Astronomy & Astrophysics

Strong lensing in UNIONS: Toward a pipeline from discovery to modeling[*]

E. Savary[1], K. Rojas[1], M. Maus[1], B. Clément[1], F. Courbin[1], R. Gavazzi[2,3], J. H. H. Chan[1], C. Lemon[1], G. Vernardos[1], R. Cañameras[4], S. Schuldt[4,5], S. H. Suyu[4,5,6], J.-C. Cuillandre[7,8], S. Fabbro[12], S. Gwyn[12], M. J. Hudson[9,10,11], M. Kilbinger[7], D. Scott[13], and C. Stone[14]

1. Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland
   e-mail: elodie.savary@epfl.ch
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 CB3 0HA, UK
4. Max-Planck-Institut für Astrophysik, Karl-Schwarzschild Str. 1, 85748 Garching, Germany
5. Technische Universität München, Physik-Department, James-Franck-Straße 1, 85748 Garching, Germany
6. Institute of Astronomy and Astrophysics, Academia Sinica, 11F of ASMAB, No. 1, Section 4, Roosevelt Road, Taipei 10617, Taiwan
7. AIM, CEA, CNRS, Université Paris-Saclay, Université de Paris, 91191 Gif-sur-Yvette, France
8. Observatoire de Paris, PSL Research University, 61 Avenue de l'Observatoire, 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

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. 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 Sersic 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, 2014, 2018; Nierenberg et al. 2013; Hezaveh et al. 2016; Despali et al. 2016; Gilman et al. 2017; 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

---

[*] The catalog of the 133 candidates and the modeling results are only available at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/666/A1

known high-redshift source, in the hope of observing signs 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, 2019; Kuijken et al. 2019); the Dark Energy Survey (DES, e.g., The Dark Energy Survey Collaboration 2005; Jacobs et al. 2019a,b), and the Hyper Suprime-Cam SSP Survey (HSC, e.g., Aihara et al. 2018; Sonnenfeld et al. 2018, 2019, 2020; Wong et al. 2018; Chan 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 Spec-

troscopic 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 resampling 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 2500 deg^2 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.
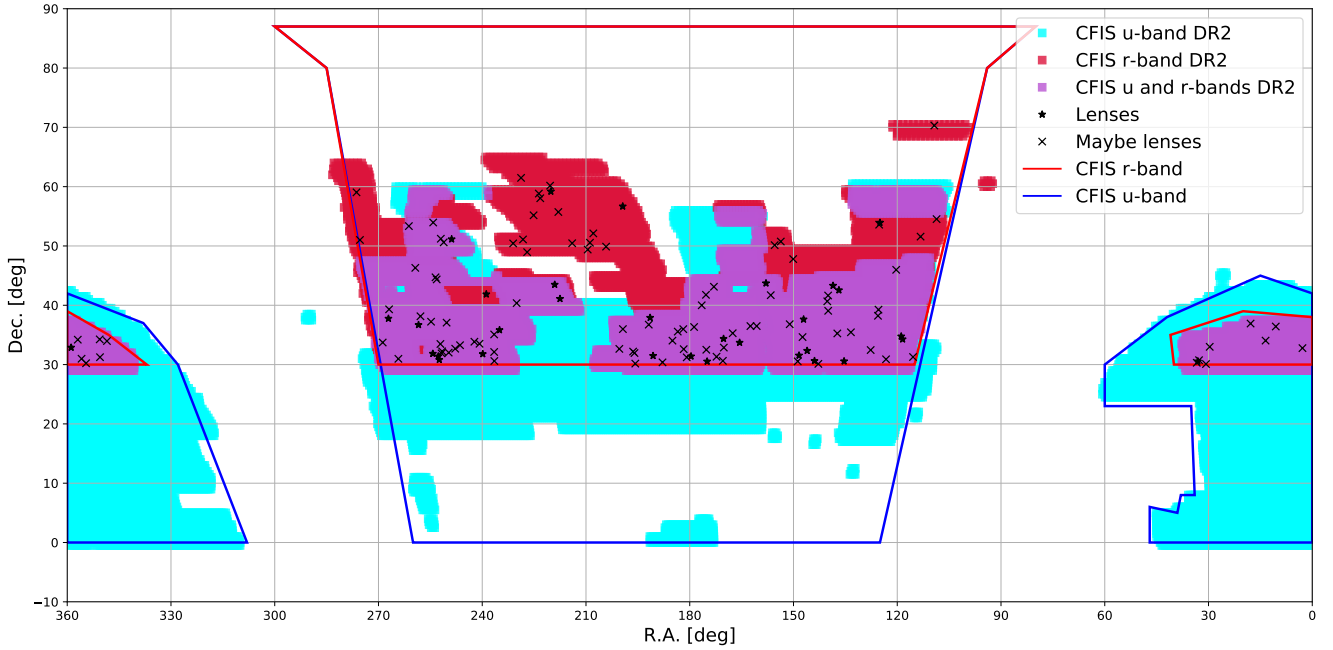
Fig. 1. Illustration of the planned and current CFIS footprint. The contours of the final CFIS-*r* footprint and the CFIS-*u* footprint are shown in red and blue, respectively. Shown are the current areas covered in the *r* band for DR2 (in red), in the *u* band (in blue), and where *r*-band and *u*-band 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.

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 the provide the necessary 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 8000 deg^2 of the northern sky in the *u* band (CFIS-*u*) and 4800 deg^2 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 2500 deg^2. CFIS-*r* has exquisite image quality with a median seeing of 0.6″, down to a depth of 24.1 (point source, 10σ rms). CFIS-*u* has a median seeing of 0.8″ to a depth of 23.6 (point source, 10σ 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*-rand 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 systems 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
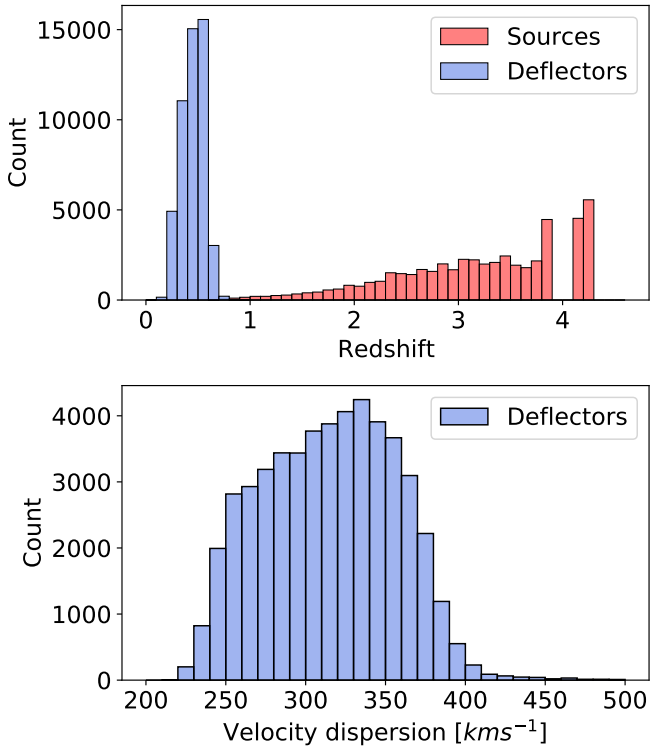
Fig. 2. Statistics of the deflectors and sources used in the lens simulations. *Upper panel*: redshift distribution of the sources and deflectors used for the lens simulations. *Lower panel*: measured velocity dispersion distribution for the deflectors.

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 ($\sigma_\star$) and redshift (z) estimates. In Fig. 2 we summarize the spectral properties of our selection, which spans the ranges $200 < \sigma_\star < 500\,\mathrm{km\,s^{-1}}$ 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 *F*814*W* 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 *F*814*W* 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 fore-

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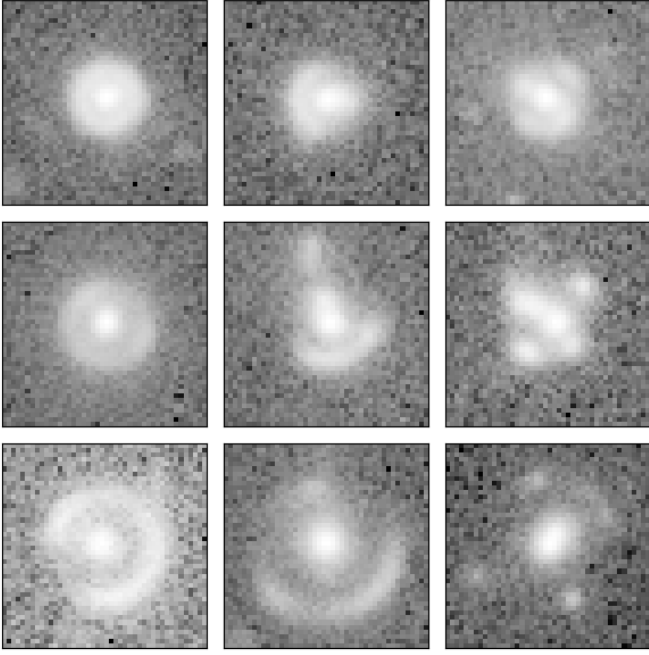
Fig. 3. Examples of simulated galaxy-scale systems of the CFIS survey. Each stamp is the sum of a real CFIS *r*-band image of a galaxy, to which a lensed HST galaxy convolved with the CFIS PSF is added. Each stamp is 8.17″ per side, and the pixel size is 0.18″.

therefore generated a set of simulated lenses using the pipeline described in Schuldt et al. (2021). 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 selected 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 0.2−0.5 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.
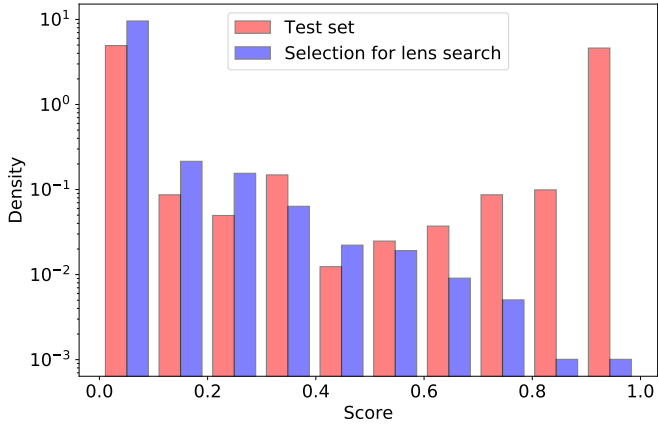
Fig. 4. Probability density distribution of the score assigned by the committee for the test set (red), containing 50% of lens simulations, and 50% of galaxies without lensing features taken from the lens search sample. In blue are shown the scores predicted by the network on the data set containing real images described in Sect. 2.1. In the lens search sample the proportion of lenses is by definition unknown.

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 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°, 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:

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

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

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 8200 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:

$$P(L|C) = \frac{P(L)P(C|L)}{P(L)P(C|L) + (1 - P(L))P(C|NL)},\qquad(3)$$
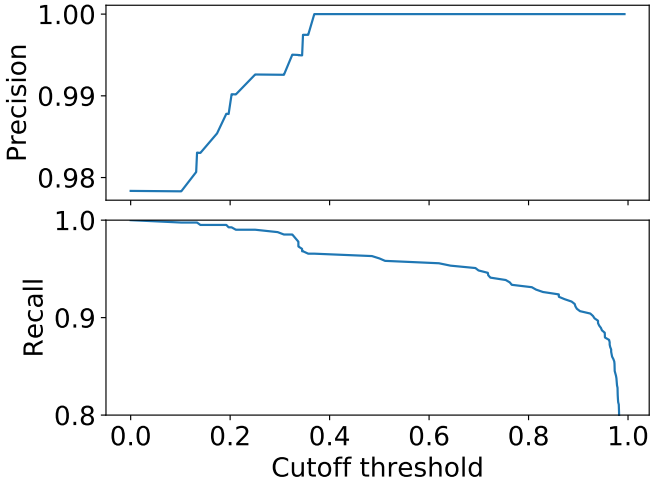
Fig. 5. Precision and recall as functions of the cutoff threshold in CNN scores applied to a test set composed of 1060 galaxies without lensing features taken from our selection for the lens search and 1060 lenses taken from our simulation set.
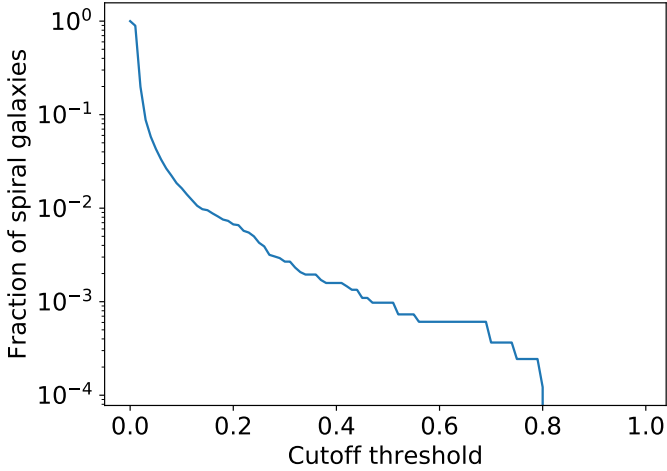
Fig. 6. Fraction of CFIS *r*-band images of 8200 Galaxy Zoo spiral (Willett et al. 2017) galaxies mislabeled as lenses as a function of the cutoff threshold on the CNN-committee score.

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, $P(L|C)$ 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, 9460 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 have a consistent visual inspection between the users and the bins, we defined common guidelines and designed a quick

and flexible `python` tool[1] 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 (4626 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 1423 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

---

[1] https://github.com/esavary/Visualisation-tool

different 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 1423 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 Figs. 7 and 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 Table of results of the cross-match and the coordinates of the candidates is available at the CDS. 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 an 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 `scarlet`[2] (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. 2021).

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 5000 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:

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

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

---

[2] https://github.com/pmelchior/scarlet

assumptions about the light profile of the lens and the source, 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 $6\sigma_{sky}$, 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
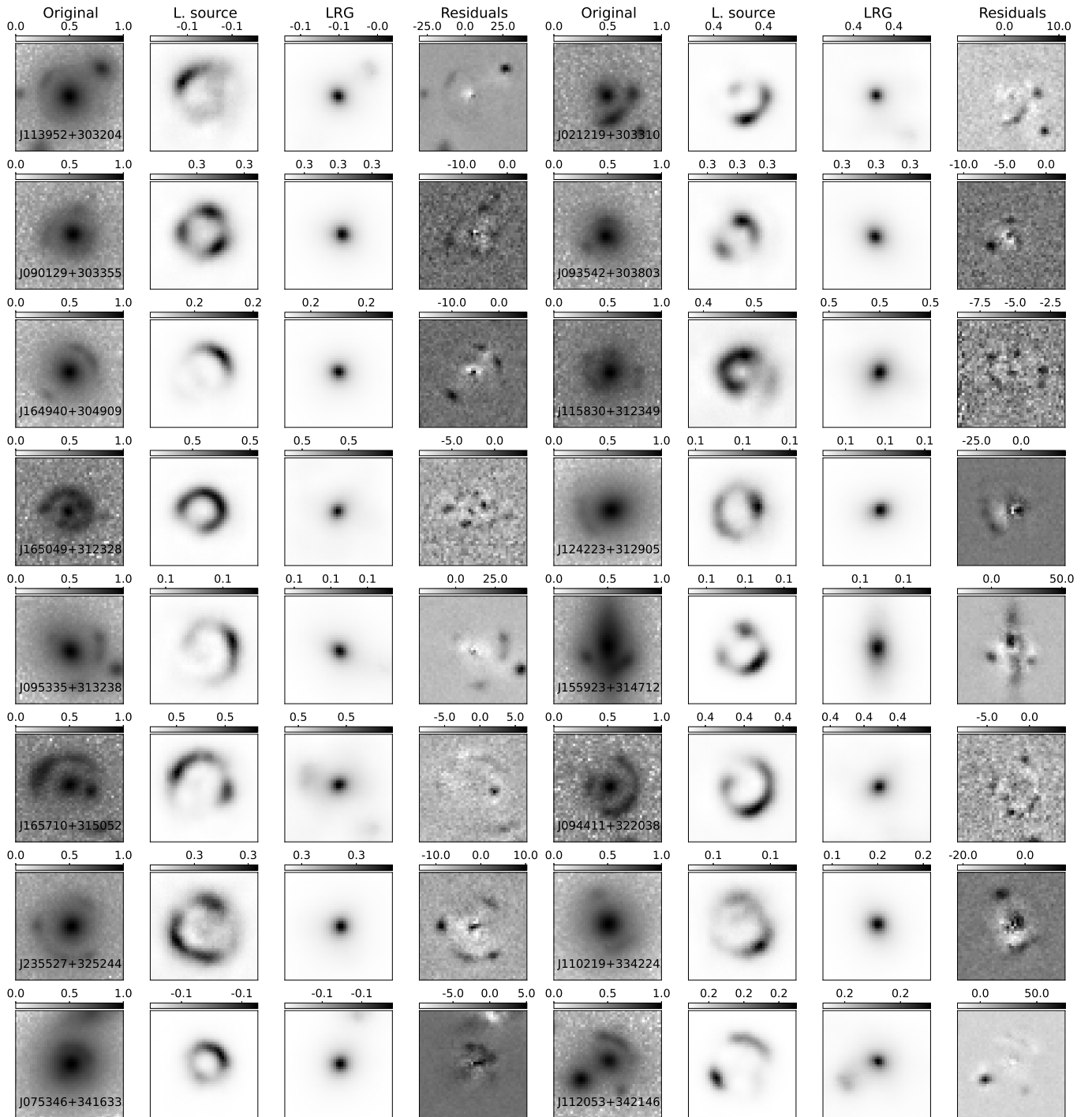
Fig. 7. Results of the auto-encoder deblending for our 32 SL candidates. Two objects are displayed in each row. For each object, we show the original image displayed using a asinh grayscale, and the lensed source and lens light deblended using the auto-encoder and the scaled residuals.

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 deflec-

tor 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
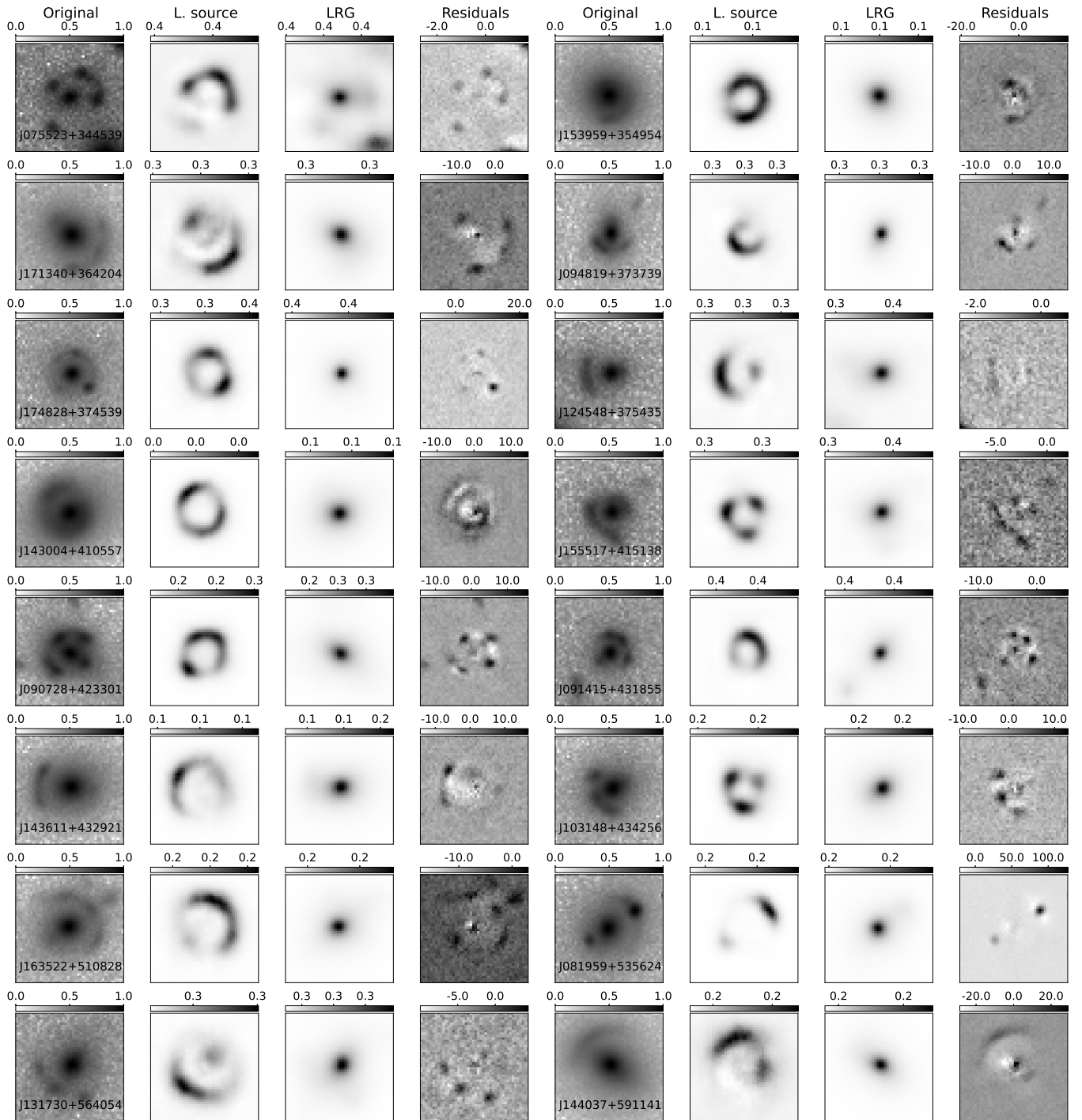
Fig. 7. continued.

as a training set for the GALSIM[5] galaxy image simulation software ([Rowe et al. 2015](#)). We show these prior distributions in $R_{\rm eff}$ and $n_{\rm s}$ 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 $1.2'' < R_{\rm E} < 2.5''$ 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 $1.2'' < R_{\rm E} < 2.5''$ in the SL sample also highlights the fact that the visual inspection predominantly selects obvious wide-separation lenses with deblended 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

---

5   https://github.com/GalSim-developers/GalSim
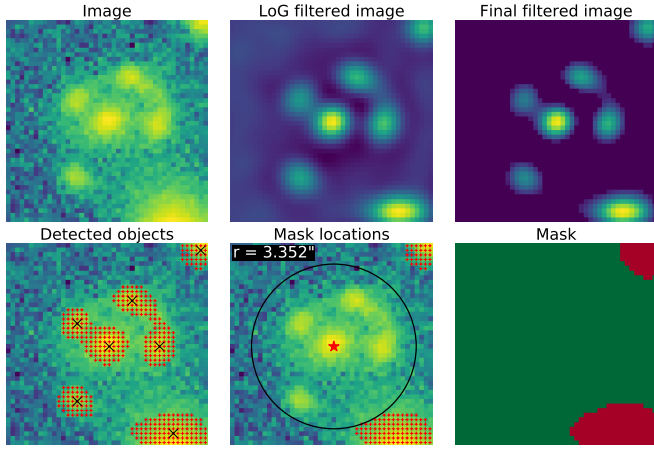
Fig. 8. Illustration of our masking procedure. *Upper row*: original CFIS image in the *r* band, followed by the image after applying the LoG filter in the *top middle panel*. This image is then thresholded and all pixels below $6\sigma_{\rm sky}$ are set to zero, as shown in the *top right panel*. *Bottom row*: detected peaks, whose centroids are marked with black crosses. *Bottom middle panel*: 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.
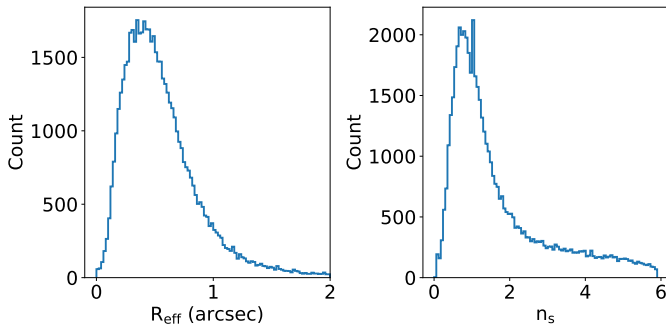
Fig. 9. Priors used for the effective radius $R_{\rm eff}$ and Sersic index $n_{\rm s}$ of the source light. The priors are derived from 56 062 galaxies from the COSMOS survey.

Fig. 10. Parameter distributions from our lens modeling results using the SIE+$\gamma_{\rm ext}$ model. *Top row*: gives the Einstein radii and external shear values of the mass model. *Middle* and *bottom rows*: effective radius and Sersic index, for the lens and source light, respectively. The models with an acceptable reduced χ^2 ($\leq$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 $\chi^2 < 3.0$.

the lens light the effective radius and Sersic index distributions peak at 2.5″ and 5.0, respectively, 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 $R_{\rm eff} \sim 0.2″$ and with Sersic index peaking at $n_{\rm s} \sim 1.0$, 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. The Table with the lists the best-fit parameters obtained from these fits is available ate the CDS. 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). 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.

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 h for a 44 pixel × 44 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 $2500\,\mathrm{deg}^2$ of CFIS in the present work since the survey is still ongoing; it is expected to reach a total area of $5000\,\mathrm{deg}^2$ 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 9460 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 validation 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 9460 objects passing our threshold due to the low prevalence of lenses in real data, as explained in Sect. 3.4. Fortunately, visually inspecting 9460 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 pixel × 44 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.

References

Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, PASJ, 70, S4

Amiaux, J., Scaramella, R., Mellier, Y., et al. 2012, in Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave, eds. M. C. Clampin, G. G. Fazio, H. A. MacEwen, J. Oschmann, & M. Jacobus, SPIE Conf. Ser., 8442, 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

Bellagamba, F., Tessore, N., & Metcalf, R. B. 2017, MNRAS, 464, 4823

Bertin, E. 2011, in Astronomical Data Analysis Software and Systems XX, eds. I. N. Evans, A. Accomazzi, D. J. Mink, & A. H. Rots, ASP Conf. Ser., 442, 435

Birrer, S., & Amara, A. 2018, Phys. Dark Univ., 22, 189

Birrer, S., Amara, A., & Refregier, A. 2015, ApJ, 813, 102

Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., & Moustakas, L. A. 2006, ApJ, 638, 703

Bolton, A. S., Burles, S., Koopmans, L. V. E., et al. 2008, ApJ, 682, 964

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

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, 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248

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

Fantin, N. J., Côté, P., McConnachie, A. W., et al. 2019, ApJ, 887, 148

Faure, C., Kneib, J.-P., Covone, G., et al. 2008, ApJS, 176, 19

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 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, Commun. Appl. Math. Comput. Sci., 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., Ravi, V., et al. 2020, ApJ, 894, 78

Huang, X., Storfer, C., Gu, A., et al. 2021, ApJ, 909, 27

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, ApJS, 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, MNRAS, 421, 1007

Kennedy, J., & Eberhart, R. 1995, Proceedings of ICNN'95 – International Conference on Neural Networks, 4, 1942

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, ApJS, 224, 24

Lanusse, F., Mandelbaum, R., Ravanbakhsh, S., et al. 2021, MNRAS, 504, 5543

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, ApJS, 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, Astron. Comput., 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., 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. 2017, MNRAS, 472, 1129

Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2019, MNRAS, 482, 807

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, ArXiv e-prints [arXiv:2109.00014]

Rowe, B. T. P., Jarvis, M., Mandelbaum, R., et al. 2015, Astron. Comput., 10, 121

Schuldt, S., Suyu, S. H., Meinhardt, T., et al. 2021, A&A, 646, A126

Shajib, A. J., Treu, T., Birrer, S., & Sonnenfeld, A. 2020, MNRAS, 503, 2380

Shu, Y., Bolton, A. S., Brownstein, J. R., et al. 2015, ApJ, 803, 71

Silverman, J. D., Kashino, D., Sanders, D., et al. 2015, ApJS, 220, 12

Sonnenfeld, A., Treu, T., Marshall, P. J., et al. 2015, ApJ, 800, 94

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., 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., & Halkola, A. 2010, A&A, 524, A94

Suyu, S. H., Hensel, S. W., McKean, J. P., et al. 2012, ApJ, 750, 10
Suyu, S. H., Bonvin, V., Courbin, F., et al. 2017, MNRAS, 468, 2590
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 [arXiv:astro-ph/0510346]
Turner, E. L., Ostriker, J. P., & Gott, J. R., III 1984, ApJ, 284, 1
Vegetti, S., & Koopmans, L. V. E. 2009, MNRAS, 392, 945
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
Vegetti, S., Koopmans, L. V. E., Auger, M. W., Treu, T., & Bolton, A. S. 2014, MNRAS, 442, 2017
Vegetti, S., Despali, G., Lovell, M. R., & Enzi, W. 2018, MNRAS, 481, 3661
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., 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

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.

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
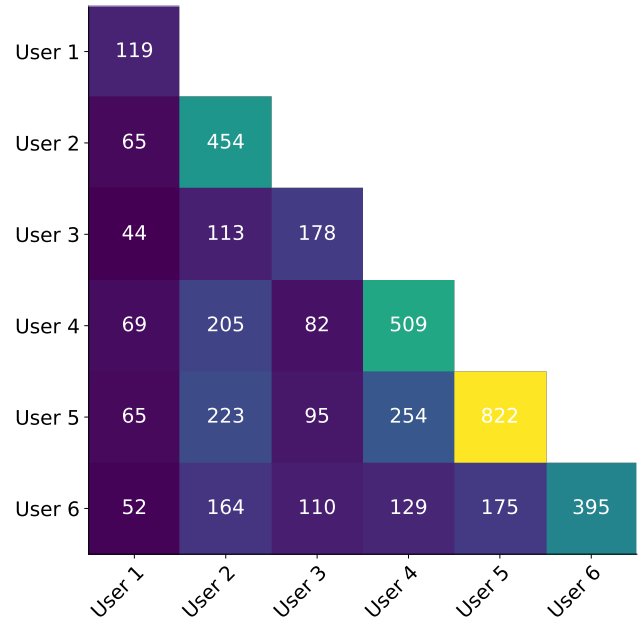
Fig. A.1. Illustration of the overlap between the visual classification of different users. The diagonal terms show the number of SL and ML labeled by each user during the first steps of the inspection, whereas the nondiagonal terms correspond to the number of ML and SL objects that are shared between 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.
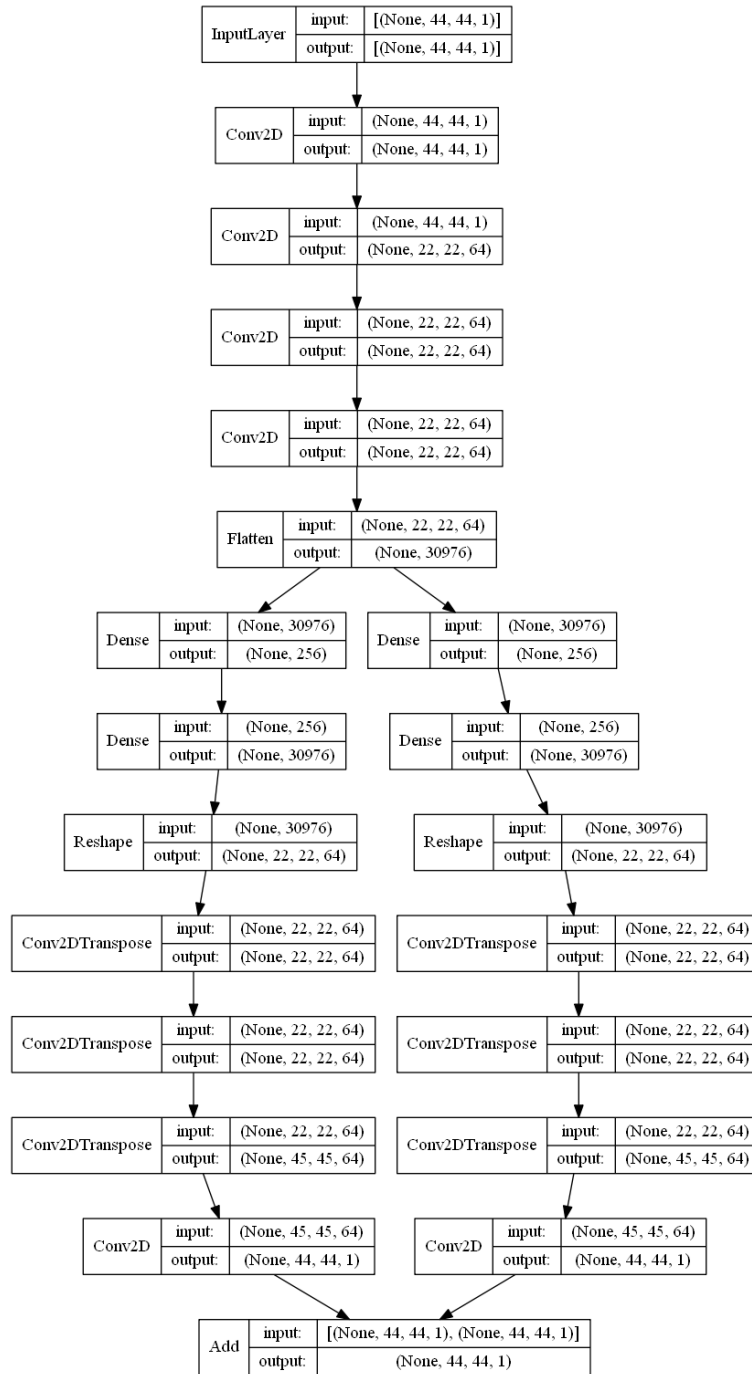
Fig. B.1. Architecture of the deblending auto-encoder. Each box represents a layer and the arrows show the connections between the different layers of the networks. The names indicated in the rectangles correspond to the different layer subclasses of the Keras API used in the model. The dimensions of the input and output of each layer are indicated in brackets. For convolutional layers the last dimension corresponds to the number of filters.

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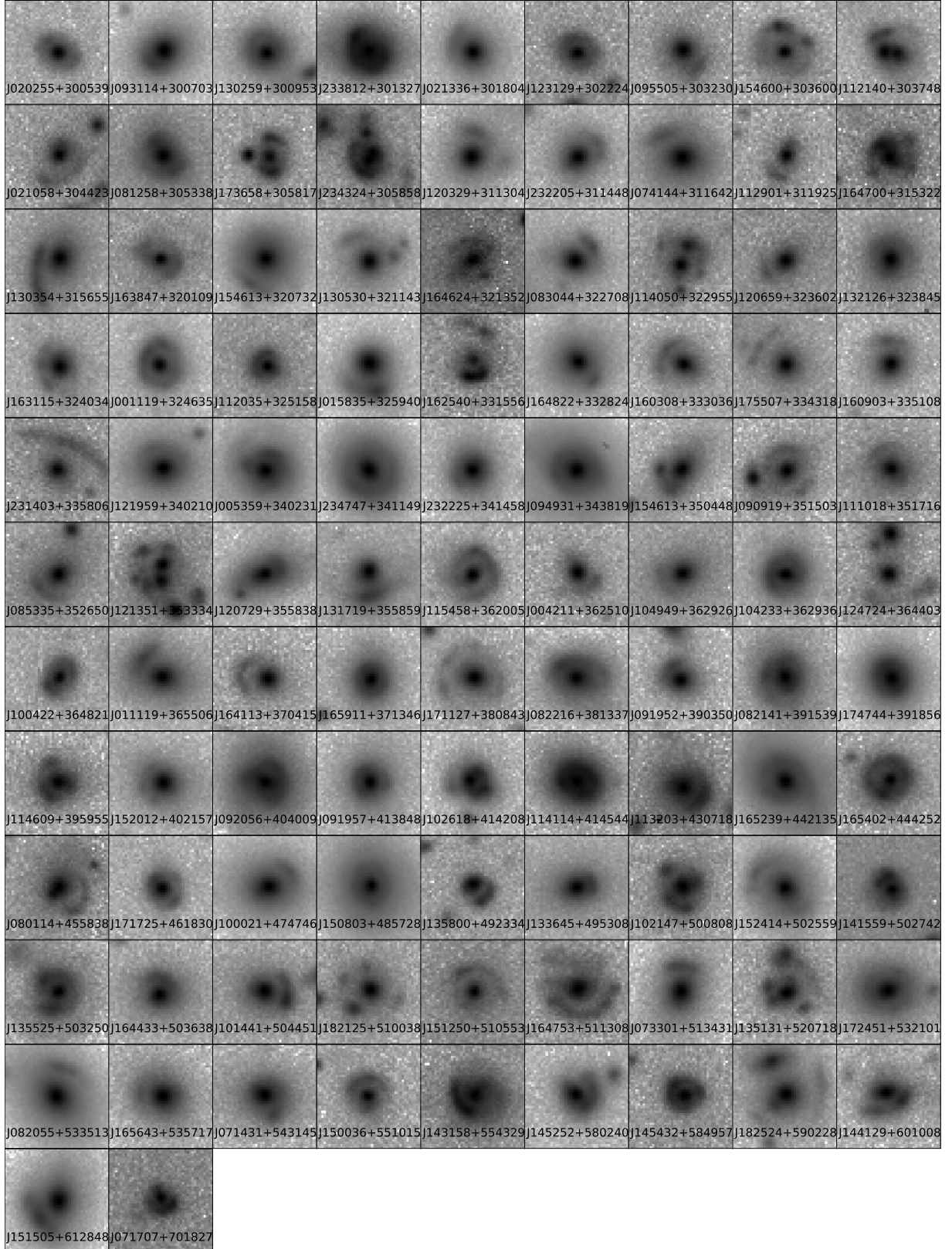
Fig. C.1. Images classified as maybe lenses (MLs) after the final visual inspection step.

Appendix D: Modeling results

We show in Figs. D.1–D.7 a mosaic with the modeling results for the 32 SL candidates.
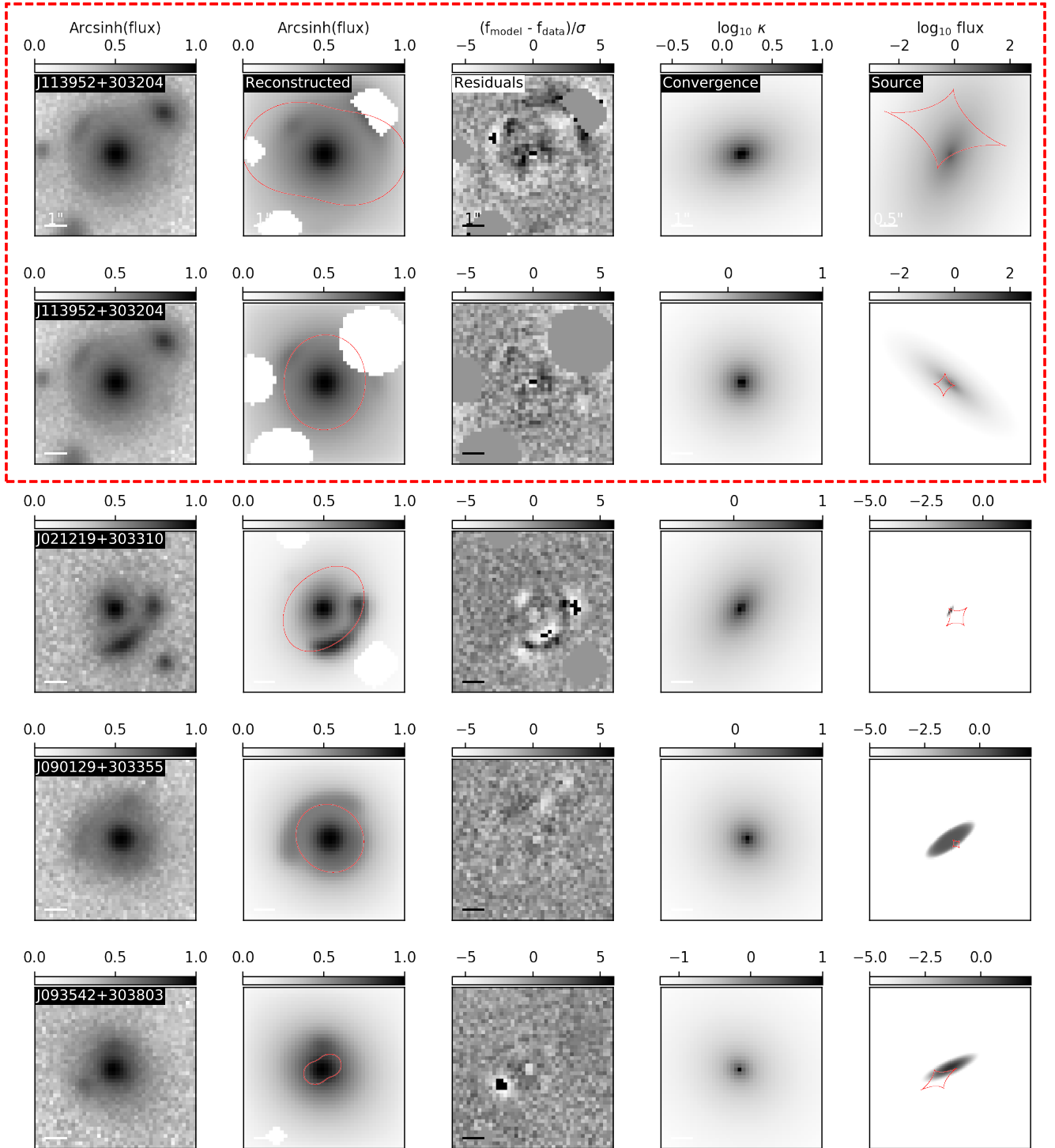
Fig. D.1. Modeling results for the first four of the 32 SL lens candidates. Shown inside the red dashed box are two modeling results for the same image, but with different masks. The top row corresponds to the results from the automated masking procedure, and in the bottom row to the results after applying 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.
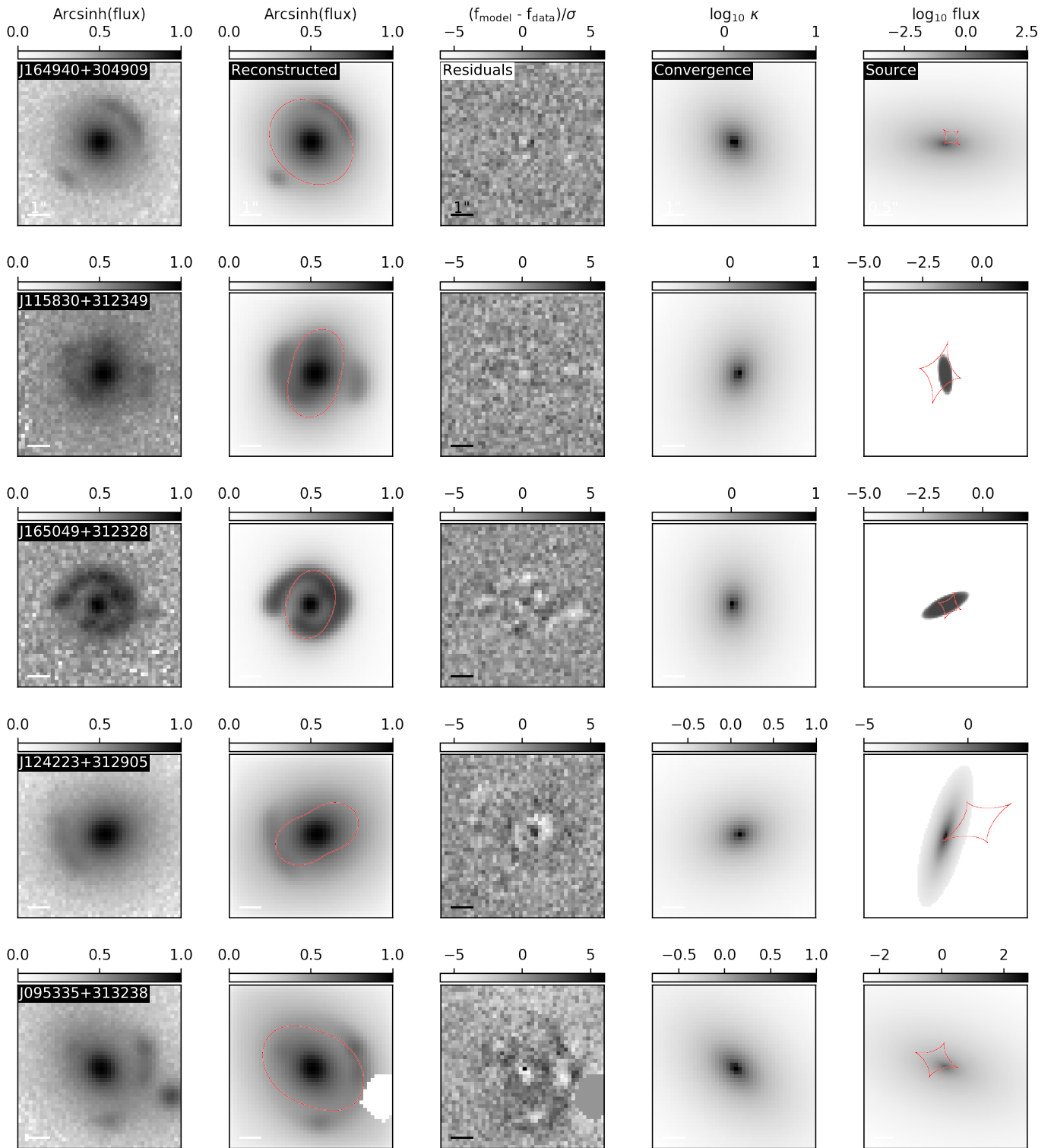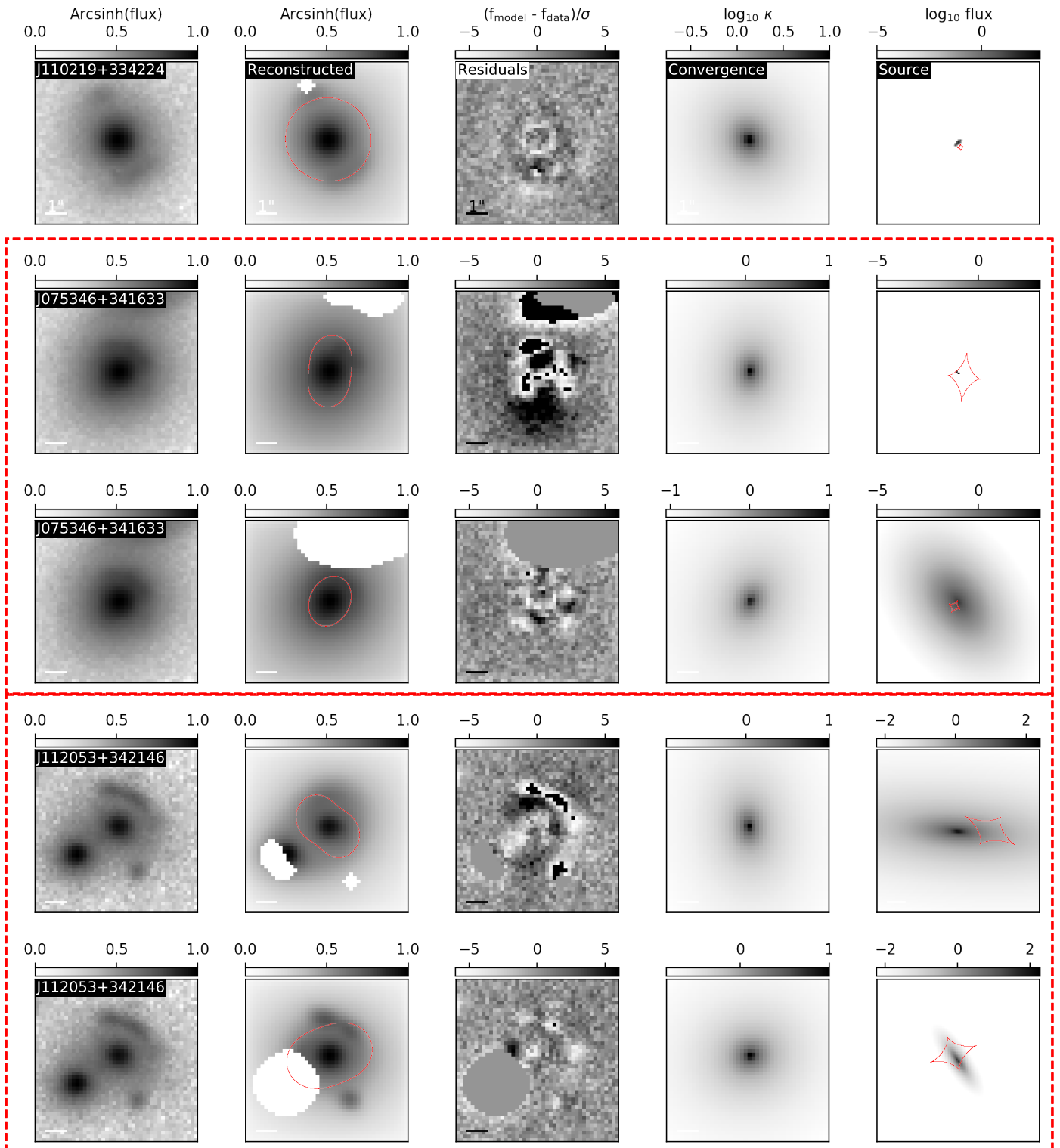
Fig. D.2. Continued from Fig. D.1.
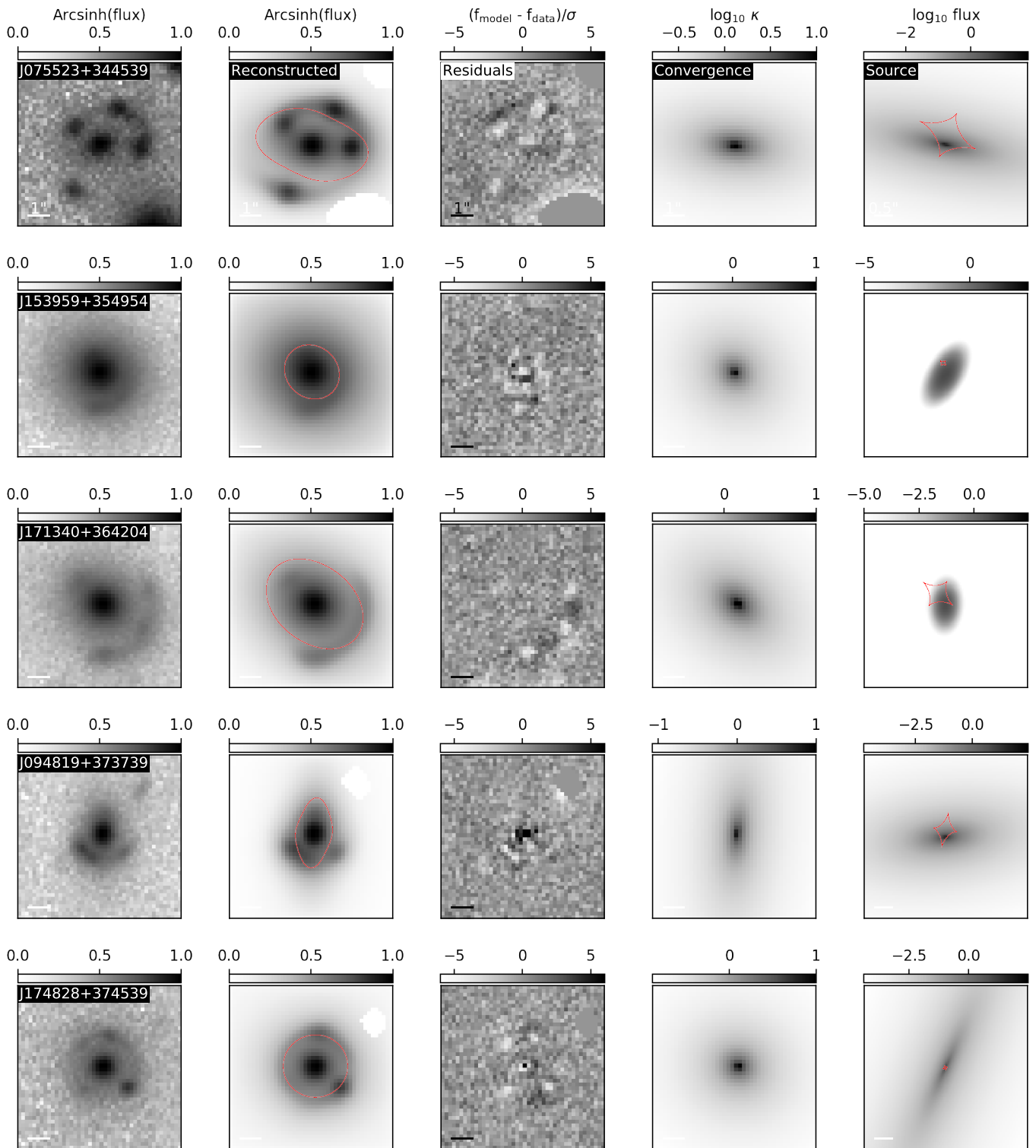
Fig. D.3. Continued from Fig. D.2.

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

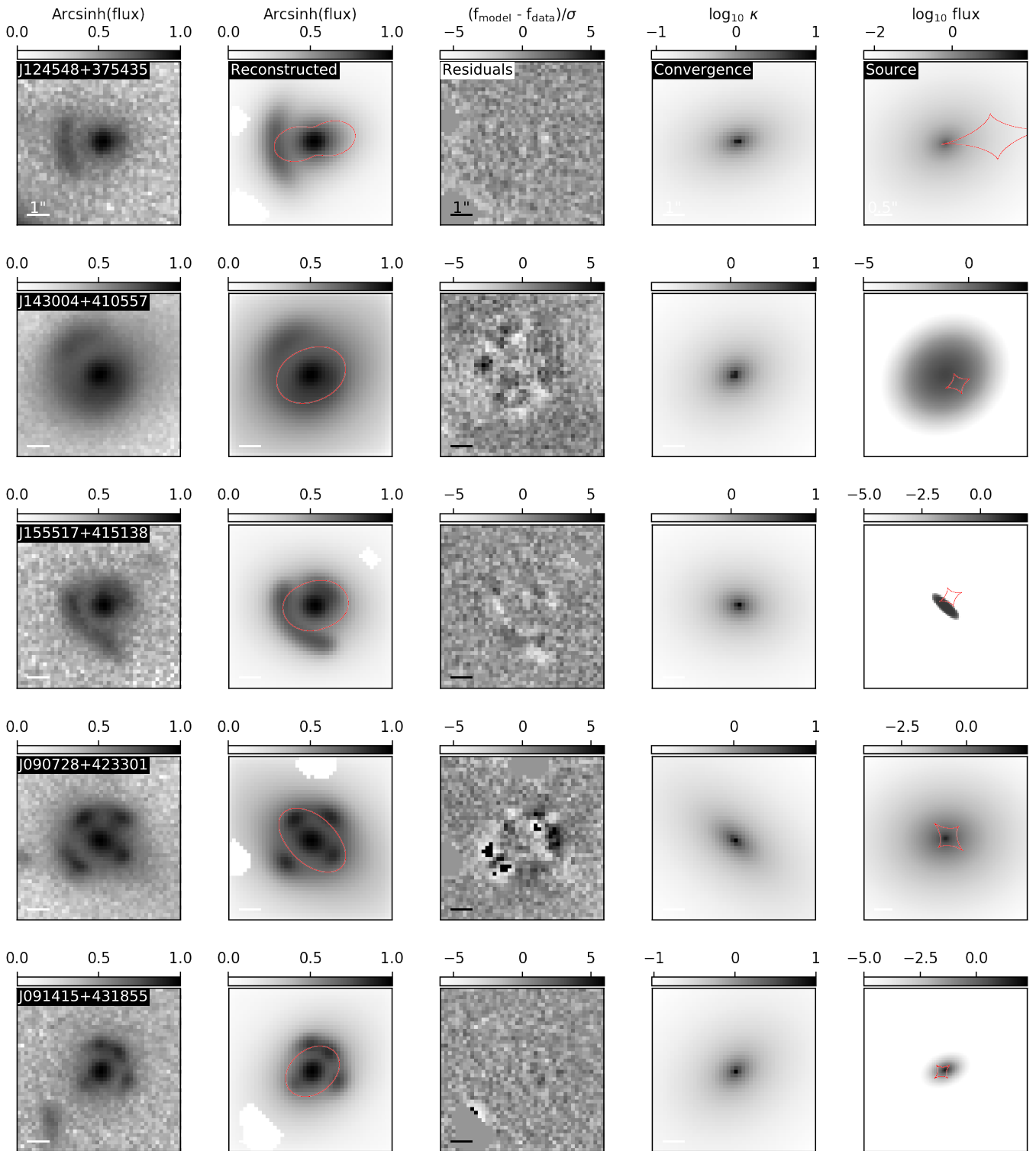**Fig. D.5.** Continued from Fig. D.4.
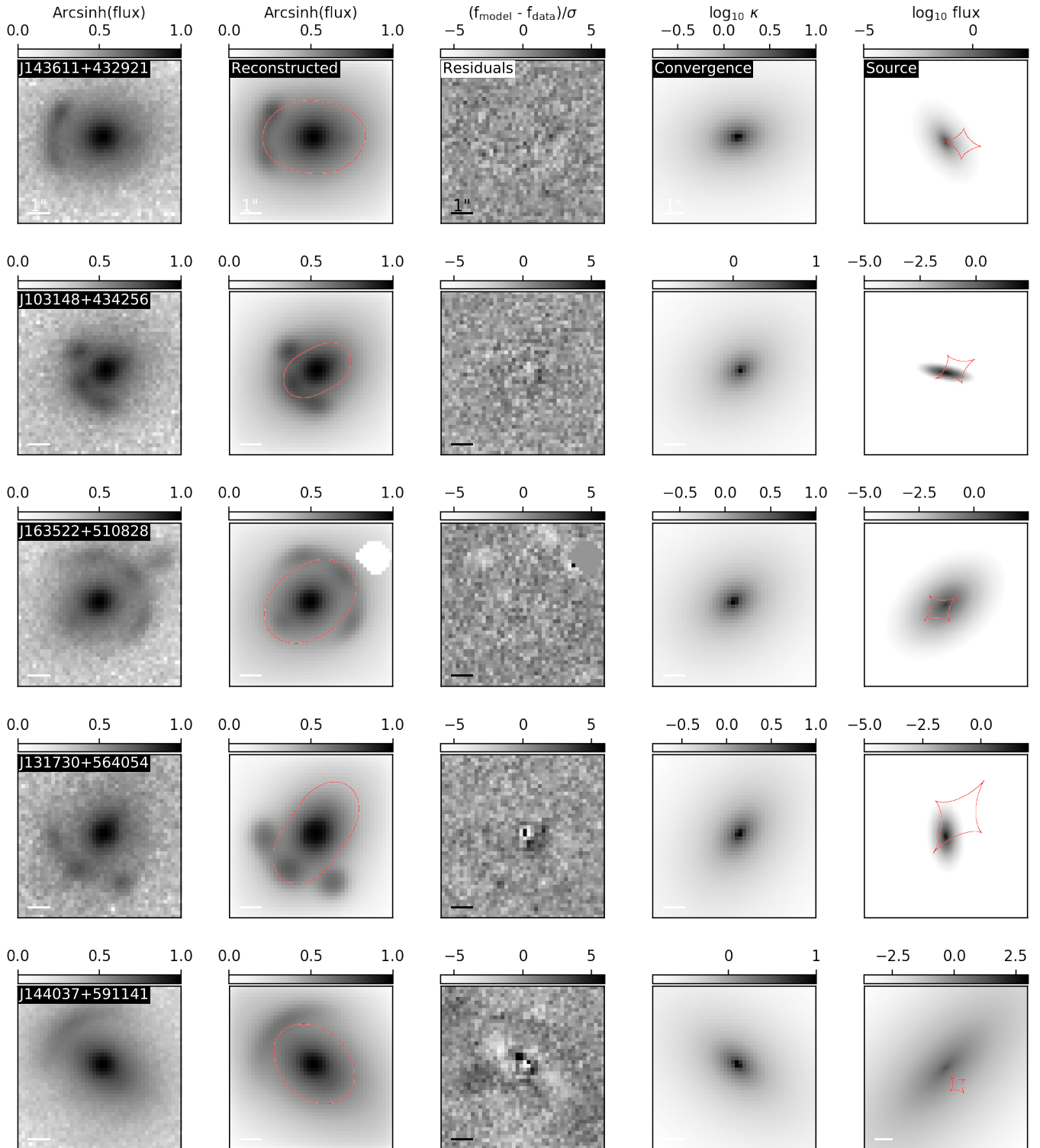
Fig. D.6. Continued from Fig. D.5.
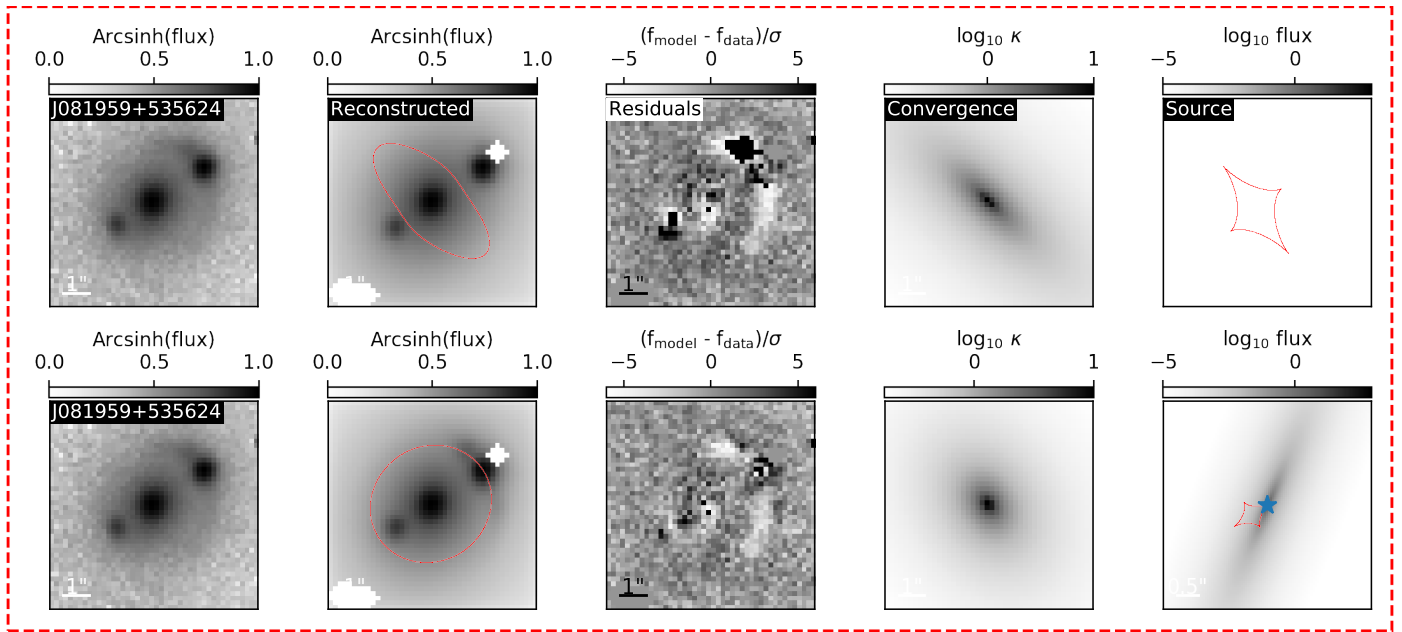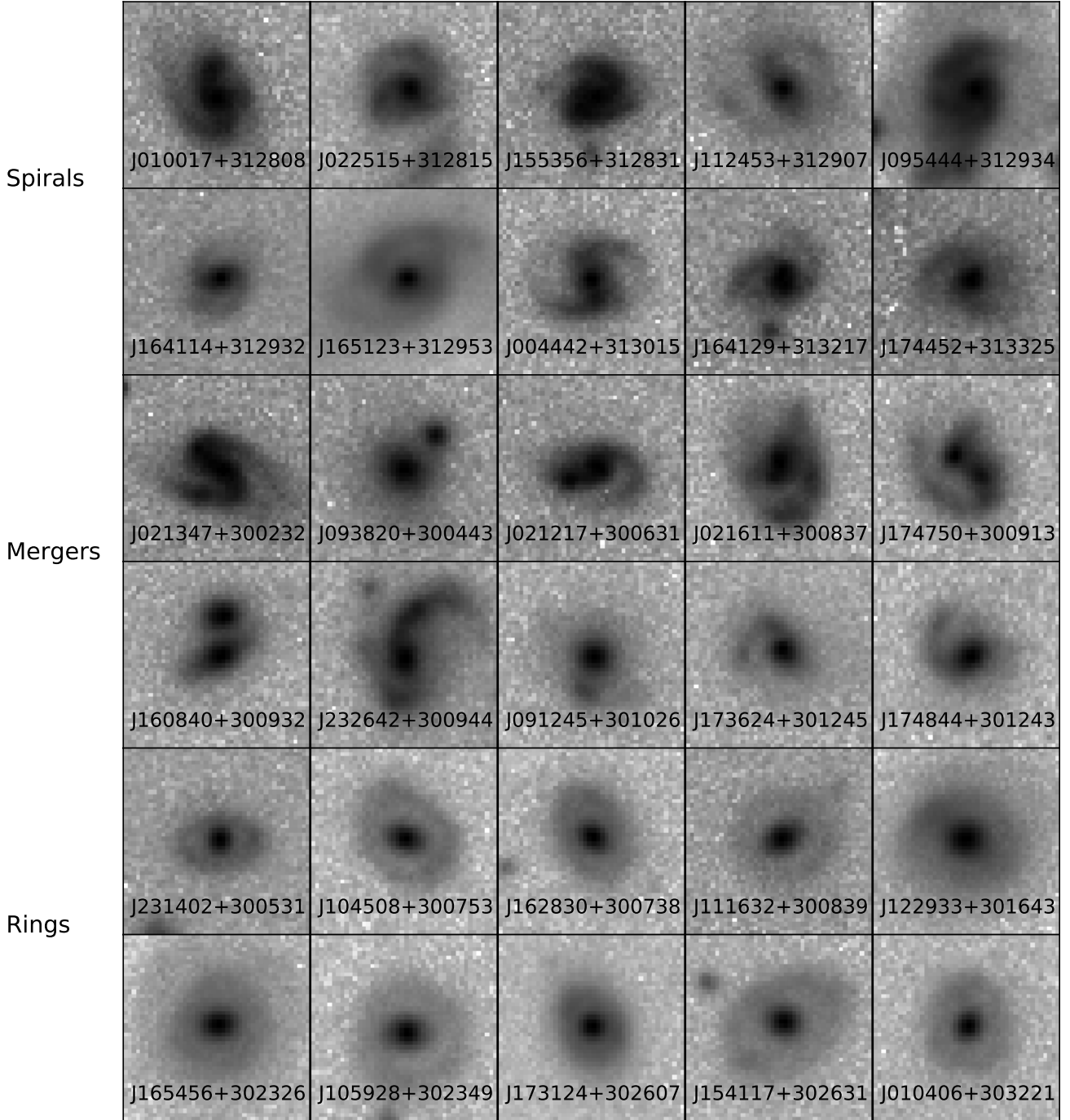
Fig. D.7. Continued from Fig. D.6.

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

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.

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.