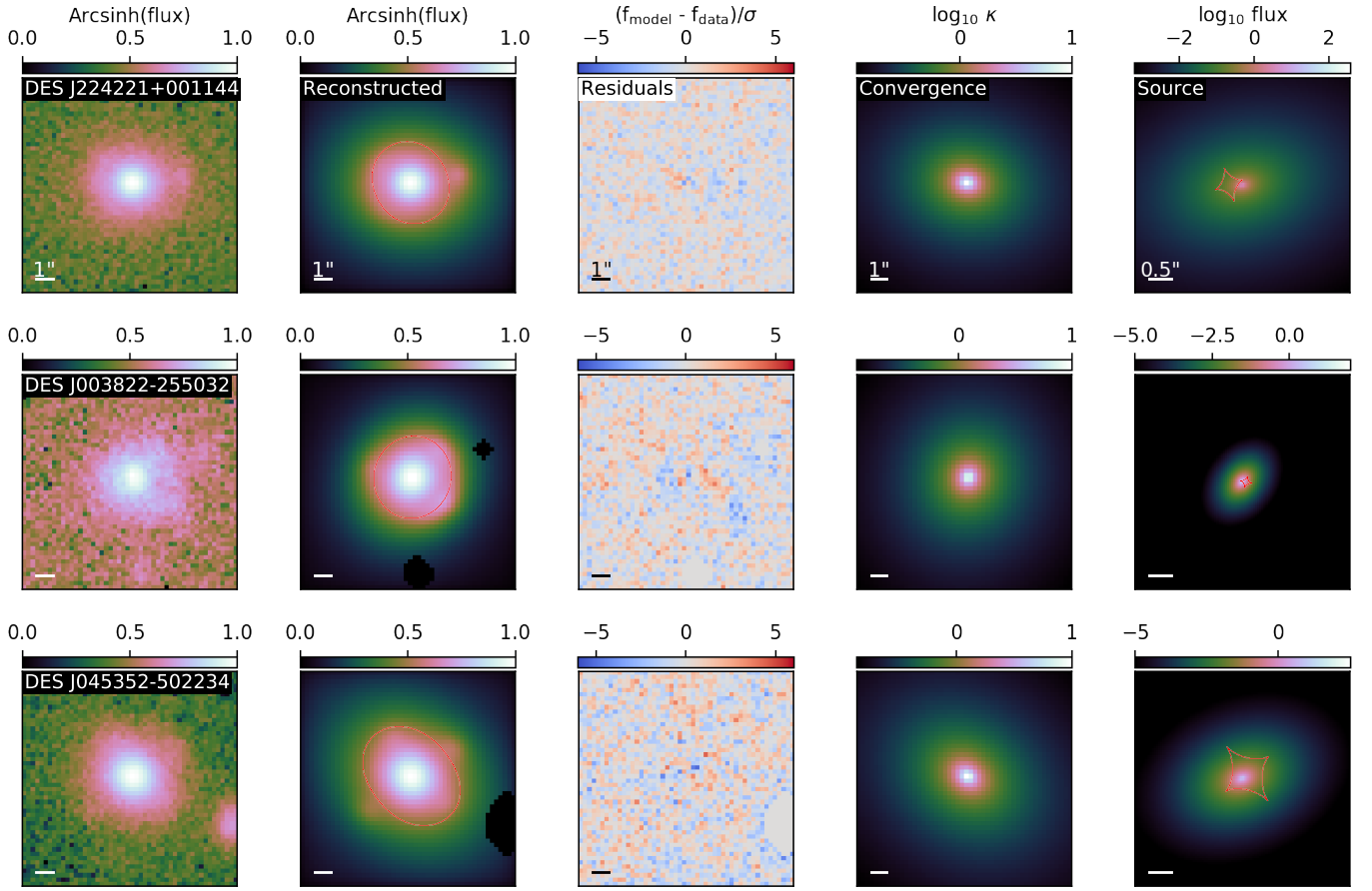


Fig. 33. continued from Figure 32

Appendix A: Comparison with [Jacobs et al. \(2019b,a\)](#)

We attempted to compare our results with ([Jacobs et al. 2019b,a](#), hereafter J19AB). These two searches use DES and obtained a total of 1256 candidates falling in three categories defined by the authors: “Definitely”, “Probably”, and “Possibly” a lens. Comparing with our work is more than challenging as our procedures are very different in color and magnitude selections, cutout size, simulated training sets, CNN architectures, and visual inspection. Being aware of these methodological differences, we compared the results for the 693 candidates in J19AB that are in our parent sample. Our CNN gave a score above 0.8 to 262 of these objects. Among these, 39 are classified as “Definitely”, 98 are “Probably”, and 125 are “Possibly” a lens, according to the J19AB classification. After our visual inspection we found that 50 are in our list of Sure lenses (splitting as 28 “Definitely”, 21 “Probably”, 1 “Possibly”, in the J19AB classification) and 83 in our maybe lens classification (8 “Definitely”, 51 “Probably”, 24 “Possibly”). This means that 129 candidates in the J19AB list that were also selected by our CNN did not pass our final visual inspection criteria. These objects are shown in Fig. A.1. Aside from a few objects in the first 3 rows on the top of the figure, we still find that no other object show sufficient evidence for strong lensing. And indeed, J19AB graded them all as “Possibly” lenses. The main difference between our work and J19AB for these 129 objects is that we discard them while J19AB still include them in their list, hence leading to a list of candidates that is larger than ours.

In addition, 505 candidates from J19AB obtained a score below 0.8 from our CNN and were not visually inspected. They split into 6 being in the “Definitely” category of J19AB, 113 as “Probably”, and 386 as “Possibly”. We display 56 examples of these objects in Fig. A.2 and see that, if we were to visually inspect them now, very few objects would pass our criteria, for the following reasons: 1- the bluish features are too close to the central galaxy that could be mistaken as a star forming galaxy, or 2- there is no evidence for lensing at all, because the lensing features are outside our stamp size or because of too low signal-to-noise. Determining more specific reasons why our CNN discarded these candidates would require a more detailed study that remains well beyond the scope of our work.

Acknowledgements. This work is supported by the Swiss National Science Foundation (SNSF) and by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (COSMICLENS: grant agreement No 787886). RC thanks the Max Planck Society for support through the Max Planck Research Group of S. H. Suyu. This project has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (LENSNOVA: grant agreement No 771776). GV has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 897124. This research has made use of the Vizier catalogue access tool, CDS, Strasbourg, France (DOI : 10.26093/cds/vizier). The original description of the Vizier service was published in [Ochsenbein et al. \(2000\)](#). This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France [Wenger et al. \(2000\)](#). This material is based upon work supported by the National Science Foundation under Cooperative Agreement 1258333 managed by the Association of Universities for Research in Astronomy (AURA), and the Department of Energy under Contract No. DE-AC02-76SF00515 with the SLAC National Accelerator Laboratory. Additional funding for *Rubin* Observatory comes from private donations, grants to universities, and in-kind support from LSSTC Institutional Members. This work rely on the following packages: [numpy \(van der Walt et al. 2011\)](#), [scipy \(Virtanen et al. 2020\)](#), [matplotlib \(Hunter 2007\)](#), [astropy \(Price-Whelan et al. 2018\)](#) and [Oscarlet_extensions](#) and [csv](#). A notebook detailing the procedure of deblending can be found in the following repository: [OLens-Deblend](#).

References

- Abbott, T. M. C., Abdalla, F. B., Allam, S., et al. 2018, *ApJS*, 239, 18
Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, *PASJ*, 70, S4
Alard, C. 2006, arXiv e-prints, astro
Auger, M. W., Treu, T., Bolton, A. S., et al. 2009, *ApJ*, 705, 1099
Aveztruz, C., Li, N., Zhu, H., et al. 2019, *ApJ*, 877, 58
Bertin, E. 2011, in *Astronomical Society of the Pacific Conference Series*, Vol. 442, *Astronomical Data Analysis Software and Systems XX*, ed. I. N. Evans, A. Accomazzi, D. J. Mink, & A. H. Rots, 435
Bertin, E. & Arnouts, S. 1996, *A&AS*, 117, 393
Biesiada, M., Piórkowska, A., & Malec, B. 2010, *MNRAS*, 406, 1055
Birrer, S. 2021, arXiv e-prints, arXiv:2104.09522
Birrer, S. & Amara, A. 2018, *Physics of the Dark Universe*, 22, 189
Birrer, S., Amara, A., & Refregier, A. 2015, *ApJ*, 813, 102
Bonvin, V., Courbin, F., Suyu, S. H., et al. 2017, *MNRAS*, 465, 4914
Cañameras, R., Schuldt, S., Suyu, S. H., et al. 2020, *A&A*, 644, A163
Cabanac, R. A., Alard, C., Dantel-Fort, M., et al. 2007, *A&A*, 461, 813
Cao, S., Biesiada, M., Gavazzi, R., Piórkowska, A., & Zhu, Z.-H. 2015, *ApJ*, 806, 185
Cao, S., Pan, Y., Biesiada, M., Godlowski, W., & Zhu, Z.-H. 2012, *J. Cosmology Astropart. Phys.*, 2012, 016
Chan, J. H. H., Suyu, S. H., Chiueh, T., et al. 2015, *ApJ*, 807, 138
Chollet, F. et al. 2015, *Keras*, <https://keras.io>
Collett, T. E. 2015, *ApJ*, 811, 20
Collett, T. E. & Auger, M. W. 2014, *MNRAS*, 443, 969
Comparat, J., Richard, J., Kneib, J.-P., et al. 2015, *A&A*, 575, A40
Davis, A. N., Huterer, D., & Krauss, L. M. 2003, *MNRAS*, 344, 1029
Diehl, H. T., Buckley-Geer, E. J., Lindgren, K. A., et al. 2017, *ApJS*, 232, 15
Falco, E. E., Shapiro, I. I., Moustakas, L. A., & Davis, M. 1997, *ApJ*, 484, 70
Fitzpatrick, M. J., Graham, M. J., Mighell, K. J., et al. 2016, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 9913, *Software and Cyberinfrastructure for Astronomy IV*, ed. G. Chiozzi & J. C. Guzman, 99130L
Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, *AJ*, 150, 150
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *Publications of the Astronomical Society of the Pacific*, 125, 306
Gavazzi, R., Marshall, P. J., Treu, T., & Sonnenfeld, A. 2014, *ApJ*, 785, 144
Gentile, F., Tortora, C., Covone, G., et al. 2021, arXiv e-prints, arXiv:2105.05602
Goodman, J. & Weare, J. 2010, *Communications in Applied Mathematics and Computational Science*, 5, 65
Griffith, R. L., Cooper, M. C., Newman, J. A., et al. 2012, *ApJS*, 200, 9
Hasinger, G., Capak, P., Salvato, M., et al. 2018, *ApJ*, 858, 77
He, K., Zhang, X., Ren, S., & Sun, J. 2015, arXiv e-prints, arXiv:1512.03385
He, Z., Er, X., Long, Q., et al. 2020, *MNRAS*, 497, 556
Honscheid, K. & DePoy, D. L. 2008, arXiv e-prints, arXiv:0810.3600
Huang, X., Storfer, C., Gu, A., et al. 2021, *ApJ*, 909, 27
Huang, X., Storfer, C., Ravi, V., et al. 2020, *ApJ*, 894, 78
Hunter, J. D. 2007, *Computing in Science & Engineering*, 9, 90
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, 873, 111
Jacobs, C., Collett, T., Glazebrook, K., et al. 2019a, *ApJS*, 243, 17
Jacobs, C., Collett, T., Glazebrook, K., et al. 2019b, *MNRAS*, 484, 5330
Jacobs, C., Glazebrook, K., Collett, T., More, A., & McCarthy, C. 2017, *MNRAS*, 471, 167
Jaelani, A. T., More, A., Oguri, M., et al. 2020, *MNRAS*, 495, 1291
Jiménez-Vicente, J., Mediavilla, E., Kochanek, C. S., & Muñoz, J. A. 2015, *ApJ*, 806, 251
Joseph, R., Courbin, F., Metcalf, R. B., et al. 2014, *A&A*, 566, A63
Joseph, R., Courbin, F., & Starck, J. L. 2016, *A&A*, 589, A2
Keeton, C. R., Kochanek, C. S., & Seljak, U. 1997, *The Astrophysical Journal*, 482, 604
Kelvin, L. S., Driver, S. P., Robotham, A. S. G., et al. 2012, *Monthly Notices of the Royal Astronomical Society*, 421, 1007
Kennedy, J. & Eberhart, R. 1995, in *Proceedings of ICNN’95 - International Conference on Neural Networks*, Vol. 4, 1942–1948 vol.4
Kochanek, C. S. & Dalal, N. 2001, arXiv e-prints, astro
Koekemoer, A. M., Aussel, H., Calzetti, D., et al. 2007, *ApJS*, 172, 196
Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJS*, 224, 24
Lanusse, F., Ma, Q., Li, N., et al. 2018, *MNRAS*, 473, 3895
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, arXiv:1110.3193
Le Fèvre, O., Tasca, L. A. M., Cassata, P., et al. 2015, *A&A*, 576, A79
Leauthaud, A., Massey, R., Kneib, J.-P., et al. 2007, *ApJS*, 172, 219
LeCun, Y., Boser, B., Denker, J. S., et al. 1989, *Neural Computation*, 1, 541
Lemon, C., Auger, M. W., McMahon, R., et al. 2020, *MNRAS*, 494, 3491
Li, R., Napolitano, N. R., Tortora, C., et al. 2020, *ApJ*, 899, 30
Lilly, S. J., Le Fèvre, O., Renzini, A., et al. 2007, *ApJS*, 172, 70
Limousin, M., Cabanac, R., Gavazzi, R., et al. 2009, *A&A*, 502, 445
LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv e-prints, arXiv:0912.0201
Maturi, M., Mizera, S., & Seidel, G. 2014, *A&A*, 567, A111

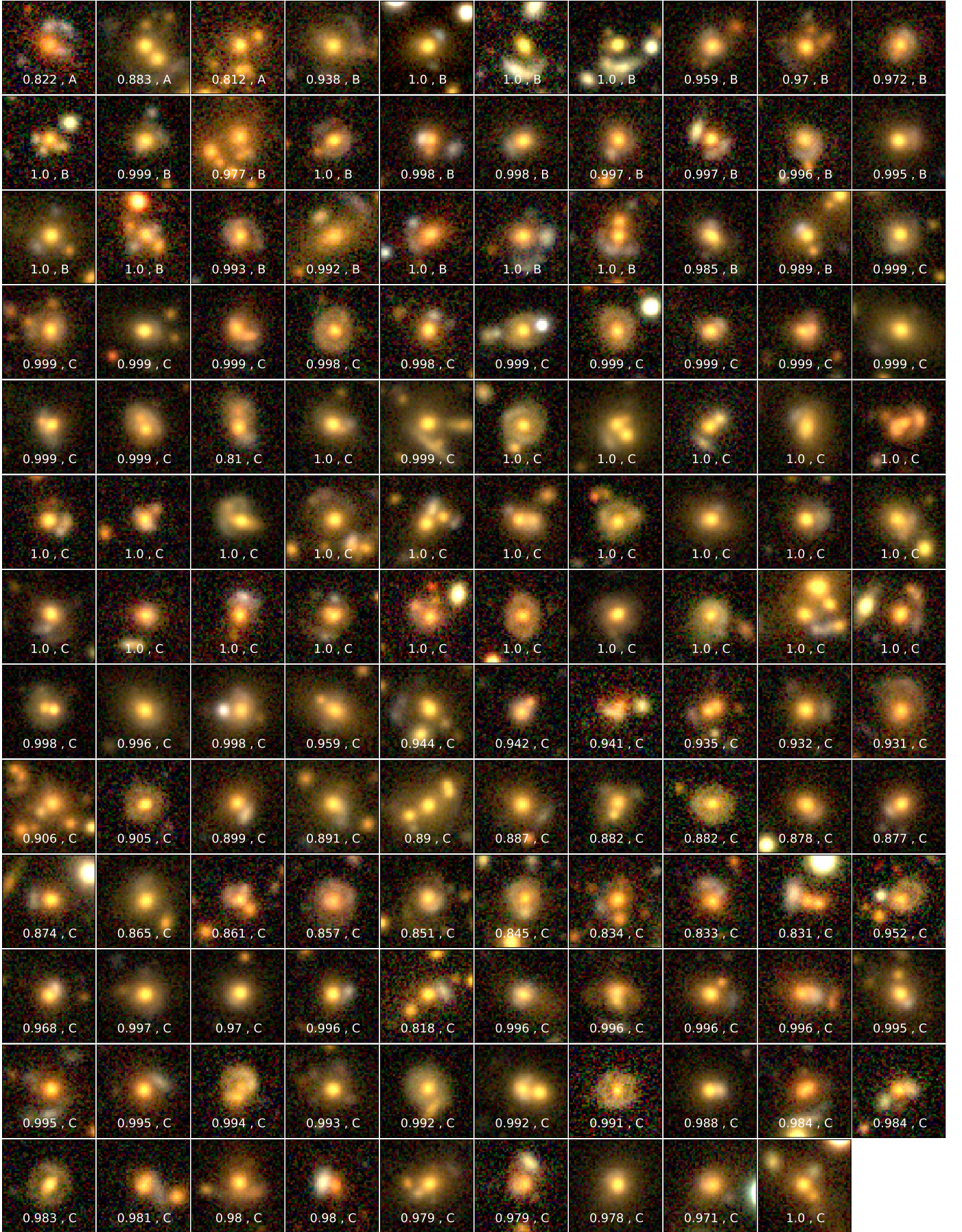


Fig. A.1. Images for the 129 candidates in J19AB that we discarded after visual inspection. Our CNN score is indicated below each object as well as the category allocated by J19AB, in which A stands for “Definitely”, B for “Probably”, and C for “Possibly” a lens.

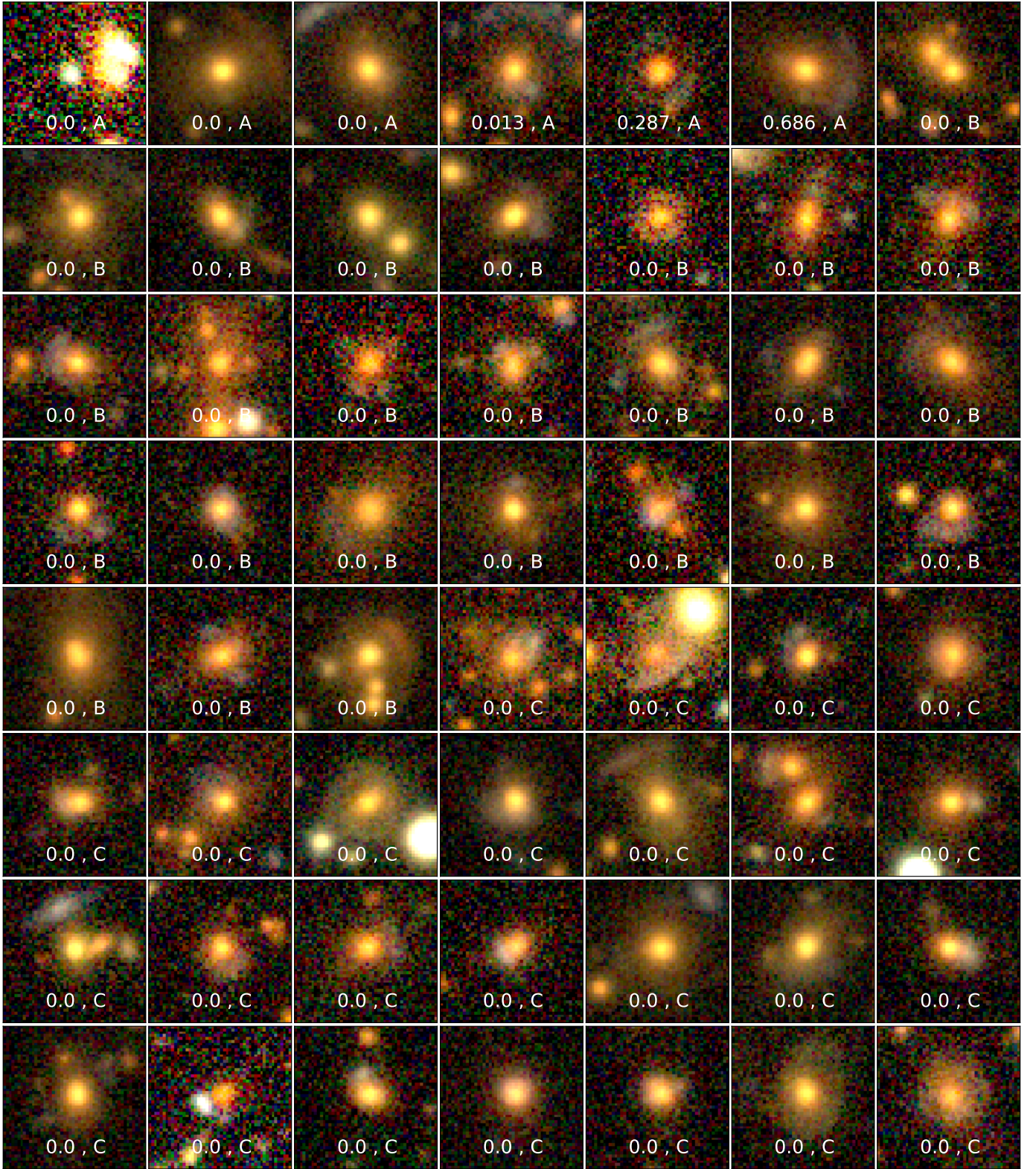


Fig. A.2. Example of cutouts of 56 candidates in J19AB that our CNN graded below 0.8. Our CNN score is indicated below each object as well as the category allocated by J19AB, in which A stands for “Definitely”, B for “Probably”, and C for “Possibly” a lens.

Melchior, P., Moolekamp, F., Jerdee, M., et al. 2018, *Astronomy and Computing*, 24, 129
 Millon, M., Galan, A., Courbin, F., et al. 2020, *A&A*, 639, A101
 More, A., Cabanac, R., More, S., et al. 2012, *ApJ*, 749, 38
 More, A., Oguri, M., Kayo, I., et al. 2016, *MNRAS*, 456, 1595
 Nightingale, J. W., Dye, S., & Massey, R. J. 2018, *MNRAS*, 478, 4738
 Nord, B., Buckley-Geer, E., Lin, H., et al. 2020, *MNRAS*, 494, 1308

Ochsenbein, F., Bauer, P., & Marcout, J. 2000, *A&AS*, 143, 23
 Oguri, M., Taruya, A., Suto, Y., & Turner, E. L. 2002, *ApJ*, 568, 488
 Paraficz, D., Courbin, F., Tramacere, A., et al. 2016, *A&A*, 592, A75
 Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2019a, *MNRAS*, 482, 807
 Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2017, *MNRAS*, 472, 1129
 Petrillo, C. E., Tortora, C., Vernardos, G., et al. 2019b, *MNRAS*, 484, 3879
 Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., et al. 2018, *AJ*, 156, 123

- Rosenblatt, M. 1957, *Proceedings of the National Academy of Science*, 43, 989
- Rowe, B. T. P., Jarvis, M., Mandelbaum, R., et al. 2015, *Astronomy and Computing*, 10, 121
- Scoville, N., Abraham, R. G., Aussel, H., et al. 2007, *ApJS*, 172, 38
- Shajib, A. J., Treu, T., Birrer, S., & Sonnenfeld, A. 2020, *arXiv e-prints*, arXiv:2008.11724
- Silverman, J. D., Kashino, D., Sanders, D., et al. 2015, *VizieR Online Data Catalog*, J/ApJS/220/12
- Starck, J.-L., Fadili, J., & Murtagh, F. 2007, *IEEE Transactions on Image Processing*, 16, 297
- Tan, M. & Le, Q. V. 2020, *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*
- Tasca, L. A. M., Le Fèvre, O., Ribeiro, B., et al. 2017, *A&A*, 600, A110
- The Dark Energy Survey Collaboration. 2005, *arXiv e-prints*, astro
- Turner, E. L., Ostriker, J. P., & Gott, J. R., I. 1984, *ApJ*, 284, 1
- van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *Computing in Science Engineering*, 13, 22
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nature Methods*, 17, 261
- Vuissoz, C., Courbin, F., Sluse, D., et al. 2007, *A&A*, 464, 845
- Walsh, D., Carswell, R. F., & Weymann, R. J. 1979, *Nature*, 279, 381
- Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, *A&AS*, 143, 9
- Willett, K. W., Galloway, M. A., Bamford, S. P., et al. 2017, *MNRAS*, 464, 4176
- Wong, K. C., Sonnenfeld, A., Chan, J. H. H., et al. 2018, *ApJ*, 867, 107
- Wong, K. C., Suyu, S. H., Chen, G. C. F., et al. 2020, *MNRAS*, 498, 1420

8.2 Discussion

In Rojas et al. (2021) , we compared our results with the work of Jacobs et al. (2019a) and Jacobs et al. (n.d.) . From the 1256 candidates of Jacobs et al. (2019a) and Jacobs et al. (n.d.) 693 were in our pre-selection. Among these, 431 were not selected by our CNN. We visually inspected the images rejected by our networks. Most images would not meet our criteria to be considered secure lenses or maybe lenses. For some candidates, we found no lensing features in our images. This may happen because our cutout size is smaller; therefore, some of the lensing features detected in Jacobs et al. (2019a) and Jacobs et al. (n.d.) may be located outside our images. A large part of the 431 objects not selected by our CNN contains blue features blended in the light of the deflector. This type of object is difficult to differentiate from a star-forming galaxy. We built our set of positive examples to avoid the presence of such objects. In addition to the lower limit on Einstein radii in the simulations, the boost of the source's magnitude allows us to unveil the shape of lensing features. Since CNN was trained mainly with simulations with blatant lensing features, it tends to give lower scores to the star-forming galaxy-like objects. Notably, a few of the candidates from Jacobs et al. (2019a) and Jacobs et al. (n.d.) missed by our CNN are doubly imaged sources. This lens type is rare in our training set due to our lens mass model and our method of positioning the source. Including more doubles in our training set may improve this issue in future search iterations. On the 262 candidates of Jacobs et al. (2019a) and Jacobs et al. (n.d.) selected by our CNN, 129 were rejected during our visual inspection. This illustrates the need to define common grading guidelines that can be used in the future version of the Master Lens Database.

In Savary et al. (2021) , we used a committee of three CNNs trained with subsets of the training set. Each subset contains a different proportion of positive examples. The ratio is drawn randomly in the range of 20 to 50 percent. This aims to mitigate the tendency of the network to learn the fraction of positive examples seen in the training set. Compared with a committee of three occurrences of Efficientnet trained with the same training set, this drastically reduced the number of images that obtained a score above 0.5 (45 823 in comparison with 9 460 candidates). To estimate the impact of the training with different fractions of lenses on the false positive rate, we evaluated 8 200 CFIS images of spiral galaxies taken from the Galaxy Zoo catalog (Willett et al., 2017) with the two versions of the committee. The committee trained with the three same training sets classified 57 spirals as lenses (score above 0.5), whereas the committee trained with a different proportion classified only eight spirals as lenses. Therefore, we think this approach can reduce the false positive rate. Nevertheless, a visual inspection of the 45 823 in the same conditions as in Savary et al. (2021) would be necessary to estimate precisely the effect of the different training sets on the false positive rate and the completeness of our sample. A visual inspection of 45 823 has a high cost in terms of human resources. For this reason, it has not been done for the moment.

Following what has been observed in Gentile et al. (2021) , we found that the CNN trained with color images could detect more candidates with a small image separation. However, due to our strict visual inspection guidelines, there were not taken into account in our final sample

of lenses. The DES candidates' sample contained significantly more galaxies with point-like companions than the CFIS sample. We think the network with color images tends to select some of these images because of the color difference between the companions and the central galaxy. However, to confirm this, a visual inspection of the candidates found with a CNN trained on one-band DES images has to be done in the future. The false positive sample of CFIS contained mainly objects mimicking the shape of lensing features and images with CCD artifacts. Methods to exclude the images with CCD artifacts must be implemented before the next iteration of the CFIS lens search.

We presented in this chapter the two first applications of our lens finding pipeline to CFIS and DES LRG images. It allowed us to find 405 lens candidates, of which 186 are new in DES and 133 lens candidates in CFIS, among which 104 were not known. In addition to the lens candidates, these two searches provided catalogs of the ring, spiral, and merger galaxies found during the visual inspection. These catalogs can be used to retrain the CNNs against the common false positives in the subsequent iterations of the searches in CFIS and DES.

These two searches demonstrated the efficiency of our pipeline to find, deblend, and model lenses in future large-scale surveys. Some improvements could, however, be necessary for the following lens searches. In particular, we must sustain efforts to decrease the false positive rate. It can be done, for example, by retraining the CNNs with the false positives found during the first iterations of the search or by combining the decision of different algorithms at the classification step.

9 Conclusion

Strong gravitational lensing is an incredible tool for a large number of applications. The main goal of this thesis is to present an adaptable pipeline for the automated detection of galaxy-galaxy gravitational lenses in ongoing and future large-scale imaging surveys. As we have seen, several significant challenges have to be overcome to build an automated lens-finding pipeline. In particular, this thesis addresses lengthy visual inspection processes, deblending issues, and realistic simulation of lenses.

The occurrence rate of strongly lensed systems is low. As a consequence, the set of known lenses is still limited. Machine learning algorithms, particularly CNNs, require large datasets to be trained. Hence, simulations of lenses must be used to train such algorithms. Mock lenses must always be used with care as CNNs can learn simulation artifacts. In this thesis, we presented a tool to simulate realistic lenses. The main difference with previous simulation software is that only the lensing effect is simulated with this tool since the source and deflector images are drawn from real images. Therefore, we limit the risk of creating synthetic features in our simulations.

The low occurrence rate of lenses has another critical consequence. Because of the base rate fallacy effect, reaching a low false positive rate on real data is extremely difficult. Follow-up observations of lens candidates require expensive telescope time. Even if we can reduce the number of false positives by improving the training methods, a visual inspection of the candidates will still be necessary for the upcoming lens searches. Therefore, we developed tools to inspect large sets of astronomical images. These tools enable us to speed up the process and limit the biases due to the use of different visualization software by different users. In the context of the visual inspection of CFIS and DES lens candidates, I contributed to drafting general guidelines that can be used to grade candidates in the future lens database.

The light of the lensed source is often blended with the light of the deflector. Therefore, we included a deblending module in our lens finding pipeline. On this occasion, I developed a new method to deblend lenses based on neural networks. This method does not rely on analytical light profiles and thus can be adapted to all kinds of deflectors and sources. It

deblended successfully the 32 SL candidates we found in the initial 2 500 square degrees of CFIS. In addition to highlighting the faint lensing structures and enabling the measurement of photometric redshifts, we can use it to initialize the lens model parameters. In the future, we may extend the applications of the auto-encoder deblending. Since it is very flexible and easy to train, it can also potentially be adapted to provide help to detect the lenses. For example, the auto-encoder-based deblending may be used as a pre-processing step before classification to highlight the potential lensing features, similar to the PCA-based deblending algorithm. Since it is a neural network-based method, a modified version could eventually even deblend and classify the images simultaneously.

The pipeline was applied to the search of galaxies lensed by LRGs in 2 500 deg² of CFIS and the first data release of DES. It allowed us to find 133 and 405 high-quality lens candidates in CFIS and DES, respectively. These candidates will be confirmed with follow-up spectroscopic observations. When confirmed, these lenses can be used, for example, to study the distribution of dark matter in the deflector. The CFIS candidates are monitored to detect the appearance of a lensed SNIa, which could allow us to measure the Hubble constant. In addition to the lens candidates, I contributed to building catalogs of the most common false positives that can be used to retrain the lens-finding algorithms. These sets of false positives can also serve to train GANs to produce larger sets of non-lens examples. The next step is to apply the pipeline to the entire CFIS survey.

I demonstrated in this thesis the ability of our automated pipeline to find, deblend, and model lenses in future large-scale surveys. In view of future large-scale surveys, we must sustain efforts to decrease the false positive rate to reduce the importance of visual inspection. It can be done, for example, by retraining the CNNs on the false positives found in the previous searches and using the information obtained with the deblending and modeling to refine the classification score obtained with the CNNs. Combining the strengths of the various methods to detect lenses may also improve the performance of the classification.

A Galaxy Image Simulation Using Progressive GANs

I present here an additional publication I contributed to during my thesis. It introduces a new data-driven method to generate mock spiral galaxy images. The method is based on GANs that trained with progressively larger images in order to improve the quality of the generated images. This method can produce large samples of self-similar images and could therefore be used in the future to produce non-lens examples for the training sets of CNNs-based lens finders. The content of this publication was presented in the form of a poster at the Astronomical Data Analysis Software and Systems (ADASS) conference 2019.

Galaxy Image Simulation Using Progressive GANs

Mohamad Dia,¹ Elodie Savary,² Martin Melchior,¹ and Frederic Courbin²

¹*Institute for Data Science, University of Applied Sciences North Western Switzerland (FHNW), 5210 Windisch, Switzerland;*

mohamad.dia@fhnw.ch, martin.melchior@fhnw.ch

²*Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland;*

elodie.savary@epfl.ch, frederic.courbin@epfl.ch

Abstract. In this work, we provide an efficient and realistic data-driven approach to simulate astronomical images using deep generative models from machine learning. Our solution is based on a variant of the generative adversarial network (GAN) with progressive training methodology and Wasserstein cost function. The proposed solution generates naturalistic images of galaxies that show complex structures and high diversity, which suggests that data-driven simulations using machine learning can replace many of the expensive model-driven methods used in astronomical data processing.

1. Introduction

Investigating the reasons behind the accelerated expansion of the universe is one of the main challenges in astronomy and modern cosmology. Future space missions, such as *Euclid*, will provide images of billions of galaxies in order to investigate the so-called *dark matter* and probe the geometry of the universe through the *gravitational lensing* effect. Due to the very large-scale of data provided by such missions, automated algorithms are needed for measurement and detection purposes. The training and calibration of such algorithms require simulated, or synthetic, images of galaxies that mimic the real observations and exhibit real morphologies.

In the case of weak lensing for instance, the accuracy of the shape measurement algorithms is very sensitive to any statistical bias induced by the Point Spread Function (PSF). Therefore, simulated images of galaxies with known ground-truth lensing are required to calibrate and detect any potential bias in the ensemble averages. Moreover, the training of automated strong lensing detector, such as deep learning architectures (Metcalf et al. 2019), requires simulated images in order to mitigate class imbalance and avoid false-positive type of error in the current datasets.

2. Model-Driven v.s. Data-Driven Galaxy Image Simulation

The current approaches to simulate images of galaxies in the cosmology literature are mostly model-driven, or rule-based, approaches. These might involve the fitting of parametric analytic profiles (size, ellipticity, brightness, etc.) to the observed galax-

ies. This approach is usually unable to reproduce all the complex morphologies. An alternative, more expensive and often infeasible, model-driven approach is to start with high-quality galaxy images as the input of the simulation pipeline followed by a model that reproduces all the data acquisition effects (Rowe et al. 2015).

Recently, several data-driven approaches have been investigated in order to generate synthetic images of galaxies via generative models used in machine learning (Regier et al. 2015; Ravanbakhsh et al. 2017), mainly variational autoencoder (VAE) (Kingma & Welling 2013) and generative adversarial network (GAN) (Goodfellow et al. 2014). Such approaches have shown some promising preliminary results in generating galaxy images. Following this data-driven approach, and motivated by the success and recent impressive improvements in GANs, we have further investigated the use of such architecture in generating galaxy images.

3. Generative Adversarial Network

Unlike most of the generative models used in machine learning, GAN represents a novel approach that learns how to sample from the data distribution without explicitly tracking the parameters of the probability distribution function via traditional maximum likelihood estimation. The GAN architecture consists of two neural networks that compete against each other in a two-player minimax game. The first network is the “generator” that is responsible of generating the data, while the second network is the “discriminator” that represents the adversarial loss function. Despite its elegant mathematical formulation and the theoretical guarantees provided by a non-parametric analysis, the initial GAN architecture suffered from some practical implementation problems.

After the invention of GAN in 2014, a plethora of work have been done to improve the training (in terms of convergence and stability) and to obtain more realistic generated data (in terms of quality and diversity). Most of this effort was made towards improving the cost function and stabilizing the training methodology, which has recently lead to unprecedented results in generating synthetic images. Based on these recent advances, we have investigated variants of GAN that use the Wasserstein distance (Gulrajani et al. 2017) and the progressive training (Karras et al. 2018) on galaxy images provided by the Galaxy-Zoo dataset (Willett et al. 2013).

4. Proposed Architecture

Following (Karras et al. 2018), we employ blocks of convolutional layers to progressively build the generator and the discriminator as mirror images of each other (see Table 1). Intuitively speaking, training a small network to generate low-resolution images that capture the large-scale structure of the galaxies is an easier task than directly training a full network to generate high-resolution images with fine details. Hence, we start by training the network to generate low-resolution images (4×4), we then progressively increase the resolution, in 4 steps until 64×64 resolution, by *smoothly* and synchronously adding blocks of convolutional layers to both the generator and discriminator. For the generator, each progress block is preceded by an up-sampling operation

while a down-sampling operation follows each progress block in the discriminator.¹ Such methodology leads to a more stable and faster training.

Table 1. Blocks of convolutional layers added progressively for both the generator and the discriminator.

Generator	Output Dimensions	Discriminator	Output Dimensions
Latent Space		RGB Reading	
Input latent vector	$1 \times 1 \times 256$	Input image	$64 \times 64 \times 3$
Conv 4×4	$4 \times 4 \times 256$	Conv 1×1	$64 \times 64 \times 16$
Conv 3×3	$4 \times 4 \times 256$		
1 st Progress		4 th Progress	
Conv 3×3	$8 \times 8 \times 128$	Conv 3×3	$64 \times 64 \times 16$
Conv 3×3	$8 \times 8 \times 128$	Conv 3×3	$64 \times 64 \times 32$
2 nd Progress		3 rd Progress	
Conv 3×3	$16 \times 16 \times 64$	Conv 3×3	$32 \times 32 \times 32$
Conv 3×3	$16 \times 16 \times 64$	Conv 3×3	$32 \times 32 \times 64$
3 rd Progress		2 nd Progress	
Conv 3×3	$32 \times 32 \times 32$	Conv 3×3	$16 \times 16 \times 64$
Conv 3×3	$32 \times 32 \times 32$	Conv 3×3	$16 \times 16 \times 128$
4 th Progress		1 st Progress	
Conv 3×3	$64 \times 64 \times 16$	Conv 3×3	$8 \times 8 \times 128$
Conv 3×3	$64 \times 64 \times 16$	Conv 3×3	$8 \times 8 \times 256$
RGB Extraction		Cost Calculation	
Conv 1×1	$64 \times 64 \times 3$	Conv 3×3	$4 \times 4 \times 256$
		Conv 4×4	$1 \times 1 \times 256$
		Conv 1×1	$1 \times 1 \times 1$

Moreover, the Wasserstein distance with gradient penalty (Gulrajani et al. 2017) is used as a cost function to mitigate the gradient problems. Furthermore, various normalization techniques are used to avoid the unhealthy competition between the generator and discriminator. In particular, we use “weight scaling” and “pixelwise feature normalization” as done in (He et al. 2015; Krizhevsky et al. 2012). In addition to that, the “mini-batch standard deviation” (Karras et al. 2018) is computed and incorporated in the cost function in order to favor diversity in the synthetic data.

4.1. Results

Our architecture is implemented in Python using PyTorch library and trained on a GPU system. The dataset is made of 6157 images of galaxies in RGB format. The images were centered at 64×64 resolution, normalized, and augmented using standard data augmentation techniques. A batch size 16 was used with 8 data loading workers.

The training was performed over a total of 100 epochs and lasted less than 24 hours. During the first 40 epochs of training, the generator and discriminator were competing to reach the minimax equilibrium and the performance was fluctuating (in terms of their loss functions). The performance stabilized after that while the image quality continued to improve. After training, the discriminator, which plays the role of an adaptive loss function, is detached from the architecture and dismissed. The generator is then able to generate galaxy images starting from a latent space made of 256 standard Gaussian i.i.d. random variables.

By changing the latent vector, we were able to obtain very diverse and high quality images of galaxies showing complex structures and morphologies (e.g. arm and disk

¹One can also use fractionally-strided and strided convolution respectively.

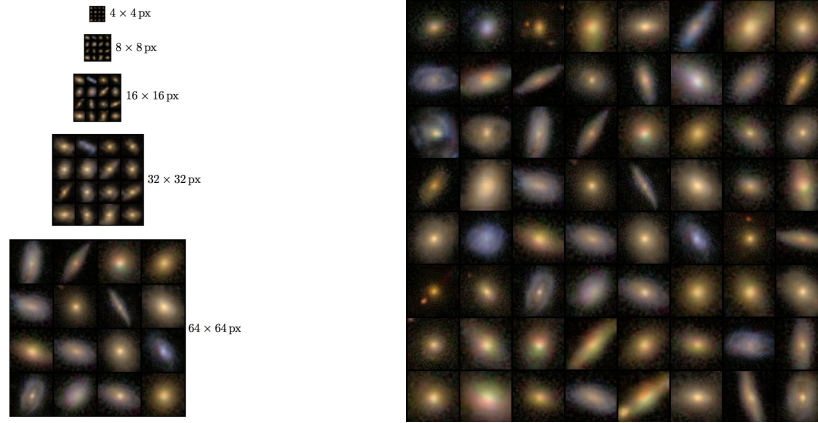


Figure 1. *Left:* Progressive increase of resolution in four steps. *Right:* Diverse set of simulated images obtained by the proposed GAN architecture.

features). Furthermore, the simulated images exhibited realistic effects (e.g. companion stars) as shown in Figure 1.

5. Future Work

We are planning to investigate the latent space of our GAN model in order to gain insight on the effect of each latent variable on the galaxies morphology. This will provide us with more control on the generation task and will permit to interpolate between the variables and perform latent space arithmetics. Furthermore, we are planning to incorporate the labels of the galaxies, when available, in a supervised or semi-supervised approach using variants of "Conditional GAN" architectures (Odena et al. 2017) in order to improve the quality of the generated images and guide the generator.

Acknowledgments. M. Dia and E. Savary would like to acknowledge funding from the SNSF (grant number 173716).

References

- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. 2014, in *NeurIPS*, 2672–2680
- Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. 2017, in *NeurIPS*, 5769–5779
- He, K., Zhang, X., Ren, S., & Sun, J. 2015, in *ICCV*, 1026–1034
- Karras, T., Aila, T., Laine, S., & Lehtinen, J. 2018, in *ICLR*
- Kingma, D. P., & Welling, M. 2013, in *ICLR*
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012, in *NeurIPS*, 1097–1105
- Metcalfe, R. B., et al. 2019, *A&A*, 625, A119
- Odena, A., Olah, C., & Shlens, J. 2017, in *ICML*, vol. 70, 2642–2651
- Ravanbakhsh, S., Lanusse, F., Mandelbaum, R., Schneider, J. G., & Póczos, B. 2017, in *AAAI*
- Regier, J., Miller, A., McAuliffe, J., Adams, R., Hoffman, M., Lang, D., Schlegel, D., & Prabhat, P. 2015, in *ICML*, vol. 37, 2095 – 2103
- Rowe, B. T. P., et al. 2015, *Astronomy and Computing*, 10, 121
- Willett, K. W., et al. 2013, *Monthly Notices of the Royal Astronomical Society*, 435, 2835

Bibliography

- Jacobs, C., Collett, T., Glazebrook, K., McCarthy, C., Qin, A. K., Abbott, T. M. C., Abdalla, F. B., Annis, J., Avila, S., Bechtol, K., Bertin, E., Brooks, D., Buckley-Geer, E., Burke, D. L., Carnero Rosell, A., Carrasco Kind, M., Carretero, J., da Costa, L. N., Davis, C., ... Zuntz, J. (2019a). Finding high-redshift strong lenses in DES using convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 484(4), 5330–5349. <https://doi.org/10.1093/mnras/stz272>
- Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (n.d.). Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, 2012.
- Jacobs, C., Collett, T., Glazebrook, K., Buckley-Geer, E., Diehl, H. T., Lin, H., McCarthy, C., Qin, A. K., Odden, C., Caso Escudero, M., Dial, P., Yung, V. J., Gaitsch, S., Pellico, A., Lindgren, K. A., Abbott, T. M. C., Annis, J., Avila, S., Brooks, D., ... DES Collaboration. (n.d.). An Extended Catalog of Galaxy-Galaxy Strong Gravitational Lenses Discovered in DES Using Convolutional Neural Networks., 243(1), Article 17, 17. <https://doi.org/10.3847/1538-4365/ab26b6>
- Bom, C. R., Fraga, B. M. O., Dias, L. O., Schubert, P., Blanco Valentin, M., Furlanetto, C., Makler, M., Teles, K., Portes de Albuquerque, M., & Metcalf, R. B. (2022). Developing a Victorious Strategy to the Second Strong Gravitational Lensing Data Challenge. *arXiv e-prints*, Article arXiv:2203.09536, arXiv:2203.09536.
- Moreno, E. A., Borzyszkowski, B., Pierini, M., Vlimant, J.-R., & Spiropulu, M. (2022). Source-agnostic gravitational-wave detection with recurrent autoencoders. *Machine Learning: Science and Technology*, 3(2), Article 025001, 025001. <https://doi.org/10.1088/2632-2153/ac5435>
- Vanzella, E., Castellano, M., Bergamini, P., Meneghetti, M., Zanella, A., Calura, F., Caminha, G. B., Rosati, P., Cupani, G., Meštrić, U., Brammer, G., Tozzi, P., Mercurio, A., Grillo, C., Sani, E., Cristiani, S., Nonino, M., Merlin, E., & Pignataro, G. V. (2022). High star cluster formation efficiency in the strongly lensed Sunburst Lyman-continuum galaxy at $z = 2.37$., 659, Article A2, A2. <https://doi.org/10.1051/0004-6361/202141590>
- Welch, B., Coe, D., Diego, J. M., Zitrin, A., Zackrisson, E., Dimauro, P., Jiménez-Teja, Y., Kelly, P., Mahler, G., Oguri, M., Timmes, F. X., Windhorst, R., Florian, M., de Mink, S. E., Avila, R. J., Anderson, J., Bradley, L., Sharon, K., Vikaeus, A., ... Broadhurst, T. (2022). A highly magnified star at redshift 6.2., 603(7903), 815–818. <https://doi.org/10.1038/s41586-022-04449-y>

- Wilde, J., Serjeant, S., Bromley, J. M., Dickinson, H., Koopmans, L. V. E., & Metcalf, R. B. (2022). Detecting gravitational lenses using machine learning: exploring interpretability and sensitivity to rare lensing configurations., *512*(3), 3464–3479. <https://doi.org/10.1093/mnras/stac562>
- Arcelin, B., Doux, C., Aubourg, E., Roucelle, C., & LSST Dark Energy Science Collaboration. (2021). Deblending galaxies with variational autoencoders: A joint multiband, multi-instrument approach., *500*(1), 531–547. <https://doi.org/10.1093/mnras/staa3062>
- Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). Emerging properties in self-supervised vision transformers. *CoRR*, *abs/2104.14294*. <https://arxiv.org/abs/2104.14294>
- Gentile, F., Tortora, C., Covone, G., Koopmans, L. V. E., Spiniello, C., Fan, Z., Li, R., Liu, D., Napolitano, N. R., Vaccari, M., & Fu, L. (2021). Lenses In VoicE (LIVE): Searching for strong gravitational lenses in the VOICE@VST survey using Convolutional Neural Networks. *arXiv e-prints*, Article arXiv:2105.05602, arXiv:2105.05602.
- Huang, X., Storfer, C., Gu, A., Ravi, V., Pilon, A., Sheu, W., Venguswamy, R., Banka, S., Dey, A., Landriau, M., Lang, D., Meisner, A., Moustakas, J., Myers, A. D., Sajith, R., Schlafly, E. F., & Schlegel, D. J. (2021). Discovering New Strong Gravitational Lenses in the DESI Legacy Imaging Surveys., *909*(1), Article 27, 27. <https://doi.org/10.3847/1538-4357/abd62b>
- Khullar, G., Gozman, K., Lin, J. J., Martinez, M. N., Matthews Acuña, O. S., Medina, E., Merz, K., Sanchez, J. A., Sisco, E. E., Kavin Stein, D. J., Sukay, E. O., Tavangar, K., Bayliss, M. B., Bleem, L. E., Brownsberger, S., Dahle, H., Florian, M. K., Gladders, M. D., Mahler, G., ... Stark, A. A. (2021). COOL-LAMPS. I. An Extraordinarily Bright Lensed Galaxy at Redshift 5.04., *906*(2), Article 107, 107. <https://doi.org/10.3847/1538-4357/abcb86>
- Lanusse, F., Mandelbaum, R., Ravanbakhsh, S., Li, C.-L., Freeman, P., & Póczos, B. (2021). Deep generative models for galaxy image simulations., *504*(4), 5543–5555. <https://doi.org/10.1093/mnras/stab1214>
- Li, R., Napolitano, N. R., Spiniello, C., Tortora, C., Kuijken, K., Koopmans, L. V. E., Schneider, P., Getman, F., Xie, L., Long, L., Shu, W., Vernardos, G., Huang, Z., Covone, G., Dvornik, A., Heymans, C., Hildebrandt, H., Radovich, M., & Wright, A. H. (2021). High-quality Strong Lens Candidates in the Final Kilo-Degree Survey Footprint., *923*(1), Article 16, 16. <https://doi.org/10.3847/1538-4357/ac2df0>
- Meneghetti, M. (2021). *Introduction to gravitational lensing* (1st ed.). Springer Nature.
- Morawski, F., Bejger, M., Cuoco, E., & Petre, L. (2021). Anomaly Detection in Gravitational Waves data using Convolutional AutoEncoders. *arXiv e-prints*, Article arXiv:2103.07688, arXiv:2103.07688.
- Rojas, K., Savary, E., Clément, B., Maus, M., Courbin, F., Lemon, C., Chan, J. H. H., Vernardos, G., Joseph, R., Cañameras, R., & Galan, A. (2021). Strong lens systems search in the Dark Energy Survey using Convolutional Neural Networks. *arXiv e-prints*, Article arXiv:2109.00014, arXiv:2109.00014.
- Savary, E., Rojas, K., Maus, M., Clément, B., Courbin, F., Gavazzi, R., Chan, J. H. H., Lemon, C., Vernardos, G., Cañameras, R., Schuldt, S., Suyu, S. H., Cuillandre, J. -, Fabbro, S.,

- Gwyn, S., Hudson, M. J., Kilbinger, M., Scott, D., & Stone, C. (2021). A search for galaxy-scale strong gravitational lenses in the Ultraviolet Near Infrared Optical Northern Survey (UNIONS). *arXiv e-prints*, Article arXiv:2110.11972, arXiv:2110.11972.
- Stein, G., Blaum, J., Harrington, P., Medan, T., & Lukic, Z. (2021). Mining for strong gravitational lenses with self-supervised learning. *arXiv e-prints*, Article arXiv:2110.00023, arXiv:2110.00023.
- Brabec, J., Komárek, T., Franc, V., & Machlica, L. (2020). On Model Evaluation under Non-constant Class Imbalance. *arXiv e-prints*, Article arXiv:2001.05571, arXiv:2001.05571.
- Cañameras, R., Schuldt, S., Suyu, S. H., Taubenberger, S., Meinhardt, T., Leal-Taixé, L., Lemon, C., Rojas, K., & Savary, E. (2020). HOLISMOKES. II. Identifying galaxy-scale strong gravitational lenses in Pan-STARRS using convolutional neural networks., *644*, Article A163, A163. <https://doi.org/10.1051/0004-6361/202038219>
- Cheng, T.-Y., Li, N., Conselice, C. J., Aragón-Salamanca, A., Dye, S., & Metcalf, R. B. (2020). Identifying strong lenses with unsupervised machine learning using convolutional autoencoder., *494*(3), 3750–3765. <https://doi.org/10.1093/mnras/staa1015>
- Collett, T. E., & Smith, R. J. (2020). A triple rollover: a third multiply imaged source at $z \approx 6$ behind the Jackpot gravitational lens., *497*(2), 1654–1660. <https://doi.org/10.1093/mnras/staa1804>
- Huang, X., Storfer, C., Ravi, V., Pilon, A., Domingo, M., Schlegel, D. J., Bailey, S., Dey, A., Gupta, R. R., Herrera, D., Juneau, S., Landriau, M., Lang, D., Meisner, A., Moustakas, J., Myers, A. D., Schlafly, E. F., Valdes, F., Weaver, B. A., ... Yèche, C. (2020). Finding Strong Gravitational Lenses in the DESI DECam Legacy Survey., *894*(1), Article 78, 78. <https://doi.org/10.3847/1538-4357/ab7ffb>
- Li, R., Napolitano, N. R., Tortora, C., Spiniello, C., Koopmans, L. V. E., Huang, Z., Roy, N., Vernardos, G., Chatterjee, S., Giblin, B., Getman, F., Radovich, M., Covone, G., & Kuijken, K. (2020). New High-quality Strong Lens Candidates with Deep Learning in the Kilo-Degree Survey., *899*(1), Article 30, 30. <https://doi.org/10.3847/1538-4357/ab9dfa>
- Planck Collaboration, Aghanim, N., Akrami, Y., Arroja, F., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A. J., Barreiro, R. B., Bartolo, N., Basak, S., Battye, R., Benabed, K., Bernard, J. -, Bersanelli, M., Bielewicz, P., Bock, J. J., Bond, J. R., ... Zonca, A. (2020). Planck 2018 results. I. Overview and the cosmological legacy of Planck., *641*, Article A1, A1. <https://doi.org/10.1051/0004-6361/201833880>
- Sonnenfeld, A., Verma, A., More, A., Baeten, E., Macmillan, C., Wong, K. C., Chan, J. H. H., Jaelani, A. T., Lee, C.-H., Oguri, M., Rusu, C. E., Veldhuis, M., Trouille, L., Marshall, P. J., Hutchings, R., Allen, C., O'Donnell, J., Cornen, C., Davis, C. P., ... Miller, G. (2020). Survey of Gravitationally-lensed Objects in HSC Imaging (SuGOHI). VI. Crowdsourced lens finding with Space Warps., *642*, Article A148, A148. <https://doi.org/10.1051/0004-6361/202038067>
- Suyu, S. H., Huber, S., Cañameras, R., Kromer, M., Schuldt, S., Taubenberger, S., Yıldırım, A., Bonvin, V., Chan, J. H. H., Courbin, F., Nöbauer, U., Sim, S. A., & Sluse, D. (2020). HOLISMOKES. I. Highly Optimised Lensing Investigations of Supernovae, Microlens-

- ing Objects, and Kinematics of Ellipticals and Spirals., 644, Article A162, A162. <https://doi.org/10.1051/0004-6361/202037757>
- Wong, K. C., Suyu, S. H., Chen, G. C. .-, Rusu, C. E., Millon, M., Sluse, D., Bonvin, V., Fassnacht, C. D., Taubenberger, S., Auger, M. W., Birrer, S., Chan, J. H. H., Courbin, F., Hilbert, S., Tihhonova, O., Treu, T., Agnello, A., Ding, X., Jee, I., . . . Meylan, G. (2020). H0LiCOW - XIII. A 2.4 per cent measurement of H_0 from lensed quasars: 5.3σ tension between early- and late-Universe probes., 498(1), 1420–1439. <https://doi.org/10.1093/mnras/stz3094>
- Aihara, H., AlSayyad, Y., Ando, M., Armstrong, R., Bosch, J., Egami, E., Furusawa, H., Furusawa, J., Goulding, A., Harikane, Y., Hikage, C., Ho, P. T. P., Hsieh, B.-C., Huang, S., Ikeda, H., Imanishi, M., Ito, K., Iwata, I., Jaelani, A. T., . . . Yamada, Y. (2019). Second data release of the Hyper Suprime-Cam Subaru Strategic Program., 71(6), Article 114, 114. <https://doi.org/10.1093/pasj/psz103>
- Bellm, E. C., Kulkarni, S. R., Graham, M. J., Dekany, R., Smith, R. M., Riddle, R., Masci, F. J., Helou, G., Prince, T. A., Adams, S. M., Barbarino, C., Barlow, T., Bauer, J., Beck, R., Belicki, J., Biswas, R., Blagorodnova, N., Bodewits, D., Bolin, B., . . . Zolkower, J. (2019). The Zwicky Transient Facility: System Overview, Performance, and First Results., 131(995), 018002. <https://doi.org/10.1088/1538-3873/aaecbe>
- Bonvin, V., Tihhonova, O., Millon, M., Chan, J. H. .-, Savary, E., Huber, S., & Courbin, F. (2019). Impact of the 3D source geometry on time-delay measurements of lensed type-Ia supernovae., 621, Article A55, A55. <https://doi.org/10.1051/0004-6361/201833405>
- Dey, A., Schlegel, D. J., Lang, D., Blum, R., Burleigh, K., Fan, X., Findlay, J. R., Finkbeiner, D., Herrera, D., Juneau, S., Landriau, M., Levi, M., McGreer, I., Meisner, A., Myers, A. D., Moustakas, J., Nugent, P., Patej, A., Schlafly, E. F., . . . Zhou, Z. (2019). Overview of the DESI Legacy Imaging Surveys., 157(5), Article 168, 168. <https://doi.org/10.3847/1538-3881/ab089d>
- Dia, M., Savary, E., Melchior, M., & Courbin, F. (2019). Galaxy image simulation using progressive gans.
- Fan, X., Wang, F., Yang, J., Keeton, C. R., Yue, M., Zabludoff, A., Bian, F., Bonaglia, M., Georgiev, I. Y., Hennawi, J. F., Li, J., McGreer, I. D., Naidu, R., Pacucci, F., Rabien, S., Thompson, D., Venemans, B., Walter, F., Wang, R., & Wu, X.-B. (2019). The Discovery of a Gravitationally Lensed Quasar at $z = 6.51$., 870(2), Article L11, L11. <https://doi.org/10.3847/2041-8213/aaeffe>
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., Abel, B., Acosta, E., Allsman, R., Alonso, D., AlSayyad, Y., Anderson, S. F., Andrew, J., Angel, J. R. P., Angeli, G. Z., Ansari, R., Antilogus, P., Araujo, C., Armstrong, R., Arndt, K. T., Astier, P., Aubourg, É., . . . Zhan, H. (2019). LSST: From Science Drivers to Reference Design and Anticipated Data Products., 873(2), Article 111, 111. <https://doi.org/10.3847/1538-4357/ab042c>
- Lu, L., Shin, Y., Su, Y., & Karniadakis, G. E. (2019). Dying relu and initialization: theory and numerical examples. *ArXiv, abs/1903.06733*.
- Masci, F. J., Laher, R. R., Rusholme, B., Shupe, D. L., Groom, S., Surace, J., Jackson, E., Monkewitz, S., Beck, R., Flynn, D., Terek, S., Landry, W., Hacopians, E., Desai, V., Howell, J., Brooke,

- T., Imel, D., Wachter, S., Ye, Q.-Z., ... Kulkarni, S. R. (2019). The Zwicky Transient Facility: Data Processing, Products, and Archive., *131*(995), 018003. <https://doi.org/10.1088/1538-3873/aae8ac>
- Metcalf, R. B., Meneghetti, M., Avestruz, C., Bellagamba, E., Bom, C. R., Bertin, E., Cabanac, R., Courbin, F., Davies, A., Decenci re, E., Flamary, R., Gavazzi, R., Geiger, M., Hartley, P., Huertas-Company, M., Jackson, N., Jacobs, C., Jullo, E., Kneib, J. .-, ... Vernardos, G. (2019). The strong gravitational lens finding challenge., *625*, Article A119, A119. <https://doi.org/10.1051/0004-6361/201832797>
- Petrillo, C. E., Tortora, C., Vernardos, G., Koopmans, L. V. E., Verdoes Kleijn, G., Bilicki, M., Napolitano, N. R., Chatterjee, S., Covone, G., Dvornik, A., Erben, T., Getman, F., Giblin, B., Heymans, C., de Jong, J. T. A., Kuijken, K., Schneider, P., Shan, H., Spiniello, C., & Wright, A. H. (2019). LinKS: discovering galaxy-scale strong lenses in the Kilo-Degree Survey using convolutional neural networks., *484*(3), 3879–3896. <https://doi.org/10.1093/mnras/stz189>
- Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., & Scolnic, D. (2019). Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond Λ CDM., *876*(1), Article 85, 85. <https://doi.org/10.3847/1538-4357/ab1422>
- Ritondale, E., Vegetti, S., Despali, G., Auger, M. W., Koopmans, L. V. E., & McKean, J. P. (2019). Low-mass halo perturbations in strong gravitational lenses at redshift $z \sim 0.5$ are consistent with CDM., *485*(2), 2179–2193. <https://doi.org/10.1093/mnras/stz464>
- Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv e-prints*, Article arXiv:1905.11946, arXiv:1905.11946.
- Birrer, S., & Amara, A. (2018). lenstronomy: Multi-purpose gravitational lens modelling software package. *Physics of the Dark Universe*, *22*, 189–201. <https://doi.org/10.1016/j.dark.2018.11.002>
- Chandra X-ray observatory. (2018). RxJ1131-1231 [<http://chandra.harvard.edu/photo/2014/rxj1131/>, accessed on 2018-07-12].
- Chatterjee, S., & Koopmans, L. V. E. (2018). The inner mass power spectrum of galaxies using strong gravitational lensing: beyond linear approximation., *474*(2), 1762–1772. <https://doi.org/10.1093/mnras/stx2674>
- Lanusse, F., Ma, Q., Li, N., Collett, T. E., Li, C.-L., Ravanbakhsh, S., Mandelbaum, R., & P czos, B. (2018). CMU DeepLens: deep learning for automatic image-based galaxy-galaxy strong lens finding., *473*(3), 3895–3906. <https://doi.org/10.1093/mnras/stx1665>
- Melchior, P., Moolekamp, F., Jerdee, M., Armstrong, R., Sun, A. .-, Bosch, J., & Lupton, R. (2018). SCARLET: Source separation in multi-band images by Constrained Matrix Factorization. *Astronomy and Computing*, *24*, Article 129, 129. <https://doi.org/10.1016/j.ascom.2018.07.001>
- Petrillo, Tortora, C., Chatterjee, S., Vernardos, G., Koopmans, L., Verdoes Kleijn, G., Napolitano, N., Covone, G., Kelvin, L., & Hopkins, A. (2018). Testing convolutional neural networks for finding strong gravitational lenses in kids.

- Pourrahmani, M., Nayyeri, H., & Cooray, A. (2018). LensFlow: A Convolutional Neural Network in Search of Strong Gravitational Lenses., *856*(1), Article 68, 68. <https://doi.org/10.3847/1538-4357/aaae6a>
- Schaefer, C., Geiger, M., Kuntzer, T., & Kneib, J. -. (2018). Deep convolutional neural networks as strong gravitational lens detectors., *611*, Article A2, A2. <https://doi.org/10.1051/0004-6361/201731201>
- Vegetti, S., Despali, G., Lovell, M. R., & Enzi, W. (2018). Constraining sterile neutrino cosmologies with strong gravitational lensing observations at redshift $z \sim 0.2$., *481*(3), 3661–3669. <https://doi.org/10.1093/mnras/sty2393>
- Bellagamba, F., Tessore, N., & Metcalf, R. B. (2017). Zooming into the Cosmic Horseshoe: new insights on the lens profile and the source shape., *464*(4), 4823–4834. <https://doi.org/10.1093/mnras/stw2726>
- Freedman, W. L. (2017). Cosmology at a crossroads. *Nature Astronomy*, *1*, Article 0121, 0121. <https://doi.org/10.1038/s41550-017-0121>
- Frontera-Pons, J., Sureau, F., Bobin, J., & Le Floch, E. (2017). Unsupervised feature-learning for galaxy SEDs with denoising autoencoders., *603*, Article A60, A60. <https://doi.org/10.1051/0004-6361/201630240>
- Goobar, A., Amanullah, R., Kulkarni, S. R., Nugent, P. E., Johansson, J., Steidel, C., Law, D., Mörtzell, E., Quimby, R., Blagorodnova, N., Brandeker, A., Cao, Y., Cooray, A., Ferretti, R., Fremling, C., Hangard, L., Kasliwal, M., Kupfer, T., Lunnan, R., ... Yaron, O. (2017). iPTF16geu: A multiply imaged, gravitationally lensed type Ia supernova. *Science*, *356*, 291–295. <https://doi.org/10.1126/science.aal2729>
- Hartley, P., Flamary, R., Jackson, N., Tagore, A. S., & Metcalf, R. B. (2017). Support vector machine classification of strong gravitational lenses., *471*(3), 3378–3397. <https://doi.org/10.1093/mnras/stx1733>
- Jacobs, C., Glazebrook, K., Collett, T., More, A., & McCarthy, C. (2017). Finding strong lenses in CFHTLS using convolutional neural networks., *471*(1), 167–181. <https://doi.org/10.1093/mnras/stx1492>
- Johnson, T. L., Rigby, J. R., Sharon, K., Gladders, M. D., Florian, M., Bayliss, M. B., Wuyts, E., Whitaker, K. E., Livermore, R., & Murray, K. T. (2017). Star Formation at $z = 2.481$ in the Lensed Galaxy SDSS J1110+6459: Star Formation Down to 30 pc Scales., *843*(2), Article L21, L21. <https://doi.org/10.3847/2041-8213/aa7516>
- Ostrovski, F., McMahon, R. G., Connolly, A. J., Lemon, C. A., Auger, M. W., Banerji, M., Hung, J. M., Koposov, S. E., Lidman, C. E., Reed, S. L., Allam, S., Benoit-Lévy, A., Bertin, E., Brooks, D., Buckley-Geer, E., Carnero Rosell, A., Carrasco Kind, M., Carretero, J., Cunha, C. E., ... Walker, A. R. (2017). VDES J2325-5229 a $z = 2.7$ gravitationally lensed quasar discovered using morphology-independent supervised machine learning., *465*(4), 4325–4334. <https://doi.org/10.1093/mnras/stw2958>
- Petrillo, Tortora, C., Chatterjee, S., Vernardos, G., Koopmans, L. V. E., Verdoes Kleijn, G., Napolitano, N. R., Covone, G., Schneider, P., Grado, A., & McFarland, J. (2017). Finding strong gravitational lenses in the Kilo Degree Survey with Convolutional Neural Networks., *472*(1), 1129–1150. <https://doi.org/10.1093/mnras/stx2052>

- Suyu, S. H., Bonvin, V., Courbin, F., Fassnacht, C. D., Rusu, C. E., Sluse, D., Treu, T., Wong, K. C., Auger, M. W., Ding, X., Hilbert, S., Marshall, P. J., Rumbaugh, N., Sonnenfeld, A., Tewes, M., Tihhonova, O., Agnello, A., Blandford, R. D., Chen, G. C. .-. . . Spiniello, C. (2017). H0LiCOW - I. H_0 Lenses in COSMOGRAIL's Wellspring: program overview., *468*(3), 2590–2604. <https://doi.org/10.1093/mnras/stx483>
- Willett, K. W., Galloway, M. A., Bamford, S. P., Lintott, C. J., Masters, K. L., Scarlata, C., Simmons, B. D., Beck, M., Cardamone, C. N., Cheung, E., Edmondson, E. M., Fortson, L. F., Griffith, R. L., Häußler, B., Han, A., Hart, R., Melvin, T., Parrish, M., Schawinski, K., . . . Smith, A. M. (2017). Galaxy Zoo: morphological classifications for 120 000 galaxies in HST legacy imaging., *464*(4), 4176–4203. <https://doi.org/10.1093/mnras/stw2568>
- Chambers, K. C., Magnier, E. A., Metcalfe, N., Flewelling, H. A., Huber, M. E., Waters, C. Z., Denneau, L., Draper, P. W., Farrow, D., Finkbeiner, D. P., Holmberg, C., Koppenhoefer, J., Price, P. A., Rest, A., Saglia, R. P., Schlafly, E. F., Smartt, S. J., Sweeney, W., Wainscoat, R. J., . . . Wyse, R. (2016). The Pan-STARRS1 Surveys. *arXiv e-prints*, Article arXiv:1612.05560, arXiv:1612.05560.
- Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., Aleksić, J., Allam, S., Amara, A., Bacon, D., Balbinot, E., Banerji, M., Bechtol, K., Benoit-Lévy, A., Bernstein, G. M., Bertin, E., Blazek, J., Bonnett, C., Bridle, S., Brooks, D., Brunner, R. J., Buckley-Geer, E., . . . Zuntz, J. (2016). The Dark Energy Survey: more than dark energy - an overview., *460*(2), 1270–1299. <https://doi.org/10.1093/mnras/stw641>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning* [<http://www.deeplearningbook.org>]. MIT Press.
- Joseph, R., Courbin, F., & Starck, J. .-. (2016). Multi-band morpho-Spectral Component Analysis Deblending Tool (MuSCADeT): Deblending colourful objects., *589*, Article A2, A2. <https://doi.org/10.1051/0004-6361/201527923>
- Mallat, S. (2016). Understanding deep convolutional networks. *Philos. Trans. A Math. Phys. Eng. Sci.*, *374*(2065), 20150203.
- Marshall, P. J., Verma, A., More, A., Davis, C. P., More, S., Kapadia, A., Parrish, M., Snyder, C., Wilcox, J., Baeten, E., Macmillan, C., Cornen, C., Baumer, M., Simpson, E., Lintott, C. J., Miller, D., Paget, E., Simpson, R., Smith, A. M., . . . Collett, T. E. (2016). SPACE WARPS - I. Crowdsourcing the discovery of gravitational lenses., *455*(2), 1171–1190. <https://doi.org/10.1093/mnras/stv2009>
- More, A., Verma, A., Marshall, P. J., More, S., Baeten, E., Wilcox, J., Macmillan, C., Cornen, C., Kapadia, A., Parrish, M., Snyder, C., Davis, C. P., Gavazzi, R., Lintott, C. J., Simpson, R., Miller, D., Smith, A. M., Paget, E., Saha, P., . . . Collett, T. E. (2016). SPACE WARPS-II. New gravitational lens candidates from the CFHTLS discovered through citizen science., *455*(2), 1191–1210. <https://doi.org/10.1093/mnras/stv1965>
- Odena, A., Olah, C., & Shlens, J. (2016). Conditional Image Synthesis With Auxiliary Classifier GANs. *arXiv e-prints*, Article arXiv:1610.09585, arXiv:1610.09585.
- Paraficz, D., Courbin, F., Tramacere, A., Joseph, R., Metcalf, R. B., Kneib, J. .-. , Dubath, P., Droz, D., Filleul, F., Ringeisen, D., & Schäfer, C. (2016). The PCA Lens-Finder: application to CFHTLS., *592*, Article A75, A75. <https://doi.org/10.1051/0004-6361/201527971>

- Planck Collaboration, Ade, P. A. R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A. J., Barreiro, R. B., Bartlett, J. G., & et al. (2016). Planck 2015 results. XIII. Cosmological parameters., 594, Article A13, A13. <https://doi.org/10.1051/0004-6361/201525830>
- Riess, A. G., Macri, L. M., Hoffmann, S. L., Scolnic, D., Casertano, S., Filippenko, A. V., Tucker, B. E., Reid, M. J., Jones, D. O., Silverman, J. M., Chornock, R., Challis, P., Yuan, W., Brown, P. J., & Foley, R. J. (2016). A 2.4% determination of the local value of the hubble constant. <https://doi.org/10.3847/0004-637X/826/1/56>
- Birrer, S., Amara, A., & Refregier, A. (2015). Gravitational Lens Modeling with Basis Sets., 813(2), Article 102, 102. <https://doi.org/10.1088/0004-637X/813/2/102>
- Collett, T. E. (2015). The Population of Galaxy-Galaxy Strong Lenses in Forthcoming Optical Imaging Surveys., 811(1), Article 20, 20. <https://doi.org/10.1088/0004-637X/811/1/20>
- Geach, J. E., More, A., Verma, A., Marshall, P. J., Jackson, N., Belles, P. .-, Beswick, R., Baeten, E., Chavez, M., Cornen, C., Cox, B. E., Erben, T., Erickson, N. J., Garrington, S., Harrison, P. A., Harrington, K., Hughes, D. H., Ivison, R. J., Jordan, C., . . . Zeballos, M. (2015). The Red Radio Ring: a gravitationally lensed hyperluminous infrared radio galaxy at $z = 2.553$ discovered through the citizen science project SPACE WARPS., 452(1), 502–510. <https://doi.org/10.1093/mnras/stv1243>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *arXiv e-prints*, Article arXiv:1502.01852, arXiv:1502.01852.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3), 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
- Sonnenfeld, A., Treu, T., Marshall, P. J., Suyu, S. H., Gavazzi, R., Auger, M. W., & Nipoti, C. (2015). The SL2S Galaxy-scale Lens Sample. V. Dark Matter Halos and Stellar IMF of Massive Early-type Galaxies Out to Redshift 0.8., 800(2), Article 94, 94. <https://doi.org/10.1088/0004-637X/800/2/94>
- Spergel, D., Gehrels, N., Baltay, C., Bennett, D., Breckinridge, J., Donahue, M., Dressler, A., Gaudi, B. S., Greene, T., Guyon, O., Hirata, C., Kalirai, J., Kasdin, N. J., Macintosh, B., Moos, W., Perlmutter, S., Postman, M., Rauscher, B., Rhodes, J., . . . Zhao, F. (2015). Wide-Field Infrared Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report. *arXiv e-prints*, Article arXiv:1503.03757, arXiv:1503.03757.
- Gavazzi, R., Marshall, P. J., Treu, T., & Sonnenfeld, A. (2014). RINGFINDER: Automated Detection of Galaxy-scale Gravitational Lenses in Ground-based Multi-filter Imaging Data., 785(2), Article 144, 144. <https://doi.org/10.1088/0004-637X/785/2/144>
- Joseph, R., Courbin, F., Metcalf, R. B., Giocoli, C., Hartley, P., Jackson, N., Bellagamba, F., Kneib, J. .-, Koopmans, L., Lemson, G., Meneghetti, M., Meylan, G., Petkova, M., & Pires, S. (2014). A PCA-based automated finder for galaxy-scale strong lenses., 566, Article A63, A63. <https://doi.org/10.1051/0004-6361/201423365>
- Kingma, D., & Welling, M. (2014). Auto-encoding variational bayes.

- Kingma, D. P., & Ba, J. (2014). Adam: a method for stochastic optimization [cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015]. <http://arxiv.org/abs/1412.6980>
- Maturi, M., Mizera, S., & Seidel, G. (2014). Multi-colour detection of gravitational arcs., 567, Article A111, A111. <https://doi.org/10.1051/0004-6361/201321634>
- Pawase, R. S., Courbin, F., Faure, C., Kokotanekova, R., & Meylan, G. (2014). A 7 deg² survey for galaxy-scale gravitational lenses with the HST imaging archive., 439(4), 3392–3404. <https://doi.org/10.1093/mnras/stu179>
- Schneider, P., & Sluse, D. (2014). Source-position transformation: an approximate invariance in strong gravitational lensing., 564, Article A103, A103. <https://doi.org/10.1051/0004-6361/201322106>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1), 1929–1958.
- Vegetti, S., Koopmans, L. V. E., Auger, M. W., Treu, T., & Bolton, A. S. (2014). Inference of the cold dark matter substructure mass function at $z = 0.2$ using strong gravitational lenses., 442(3), 2017–2035. <https://doi.org/10.1093/mnras/stu943>
- Wuyts, E., Rigby, J. R., Gladders, M. D., & Sharon, K. (2014). A Magnified View of the Kinematics and Morphology of RCSGA 032727-132609: Zooming in on a Merger at $z = 1.7$., 781(2), Article 61, 61. <https://doi.org/10.1088/0004-637X/781/2/61>
- de Jong, J. T. A., Kuijken, K., Applegate, D., Begeman, K., Belikov, A., Blake, C., Bout, J., Boxhoorn, D., Buddelmeijer, H., Buddendiek, A., Cacciato, M., Capaccioli, M., Choi, A., Cordes, O., Covone, G., Dall’Ora, M., Edge, A., Erben, T., Franse, J., ... Vriend, W. -. (2013). The Kilo-Degree Survey. *The Messenger*, 154, 44–46.
- Schneider, P., & Sluse, D. (2013). Mass-sheet degeneracy, power-law models and external convergence: Impact on the determination of the Hubble constant from gravitational lensing., 559, Article A37, A37. <https://doi.org/10.1051/0004-6361/201321882>
- Vieira, J. D., Marrone, D. P., Chapman, S. C., De Breuck, C., Hezaveh, Y. D., Weiß, A., Aguirre, J. E., Aird, K. A., Aravena, M., Ashby, M. L. N., Bayliss, M., Benson, B. A., Biggs, A. D., Bleem, L. E., Bock, J. J., Bothwell, M., Bradford, C. M., Brodwin, M., Carlstrom, J. E., ... Williamson, R. (2013). Dusty starburst galaxies in the early Universe as revealed by gravitational lensing., 495(7441), 344–347. <https://doi.org/10.1038/nature12001>
- Weinberg, D. H., Mortonson, M. J., Eisenstein, D. J., Hirata, C., Riess, A. G., & Rozo, E. (2013). Observational probes of cosmic acceleration., 530(2), 87–255. <https://doi.org/10.1016/j.physrep.2013.05.001>
- Amiaux, J., Scaramella, R., Mellier, Y., Altieri, B., Burigana, C., Da Silva, A., Gomez, P., Hoar, J., Laureijs, R., Maiorano, E., Magalhães Oliveira, D., Renk, F., Saavedra Criado, G., Tereno, I., Auguères, J. L., Brinchmann, J., Cropper, M., Duvet, L., Ealet, A., ... Warren, S. (2012). Euclid mission: building of a reference survey. In M. C. Clampin, G. G. Fazio, H. A. MacEwen, & J. Oschmann Jacobus M. (Eds.), *Space telescopes and instrumentation 2012: optical, infrared, and millimeter wave* (84420Z). <https://doi.org/10.1117/12.926513>

- Brownstein, J. R., Bolton, A. S., Schlegel, D. J., Eisenstein, D. J., Kochanek, C. S., Connolly, N., Maraston, C., Pandey, P., Seitz, S., Wake, D. A., Wood-Vasey, W. M., Brinkmann, J., Schneider, D. P., & Weaver, B. A. (2012). The BOSS Emission-Line Lens Survey (BELLS). I. A Large Spectroscopically Selected Sample of Lens Galaxies at Redshift ~ 0.5 , 744(1), Article 41, 41. <https://doi.org/10.1088/0004-637X/744/1/41>
- Courbin, F., Faure, C., Djorgovski, S. G., R  rat, F., Tewes, M., Meylan, G., Stern, D., Mahabal, A., Boroson, T., Dheeraj, R., & Sluse, D. (2012). Three quasi-stellar objects acting as strong gravitational lenses., 540, Article A36, A36. <https://doi.org/10.1051/0004-6361/201118015>
- Deng, L. (2012). The mnist database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, 29(6), 141–142.
- Gavazzi, R., Treu, T., Marshall, P. J., Brault, F., & Ruff, A. (2012). The SL2S Galaxy-scale Gravitational Lens Sample. I. The Alignment of Mass and Light in Massive Early-type Galaxies at $z = 0.2\text{--}0.9$., 761(2), Article 170, 170. <https://doi.org/10.1088/0004-637X/761/2/170>
- Inada, N., Oguri, M., Shin, M.-S., Kayo, I., Strauss, M. A., Morokuma, T., Rusu, C. E., Fukugita, M., Kochanek, C. S., Richards, G. T., Schneider, D. P., York, D. G., Bahcall, N. A., Frieman, J. A., Hall, P. B., & White, R. L. (2012). The Sloan Digital Sky Survey Quasar Lens Search. V. Final Catalog from the Seventh Data Release., 143(5), Article 119, 119. <https://doi.org/10.1088/0004-6256/143/5/119>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012a). Imagenet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, & K. Weinberger (Eds.), *Advances in neural information processing systems*. Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012b). Imagenet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, & K. Weinberger (Eds.), *Advances in neural information processing systems*. Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>
- More, A., Cabanac, R., More, S., Alard, C., Limousin, M., Kneib, J. .-, Gavazzi, R., & Motta, V. (2012). The CFHTLS-Strong Lensing Legacy Survey (SL2S): Investigating the Group-scale Lenses with the SARCS Sample., 749(1), Article 38, 38. <https://doi.org/10.1088/0004-637X/749/1/38>
- Moustakas, L. (2012). The Master Lens Database and The Orphan Lenses Project.
- Schwamb, M. E., Lintott, C. J., Fischer, D. A., Giguere, M. J., Lynn, S., Smith, A. M., Brewer, J. M., Parrish, M., Schawinski, K., & Simpson, R. J. (2012). Planet Hunters: Assessing the Kepler Inventory of Short-period Planets., 754(2), Article 129, 129. <https://doi.org/10.1088/0004-637X/754/2/129>
- Suyu, S. H., Hensel, S. W., McKean, J. P., Fassnacht, C. D., Treu, T., Halkola, A., Norbury, M., Jackson, N., Schneider, P., Thompson, D., Auger, M. W., Koopmans, L. V. E., & Matthews, K. (2012). Disentangling Baryons and Dark Matter in the Spiral Gravitational Lens B1933+503., 750(1), Article 10, 10. <https://doi.org/10.1088/0004-637X/750/1/10>

- Vegetti, S., Lagattuta, D. J., McKean, J. P., Auger, M. W., Fassnacht, C. D., & Koopmans, L. V. E. (2012). Gravitational detection of a low-mass dark satellite galaxy at cosmological distance., *481*(7381), 341–343. <https://doi.org/10.1038/nature10669>
- Laureijs, R., Amiaux, J., Arduini, S., Auguères, J. .-, Brinchmann, J., Cole, R., Cropper, M., Dabin, C., Duvet, L., Ealet, A., Garilli, B., Gondoin, P., Guzzo, L., Hoar, J., Hoekstra, H., Holmes, R., Kitching, T., Maciaszek, T., Mellier, Y., ... Zucca, E. (2011). Euclid Definition Study Report. *arXiv e-prints*, Article arXiv:1110.3193, arXiv:1110.3193.
- Linder, E. V. (2011). Lensing time delays and cosmological complementarity., *84*(12), Article 123529, 123529. <https://doi.org/10.1103/PhysRevD.84.123529>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.
- Treu, T., Dutton, A. A., Auger, M. W., Marshall, P. J., Bolton, A. S., Brewer, B. J., Koo, D. C., & Koopmans, L. V. E. (2011). The SWELLS survey - I. A large spectroscopically selected sample of edge-on late-type lens galaxies., *417*(3), 1601–1620. <https://doi.org/10.1111/j.1365-2966.2011.19378.x>
- Boroson, T. A., & Lauer, T. R. (2010). Exploring the Spectral Space of Low Redshift QSOs., *140*(2), 390–402. <https://doi.org/10.1088/0004-6256/140/2/390>
- Kravtsov, A. (2010). The Dark Matter Annihilation Signal from Dwarf Galaxies and Subhalos. *Advances in Astronomy*, *2010*, Article 281913, 281913. <https://doi.org/10.1155/2010/281913>
- Oguri, M., & Marshall, P. J. (2010). Gravitationally lensed quasars and supernovae in future wide-field optical imaging surveys., *405*(4), 2579–2593. <https://doi.org/10.1111/j.1365-2966.2010.16639.x>
- Suyu, S. H., & Halkola, A. (2010). The halos of satellite galaxies: the companion of the massive elliptical lens SL2S J08544-0121., *524*, Article A94, A94. <https://doi.org/10.1051/0004-6361/201015481>
- Syget, J. E., Tu, H., Fort, B., & Gavazzi, R. (2010). A search for edge-on galaxy lenses in the CFHT Legacy Survey., *517*, Article A25, A25. <https://doi.org/10.1051/0004-6361/200913977>
- Vegetti, S., Koopmans, L. V. E., Bolton, A., Treu, T., & Gavazzi, R. (2010). Detection of a dark substructure through gravitational imaging., *408*(4), 1969–1981. <https://doi.org/10.1111/j.1365-2966.2010.16865.x>
- Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P.-A. (2010). Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.*, *11*, 3371–3408.
- Barnabè, M., Czoske, O., Koopmans, L. V. E., Treu, T., Bolton, A. S., & Gavazzi, R. (2009). Two-dimensional kinematics of SLACS lenses - II. Combined lensing and dynamics analysis of early-type galaxies at $z = 0.08\text{--}0.33$., *399*(1), 21–36. <https://doi.org/10.1111/j.1365-2966.2009.14941.x>

- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: a large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
- Keeton, C. R., & Moustakas, L. A. (2009). A New Channel for Detecting Dark Matter Substructure in Galaxies: Gravitational Lens Time Delays., *699*(2), 1720–1731. <https://doi.org/10.1088/0004-637X/699/2/1720>
- Li, N., & Chen, D.-M. (2009). Cusp-core problem and strong gravitational lensing. *Research in Astronomy and Astrophysics*, *9*(11), 1173–1184. <https://doi.org/10.1088/1674-4527/9/11/001>
- Orban de Xivry, G., & Marshall, P. (2009). An atlas of predicted exotic gravitational lenses., *399*(1), 2–20. <https://doi.org/10.1111/j.1365-2966.2009.14925.x>
- Vegetti, S., & Koopmans, L. V. E. (2009). Bayesian strong gravitational-lens modelling on adaptive grids: objective detection of mass substructure in Galaxies., *392*(3), 945–963. <https://doi.org/10.1111/j.1365-2966.2008.14005.x>
- Blackburne, J. A., Wisotzki, L., & Schechter, P. L. (2008). HE 1113-0641: The Smallest-Separation Quadruple Lens Identified by a Ground-Based Optical Telescope., *135*(1), 374–379. <https://doi.org/10.1088/0004-6256/135/1/374>
- Faure, C., Kneib, J.-P., Covone, G., Tasca, L., Leauthaud, A., Capak, P., Jahnke, K., Smolcic, V., de la Torre, S., Ellis, R., Finoguenov, A., Koekemoer, A., Le Fevre, O., Massey, R., Mellier, Y., Refregier, A., Rhodes, J., Scoville, N., Schinnerer, E., ... Walcher, J. (2008). First Catalog of Strong Lens Candidates in the COSMOS Field., *176*(1), 19–38. <https://doi.org/10.1086/526426>
- Jackson, N. (2008). Gravitational lenses and lens candidates identified from the COSMOS field., *389*(3), 1311–1318. <https://doi.org/10.1111/j.1365-2966.2008.13629.x>
- Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th International Conference on Machine Learning*, 1096–1103. <https://doi.org/10.1145/1390156.1390294>
- Cabanac, R. A., Alard, C., Dantel-Fort, M., Fort, B., Gavazzi, R., Gomez, P., Kneib, J. P., Le Fèvre, O., Mellier, Y., Pello, R., Soucail, G., Sygnet, J. F., & Valls-Gabaud, D. (2007). The CFHTLS strong lensing legacy survey. I. Survey overview and T0002 release sample., *461*, 813–821. <https://doi.org/10.1051/0004-6361:20065810>
- Chen, J., Rozo, E., Dalal, N., & Taylor, J. E. (2007). Astrometric Perturbations in Substructure Lensing., *659*(1), 52–68. <https://doi.org/10.1086/512002>
- Estrada, J., Annis, J., Diehl, H. T., Hall, P. B., Las, T., Lin, H., Makler, M., Merritt, K. W., Scarpine, V., Allam, S., & Tucker, D. (2007). A Systematic Search for High Surface Brightness Giant Arcs in a Sloan Digital Sky Survey Cluster Sample., *660*(2), 1176–1185. <https://doi.org/10.1086/512599>
- Jiang, G., & Kochanek, C. S. (2007). The baryon fractions and mass-to-light ratios of early-type galaxies. *The Astrophysical Journal*, *671*(2), 1568–1578. <https://doi.org/10.1086/522580>
- Marshall, P. J., Treu, T., Melbourne, J., Gavazzi, R., Bundy, K., Ammons, S. M., Bolton, A. S., Burles, S., Larkin, J. E., Le Mignant, D., Koo, D. C., Koopmans, L. V. E., Max, C. E.,

- Moustakas, L. A., Steinbring, E., & Wright, S. A. (2007). Superresolving Distant Galaxies with Gravitational Telescopes: Keck Laser Guide Star Adaptive Optics and Hubble Space Telescope Imaging of the Lens System SDSS J0737+3216., *671*(2), 1196–1211. <https://doi.org/10.1086/523091>
- Miranda, M., & Macciò, A. V. (2007). Constraining warm dark matter using QSO gravitational lensing., *382*(3), 1225–1232. <https://doi.org/10.1111/j.1365-2966.2007.12440.x>
- Seidel, G., & Bartelmann, M. (2007). Arcfinder: an algorithm for the automatic detection of gravitational arcs., *472*(1), 341–352. <https://doi.org/10.1051/0004-6361:20066097>
- Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., & Moustakas, L. A. (2006). The Sloan Lens ACS Survey. I. A Large Spectroscopically Selected Sample of Massive Early-Type Lens Galaxies., *638*(2), 703–724. <https://doi.org/10.1086/498884>
- Davis, J., & Goadrich, M. (2006). The relationship between precision-recall and roc curves. *Proceedings of the 23rd International Conference on Machine Learning*, 233–240. <https://doi.org/10.1145/1143844.1143874>
- Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504–507.
- Oguri, M., Inada, N., Pindor, B., Strauss, M. A., Richards, G. T., Hennawi, J. F., Turner, E. L., Lupton, R. H., Schneider, D. P., Fukugita, M., & Brinkmann, J. (2006). The Sloan Digital Sky Survey Quasar Lens Search. I. Candidate Selection Algorithm., *132*(3), 999–1013. <https://doi.org/10.1086/506019>
- Schneider, P. (2006). Weak gravitational lensing. *33rd Advanced Saas Fee Course on Gravitational Lensing: Strong, Weak, and Micro*, 269–451. https://doi.org/10.1007/978-3-540-30310-7_3
- Koopmans, L. V. E. (2005). Gravitational imaging of cold dark matter substructures., *363*(4), 1136–1144. <https://doi.org/10.1111/j.1365-2966.2005.09523.x>
- Drummond, C., & Holte, R. (2004). What roc curves can't do (and cost curves can).
- Lenzen, F., Schindler, S., & Scherzer, O. (2004). Automatic detection of arcs and arclets formed by gravitational lensing. *A&A*, *416*(1), 391–401. <https://doi.org/10.1051/0004-6361:20034619>
- Lotz, J. M., Primack, J., & Madau, P. (2004). A New Nonparametric Approach to Galaxy Morphological Classification., *128*(1), 163–182. <https://doi.org/10.1086/421849>
- Ng, A. Y. (2004). Feature selection, l1 vs. l2 regularization, and rotational invariance. *Proceedings of the Twenty-First International Conference on Machine Learning*, 78. <https://doi.org/10.1145/1015330.1015435>
- Treu, T., & Koopmans, L. V. E. (2004). Massive Dark Matter Halos and Evolution of Early-Type Galaxies to $z \sim 1$., *611*(2), 739–760. <https://doi.org/10.1086/422245>
- Wisotzki, L., Schechter, P. L., Chen, H. .-, Richstone, D., Jahnke, K., Sánchez, S. F., & Reimers, D. (2004). HE 0047-1756: A new gravitationally lensed double QSO., *419*, L31–L34. <https://doi.org/10.1051/0004-6361:20040131>
- Conselice, C. J. (2003). The Relationship between Stellar Light Distributions of Galaxies and Their Formation Histories., *147*(1), 1–28. <https://doi.org/10.1086/375001>

- Joye, W. A., & Mandel, E. (2003). New Features of SAOImage DS9. In H. E. Payne, R. I. Jerzjewski, & R. N. Hook (Eds.), *Astronomical data analysis software and systems xii* (p. 489).
- Koopmans, L. V. E., & Treu, T. (2003). The Structure and Dynamics of Luminous and Dark Matter in the Early-Type Lens Galaxy of 0047-281 at $z = 0.485$, 583(2), 606–615. <https://doi.org/10.1086/345423>
- Myers, S. T., Jackson, N. J., Browne, I. W. A., de Bruyn, A. G., Pearson, T. J., Readhead, A. C. S., Wilkinson, P. N., Biggs, A. D., Blandford, R. D., Fassnacht, C. D., Koopmans, L. V. E., Marlow, D. R., McKean, J. P., Norbury, M. A., Phillips, P. M., Rusin, D., Shepherd, M. C., & Sykes, C. M. (2003). The Cosmic Lens All-Sky Survey - I. Source selection and observations., 341(1), 1–12. <https://doi.org/10.1046/j.1365-8711.2003.06256.x>
- Wisotzki, L., Becker, T., Christensen, L., Helms, A., Jahnke, K., Kelz, A., Roth, M. M., & Sanchez, S. F. (2003). Integral-field spectrophotometry of the quadruple QSO HE 0435-1223: Evidence for microlensing., 408, 455–463. <https://doi.org/10.1051/0004-6361:20031004>
- Blain, A. W., & Phillips, T. G. (2002). The 60- μm extragalactic background radiation intensity, dust-enshrouded active galactic nuclei and the assembly of groups and clusters of galaxies., 333(1), 222–230. <https://doi.org/10.1046/j.1365-8711.2002.05398.x>
- Dalal, N., & Kochanek, C. S. (2002). Direct Detection of Cold Dark Matter Substructure., 572(1), 25–33. <https://doi.org/10.1086/340303>
- Metcalf, R. B., & Zhao, H. (2002). Flux Ratios as a Probe of Dark Substructures in Quadruple-Image Gravitational Lenses., 567(1), L5–L8. <https://doi.org/10.1086/339798>
- Wisotzki, L., Schechter, P. L., Bradt, H. V., Heinmüller, J., & Reimers, D. (2002). HE 0435-1223: A wide separation quadruple QSO and gravitational lens., 395, 17–23. <https://doi.org/10.1051/0004-6361:20021213>
- Bartelmann, M., & Schneider, P. (2001). Weak gravitational lensing., 340(4-5), 291–472. [https://doi.org/10.1016/S0370-1573\(00\)00082-X](https://doi.org/10.1016/S0370-1573(00)00082-X)
- Axelsson, S. (2000). The base-rate fallacy and the difficulty of intrusion detection. *ACM Trans. Inf. Syst. Secur.*, 3(3), 186–205. <https://doi.org/10.1145/357830.357849>
- Bershady, M. A., Jangren, A., & Conselice, C. J. (2000). Structural and Photometric Classification of Galaxies. I. Calibration Based on a Nearby Galaxy Sample., 119(6), 2645–2663. <https://doi.org/10.1086/301386>
- Caruana, R., Lawrence, S., & Giles, C. (2000). Overfitting in neural nets: backpropagation, conjugate gradient, and early stopping. In T. Leen, T. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems*. MIT Press. <https://proceedings.neurips.cc/paper/2000/file/059fdcd96baeb75112f09fa1dcc740cc-Paper.pdf>
- Wisotzki, L., Christlieb, N., Bade, N., Beckmann, V., Köhler, T., Vanelle, C., & Reimers, D. (2000). The Hamburg/ESO survey for bright QSOs. III. A large flux-limited sample of QSOs., 358, 77–87.
- Blain, A. W., Kneib, J. .-, Ivison, R. J., & Smail, I. (1999). Deep Counts of Submillimeter Galaxies., 512(2), L87–L90. <https://doi.org/10.1086/311879>

- Mao, S., & Schneider, P. (1998). Evidence for substructure in lens galaxies?, 295(3), 587–594. <https://doi.org/10.1046/j.1365-8711.1998.01319.x>
- Mitchell, T. M. (1997). *Machine learning, international edition*. McGraw-Hill. <https://www.worldcat.org/oclc/61321007>
- Narayan, R., & Bartelmann, M. (1996). Lectures on Gravitational Lensing. *arXiv e-prints*, Article astro-ph/9606001, astro-ph/9606001.
- Lecun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time-series. In M. Arbib (Ed.), *The handbook of brain theory and neural networks*. MIT Press.
- Kormann, R., Schneider, P., & Bartelmann, M. (1994). Isothermal elliptical gravitational lens models., 284, 285–299.
- Kassiola, A., & Kovner, I. (1993). Elliptic Mass Distributions versus Elliptic Potentials in Gravitational Lenses., 417, 450. <https://doi.org/10.1086/173325>
- Narayan, R., & Wallington, S. (1993). Magnification bias and gravitational lensing statistics. In J. Surdej, D. Fraipont-Caro, E. Gosset, S. Refsdal, & M. Remy (Eds.), *Liege international astrophysical colloquia* (p. 217).
- Wisotzki, L., Koehler, T., Kayser, R., & Reimers, D. (1993). The new double QSO HE 1104_1805 : gravitational lens with microlensing or binary quasar ?, 278, L15–L18.
- Krogh, A., & Hertz, J. (1991). A simple weight decay can improve generalization. In J. Moody, S. Hanson, & R. Lippmann (Eds.), *Advances in neural information processing systems*. Morgan-Kaufmann. <https://proceedings.neurips.cc/paper/1991/file/8eefcfd5990e441f0fb6f3fad709e2Paper.pdf>
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)*, 2(4), 303–314. <https://doi.org/10.1007/BF02551274>
- Bourlard, H., & Kamp, Y. (1988). Auto-association by multilayer perceptrons and singular value decomposition. *Biological Cybernetics*, 59(4-5), 291–294. <https://doi.org/10.1007/bf00332918>
- Magain, P., Surdej, J., Swings, J. .-, Borgeest, U., & Kayser, R. (1988). Discovery of a quadruply lensed quasar: the 'clover leaf H1413 + 117., 334(6180), 325–327. <https://doi.org/10.1038/334325a0>
- Surdej, J., Magain, P., Swings, J. .-, Borgeest, U., Courvoisier, T. J. .-, Kayser, R., Kellermann, K. I., Kuhr, H., & Refsdal, S. (1987). A new case of gravitational lensing., 329(6141), 695–696. <https://doi.org/10.1038/329695a0>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors., 323(6088), 533–536. <https://doi.org/10.1038/323533a0>
- Huchra, J., Gorenstein, M., Kent, S., Shapiro, I., Smith, G., Horine, E., & Perley, R. (1985). 2237+0305 : a new and unusual gravitational lens., 90, 691–696. <https://doi.org/10.1086/113777>
- Burke, W. L. (1981). Multiple Gravitational Imaging by Distributed Masses., 244, L1. <https://doi.org/10.1086/183466>

- Rubin, V. C., Ford, J., W. K., & Thonnard, N. (1980). Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 ($R=4\text{kpc}$) to UGC 2885 ($R=122\text{kpc}$)., 238, 471–487. <https://doi.org/10.1086/158003>
- Weymann, R. J., Latham, D., Angel, J. R. P., Green, R. F., Liebert, J. W., Turnshek, D. A., Turnshek, D. E., & Tyson, J. A. (1980). The triple QSO PG1115 + 08: another probable gravitational lens., 285(5767), 641–643. <https://doi.org/10.1038/285641a0>
- Walsh, D., Carswell, R. F., & Weymann, R. J. (1979). 0957+561 A, B: twin quasistellar objects or gravitational lens?, 279, 381–384. <https://doi.org/10.1038/279381a0>
- Rubin, V. C., Ford, J., W. K., & Thonnard, N. (1978). Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa -> Sc., 225, L107–L111. <https://doi.org/10.1086/182804>
- Fukushima, K. (1975). Cognitron: A self-organizing multilayer neural network. *Biological Cybernetics*, 20, 121–136.
- Weinberg, S. (1972). *Gravitation and cosmology*. John Wiley; Sons.
- Refsdal, S. (1964). On the possibility of determining Hubble's parameter and the masses of galaxies from the gravitational lens effect., 128, 307. <https://doi.org/10.1093/mnras/128.4.307>
- Bellman, R. (1957). *Dynamic programming* (1st ed.). Princeton University Press.
- Rosenblatt, F. (1957). *The perceptron - a perceiving and recognizing automaton* (tech. rep. No. 85-460-1). Cornell Aeronautical Laboratory. Ithaca, New York.
- Zwicky, F. (1937a). Nebulae as Gravitational Lenses. *Physical Review*, 51(4), 290–290. <https://doi.org/10.1103/PhysRev.51.290>
- Zwicky, F. (1937b). On the Probability of Detecting Nebulae Which Act as Gravitational Lenses. *Physical Review*, 51(8), 679–679. <https://doi.org/10.1103/PhysRev.51.679>
- Einstein, A. (1936). Lens-like action of a star by the deviation of light in the gravitational field. *Science*, 84(2188), 506–507.
- Zwicky, F. (1933). Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6, 110–127.
- Chwolson, O. (1924). Über eine mögliche form fiktiver doppelsterne. *Astronomische Nachrichten*, 221, 329.
- FR.S., K. P. (1901). Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559–572. <https://doi.org/10.1080/14786440109462720>



Route du Bois 55B
1024 Ecublens
+41 79 472 29 93
elodie.savary@outlook.com

Computer skills

Programming in Python, Java, C++, C, Matlab

Machine Learning and data analysis with Tensorflow, keras, Pytorch, Sklearn

Cluster computing (Slurm Workload Manager)

Unix/Linux

Awards

EPFL - [Award of Merit](#)

Languages

English: B2-C1, Professional working proficiency

German: B2, worked in a Swiss-German factory

Italian: B1

French: Native speaker

Personal interests

Playing board games with friends, running

Elodie Savary

Passionate machine learning specialist with experience in computer vision, image processing and generative models. As a fast learner, I can adapt quickly to a large variety of domains. I strive to place my problem solving skills at the service of new challenges.

Professional Experience

PHD RESEARCHER - 10/2018 to now, EPFL, Lausanne, Switzerland

- Developed a pipeline based on convolutional neural networks to detect the presence of gravitational lensing features in astrophysical images which was applied successfully on images from the [CFIS](#) and [DES](#) surveys.
- Designed a new unsupervised machine learning algorithm to separate the light of different components in gravitational lens images.
- Implemented a visualization tool which is now used by multiple research groups to save time and reduce biases during the visual inspection of large sets of astrophysical images.
- Worked in an international collaboration ([Euclid Consortium](#)) on the development of the data processing pipeline for the Euclid mission.
- Supervised 7 master students during their master thesis and semester projects.

MATHEMATICS AND PHYSICS INSTRUCTOR - 09/2017 and 09/2018, Laprerenetree, Lausanne, Switzerland

- Delivered mathematics and physics lectures to classes of 30 students in the context of a preparatory week for future EPFL students.

APPRENTICE IN PRECISION MECHANICS - 08/2005 to 07/2009, Wifag-Polytype, Fribourg, Switzerland

- Assembled printing machines for plastic food containers within tight deadlines.

Academic projects

SPECIALISATION SEMESTER PROJECT:

- Implemented different generative models to simulate images of galaxies (Variational auto-encoder, Wasserstein-GAN).

PRACTICAL WORK PROJECT:

- Applied a machine learning method (Boosted trees) to detect signs of gravitational lensing in the spectra of galaxies.

Education

PHD IN ASTROPHYSICS -10/2018 to now

EPFL, Lausanne, Switzerland

- *Thesis* : «Teaching machines how to find strongly lensed galaxies in cosmological sky surveys»

MASTER IN PHYSICS - 09/2016- 07/2018

EPFL, Lausanne, Switzerland

BACHELOR IN PHYSICS – 09/2011 to 07/2016

EPFL, Lausanne, Switzerland

INTRODUCTORY MATHEMATICS COURSE – 09/2009 to 07/2011

EPFL, Lausanne, Switzerland

FEDERAL VET DIPLOMA IN PRECISION MECHANICS WITH INTEGRATED FEDERAL VOCATIONAL BACCALAUREATE – 08/2005 to 07/2009

Wifag-Polytype / EPAI , Fribourg, Switzerland

Publications

The complete list can be found on [ADS](#). My main contributions are:

- Savary et al. 2022 [A search for galaxy-scale strong gravitational lenses in the Ultraviolet Near Infrared Optical Northern Survey \(UNIONS\)](#)
- Rojas et al. 2022 [Strong lens systems search in the Dark Energy Survey using Convolutional Neural Networks](#)
- Cañameras et al. 2020 [HOLISMOKES. II. Identifying galaxy-scale strong gravitational lenses in Pan-STARRS using convolutional neural networks](#)
- Dia et al. 2020 [Galaxy Image Simulation Using Progressive GANs](#)
- Bonvin et al. 2019 [Impact of the 3D source geometry on time-delay measurements of lensed type-Ia supernovae](#)

Civil Status

Swiss national, born on 3 March 1989