



Publication Year	2019
Acceptance in OA	2021-01-07T12:46:11Z
Title	Testing convolutional neural networks for finding strong gravitational lenses in KiDS
Authors	Petrillo, C. E., TORTORA, CRESCENZO, Chatterjee, S., Vernardos, G., Koopmans, L. V. E., Verdoes Kleijn, G., NAPOLITANO, NICOLA ROSARIO, Covone, G., Kelvin, L. S., Hopkins, A. M.
Publisher's version (DOI)	10.1093/mnras/sty2683
Handle	http://hdl.handle.net/20.500.12386/29539
Journal	MONTHLY NOTICES OF THE ROYAL ASTRONOMICAL SOCIETY
Volume	482

Testing convolutional neural networks for finding strong gravitational lenses in KiDS

C. E. Petrillo,¹★ C. Tortora,¹ S. Chatterjee,¹ G. Vernardos,¹ L. V. E. Koopmans,¹
G. Verdoes Kleijn,¹ N. R. Napolitano,² G. Covone,^{3,4} L. S. Kelvin⁵ and
A. M. Hopkins⁶

¹Kapteyn Astronomical Institute, University of Groningen, Postbus 800, NL-9700 AV, Groningen, The Netherlands

²INAF - Osservatorio Astronomico di Capodimonte, Salita Moiariello, 16, I-80131 Napoli, Italy

³Dipartimento di Fisica ‘E. Pancini’, Università di Napoli Federico II, Compl. Univ. Monte S. Angelo, I-80126 Napoli, Italy

⁴INFN, Sezione di Napoli, Compl. Univ. Monte S. Angelo, I-80126 Napoli, Italy

⁵Astrophysics Research Institute, Liverpool John Moores University, IC2, Liverpool Science Park, 146 Brownlow Hill, Liverpool L3 5RF, UK

⁶Australian Astronomical Optics, Macquarie University, 105 Delhi Rd, North Ryde, NSW 2113, Australia

Accepted 2018 September 16. Received 2018 September 19; in original form 2018 July 13

ABSTRACT

Convolutional Neural Networks (ConvNets) are one of the most promising methods for identifying strong gravitational lens candidates in survey data. We present two ConvNet lens-finders that we have trained with a dataset composed of real galaxies from the Kilo Degree Survey (KiDS) and simulated lensed sources. One ConvNet is trained with single r -band galaxy images, hence basing the classification mostly on the morphology. While the other ConvNet is trained on $g-r-i$ composite images, relying mostly on colours and morphology. We have tested the ConvNet lens-finders on a sample of 21 789 luminous red galaxies (LRGs) selected from KiDS and we have analysed and compared the results with our previous ConvNet lens-finder on the same sample. The new lens-finders achieve a higher accuracy and completeness in identifying gravitational lens candidates, especially the single-band ConvNet. Our analysis indicates that this is mainly due to improved simulations of the lensed sources. In particular, the single-band ConvNet can select a sample of lens candidates with ~ 40 per cent purity, retrieving three out of four of the confirmed gravitational lenses in the LRG sample. With this particular setup and limited human intervention, it will be possible to retrieve, in future surveys such as *Euclid*, a sample of lenses exceeding in size the total number of currently known gravitational lenses.

Key words: gravitational lensing; strong – methods: statistical – galaxies: elliptical and lenticular, cD.

1 INTRODUCTION

Strong gravitational lensing is a phenomenon that originates when light rays propagating from a background source galaxy are deflected, on their way towards the observer, by the gravitational field of a foreground galaxy, creating multiple images, arcs, and/or rings around the foreground galaxy. Strong gravitational lensing is a unique probe for studying the (dark) matter distribution of galaxies and providing cosmological constraints. E.g. gravitational lenses have been used to measure the Hubble constant through time delays of lensed quasar images (e.g. Suyu et al. 2010, 2017; Bonvin et al. 2017) and to constrain the dark energy equation of state (e.g.

Biesiada, Piórkowska & Malec 2010; Cao, Covone & Zhu 2012; Collett & Auger 2014; Cao et al. 2015). Gravitational lensing also allows measuring the fraction of dark matter in the central regions of galaxies (Gavazzi et al. 2007; Jiang & Kochanek 2007; Cardone et al. 2009; Covone et al. 2009; Auger et al. 2010; Cardone & Tortora 2010; Grillo et al. 2010; Tortora et al. 2010; More et al. 2011; Ruff et al. 2011; Sonnenfeld et al. 2015) and to constrain the slope of the inner mass density profile (e.g. Treu & Koopmans 2002a,b; Koopmans & Treu 2003; Koopmans et al. 2006; More et al. 2008; Barnabè et al. 2009; Koopmans et al. 2009; Cao et al. 2016; Mukherjee et al. 2018). Moreover, studying gravitational lenses can constrain the stellar initial mass function (e.g. Möller, Kitzbichler & Natarajan 2007; Ferreras et al. 2010; Treu et al. 2010; Spiniello et al. 2011; Brewer et al. 2012; Posacki et al. 2015; Sonnenfeld et al. 2015, 2018b; Spiniello et al. 2015). Strong lensing also works as a

★ E-mail: petrillo@astro.rug.nl

‘cosmic telescope’, producing a magnified view of background objects otherwise not observable (e.g. Impellizzeri et al. 2008; Swinbank et al. 2009; Richard et al. 2011; Deane et al. 2013; Treu et al. 2015; Mason et al. 2017; Salmon et al. 2017; Kelly et al. 2018). Discovering new gravitational lenses allows placement of more precise constraints on the above-mentioned quantities (see e.g. Vegetti & Koopmans 2009; Barnabè et al. 2011; Li et al. 2016). For a comprehensive review of the scientific applications of strong lensing, see, e.g. Schneider, Ehlers & Falco (1992), Schneider, Kochanek & Wambsganss (2006), and Treu (2010).

Originally, gravitational lenses were found serendipitously in astronomical surveys, while currently they are considered as an important class of objects to systematically search in large sky surveys. The most successful campaign aiming at building a homogeneous dataset of strong gravitational lenses was the Sloan Lens ACS Survey (SLACS; Bolton et al. 2008) with more than 100 observed lenses that were identified by analyzing spectra from the Sloan Digital Sky Survey (SDSS; Eisenstein et al. 2011) that exhibited the imprint of two galaxies at two different redshifts. On-going optical surveys such as the Hyper Suprime-Cam survey (HSC; Miyazaki et al. 2012), the Kilo Degree Survey (KiDS; de Jong et al. 2015), and the Dark Energy Survey (DES; The Dark Energy Survey Collaboration 2005) are expected to provide in the coming years thousands of new lenses (see Collett 2015; Petrillo et al. 2017) and have already provided new lens candidates (Diehl et al. 2017; Petrillo et al. 2017; Sonnenfeld et al. 2018a). The future is bright also in the submillimetre wavelength, where *Herschel* (Negrello et al. 2010) and the South Pole Telescope (Carlstrom et al. 2011), coupled with the Atacama Large Millimetre/submillimetre Array, are providing several hundreds of new lens candidates (Vieira et al. 2013; Negrello et al. 2017). However, it is the next decade that holds a treasure trove of new gravitational lenses. It has been estimated that samples of $\sim 10^5$ strong lenses (Oguri & Marshall 2010; Pawase et al. 2014; Collett 2015; McKean et al. 2015) will be observed by *Euclid* (Laureijs et al. 2011), the Large Synoptic Survey Telescope (LSST; LSST Science Collaboration et al. 2009) and the Square Kilometre Array.¹

The huge number of possible new candidates, together with the difficulty of identifying them in the enormous volume of survey data, drives the growing effort in developing automatic lens-finders. Most are based on the identification of arc-like features (e.g. Lenzen, Schindler & Scherzer 2004; Horesh et al. 2005; Alard 2006; Estrada et al. 2007; Seidel & Bartelmann 2007; Kubo & Dell’Antonio 2008; More et al. 2012; Maturi, Mizera & Seidel 2014). Other approaches, such as described by Gavazzi et al. (2014) and Joseph et al. (2014), focus on the analysis of the residuals after subtracting the candidate lens galaxies from the astronomical images. Both methods have been used to find lens candidates in the Canada–France–Hawaii Telescope Legacy Survey (CFHTLS²) by Sonnenfeld et al. (2013), Gavazzi et al. (2014) and Paraficz et al. (2016). Instead, the algorithm developed by Chan et al. (2015) is specialized in identifying lensed quasars and together with the algorithm YATTALENS (Sonnenfeld et al. 2018a) has been applied to find lens candidates in HSC (Sonnenfeld et al. 2018a). Another approach, as in Brault & Gavazzi (2015), is modelling the probability that the targets are actual lenses. Stapelberg, Carrasco & Maturi (2017) applied the same strategy to clusters and groups of galaxies. Gravitational lenses have been identified also with citizen-science experiment approaches with the

Space Warps project (Marshall et al. 2016; More et al. 2016) where non-professional volunteers can classify galaxy images with the help of a web applet.³ Most recently, Petrillo et al. (2017) and Jacobs et al. (2017) have used Convolutional Neural Network (ConvNets) for finding lens candidates in KiDS and CFHTLS, respectively. Finally, Hartley et al. (2017) have used a technique based on support vector machines (SVMs) and applied it to KiDS. Instead, Spiniello et al. (2018) focused on the search of lensed quasars in KiDS using three different morphology based methods.

In this paper, we present and test our latest ConvNet lens-finders, improving on the work of Petrillo et al. (2017). ConvNets (Fukushima 1980; LeCun et al. 1998) are the state of the art and often the standard choice among machine learning algorithms for pattern recognition in digital images. The winners of the ImageNet Large Scale Visual Recognition Competition (ILSVRC; Russakovsky et al. 2015; the most important image classification competition) in recent years have all been groups utilizing ConvNets. The advantage of the latter method with respect to other pattern recognition algorithms is that the features are not hand-crafted but are themselves extracted automatically during the training procedure, thus the algorithm decides which features are most representative for classifying the images. The theoretical basis of ConvNets was developed in the 1980s and the 1990s. However, only recently ConvNets have started to outperform other algorithms thanks to the advent of large labelled datasets, improved algorithms and faster training, especially on graphics processing units. The interested reader is referred to the appendix for a brief introduction on ConvNets and to the reviews by Schmidhuber (2015), LeCun, Bengio & Hinton (2015), and Guo et al. (2016) for a more detailed introduction.

ConvNets have been used recently in many astronomical problems, e.g. galaxy morphology classification (Dieleman, Willett & Dambre 2015; Huertas-Company et al. 2015), estimation of photometric redshifts (Hoyle 2016; D’Isanto & Polsterer 2018), spectra classification (Hála 2014; Tao et al. 2018), identifying exoplanets (Shallue & Vanderburg 2018), transient detection (Cabrera-Vives et al. 2017), galaxy surface brightness estimation (Tuccillo et al. 2018), strong lensing parameters estimation (Hezaveh, Levasseur & Marshall 2017), and star/galaxy separation (Kim & Brunner 2017).

More importantly, Metcalf et al. (2018) presented the results of a large international *challenge* in which various methods of identifying simulated gravitational lenses were tested blindly. This challenge, the first of a series, sets out to prepare the community for finding lenses in the data of ESA’s *Euclid* mission. Its large data volume requires fast and efficient algorithms to identify strong gravitational lenses. However, the methods were also tested on simulated KiDS data. ConvNets and SVMs were recognized to be the most promising methods, among many different methods tested in the challenge.

The ConvNet lens-finders presented in this paper will be applied on ~ 900 square degrees of the KiDS survey in a forthcoming paper with the purpose of starting a systematic census of strong lenses named ‘LinKS’ (Lenses in KiDS Survey). The paper is organized as follows. In Section 2, we illustrate our lens-finding ConvNet-based algorithms and how the training dataset is built. In Section 3, we evaluate the performances of the ConvNets. In Section 4, we apply the lens-finders to $\sim 22\,000$ extracted from ~ 255 square degrees of KiDS for testing the algorithms on real data. Finally, in Section 5, we provide a summary and the main conclusions of this work.

¹<https://www.skatelescope.org/>

²<http://www.cfht.hawaii.edu/Science/CFHLS/>

³<https://spacewarps.org/>

2 TRAINING THE CONVNETS TO FIND STRONG LENSES

A ConvNet can be seen as a sequence of non-linear functions, called layers, that create, starting from an input image, a series of increasingly abstract representations of the input called feature maps. The final layer of the ConvNet converts the input feature maps into a set of numbers that represent the outcome of the classification. Hence, a ConvNet maps an image on to a single or few numbers. In our case, the output is a single number, denoted by p , which can vary between 0 and 1, and it is related to the probability that the input image is a lens (see Saerens, Latinne & Decaestecker 2002 for a detailed discussion). The parameters of the non-linear functions are obtained during the so called training phase where labelled images are fed to the ConvNet. In more detail, the parameters are derived by minimizing a loss function that expresses the difference between the label values of the images (1 for lenses, 0 for non-strong-lensing systems) and the output p of the ConvNet. Although in Petrillo et al. (2017), we have used a similar set-up, the aim of this work is to improve the performance of our previous lens-finder.

Currently, we use a ConvNet with a *ResNet*-type architecture that has 18 layers, exactly as described in He et al. (2015b). *ResNet*-type architectures are often the preferred choice in image classification tasks because of their faster convergence and higher classification accuracy with respect to other architectures. Moreover, *ResNet* architectures have already been tested successfully on identifying simulated lenses and they have proven to be one of the best architecture for this task (Lanusse et al. 2018; Metcalf et al. 2018; Schaefer et al. 2018). We train two different ConvNets with the same architecture except that one takes in input RGB-images composed with HUMVI⁴ (Marshall et al. 2016), while the other takes single r -band images as input. We choose the r band as single-band input because the KiDS observing strategy reserves the best seeing conditions for this band (which is used for the weak lensing studies; Kuijken et al. 2015; Hildebrandt et al. 2017). The technical details of the ConvNet and of the training procedure are described in Appendix A together with a brief introduction on ConvNets.

To produce the data used to train and validate the ConvNets, we adopt a hybrid approach similarly as done in Petrillo et al. (2017), Jacobs et al. (2017) and Pourrahmani, Nayyeri & Cooray (2018), creating mock images of strong gravitational lenses using images of real galaxies from KiDS and superimposing simulated lensed images. We adopt this approach because we do not have a sample of genuine KiDS lenses large enough to train a ConvNet (usually of the order of 10^6).

2.1 Data

In this section, we describe the dataset used to train the ConvNets, which is composed of real KiDS galaxies and simulated lensed sources.

2.1.1 Luminous red galaxies

We use the sample of luminous red galaxies (LRGs; Eisenstein et al. 2001) presented in Petrillo et al. (2017). We choose to focus on massive early-type galaxies, because it has been estimated that these

galaxies form ~ 80 per cent of the lens-galaxy population (Turner, Ostriker & Gott 1984; Fukugita et al. 1992; Kochanek 1996; Chae 2003; Oguri 2006; Möller et al. 2007). Spiral galaxies form the other ~ 20 per cent but are much harder to identify. This training sample of LRGs is a random subset of 6554 galaxies from a parent sample of 21 789 selected from 255 square degrees of KiDS DR3 (de Jong et al. 2017) with the following criteria (see Petrillo et al. 2017 for more details).

(i) The low- z ($z < 0.4$) LRG colour-magnitude selection of Eisenstein et al. (2001), adapted to include more sources, both fainter and bluer:

$$\begin{aligned} r &< 20 \\ |c_{\text{perp}}| &< 0.2 \\ r &< 14 + c_{\text{par}}/0.3, \end{aligned} \quad (1)$$

where

$$\begin{aligned} c_{\text{par}} &= 0.7(g - r) + 1.2[(r - i) - 0.18] \\ c_{\text{perp}} &= (r - i) - (g - r)/4.0 - 0.18. \end{aligned}$$

(ii) A source size in the r band larger than the average full width at half-maximum (FWHM) of the point spread function (PSF) of the respective tiles, times an empirical factor to maximize the separation between stars and galaxies.

2.1.2 Contaminants

Moreover, we have used a set of ~ 6000 KiDS sources to train the ConvNets to recognize sources that would likely be incorrectly classified as lenses otherwise, either because they can resemble lensing features or they are ‘ghosts’, i.e. they are undetected, at least significantly, in the LRGs sample discussed in Section 2.1.1.

(i) ~ 2000 sources wrongly classified as lenses in previous tests with ConvNets identified by the authors. This is done to teach the ConvNets not to replicate previous mistakes.

(ii) ~ 3000 randomly extracted KiDS sources with r -band magnitude brighter than 21. To provide the network with general true negatives.

(iii) ~ 1000 KiDS sources visually classified as spiral galaxies from an on-going new project of GalaxyZoo (Willett et al. 2013, Kelvin et al., in preparation). This is done to decrease the false positives due to spiral features. To select the galaxies, we used a preliminary reduced version of the GAMA-KiDS Galaxy Zoo catalogue for the GAMA09 9 h region (see Driver et al. 2011 for further details). This catalogue contains $\sim 10^4$ sources out to a redshift of $z = 0.15$. We select galaxies for which a large majority of people replied to the question ‘Is the galaxy in the centre of the image simply smooth and rounded, or does it have features?’ with ‘it has features’.⁵

There is a non-zero probability that among the contaminants and the LRGs described in the previous section, there are actual gravitational lenses. We can estimate that the percentage would be of the order of 10^{-2} among the contaminants and ~ 1 per cent among the LRGs (Petrillo et al. 2017). Thus, even if real lenses are actually in the training sample, with such a small percentage they would not contaminate the training procedure.

⁵The actual selection is done by selecting sources from the catalogue with a value of the attribute `features.features_frac` larger than 0.6.

⁴<https://github.com/drphilmarshall/HumVI>

Table 1. Range of parameter values adopted for simulating the lensed sources. The parameters are drawn uniformly, except for *Einstein* and effective radius, as indicated. See Section 2.1.3 for further details.

Parameter	Range	Unit
Lens (SIE)		
<i>Einstein</i> radius	1.0–5.0 (log)	arcsec
Axis ratio	0.3–1.0	–
Major-axis angle	0.0–180	degree
External shear	0.0–0.05	–
External-shear angle	0.0–180	degree
Main source (Sérsic)		
Effective radius (R_{eff})	0.2–0.6 (log)	arcsec
Axis ratio	0.3–1.0	–
Major-axis angle	0.0–180	degree
Sérsic index	0.5–5.0	–
Sérsic blobs (1 up to 5)		
Effective radius	(1–10 per cent) R_{eff}	arcsec
Axis ratio	1.0	–
Major-axis angle	0.0	degree
Sérsic index	0.5–5.0	–

2.1.3 Mock lensed-sources

We simulate 10^6 lensed images of 101×101 pixels, using the same spatial resolution of KiDS (~ 0.2 arcsec per pixel), corresponding to a 20×20 arcsec field of view. To produce more realistic lensing systems, we add more complexity both in the source and in the lens plane with respect to the simulations in Petrillo et al. (2017). The distribution of the lens and source parameters that we choose for simulating the lensed images are chosen to create a wide range of realistic lensed images. They are not meant to statistically represent a real lens population, since the training set has to be populated sufficiently densely in the parameter space to allow the ConvNets to learn all the possible configurations and to recognize lenses that are rare in a real distribution (or currently even unknown). To ensure this, a more homogeneous distribution of the parameters is advantageous in order not to overtrain on the most common lens configurations.

We proceed in the following way, we sample the parameters of the singular isothermal ellipsoid (SIE; Kormann, Schneider & Bartelmann 1994) and Sérsic (1968) source models as listed in Table 1. The values of the lens *Einstein* radius and the source effective radius are drawn from a logarithmic distribution, while the remaining parameters, listed in Table 1, are drawn from a uniform distribution. In this way, our simulation sample contains a higher fraction of smaller rings and arcs compared to Petrillo et al. (2017) for making the new ConvNets more sensitive to this kind of objects with respect to the old one. The source positions are chosen uniformly within the radial distance of the tangential caustics plus one effective radius of the source Sérsic profile. This leads our training set to be mostly composed of high-magnification rings, arcs, quads, folds, and cusps rather than doubles (Schneider et al. 1992). To add complexity in the lensed sources, besides the Sérsic profile, we add between 1 and 5 small circular Sérsic blobs. The centres of these blobs are drawn from a Gaussian probability distribution function (PDF) around the main Sérsic source. The width of the standard deviation of the PDF is the same as the effective radius of the main Sérsic profile. The sizes of the blobs are chosen uniformly within 1–10 per cent of the effective radius of the main Sérsic profile. The Sérsic indices of

the blobs are drawn using the same prescription as for the main central source. The amplitudes of the blobs are also chosen from a uniform distribution in such a way that the ratio of the amplitude of an individual blob to the amplitude of the main Sérsic profile is at most 20 per cent.

Moreover, we add Gaussian random field (GRF) fluctuations to the lens potential, which, to a first-order approximation, make the lens sample more realistic by adding small scale substructures (Chatterjee & Koopmans 2018). The GRF realizations we added in our simulations all follow a power-law power-spectrum with a fixed exponent -6 , which is to the first order a good approximation of substructures in lens plane in the Λ cold dark matter paradigm (Hezaveh et al. 2014). The variances of the realizations are drawn from a logarithmic distribution between 10^{-4} and 10^{-1} about mean zero in the units of square of the lensing potential. This yields both structured sources and lenses that are not perfect SIE.

For each source, a realistic colour is simulated to create images in g , r , i bands. In order to produce realistic 3-band images, we extract magnitudes from ‘COSMOS’ models in LE PHARE (Arnouts et al. 1999; Ilbert et al. 2006). This library of spectra, consists of 31 models, used for COSMOS photo- z (Ilbert et al. 2009). The basic ‘COSMOS’ library is composed of eight templates for elliptical/S0 galaxies, 11 for spiral types, and 12 for galaxies with star-burst ages ranging from 0.03 to 3 Gyr, allowing us to span a wide range of galaxy types and colours. In order to simulate the typical blue arcs observed in most of the observed lenses, we choose models bluer than S0 and calculate observer-frame magnitudes in the three KiDS wavebands g , r , and i for model spectra redshifted up to a redshift of $z = 3$ with a 0.1 binning. Moreover, to populate the magnitude-space more uniformly, we perturb the three magnitudes adding to each of them a random number uniformly extracted from the range $[-0.1, 0.1]$ mag. We also take into account dust extinction by considering a colour excess $E(B - V)$, we extract it from a normal distribution with $\sigma = 0.1$ and mean 0 considering only the positives values. In this way, we obtain a small extinction correction in order to avoid very red lensed sources that, in the real universe, are much rarer than blue ones. We adopt a typical extinction curve with $R_V = 3.1$, using the relation $A_x = R_x E(B - V)$ where x represents the value for the g , r , and i SDSS-filters that can be found in table 2 of Yuan, Liu & Xiang (2013). Finally, we convolve the three images with an average KiDS-DR3 PSF for each different band: with an FWHM of ~ 0.86 arcsec for g , ~ 0.68 arcsec for r , and ~ 0.81 arcsec for i (de Jong et al. 2017).

2.2 Creating the training set

The data presented above are used to build the training set that is composed of mock strong-lensing systems (labelled with a 1) and non-strong-lensing systems (labelled with a 0), i.e. objects without lensing features. In the following, we outline the procedure used to build the two kinds of objects in the training set.

Mock strong-lensing systems. To create mock strong-lensing systems, we carry out the following procedure.

(i) We randomly choose a mock lensed source (Section 2.1.3) and an LRG (Section 2.1.1); we rescale the brightness of the simulated source to the peak brightness of the LRG in the r -band multiplied by a factor α randomly drawn from the interval $[0.02, 0.3]$. This accounts for the typical lower brightness of the lensing features with respect to the lens galaxies;

(ii) we stack the LRG and the mock source for each one of the three bands;



Figure 1. Examples of RGB images of simulated strong lens galaxies used to train the ConvNets. The lens galaxies are observed KiDS galaxies, while the lensed sources are simulated, as described in Section 2.1.3.

(iii) for the single-band images, we clip the negative values of the pixels to zero and performing a square-root stretch of the image to emphasize lower luminosity features. Instead, we create 3-band images with HUMVI that operates an arcsinh stretch of the image following the Lupton et al. (2004) composition algorithm;

(iv) finally, we normalize the resulting images by the galaxy peak brightness (only for single-band images).

Some examples of mock strong-lensing systems obtained in this way are shown in Fig. 1.

Non-strong-lensing systems. To create the non-strong-lensing system sample, we carry out the following procedure:

(i) we choose a random galaxy from either the LRG sample (with a probability of 20 per cent) or from the contaminant sample (80 per cent probability);

(ii) we clip the negative values of the pixels to zero and performing a square-root stretch of the images. We create 3-band images with HUMVI;

(iii) we normalize the images by the galaxy peak brightness (only for single-band images).

Finally, we augment the images, which is a standard procedure in machine learning (see e.g. Simard, Steinkraus & Platt 2003). It is used to avoid overfitting by expanding artificially the training set through different transformations of the images. Before feeding the images to the ConvNets, we apply the following transformations:

- (i) a random rotation between 0 and 2π ;
- (ii) a random shift in both x and y direction between -4 and $+4$ pixels;
- (iii) a 50 per cent probability of horizontally flipping the image;
- (iv) a rescaling with a scale factor sampled log-uniformly between 1 and 1.1.

All transformations are applied to both the mock strong-lensing systems and the non-strong-lensing systems. The final set of inputs of the ConvNets are postage stamps of 101 times 101 pixels that correspond to $\sim 20 \times 20$ arcsec. The images are produced in real-time during the training phase. For more details on the training phase see Appendix A.

3 ANALYSIS

After the training is completed, the ConvNets must be tested in order to assess whether the training was successful. In this section, we define the metric for evaluating the results and evaluate the performances of the ConvNets on a dataset composed by non-lenses and mock lenses in comparison to Petrillo et al. (2017).

3.1 Performance metric

To evaluate the performances of the ConvNets, we use:

(i) the true-positive rate (TPR), which measures the fraction of positive objects (in our case the lenses) detected by the algorithm. It is given by the ratio between the number of real positive (the number of real lenses that algorithm finds) and the sum of the latter and the number of false negatives (the lenses that the algorithm does not find):

$$\text{TPR} = \frac{N_{\text{Truepositives}}}{N_{\text{Truepositives}} + N_{\text{Falsenegatives}}} \in [0, 1]; \quad (2)$$

(ii) the false-positive rate (FPR), which measures the fraction of negative objects (non-strong-lensing systems) misclassified as positives (lenses). It is given by the ratio between the number of false positive (the number of non-lenses that algorithm misclassifies as lenses) and the sum of the latter and the number of true negatives (the non-lenses that the algorithm classifies correctly)

$$\text{FPR} = \frac{N_{\text{Falsepositives}}}{N_{\text{Truenegatives}} + N_{\text{Falsepositives}}} \in [0, 1]; \quad (3)$$

(iii) these two quantities can be used to build receiver operating characteristic (ROC) curves that allow us to check at a glance the degree of completeness and contamination of a binary classifier. ROC curves are created by plotting TPR as a function of FPR varying the threshold of detection for p between 0 and 1. This allows us to tune the value for the threshold for p in order to get the desired amount of TPR and FPR for a given classification problem. In our case, p is the output of the ConvNet and we can tune the p -threshold depending how many lens candidates we desire and what level of contamination is deemed to be acceptable.

3.2 Performance

The ROC curves for a test-set composed of 5000 mock strong-lensing systems and 5000 non-strong-lensing systems created, as described in Section 2, are shown in Fig. 2. In general, the 3-bands ConvNet has a better performance than the 1-band ConvNet, retrieving more mock strong-lensing systems than the 1-band ConvNet. On the contrary, the 1-band ConvNet is less contaminated by false positives at higher values of the threshold for p . Since gravitational lenses are rare events, it is important to keep a low value of FPR. Otherwise a candidate sample selected from real data would be dominated by false positives and a large amount of time would be needed to discard them through a visual inspection. In Fig. 3, we show, for a fiducial value for the threshold of the detection $p = 0.8$, the percentage of false negatives (i.e. the percentage of lenses that have been misclassified) as a function of the *Einstein* radius, R_E , and the source over lens-galaxy brightness contrast, α , defined in Section 2.2. Lenses with small *Einstein* radii and low-contrast

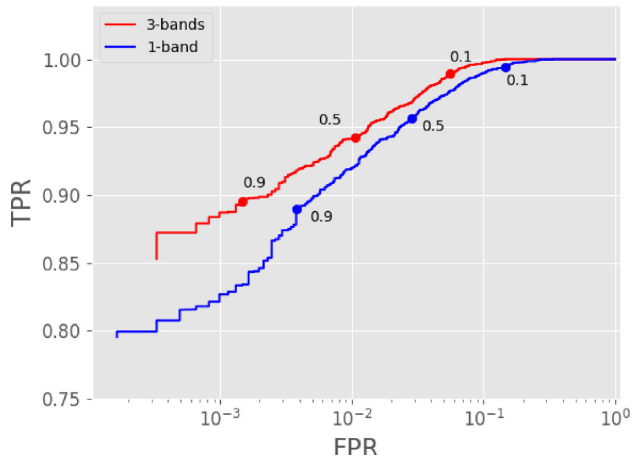


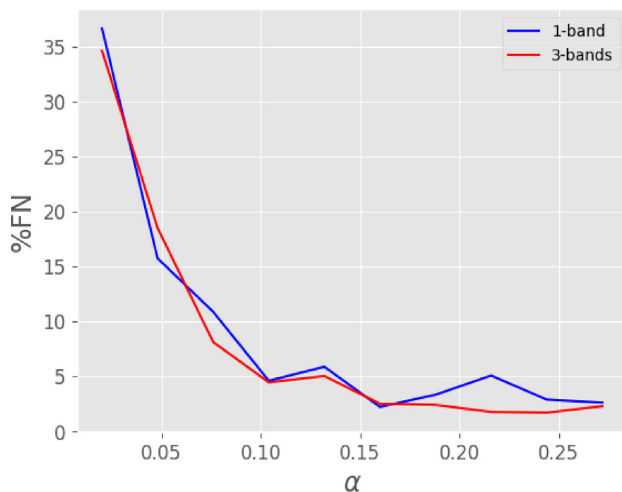
Figure 2. ROC curves for the 1-band (blue) and 3-band (red) ConvNet. Each point of the curves is the TPR versus FPR for different values of threshold for p (decreasing from left to right; some values are shown on the curve for reference).

lensed images are, as expected, the ones with which the ConvNets struggle the most. This suggests that our mock lens training samples currently covers the range in which lenses are found most easily. Smaller lenses are effectively smeared to an unrecognizable configuration by the PSF and fainter lensed sources will be too noisy to detect. In Fig. 3, we also see that the accuracy decreases for larger *Einstein* radii, possibly due to the fact that we covered the *Einstein* radius with logarithmic distribution, focusing slightly more on the small-separation systems, and secondly because their lensed images are more likely to blend in to the local environment and therefore harder to distinguish from nearby galaxies by the ConvNets.

However, because our goal is to efficiently select true strong-lenses in real astronomical observations, therefore it is necessary to assess the TPR and FPR when the ConvNets are applied to real data where the performance might be worse than on a simulated dataset.

4 APPLICATION TO REAL DATA

Testing the ConvNets on real data is fundamental, since the algorithms have been trained on a mixture of real and simulated data.



It is not trivial how the method will perform on a completely real dataset, since the domain of application is slightly different with respect to the domain where the ConvNets have been trained on. Ideally, the morphologies in the ensemble of simulated strong-lens systems and non-strong-lens systems would be a fair representation of all morphologies observed in their equivalents in real observations.

Hence, to properly analyse the performances of the ConvNets, we apply them to the full LRG sample composed of 21 789 galaxies extracted from 255 square degrees as described in Section 2.1.1. Using the same LRG sample of Petrillo et al. (2017) allows us to assess whether there has been any improvement with respect to our previous work. For each galaxy image, we opt to obtain an average prediction given by the average of the p 's for the original image and the images obtained operating a rotation of 90° , 180° , and 270° , respectively. Generally, this procedure allows to increase the accuracy of the classifications.

4.1 Results on the LRG sample

In Fig. 4, we show the number of lens candidates detected varying the threshold p . The 1-band ConvNet detects more lens candidates compared to the 3-band one for any given threshold for p . For each of the three ConvNets, it holds that the lower the threshold in p is set, the more candidates will have to be inspected visually. In other words, one wants to set as the threshold to an as low as possible value that yields both a sufficiently large sample of candidates and a sufficiently high TPR for the purpose of the scientific project. In Petrillo et al. (2017), we used visual inspection to select visually promising strong-lens candidates within the sample of systems assigned with a $p > 0.5$ by the ConvNet. This sample contains 56 candidates. Moreover, in Petrillo et al. (2017), we selected a subsample of 22 candidates based on the agreement between their expected *Einstein* radii, computed from the stellar mass or the velocity dispersion of the candidate lens galaxies, and the actual galaxy-image configurations. This does not guarantee that the 22 candidates are actual lenses but it allows us to exclude the cases with more implausible configurations. Fig. 5 compares the p -values for these two samples assigned by the two new ConvNets to those assigned by Petrillo et al. (2017) ConvNet.

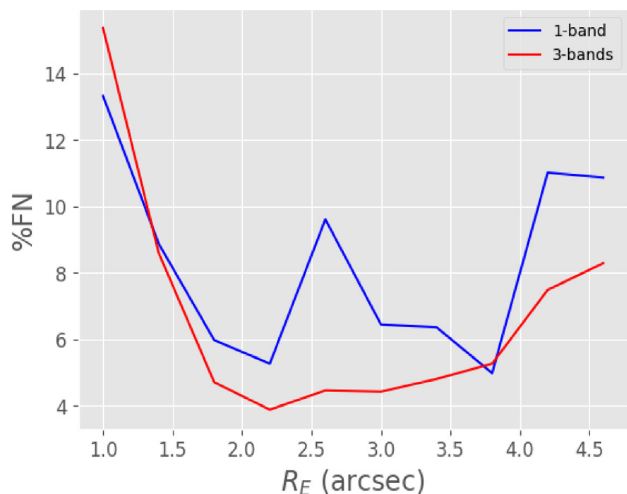


Figure 3. Percentage of false negatives (i.e. the percentage of lenses that have been misclassified) for bins of R_E and α defined as the ratio between the peak brightness of the lens galaxy and the lensed source.

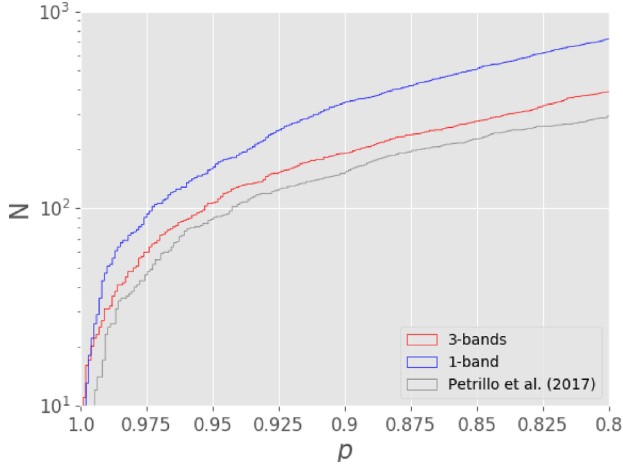


Figure 4. Number of detections as a function of the threshold for p for the 1-band (blue) and the 3-bands ConvNet (red) compared to the ConvNet of Petrillo et al. (2017, grey).

We note that the p -values of the new ConvNets have a noticeable peak at high values that becomes even more pronounced considering only the 22 candidates. In particular, the single-band ConvNet selects high-confidence candidates assigning high values of p . This is a fair improvement of the performance of the algorithm since there is a larger clustering of the higher visually ranked candidates

towards the high p -values. Instead in Fig. 6, we show the subset of the 56 galaxies that the new ConvNets classify with $p < 0.5$. For the 1-band finder, there are no clear candidates that would be lost; maybe a couple of galaxies could be considered acceptable lens candidates, while for the rest a low p -value it is the ideally desired output. In particular, three candidates (third, fourth, and fifth galaxy in Fig. 6, which have been selected as lenses in Petrillo et al. 2017 by visual inspection but after a more careful analysis have been revealed as false positives (likely a merger and two ring galaxies) are classified as non-lenses. Thus, the new finder does cumulatively a better job in excluding contaminants and selecting lens candidates. Instead, the 3-bands lose some acceptable candidates, but more importantly misidentifies a known gravitational lens and a clear good candidate (first and second galaxy in Fig. 6). This needs further investigation, thus, in the following subsection, we analyse the behaviour of the two lens-finders on a small sample composed by real lenses and striking lens candidates.

4.2 Application to a small sample of clear lens candidates

Additional insights on ConvNet performance can be obtained from inspecting the results on a set of real lenses and striking lens candidates. We gather a set of six galaxies composed as follows (see Fig. 7).

- (i) The four confirmed lenses known in literature that are present in our LRG sample: J085446 – 012137 (Limousin et al. 2010),

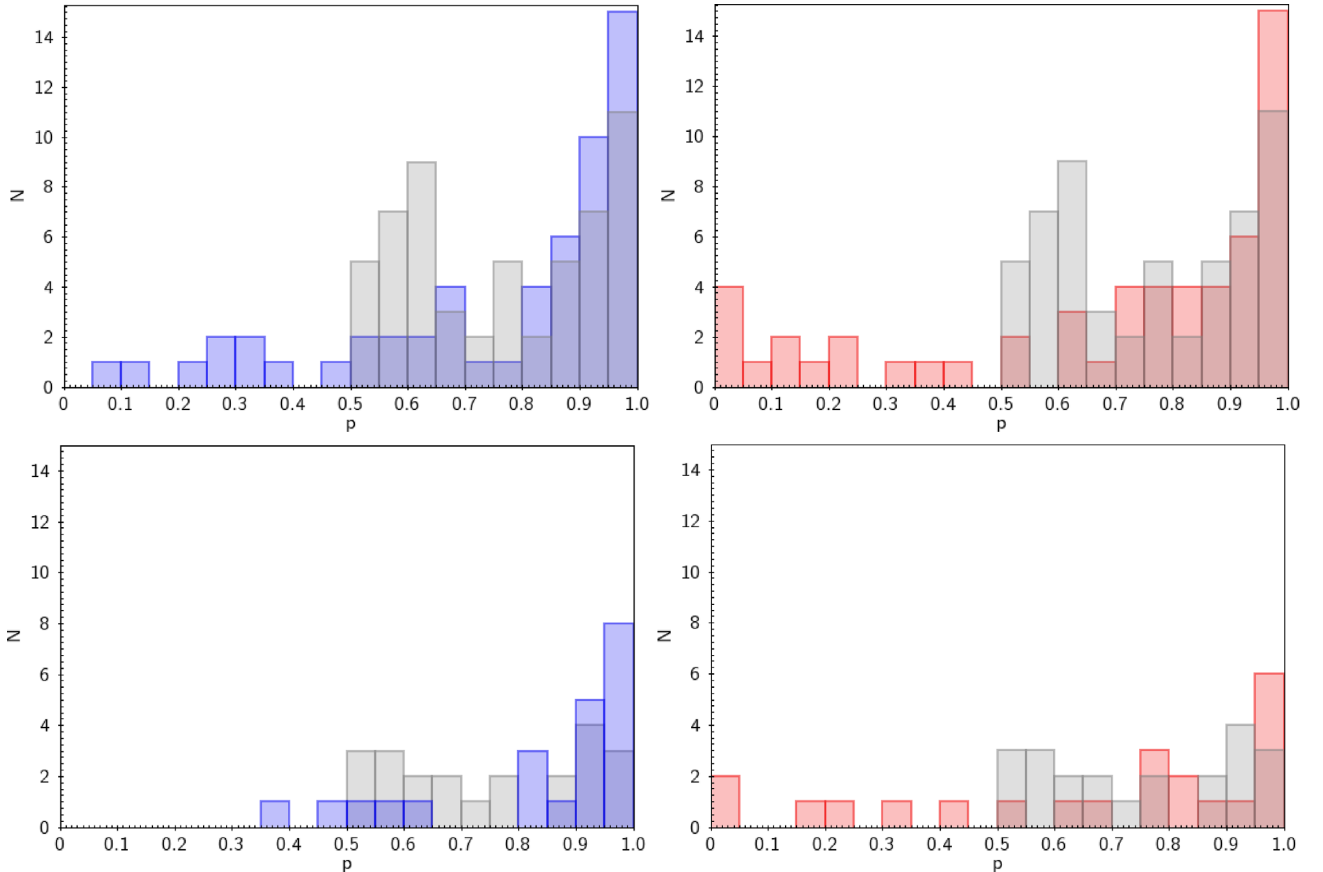


Figure 5. The first row shows the distribution of the scores for the 56 candidates with $p > 0.5$ selected in Petrillo et al. (2017). The scores of the 1-band (blue) and 3-bands ConvNet (red) are compared with those of Petrillo et al. (2017, grey). The second row shows the same for a subsample of 22 candidates selected as described in Section 4.1.

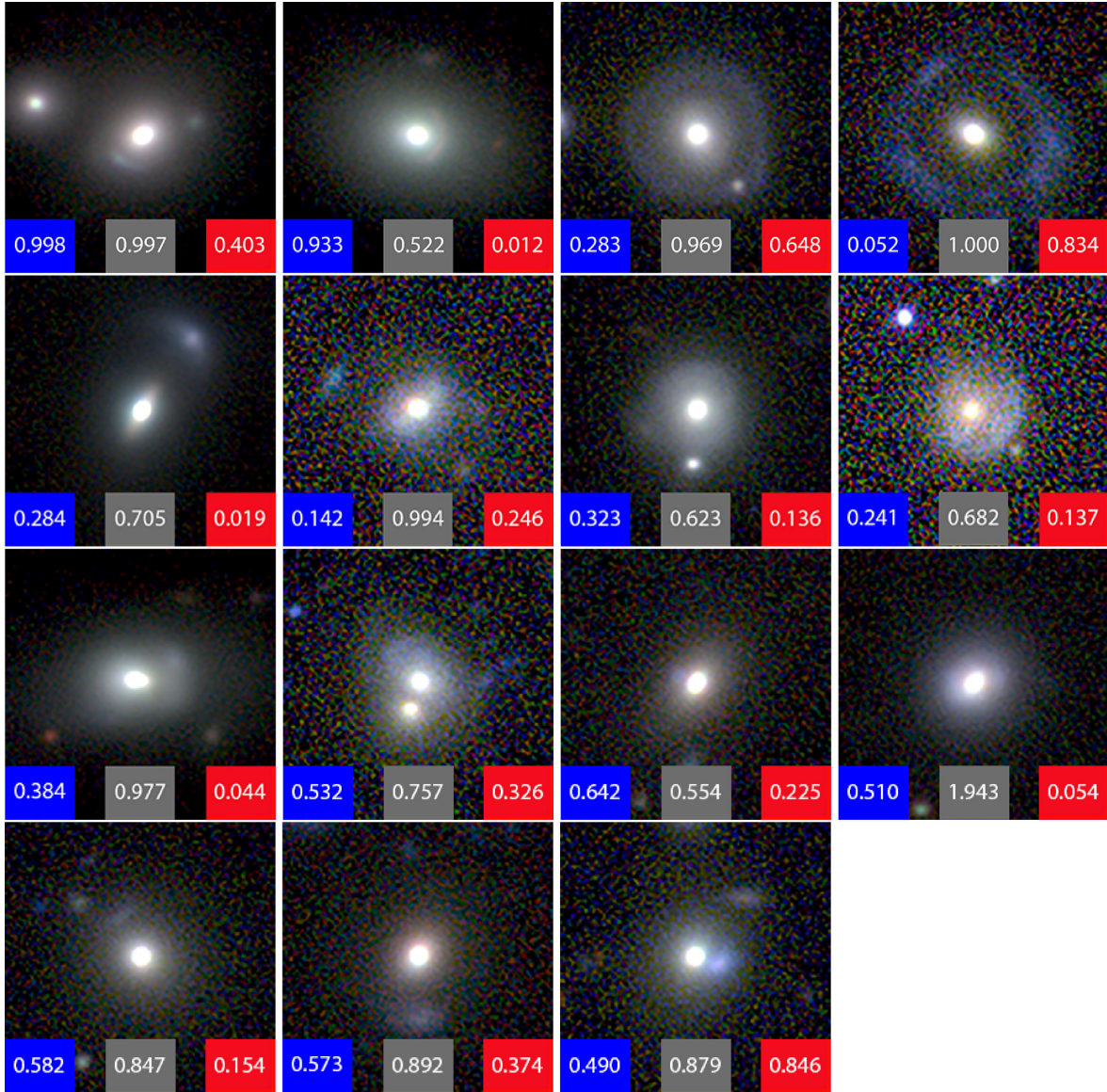


Figure 6. Sources classified by the new ConvNets with $p < 0.5$ that were classified in Petrillo et al. (2017) with $p > 0.5$. On each image, the score assigned by the single-band ConvNet (blue), the multiband ConvNet (red), and Petrillo et al. (2017)’s ConvNet (grey) are reported.

J114330 – 014427 and J1403 + 0006 (Bolton et al. 2008), J115252 + 004733 (More et al. 2017).

(ii) Three extremely likely lenses found in Petrillo et al. (2017), i.e. KSL713, KSL327, and KSL427.

The p -values for each of these galaxies are shown in Fig. 7 and Table 2. It is immediately noticeable that the 1-band ConvNet is the best-performing one: the three extremely likely lenses and three of the four known lenses are identified correctly with a very high p -value. Instead, the 3-bands ConvNet gives more weight to the source colours: it identifies easily blue-features but it struggles with other colours. This could be due to the larger number of training examples with blue colours.

Moreover, both the new ConvNets presented in this paper are able to pick up the quad lens J115252 + 004733 not selected in Petrillo et al. (2017). This piece of improvement is possibly due to the improved realism of the mock lensed sources adopted for this new

ConvNets, collecting a larger variety of configurations with respect to Petrillo et al. (2017). To validate this hypothesis, we train another 1-band ConvNet with the same mock lensed sources of Petrillo et al. (2017). This new ConvNet does not detect the quad (see Table 2), thus confirming the idea that the realism of the simulations plays a key-role in the performance of the lens-finder.

To further understand the role of the simulations, we also train a ConvNet with the same configuration of Petrillo et al. (2017) but with the simulated strong-lensing systems produced for this work (see Section 2.1.3). Also in this case, the quad is detected (see Table 2) even if the performance is worse than the new 1-band ConvNet. This could be due to the different architecture of the two ConvNets.

To conclude, despite the limited size of the control sample presented in this section, the 1-band ConvNet is the one that has generally the best performance under all conditions and it seems the best one to systematically apply to real data, both in terms of purity

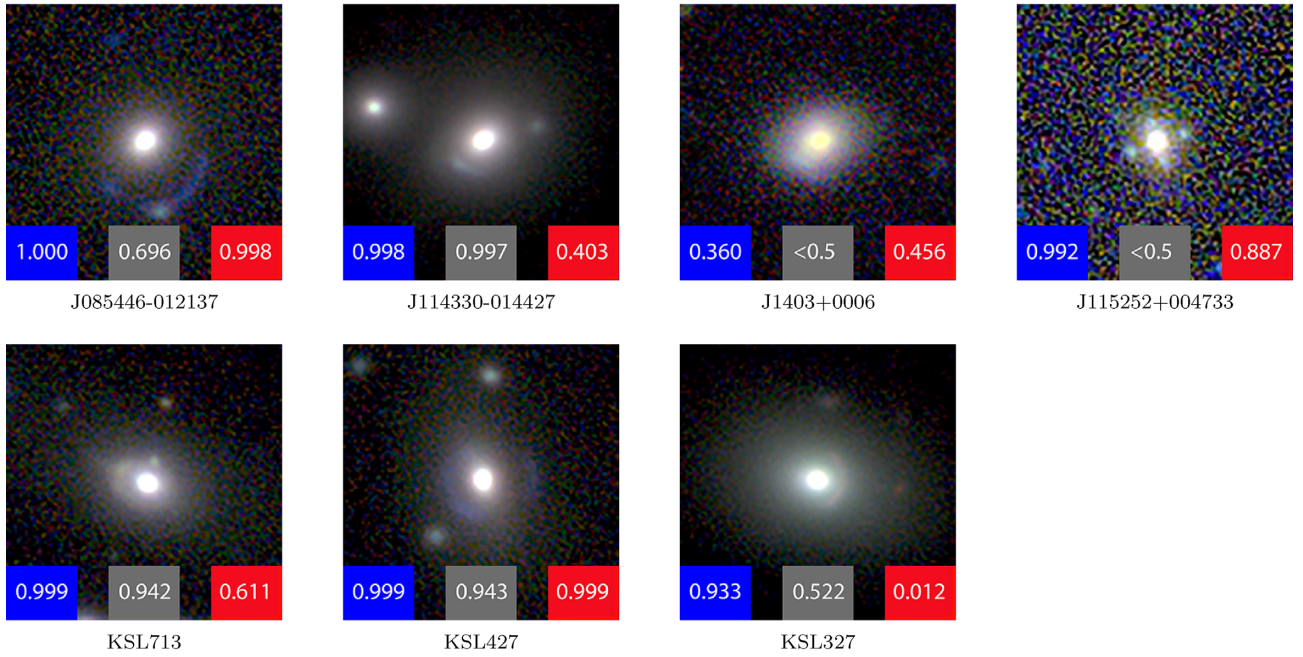


Figure 7. RGB images of the previously known lenses in our test-sample (first row) and the extremely likely lenses found by Petrillo et al. (2017) in the same test-sample (second row), this sample is discussed in Section 4.2. The scores assigned by the single-band ConvNet (blue), the multi-band ConvNet (red), and Petrillo et al. (2017)’s ConvNet (grey) are reported on each image.

Table 2. Scores given by ConvNets on the sample described in Section 4.2. The 1-band and 3-band ConvNets presented in this work are compared with Petrillo et al. (2017, old ConvNet). In the fifth column are shown the scores for a ConvNet with the same architecture of the 1-band ConvNet but trained with the mock lensed sources used in Petrillo et al. (2017) (1-band/old mocks). In the sixth column are reported the scores for a ConvNet with the same architecture of Petrillo et al. (2017) but trained with the same training set of this work (old ConvNet/new mocks).

ID	1-band	3-bands	old ConvNet	1-band/old mocks	old ConvNet/new mocks
J085446 – 012137	1.000	0.998	0.696	0.148	0.996
J114330 – 014427	0.998	0.403	0.997	0.947	0.605
J1403 + 0006	0.360	0.456	<0.5	0.038	0.419
J115252 + 004733	0.992	0.887	<0.5	0.321	0.904
KSL713	0.999	0.611	0.942	0.990	0.942
KSL427	0.999	0.999	0.943	0.982	0.752
KSL327	0.933	0.012	0.522	0.587	0.479

and completeness. The 3-band set-up, instead, is generally biased towards bluer colours and sometimes has failed to select known or very promising strong-lensing systems.

4.3 Visual inspection

To further analyse the performances of the ConvNets, we visually inspect the sources in the LRG sample (see Section 4) selected by the ConvNets with a threshold $p > 0.8$. To compare the results, we visually inspect the sources selected with the same threshold, $p > 0.8$, by the ConvNet in Petrillo et al. (2017). This yields 721 sources to inspect for the 1-band ConvNet, 390 for the 3-band ConvNet, and 292 sources for the ConvNet of Petrillo et al. (2017) ConvNet. The visual inspection is carried out by three observers who have three possible choices: ‘Yes, it is a lens’, ‘Maybe it is a lens’, ‘No, it is not a lens’. For each source, the classifiers visualize the g , r , and i fits files and a $g-r-i$ composed image with the software STIFF⁶ (Bertin 2012).

⁶<http://www.astromatic.net/software/stiff>

After the visual inspection, we build ROC curves assigning a ground truth for each source from the result of the visual inspection. ROC curves are shown in Fig. 8 for two different ground truths: one where we assign the label ‘lens’ to the sources that received at least two times *Maybe*, and another where ‘lenses’ are the sources rated with three times *Maybe* or at least one *Yes* and one *Maybe*. In this way, we have two different ground truths: one where also dubious lens candidates are labelled as ‘lenses’, and one more conservative where only more solid lens candidates are labelled as ‘lenses’. In Fig. 8, we also show the absolute number of true positives versus false positives as a function of the threshold p for the same ground truths. The ConvNets presented in this paper have both higher purity and completeness with respect to Petrillo et al. (2017) and a larger number of lens candidates retrieved. Moreover, the 1-band ConvNet has higher purity and completeness, especially for the more conservative sample, for sources with higher values of p as also the results of Section 4.2 seems to indicate. In addition, in Fig. 9 we show for the conservative classification the number of candidates detected by the ConvNets varying the threshold for p and how many of these are false positives. The true positives and

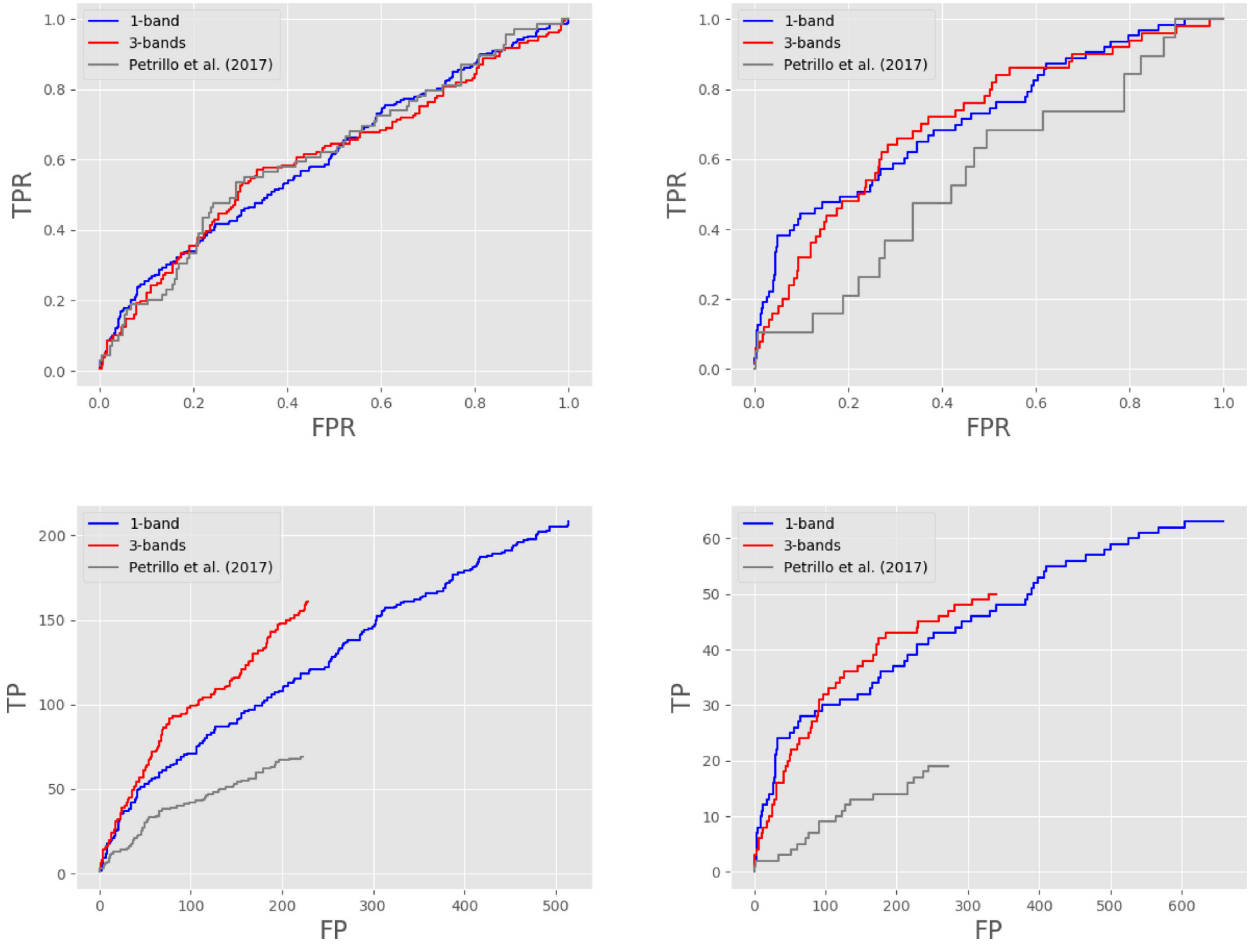


Figure 8. In the first row, we show ROC curves of the three ConvNets built for a real sample of LRGs as described in Section 4.3. On the left is shown the result for a sample including also more dubious candidates; on the right for a sample with solid candidates only. On the bottom row, we plot the absolute number of true positive versus false positives for the same samples.

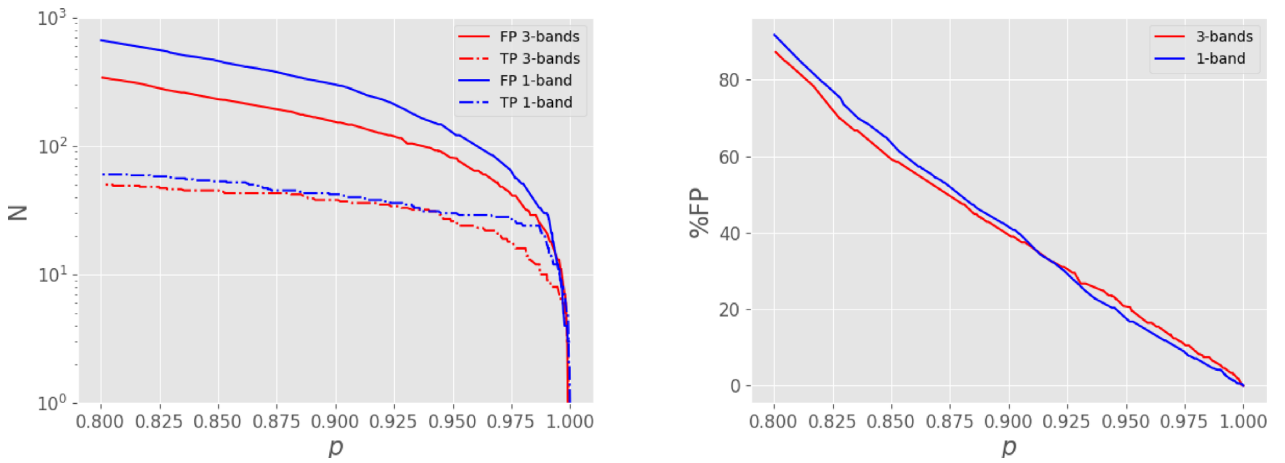


Figure 9. Results of the more conservative visual classification of the candidates selected by the ConvNets (Section 4.3). On the left: number of false positives and true negatives for the 1-band and 3-bands ConvNet as a function of the threshold of detection p . On the right: percentage of false positives as a function of the threshold of detection p for both the ConvNets.

false positives do not grow at the same pace as a function of the threshold p of detection. Moreover, the percentage of false positives is similar for the two new ConvNets. From the visual inspection analysis, we conclude that the 1-band ConvNet would be the best choice if one wants to have a small sample of good lens candidates with few contaminants and thus less images to inspect visually.

4.4 Prospects for *Euclid*

We have employed about 1 h to visually inspect the 721 targets selected by the 1-band ConvNet, selecting 60 lens candidates in the more conservative case (see previous subsection). For current surveys, investing such an amount of time in visual inspection is still reasonable for obtaining a more complete sample of lens candidates. However, for future surveys, the time dedicated to visually inspect the targets must be minimized considering the extremely higher number of resolved targets in the survey data. Nevertheless, one can choose a higher threshold of detection for the ConvNet and spend considerably less time in visually inspecting the targets. For example, if we select a threshold $p = 0.9919$ for the 1-band ConvNet, which is the threshold to retrieve three of the four known lenses presented in Section 4.2, we obtain a total of 35 sources selected by the 1-band ConvNet. These 35 sources are composed of 14 positives and 21 negatives, considering the more conservative result from the visual classification. This translates to roughly ~ 40 per cent purity, assuming that our visual inspection is accurate enough. By using the code LENSPOP by Collett (2015), we estimate that there are, in 255 square degree of KiDS, ~ 125 lenses with an *Einstein* radius larger than 1 arcsec and with our LRG colour cut selection (which corresponds to a cut in redshift of $z \lesssim 0.5$, as shown in fig. 8 in Petrillo et al. 2017). Hence, we can very roughly say that a candidate sample selected by the 1-band ConvNet with the fiducial threshold $p = 0.9919$ is ~ 40 per cent pure and ~ 11 per cent (i.e. 14 out of 125) complete by considering a population similar to that on which we trained our ConvNets. To translate this result in a prediction for *Euclid*, assuming that ConvNets will perform at least as well on the same domain on *Euclid* data as they do on KiDS data, we can proceed in the following way. Collett (2015) predicted that there will be $\sim 170\,000$ discoverable lenses in *Euclid*. If we consider only the lenses with an *Einstein* radius larger than 1 arcsec and with a redshift $z < 0.5$, the number reduces to $\sim 20\,000$. Therefore, with the fiducial threshold of $p = 0.9919$, it should be possible to select, from a sample of *Euclid* LRGs, ~ 2200 good gravitational lens candidates, by visually inspecting ~ 5500 galaxies. Considering that an observer needs ~ 5 s to inspect a lens candidate, which would imply that ~ 8 h of visual inspection would be already enough to obtain a sample of lens candidates far exceeding in size to the total number of currently known gravitational lenses. Nevertheless, this is an extremely conservative estimate, since *Euclid* data will have better spatial sampling (0.1 arcsec pixel $^{-1}$), PSF FWHM (~ 0.2 arcsec), and thus image quality, allowing algorithms and humans to better identify lensing features. Moreover, it will be possible to train the algorithms on a wider parameter space, and thus retrieving a larger number of lenses. We conclude however that current ConvNets, without much adaptation, can yield already enormous numbers of lenses from *Euclid* data without much human effort.

5 DISCUSSION AND CONCLUSIONS

Automatic lens-finders have become a standard method for finding gravitational lens candidates in survey data. They are crucial to cope with the large set of an estimated $\sim 10^6$ gravitational lens candidates

that will be produced by upcoming surveys such as *Euclid* and LSST. Therefore, it is important to build and validate such lens-finders in order to be ready for this anticipated data-avalanche.

It has been shown that ConvNets are one of the most promising methods for finding lens candidates (see, e.g. Metcalf et al. 2018). ConvNets achieve outstanding results when the data used to train the algorithm are very similar to the data where the algorithm is intended to be applied. However, this is not necessarily the case when the domain of application is different from the domain where the algorithm is trained on. There is an active field of research investigating how to adapt the algorithms, trained on one domain to other domains (see, e.g. Csurka 2017 for a review). In all published cases, ConvNet lens-finders are trained using simulated mock lensed sources. Moreover, in many cases, the lens-finders are tested only on fully simulated datasets. This does not ensure that the lens-finders will perform in a satisfactory way on real survey data as shown in this paper. It is important to conduct a thorough analysis of the performances on real data. Optimally, one would like to build a benchmark where all the lens-finders can be tested and compared against. Data from surveys such as KiDS and DES, with their identified lens candidates, could be used to build such a benchmark.

In this paper, we have tested two lens-finders based on a ConvNet for selecting strong gravitational lens candidates in the KiDS. One finder just uses r -band images while the other uses RGB images composed with g , r , i images. To train the algorithms, we have generated a large sample of simulated lensing features on top of real colour-magnitude selected galaxies from KiDS. Both the lens-finders are able to identify real lenses and good lens candidates in the survey. The performance of the two lens-finders is similar but the three-bands finder seems to under-perform when the lensed sources do not exhibit a bluer colour. This is most likely due to the fact that the mock lensed sources in the training set have mostly blue colours. Although genuine lensed sources are usually blue, this could select against non-blue sources. One way of dealing with this issue could be populating the mock lensed sources with a wider selection of colours and, in addition, dividing the training set in different classes of colours. This would help not only the ConvNet to explicitly classify sources with different colours but could also improve the general classification given the new information provided. In any case, the power of the single-band set-up is particularly encouraging in view of the *Euclid* mission data that will rely especially on the VIS band.

In addition, we have tested and compared the lens-finders with a similar one presented in Petrillo et al. (2017). The lens-finders presented in this work have a better performance, i.e. they have both a better completeness and purity (Section 4.3) and also they tend to classify more probable lens candidates with higher output values (see Section 4.2). This is a fair improvement of the performance which implies that selecting candidates with high output values we will have a purer sample that turns out to be convenient for visual inspection and/or spectroscopic follow-ups. Indeed, the larger the future survey will become (see e.g. *Euclid* and LSST), the more prohibiting the visual inspection of candidates will be, hence the purity of machine learning tools will be crucial for the automatic selection of sources to be set on the queue of spectrographs for their spectroscopic confirmations. The differences between this work and Petrillo et al. (2017) can be summarized in three points:

- (i) more complex mock lensed source simulations;
- (ii) a modified ConvNet architecture;
- (iii) a slightly larger training set.

These differences have contributed to the improvement of the performance, but our analysis presented in Sections 4.2 indicates that the main reason is the improved mock lensed sources simulations. In this work, the simulated sample is more populated with smaller sized lensed source galaxies and the sources exhibit more complexity in their structure, i.e. the presence of substructures in the lensed sources and a GRF perturbing the lens potential.

The ConvNet lens-finders can be tuned in terms of completeness and purity according to the specific needs of the science project. If one wants to have a more complete sample of lenses, a low threshold of detection can be chosen and the lower purity can be corrected by visually inspecting the candidates, something still feasible for current surveys. On the other hand, a purer (but less complete) sample of candidates can be obtained choosing a higher threshold. We have shown in Section 4.4 that by using a high threshold for detection will be already enough to retrieve in *Euclid*, a sample of lenses exceeding in size the total number of currently known gravitational lenses.

A series of possible improvements can be applied to the lens-finders. As we have shown, the performance strongly depends on the composition of the training set. Hence, making the lens simulations more realistic and using real gravitational lens candidates in the training set would probably improve the quality of the classification. Also enlarging the training set with more KiDS galaxies would probably help, as well as adding more labels for helping the ConvNets to discriminate between different classes of galaxies (e.g. adding labels for ellipticals and spirals). Moreover, particular care can be put in producing lens simulations where the S/N is always high enough for the lensing features to be recognizable by an observer. Another possibility would be to specialize the ConvNets in recognizing different kinds of sources (e.g. large/small-separation systems or extended/compact-sources). This could be obtained by either training different ConvNets with specialized training sets or using a single ConvNet trained with a unique training set but with multiple labels rather than with a binary classification scheme. Instead, on the algorithm side, a trivial improvement could be the so called *ensemble averaging*, i.e. averaging the output of different ConvNets in order to possibly reduce the statistical noise of the classification. An approach experimented by e.g. Schaefer et al. (2018) for identifying strong lens simulations.

Finally, in a forthcoming paper we will apply the algorithms to ~ 900 square degrees of the KiDS survey starting a systematic census of strong gravitational lens candidates named ‘*Links*’ (Lenses in the KiDS).

ACKNOWLEDGEMENTS

CEP thanks Leon Doddema, Ewout Helmich, Jelte de Jong, and Kateryna Frantseva for help and support. CEP, CT, GV, and LVEK are supported through an NWO-VICI grant (project number 639.043.308). The authors thank the anonymous referee for the insightful report. SC has been financially supported by a grant (project number 614.001.206) from the Netherlands Organization for Scientific Research (NWO). GVK acknowledges financial support from the Netherlands Research School for Astronomy (NOVA) and Target. Target is supported by Samenwerkingsverband Noord Nederland, European fund for regional development, Dutch Ministry of economic affairs, Pieken in de Delta, Provinces of Groningen, and Drenthe. NRN acknowledges financial support from the European Union Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 721463 to the SUNDIAL ITN network. GAMA is a joint European-Australasian

project based around a spectroscopic campaign using the Anglo-Australian Telescope. The GAMA input catalogue is based on data taken from the SDSS and the UKIDSS. Complementary imaging of the GAMA regions is being obtained by a number of independent survey programmes including *GALEX* MIS, VST KiDS, VISTA VIKING, *WISE*, Herschel-ATLAS, GMRT, and ASKAP, providing UV to radio coverage. GAMA is funded by the STFC (UK), the ARC (Australia), the AAO, and the participating institutions. The GAMA website is www.gama-survey.org. Based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 177.A-3016, 177.A-3017, and 177.A-3018, and on data products produced by Target/OmegaCEN, INAF-OACN, INAF-OAPD, and the KiDS production team, on behalf of the KiDS consortium. OmegaCEN and the KiDS production team acknowledge support by NOVA and NWO-M grants. Members of INAF-OAPD and INAF-OACN also acknowledge the support from the Department of Physics and Astronomy of the University of Padova, and of the Department of Physics of University of Federico II (Naples). This publication has been made possible by the participation in the Galaxy Zoo project of more than 20 000 volunteers from around the world, with almost 2 million classifications provided. Their contributions are individually acknowledged at <http://authors.galaxyzoo.org/>. The data are generated via the Zooniverse.org platform, development of which is funded by generous support, including a Global Impact Award from Google, and by a grant from the Alfred P. Sloan Foundation.

REFERENCES

- Abadi M. et al., 2016, preprint ([arXiv:1603.04467](https://arxiv.org/abs/1603.04467))
- Alard C., 2006, preprint ([astro-ph/0606757](https://arxiv.org/abs/astro-ph/0606757))
- Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, *MNRAS*, 310, 540
- Auger M. W., Treu T., Bolton A. S., Gavazzi R., Koopmans L. V. E., Marshall P. J., Moustakas L. A., Burles S., 2010, *ApJ*, 724, 511
- Barnabè M., Czoske O., Koopmans L. V. E., Treu T., Bolton A. S., Gavazzi R., 2009, *MNRAS*, 399, 21
- Barnabè M., Czoske O., Koopmans L. V. E., Treu T., Bolton A. S., 2011, *MNRAS*, 415, 2215
- Bertin E., 2012, in Ballester P., Egret D., Lorente N. P. F., eds, ASP Conf. Ser. Vol. 461, Astronomical Data Analysis Software and Systems XXI. Astron. Soc. Pac., San Francisco, p. 26263
- Biesiada M., Piórkowska A., Malec B., 2010, *MNRAS*, 406, 1055
- Bolton A. S., Burles S., Koopmans L. V. E., Treu T., Gavazzi R., Moustakas L. A., Wayth R., Schlegel D. J., 2008, *ApJ*, 682, 964
- Bonvin V. et al., 2017, *MNRAS*, 465, 4914
- Brault F., Gavazzi R., 2015, *A&A*, 577, A85
- Brewer B. J. et al., 2012, *MNRAS*, 422, 3574
- Cabrera-Vives G., Reyes I., Förster F., Estévez P. A., Maureira J.-C., 2017, *ApJ*, 836, 97
- Cao S., Covone G., Zhu Z.-H., 2012, *ApJ*, 755, 31
- Cao S., Biesiada M., Gavazzi R., Piórkowska A., Zhu Z.-H., 2015, *ApJ*, 806, 185
- Cao S., Biesiada M., Yao M., Zhu Z.-H., 2016, *MNRAS*, 461, 2192
- Cardone V. F., Tortora C., 2010, *MNRAS*, 409, 1570
- Cardone V. F., Tortora C., Molinaro R., Salzano V., 2009, *A&A*, 504, 769
- Carlstrom J. E. et al., 2011, *PASP*, 123, 568
- Chae K.-H., 2003, *MNRAS*, 346, 746
- Chan J. H. H., Suyu S. H., Chiueh T., More A., Marshall P. J., Coupon J., Oguri M., Price P., 2015, *ApJ*, 807, 138
- Chatterjee S., Koopmans L. V. E., 2018, *MNRAS*, 474, 1762
- Collett T. E., 2015, *ApJ*, 811, 20
- Collett T. E., Auger M. W., 2014, *MNRAS*, 443, 969
- Chollet F. et al., 2015, Keras. Available at: <https://keras.io>
- Covone G. et al., 2009, *ApJ*, 691, 531

- Csurka G., 2017, preprint ([arXiv:1702.05374](https://arxiv.org/abs/1702.05374))
- D’Isanto A., Polsterer K. L., 2018, *A&A*, 609, A111
- de Jong J. T. A. et al., 2015, *A&A*, 582, A62
- de Jong J. T. A. et al., 2017, *A&A*, 604, A134
- Deane R. P., Rawlings S., Garrett M. A., Heywood I., Jarvis M. J., Klöckner H.-R., Marshall P. J., McKean J. P., 2013, *MNRAS*, 434, 3322
- Diehl H. T. et al., 2017, *ApJS*, 232, 15
- Dieleman S., Willett K. W., Dambre J., 2015, *MNRAS*, 450, 1441
- Driver S. P. et al., 2011, *MNRAS*, 413, 971
- Eisenstein D. J. et al., 2001, *AJ*, 122, 2267
- Eisenstein D. J. et al., 2011, *AJ*, 142, 72
- Estrada J. et al., 2007, *ApJ*, 660, 1176
- Ferreras I., Saha P., Leier D., Courbin F., Falco E. E., 2010, *MNRAS*, 409, L30
- Fukugita M., Futamase T., Kasai M., Turner E., 1992, *ApJ*, 393, 3
- Fukushima K., 1980, *Biol. cybern.*, 36, 193
- Gavazzi R., Treu T., Rhodes J. D., Koopmans L. V. E., Bolton A. S., Burles S., Massey R. J., Moustakas L. A., 2007, *ApJ*, 667, 176
- Gavazzi R., Marshall P. J., Treu T., Sonnenfeld A., 2014, *ApJ*, 785, 144
- Grillo C., Eichner T., Seitz S., Bender R., Lombardi M., Gobat R., Bauer A., 2010, *ApJ*, 710, 372
- Guo Y., Liu Y., Oerlemans A., Lao S., Wu S., Lew M. S., 2016, *Neurocomputing*, 187, 27
- Hála P., 2014, preprint ([arXiv:1412.8341](https://arxiv.org/abs/1412.8341))
- Hartley P., Flamary R., Jackson N., Tagore A. S., Metcalf R. B., 2017, *MNRAS*, 471, 3378
- He K., Zhang X., Ren S., Sun J., 2015a, Proceedings of the IEEE International Conference on Computer Vision. p. 1026
- He K., Zhang X., Ren S., Sun J., 2015b, CoRR. abs/1512.03385
- Hezaveh Y., Dalal N., Holder G., Kisner T., Kuhlen M., Perreault Levasseur L., 2016, *J. Cosmol. Astropart. Phys.*, 11, 048
- Hezaveh Y. D., Levasseur L. P., Marshall P. J., 2017, *Nature*, 548, 555
- Hildebrandt H. et al., 2017, *MNRAS*, 465, 1454
- Horesh A., Ofek E. O., Maoz D., Bartelmann M., Meneghetti M., Rix H.-W., 2005, *ApJ*, 633, 768
- Hoyle B., 2016, *Astron. Comput.*, 16, 34
- Huertas-Company M. et al., 2015, *ApJS*, 221, 8
- Ilbert O. et al., 2006, *A&A*, 457, 841
- Ilbert O. et al., 2009, *ApJ*, 690, 1236
- Impellizzeri C. M. V., McKean J. P., Castangia P., Roy A. L., Henkel C., Brunthaler A., Wucknitz O., 2008, *Nature*, 456, 927
- Jacobs C., Glazebrook K., Collett T., More A., McCarthy C., 2017, *MNRAS*, 471, 167
- Jiang G., Kochanek C. S., 2007, *ApJ*, 671, 1568
- Joseph R. et al., 2014, *A&A*, 566, A63
- Kelly P. L. et al., 2018, *Nat. Astron.*, 2, 334
- Kim E. J., Brunner R. J., 2017, *MNRAS*, 464, 4463
- Kingma D., Ba J., 2014, preprint ([arXiv:1412.6980](https://arxiv.org/abs/1412.6980))
- Kochanek C., 1996, *ApJ*, 473, 595
- Koopmans L. V. E., Treu T., 2003, *ApJ*, 583, 606
- Koopmans L. V. E., Treu T., Bolton A. S., Burles S., Moustakas L. A., 2006, *ApJ*, 649, 599
- Koopmans L. V. E. et al., 2009, *ApJ*, 703, L51
- Kormann R., Schneider P., Bartelmann M., 1994, *A&A*, 284, 285
- Kubo J. M., Dell’Antonio I. P., 2008, *MNRAS*, 385, 918
- Kuijken K. et al., 2015, *MNRAS*, 454, 3500
- Lanusse F., Ma Q., Li N., Collett T. E., Li C.-L., Ravanbakhsh S., Mandelbaum R., Póczos B., 2018, *MNRAS*, 473, 3895
- Laureijs R. et al., 2011, preprint ([arXiv:1105.3486](https://arxiv.org/abs/1105.3486))
- LeCun Y., Bottou L., Bengio Y., Haffner P., 1998, *Proc. IEEE*, 86, 2278
- LeCun Y., Bengio Y., Hinton G., 2015, *Nature*, 521, 436
- Lenzen F., Schindler S., Scherzer O., 2004, *A&A*, 416, 391
- Li R., Frenk C. S., Cole S., Gao L., Bose S., Hellwing W. A., 2016, *MNRAS*, 460, 363
- Limousin M. et al., 2010, *A&A*, 524, A95
- LSST Science Collaboration et al., 2009, preprint ([arXiv:0912.0201](https://arxiv.org/abs/0912.0201))
- Lupton R., Blanton M. R., Fekete G., Hogg D. W., O’Mullane W., Szalay A., Wherry N., 2004, *PASP*, 116, 133
- McKean J. et al., 2015, in Proc. Sci., Strong Gravitational Lensing with the SKA. SISSA, Trieste, PoS(AASKA)084
- Marshall P. J. et al., 2016, *MNRAS*, 455, 1171
- Mason C. A. et al., 2017, *ApJ*, 838, 14
- Maturi M., Mizera S., Seidel G., 2014, *A&A*, 567, A111
- Metcalf R. B. et al., 2018, preprint ([arXiv:1802.03609](https://arxiv.org/abs/1802.03609))
- Miyazaki S. et al., 2012, Proc. SPIE Vol. 84460, Ground-based and Airborne Instrumentation for Astronomy IV. Astronomical Telescopes + Instrumentation, Amsterdam, the Netherlands, p. 84460Z
- Möller O., Kitzbichler M., Natarajan P., 2007, *MNRAS*, 379, 1195
- More A., McKean J. P., Muxlow T. W. B., Porcas R. W., Fassnacht C. D., Koopmans L. V. E., 2008, *MNRAS*, 384, 1701
- More A., Jahnke K., More S., Gallazzi A., Bell E. F., Barden M., Häußler B., 2011, *ApJ*, 734, 69
- More A., Cabanac R., More S., Alard C., Limousin M., Kneib J.-P., Gavazzi R., Motta V., 2012, *ApJ*, 749, 38
- More A. et al., 2016, *MNRAS*, 455, 1191
- More A. et al., 2017, *MNRAS*, 465, 2411
- Mukherjee S. et al., 2018, *MNRAS*, 479, 4108
- Negrello M. et al., 2010, *Science*, 330, 800
- Negrello M. et al., 2017, *MNRAS*, 465, 3558
- Oguri M., 2006, *MNRAS*, 367, 1241
- Oguri M., Marshall P. J., 2010, *MNRAS*, 405, 2579
- Paraficz D. et al., 2016, *A&A*, 592, A75
- Pawase R. S., Faure C., Courbin F., Kokotanekova R., Meylan G., 2014, *MNRAS*, 439, 3392
- Petrillo C. E. et al., 2017, *MNRAS*, 472, 1129
- Posacki S., Cappellari M., Treu T., Pellegrini S., Ciotti L., 2015, *MNRAS*, 446, 493
- Pourrahmani M., Nayyeri H., Cooray A., 2018, *ApJ*, 856, 68
- Richard J., Kneib J.-P., Ebeling H., Stark D. P., Egami E., Fiedler A. K., 2011, *MNRAS*, 414, L31
- Ruff A. J., Gavazzi R., Marshall P. J., Treu T., Auger M. W., Brault F., 2011, *ApJ*, 727, 96
- Russakovsky O. et al., 2015, *Int. J. Comput. Vis.*, 115, 211
- Saerens M., Latinne P., Decaestecker C., 2002, *IEEE Trans. Neural Netw.*, 13, 1205
- Salmon B. et al., 2017, preprint ([arXiv:1710.08930](https://arxiv.org/abs/1710.08930))
- Schaefer C., Geiger M., Kuntzer T., Kneib J., 2018, *A&A*, 611, A2
- Schmidhuber J., 2015, *Neural Netw.*, 61, 85
- Schneider P., Kochanek C. S., Wambsganss J., 2006, in Meylan G., Jetzer P., North P., eds, Gravitational Lensing: Strong, Weak and Micro. Springer-Verlag, Berlin, Heidelberg, p. 1
- Schneider P., Ehlers J., Falco E., 1992, Gravitational Lenses. Springer-Verlag, Berlin
- Seidel G., Bartelmann M., 2007, *A&A*, 472, 341
- Sérsic J. L., 1968, Atlas de Galaxias Australes. Cordoba, Argentina: Observatorio Astronómico
- Shallue C. J., Vanderburg A., 2018, *AJ*, 155, 94
- Simard P. Y., Steinkraus D., Platt J. C., 2003, Proceedings of the Seventh International Conference on Document Analysis and Recognition - Volume 2. ICDAR ’03. IEEE Computer Society, Washington, p. 958
- Sonnenfeld A., Gavazzi R., Suyu S. H., Treu T., Marshall P. J., 2013, *ApJ*, 777, 97
- Sonnenfeld A., Treu T., Marshall P. J., Suyu S. H., Gavazzi R., Auger M. W., Nipoti C., 2015, *ApJ*, 800, 94
- Sonnenfeld A. et al., 2018a, *PASJ*, 70, S29
- Sonnenfeld A., Leauthaud A., Auger M. W., Gavazzi R., Treu T., More S., Komiyama Y., 2018b, *MNRAS*, 481, 164
- Spiniello C., Koopmans L., Trager S. C., Czoske O., Treu T., 2011, *MNRAS*, 417, 3000
- Spiniello C., Barnabè M., Koopmans L. V. E., Trager S. C., 2015, *MNRAS*, 452, L21
- Spiniello C. et al., 2018, *MNRAS*, 480, 1163
- Stapelberg S., Carrasco M., Maturi M., 2017, preprint ([arXiv:1709.09758](https://arxiv.org/abs/1709.09758))
- Suyu S. H., Marshall P. J., Auger M. W., Hilbert S., Blandford R. D., Koopmans L. V. E., Fassnacht C. D., Treu T., 2010, *ApJ*, 711, 201
- Suyu S. H. et al., 2017, *MNRAS*, 468, 2590

- Swinbank A. M. et al., 2009, *MNRAS*, 400, 1121
 Tao Y., Zhang Y., Cui C., Zhang G., 2018, preprint ([arXiv:1801.04839](https://arxiv.org/abs/1801.04839))
 The Dark Energy Survey Collaboration, 2005, preprint ([astro-ph/0510346](https://arxiv.org/abs/astro-ph/0510346))
 Tortora C., Napolitano N., Romanowsky A., Jetzer P., 2010, *ApJ*, 721, L1
 Treu T., 2010, *ARA&A*, 48, 87
 Treu T., Koopmans L. V. E., 2002a, *MNRAS*, 337, L6
 Treu T., Koopmans L. V. E., 2002b, *ApJ*, 575, 87
 Treu T., Auger M. W., Koopmans L. V. E., Gavazzi R., Marshall P. J., Bolton A. S., 2010, *ApJ*, 709, 1195
 Treu T. et al., 2015, *ApJ*, 812, 114
 Tuccillo D., Huertas-Company M., Decenci re E., Velasco-Forero S., Dom nguez S nchez H., Dimauro P., 2018, *MNRAS*, 475, 894
 Turner E. L., Ostriker J. P., Gott J. R., III, 1984, *ApJ*, 284, 1
 Vegetti S., Koopmans L. V. E., 2009, *MNRAS*, 400, 1583
 Vieira J. et al., 2013, *Nature*, 495, 344
 Willett K. W. et al., 2013, *MNRAS*, 435, 2835
 Yuan H. B., Liu X. W., Xiang M. S., 2013, *MNRAS*, 430, 2188
 Zeiler M. D., Fergus R., 2013, preprint ([arXiv:1311.2901](https://arxiv.org/abs/1311.2901))

APPENDIX A: TECHNICAL DETAILS

A1 Convolutional Neural Networks

The input data of a ConvNet have a topological structure, e.g. the images of cutout around galaxies that we use in this paper. The input images can be seen as a set of matrices $\mathbf{X}_k$ with $k = 1, 2, \dots, K$ (in the case of our 3-bands ConvNet $K = 3$). The convolution property of the network is given by the convolutional layer that is composed by n kernels that produce n feature maps Y_n as in the following equation:

$$\mathbf{Y} = \sigma \left(\sum_{k=1}^K \mathbf{W}_k * \mathbf{X}_k + \mathbf{B} \right), \quad (\text{A1})$$

where $*$ is the convolution operator, σ is a non-linear function (the so-called activation function), $\mathbf{W}_k$ are the K weight arrays with $k = 1, 2, \dots, K$, representing a filter with its bias given by the constant array $\mathbf{B}$. The weights $\mathbf{W}$ and the biases $\mathbf{B}$ are the parameters that are determined during the training of the algorithm. Convolutional layers are sequentially stacked such as the input of the deeper layers are the feature maps of the preceding layer. There are far fewer parameters to be determined in a ConvNet as compared to a neural network because convolutional layers are only locally connected to its inputs as opposed to the fully connected layers of general neural networks.

It has been shown that ConvNets work better for image classification tasks than classical neural networks. One could expect this result from the following characteristics of the algorithm. ConvNets read images preserving their structure as a 2d matrix (3d in the case of multiband images) rather than transforming them in one-dimensional vectors where the topological information is lost. This turns out to be important because the relevant features to classify an image (e.g. a human face) are normally localized regions of pixels (e.g. eyes, mouth, etc.) and their spatial relations (e.g. the distance between eyes and mouth). The convolution operation of the filters with the image localizes where those patterns of pixels are in the image. The feature maps produced with this operation highlight those predictive features in the images. During the training phase, the optimal weights to be used in the filters are learned, which in

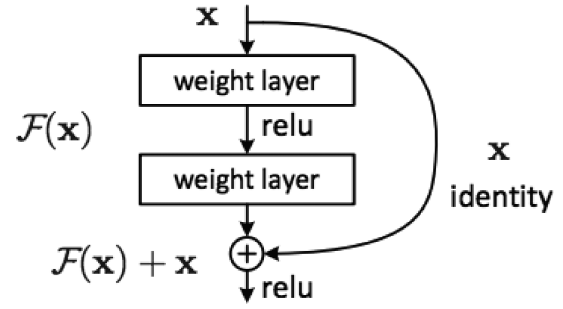


Figure A1. A building block of a ResNet. Image from He et al. (2015b). *Relu* is the non-linear function used between the convolutional layers; *identity* is the identity function that maps the input x into itself.

turn means that the ConvNet learns which are the features useful for classifying the images. What usually happens is that the filters learn to extract simple features in the first layers (usually they act as edge detectors), while deeper filters learn more complicated and abstract features. See Zeiler & Fergus (2013) for a thorough analysis.

A2 Implementation

The ConvNets used in this paper are implemented in PYTHON 3.6 using KERAS (Chollet et al. 2015) with TENSORFLOW backend (Abadi et al. 2016). The training phase took ~ 12 h on a GeForce GTX 1080. Once trained, the ConvNets take ~ 0.019 s to classify one galaxy. For our ConvNets, we use a ResNet architecture with 18 layers as described in He et al. (2015b). The building block of a ResNet (see Fig. A1) is characterized by a shortcut connection between the input and the output of a stack of a few convolution layers. In this way, the input x is mapped into the *residual* mapping $F(x) + x$. A residual mapping should be easier to optimize than an unreferenced mapping. For example, in the case where an optimal mapping was the identity function, it would be easier to fit $F(x) = 0$ to satisfy the requirement $F(x) + x = x$ than to satisfy the equality $F(x) = x$.

The ConvNets in this paper are trained, as in Petrillo et al. (2017), by minimizing a loss function of the targets t (1 for lenses and 0 for non-strong-lensing systems) and the predictions p (the output of the sigmoid unit of the ConvNet). The loss function that we minimize is the binary cross-entropy, a common choice in two-class classification problems

$$L = -t \log p - (1 - t) \log(1 - p). \quad (\text{A2})$$

The minimization is done via mini-batch stochastic gradient descent and ADAM updates (Kingma & Ba 2014) using a learning rate of 0.0001. We use a batch size of 64 images and perform 80 000 gradient updates, which corresponds to about six million examples. Each batch is composed by 32 lens and 32 non-lens examples. The weights of the ConvNets are initialized, as discussed in He et al. (2015a), from a random normal distribution with variance $2/n$ where n is the number of inputs of the unit. The initial values of the biases are set to zero.

This paper has been typeset from a $\text{\LaTeX}$ file prepared by the author.