SPACE WARPS– II. New gravitational lens candidates from the CFHTLS discovered through citizen science

Anupreeta More,[1]★ Aprajita Verma,[2]★ Philip J. Marshall,[2,3]★ Surhud More,[1]
Elisabeth Baeten,[4] Julianne Wilcox,[4] Christine Macmillan,[4] Claude Cornen,[4]
Amit Kapadia,[5] Michael Parrish,[5] Chris Snyder,[5] Christopher P. Davis,[3]
Raphael Gavazzi,[6] Chris J. Lintott,[2] Robert Simpson,[2] David Miller,[4]
Arfon M. Smith,[4] Edward Paget,[4] Prasenjit Saha,[7] Rafael Küng[7]
and Thomas E. Collett[8]

[1]*Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan*
[2]*Department of Physics, University of Oxford, Keble Road, Oxford OX1 3RH, UK*
[3]*Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94035, USA*
[4]*Zooniverse, c/o Astrophysics Department, University of Oxford, Oxford OX1 3RH, UK*
[5]*Adler Planetarium, Chicago, IL 60605, USA*
[6]*Institut dAstrophysique de Paris, UMR7095 CNRS Université Pierre et Marie Curie, 98bis bd Arago, F-75014 Paris, France*
[7]*Department of Physics, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland*
[8]*Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth P01 3FX, UK*

ABSTRACT

We report the discovery of 29 promising (and 59 total) new lens candidates from the Canada–France–Hawaii Telescope Legacy Survey (CFHTLS) based on about 11 million classifications performed by citizen scientists as part of the first SPACE WARPS lens search. The goal of the blind lens search was to identify lens candidates missed by robots (the RINGFINDER on galaxy scales and ARCFINDER on group/cluster scales) which had been previously used to mine the CFHTLS for lenses. We compare some properties of the samples detected by these algorithms to the SPACE WARPS sample and find them to be broadly similar. The image separation distribution calculated from the SPACE WARPS sample shows that previous constraints on the average density profile of lens galaxies are robust. SPACE WARPS recovers about 65 per cent of known lenses, while the new candidates show a richer variety compared to those found by the two robots. This detection rate could be increased to 80 per cent by only using classifications performed by expert volunteers (albeit at the cost of a lower purity), indicating that the training and performance calibration of the citizen scientists is very important for the success of SPACE WARPS. In this work we present the SIMCT pipeline, used for generating *in situ* a sample of realistic simulated lensed images. This training sample, along with the false positives identified during the search, has a legacy value for testing future lens-finding algorithms. We make the pipeline and the training set publicly available.

Key words: gravitational lensing: strong – methods: statistical.

1 INTRODUCTION

The last few decades have seen a rise in the discoveries of strong gravitational lenses owing to the plethora of interesting applications lenses have in astrophysics and cosmology. Strong lenses are routinely used to probe the dark matter distribution from galaxy (e.g.

Koopmans et al. 2006; Barnabè et al. 2009; Leier et al. 2011; Sonnenfeld et al. 2015) to group and cluster scales (e.g. Limousin et al. 2008; Zitrin et al. 2011; More et al. 2012; Oguri et al. 2012; Newman et al. 2013), to study distant young galaxies by using the lensing magnification as a natural telescope (e.g. Zitrin & Broadhurst 2009; Zheng et al. 2012; Whitaker et al. 2014), to test the cosmological model by constraining cosmological parameters such as the Hubble constant and the dark energy equation of state (e.g. Suyu & Halkola 2010; Collett et al. 2012; Collett & Auger 2014; Sereno & Paraficz 2014), and many more. Strong lenses are rare,

because a foreground massive object needs to be sufficiently aligned with a distant background source to produce multiple images. Nevertheless, systematic lens searches have led to the discovery of over 500 lenses to date.[1]

The search for gravitational lenses is a needle-in-a-haystack problem. Several automated lens-finding algorithms have been developed so far (e.g. Lenzen, Schindler & Scherzer 2004; Alard 2006; Seidel & Bartelmann 2007; More et al. 2012; Brault & Gavazzi 2015; Gavazzi et al. 2014), but they cannot simultaneously capture the myriad types of lenses that are known to exist. For example, the lensed images of background galaxies show variety in their surface brightness distributions, colours, light profiles, shapes, structures and angular image separations. Moreover, many lensed images appear similar to features found commonly in galaxies (such as spiral arms) or to artefacts in astronomical images (scattered light around stars). Almost all lens-finding algorithms find it difficult to distinguish these from the real lenses and thus suffer from a high rate of false positive detections. To mitigate this problem, algorithms are often restricted to detect a very narrow class of lens systems. However, even after such restrictions, robotic lens searches have to always rely on visual screening to produce a sample of plausible lens candidates.

Recognizing patterns is one of the strengths of the human brain. Humans are also capable of dealing with multitiered complex web of questions before arriving at a conclusion, a process which may not be always possible to automate. The algorithm by which our brains process a task is extremely malleable, self-learning and self-evolving. Therefore it has a huge potential for the discovery of exotic objects which do not quite fit a set criteria, but are still very likely to be objects of interest. The lens-finding algorithms are not yet advanced enough to produce better performance than visual classifications. Consequently, as we enter the era of large area imaging surveys spanning thousands of square degrees, the participation of a large community of volunteers to help with the visual identification of lenses would be very beneficial for the lensing community. Now seems the perfect time to investigate the potential of citizen science.

GALAXY ZOO, one of the most successful citizen science projects in astronomy, addressed the problem of how to classify large numbers of galaxies by their morphology (Lintott et al. 2008). From these early results to several new unexpected and interesting discoveries, such as that of green pea galaxies (Cardamone et al. 2009; Jaskot & Oey 2013) and Hanny's Voorwerp (Lintott et al. 2009; Keel et al. 2012), GALAXY ZOO has been able to start to realize the potential of citizen scientists. Since then, both astronomy and non-astronomy projects have been launched under the citizen science web portal Zooniverse (http://zooniverse.org). The task of finding gravitational lenses is significantly challenging, given that the lens systems show such complexity and that they are rare. To add to the challenge, not many citizen scientists are expected to be aware of the phenomenon of gravitational lensing, and the resulting characteristic image configurations. With these significant challenges at hand, we designed the SPACE WARPS project to enable the discovery of lenses through citizen science (learning from previous experience in serendipitous identification of lens candidates in GALAXY ZOO). In a companion paper (Marshall et al. 2015, hereafter Paper I), we describe the design of SPACE WARPS and how the entire system functions as a discovery service. In this paper (Paper II), we describe our first lens search using data from the Canada–France–Hawaii

Telescope Legacy Survey (CFHTLS[2]). In Küng et al. (2015), we describe the design of a collaborative mass modelling tool that can be used by citizen scientists.

This paper is organized as follows. In Section 2, we introduce the CFHTLS imaging data and the previously published lens samples from the CFHTLS. We generated a training sample, consisting of simulated lenses, duds and impostors, in order to aid the SPACE WARPS volunteers in the process of finding lenses. We give details of this training sample in Section 3 and Section 3.4. In Section 4, we briefly describe how the classifications of images from the volunteers are turned into a catalogue of plausible candidates (for further details, see Paper I). In Section 5, we present the new lens candidates from SPACE WARPS and compare it to the lens samples produced by past robotic searches of the CFHTLS. Next, we discuss what kind of lenses are detected or missed by the algorithms and SPACE WARPS in Section 6. Our conclusions are given in Section 7.

2 DATA

2.1 The CFHT Legacy Survey

The CFHTLS is a photometric survey in five optical bands ($u^*g'r'i'z'$) carried out with the wide-field imager MegaPrime which has a 1 deg^2 field-of-view and a pixel size of 0.186 arcsec (Gwyn 2012). The CFHTLS Wide covers a total non-overlapping area of 160 deg^2 on the sky and consists of four fields W1, W2, W3 and W4. The field W1 has the largest sky coverage of 63.7 deg^2. The fields W2 and W4 have similar sky coverages of 22.6 and 23.3 deg^2, respectively.[3] The field W3 has a sky coverage of 44.2 deg^2 and is more than twice as large as W2 and W4.

The CFHTLS imaging is very homogeneous and has good image quality. Most of the lensed arcs are much brighter in the g band, so deep imaging in this band is desirable. The limiting magnitude is 25.47 for the g band which goes the deepest among all of the five bands. The mean seeing in the g band is 0.78 arcsec. The zero-point to convert flux to AB magnitude for all bands is 30. These characteristics make CFHTLS ideal to do visual inspection for finding lenses. We use the stacked images from the final T0007 release taken from the TERAPIX website (see footnote 3) for this work.

We note that the CFHTLS is a niche survey with a unique combination of wide imaging with deep sensitivity. It is a precursor to the ongoing wide imaging surveys such as the Dark Energy Survey (DES), Kilo Degree Survey and the Hyper Suprime-Cam (HSC) survey and other planned future surveys such as the Large Synoptic Survey Telescope (LSST) survey. The search for lenses with SPACE WARPS in the CFHTLS is an important step to learn lessons and prepare for lens searches in these larger imaging surveys.

2.2 Previously published lens samples from the CFHTLS

The CFHTLS has been searched for lenses using various lens-finding methods and algorithms. Here, we give a brief summary of previously published lens samples in the chronological order.

From the early release of the CFHTLS (T0002) covering 28 deg^2, Cabanac et al. (2007) used an arc-finding algorithm (Alard 2006) to find arcs in galaxies, groups and clusters. They found about 40 lens candidates with quality grades from low to high.

---

[1] http://admin.masterlens.org/index.php

[2] http://www.cfht.hawaii.edu/Science/CFHTLS/

[3] These numbers are estimated from http://terapix.iap.fr/cplt/T0007/doc/T0007-doc.pdf.

In the thesis dissertation of Thanjavur (2009), nine promising and two low-probability candidates were reported as having been discovered serendipitously. These detections were made during the visual inspection of the CFHTLS images as part of data reduction procedures for the Weak Lensing survey (Benjamin et al. 2007).

Sygnet et al. (2010) carried out a search for edge-on disc galaxy lenses in the CFHTLS Wide. They identified galaxies, using SEXTRACTOR, which had $18 < i < 21$ and inclination angle $<25°$. After applying few more selection criteria and visual inspection, they found about 3 promising and a total of 18 lens candidates.

The ARCFINDER (More et al. 2012) was used for finding blue arc-like features in the entire CFHTLS imaging without any pre-selection on the type of the lensing object. This ARCFINDER, an improved version of the algorithm by Alard (2006), measures the second-order moments of the flux distribution in pixels within small regions to estimate the direction and extent of local elongation of features. Pixels with high values of elongation are connected to form an arc candidate. Finally, a set of thresholds on arc properties such as the area, length, width, curvature and surface brightness are used to select arc-like candidates. The search was carried out in the *g* band which is the most efficient wavelength to find typical lensed features. This sample, called SARCS, has 55 promising and a total of 127 lens candidates which are selected from both CFHTLS Wide and Deep fields. The SARCS sample consists of some galaxy-scale candidates and mostly groups/cluster scale lens candidates. This is because more massive systems produce arcs or lensed images with large image separation from the lensing galaxy which are easier to detect compared to the galaxy scales. In the absence of a large systematically followed up verified sample of candidates, we choose the most promising 26 systems as our bona fide lens sample from the CFHTLS Wide. The total number of lens candidates in the CFHTLS Wide alone is 108.

In Elyiv et al. (2013), the authors visually inspected a sample of 5500 optical counterparts of X-ray point-like sources identified in the *XMM*-LSS imaging of the CFHTLS *W*1 field. The goal was to find instances of lensed quasars. Their sample consists of a total of 18 candidates, of which 3 candidates were found to be promising.

Gavazzi et al. (2014) used their RINGFINDER code to find compact rings or arcs around centres of isolated and massive early-type galaxies. RINGFINDER subtracts the point spread function (PSF)-matched *i*-band images from the *g*-band images, and looks for excess flux in the bluer *g* band. An object detector measures the properties of these residual blue features, and candidates which meet the length–width ratio and tangential alignment criteria are then visually inspected to form the final sample. Gavazzi et al. (2014) pre-selected ∼638 000 targets as either photometrically-classified early-type galaxies, or objects selected to have red centres and blue outer parts, from the T0006 CFHTLS data release catalogues. A total of 14 370 galaxies were found to show detectable blue residuals, and 2524 were visually inspected, having passed the automatic feature selection process. This led to a total of 330 lens candidates out of which 42 were deemed good quality (q_flag = 3) and 288 medium quality (q_flag = 2) candidates. In addition to the main well-defined sample of Gavazzi et al. (2014), a further 71 candidates were reported to have been detected by earlier versions of the RINGFINDER, or from the CFHTLS Deep. From the main sample of 'RINGFINDER candidates', the SL2S team found, during their follow-up campaign, 33 confirmed lenses (Sonnenfeld et al. 2013b,a).

The work by Maturi, Mizera & Seidel (2014) used the arc-finding code of Seidel & Bartelmann (2007) and colour properties of typical arcs to optimize arc detection. This new approach was tested on the CFHTLS-Archive-Research Survey (Erben et al. 2009) which

covers an area of 37 deg^2 only, and this entire image set was also visually inspected by the authors to estimate the completeness and purity of their robotic search. They found 29 candidates with the robotic search alone and 41 candidates through pure visual inspection – some of which were known from previous searches. Most of these candidates are medium–low probability.[4]

The RINGFINDER and the ARCFINDER searches are the only searches that make use of a lens-finding algorithm and that have been run on the entire CFHTLS imaging data set. Thus, we considered these to be our reference sample of known lenses from robotic searches. For the purposes of transparency and to help with the training, the volunteers participating in SPACE WARPS-CFHTLS lens search were made aware of these two known lens samples. Images containing the systems from the RINGFINDER and the ARCFINDER samples were labelled as 'known lens candidates' in the SPACE WARPS discussion forum, Talk,[5] where volunteers have the opportunity to discuss their findings with fellow volunteers and the science team. In this paper, we refer to the sample of 330 RINGFINDER and 108 ARCFINDER lens candidates as the sample of 'known lens candidates' and the sample of confirmed (or most promising) 33 RINGFINDER and 26 ARCFINDER as the sample of 'known lenses'. Note that the 'known lens' sample is a subset of the 'known lens candidates' sample. Also, note that the lens candidates from the other papers listed above were not included in our reference 'known' sample and were not labelled as such in Talk. However, we did exclude these candidates when compiling the list of new SPACE WARPS lens candidates, as described in Section 5.2.

2.3 Image presentation in SPACE WARPS

In order to perform a blind lens search over the entire CFHTLS Wide, we present the volunteers with cut-outs of images selected from the survey region. We briefly describe the image presentation here for completeness; more information can be found in Paper I. We use the *g*-, *r*- and *i*-band imaging from CFHTLS which is most useful for visual identification of lenses. We made colour composite images using the publicly available code, HUMVI[6] following the prescription of Lupton et al. (2004). The colour scales were chosen to maximize the contrast between faint extended objects and bright early-type galaxies. These parameters were then fixed during the production of all the tiles, in order to allow straightforward comparison between one image and another, and for intuition to be built up about the appearance of stars and galaxies across the survey.

We extracted contiguous cut-outs of size 82 arcsec (440 pixels), including overlapping region of 10 arcsec (54 pixels) between the neighbouring cut-outs. This resulted in a catalogue of some 430 000 cut-outs for the entire CFHTLS Wide region. The size of the individual cut-out was determined by optimizing factors such as the typical angular scales of gravitational lenses, the number of objects seen in a single cut-out and the total number of image cut-outs in the survey. If a lens candidate happens to be too close to the edge of a cut-out, then the overlap between neighbouring cut-outs allows a volunteer to get a clearer view of the same candidate in at least one of the cut-outs. We note that since the images are shown randomly, a volunteer may not necessarily come across the neighbouring cut-out

---

[4] http://www.ita.uni-heidelberg.de/~maturi/Public/arcs

[5] http://talk.spacewarps.org/

[6] The open source colour image composition code used in this work is available from http://github.com/drphilmarshall/HumVI.

unless they classify a large number of images. This is not a problem since our user base is extremely large and we receive multiple classifications for the same cut-out.

3 TRAINING SAMPLE

The simulated lenses are important to train citizen scientists who may be new to the task of finding lenses, but they are also crucial for analysing the classifications performed by the citizen scientists (more details can be found in Paper I, but see Section 4 below for a brief summary). In this section, we describe the framework used for generating the simulated lens sample, give details of the sample itself along with some of its known limitations and also describe the sample of duds and impostors.

3.1 Methodology to simulate lenses

For the purpose of generating simulated lens systems, we divide them into two main categories (a) galaxy-scale lenses (b) and group- or cluster-scale lenses. We further subdivide galaxy-scale lenses based on the type of the background sources, namely galaxies and quasars. We do not simulate group-scale quasar lenses as they are expected to be even more rare. We now describe our procedure to generate these different types of lens systems.

3.1.1 Galaxy-scale lenses

We begin by considering all elliptical galaxies at $z < 1$ in our parent CFHTLS catalogue (Gavazzi et al. 2014) as potential lens candidates for the simulated sample. To avoid using a known lens galaxy for our simulation purpose, we exclude all those galaxies whose positions match with the lensing galaxies from the SARCS samples within 2 arcsec.[7]

For each galaxy, the average number of source objects (either quasars or galaxies) above a minimum luminosity L_{min} in the background that may get lensed can be calculated as

$$N_{src} = \int_{z_l}^{\infty} n_{src}(> L_{min}, z_s)\sigma_{lens}(\sigma_v, z_l, z_s, q)\frac{dV}{dz_s}dz_s, \quad (1)$$

where

$$n_{src}(> L_{min}, z_s) = \int_{L_{min}}^{\infty} \Phi(L', z_s)dL'. \quad (2)$$

Here, $\Phi(L', z_s)$ denotes the source luminosity function per unit comoving volume, σ_{lens} denotes the angular lens cross-section, which depends upon the lens redshift (z_l), source redshift (z_s), the lens velocity dispersion σ_v as well as the projected axis ratio of the lens ellipticity, q.

In order to calculate the lensing cross-section, we first calculate the luminosity of each potential lensing galaxy using the photometric redshifts (z_l) from the parent galaxy catalogue. Next, we use the L–σ scaling relation from the bright sample of (Parker et al. 2005) given by

$$\sigma_v = 142 \left(\frac{L}{L_*}\right)^{1/3} \text{ km s}^{-1}. \quad (3)$$

---

[7] Due to inaccuracies and uncertainties in measurements of the centres of some of the lens candidates, some simulated lensed images were superposed on the galaxies of known lens candidates. This issue was overcome by presenting the same CFHTLS images with and without the simulated lenses to the volunteers.

Table 1. Thresholds used in the selection of the simulated lenses.

Name	Gal–Gal (Grp–Gal) Min	Max	Gal–Qua Min	Max
Source redshift	1.0	4.0	1.0	5.9
Source flux	21.0	25.5	21.0	25.5
Source ellipticity	0.1	0.6	–	–
Source PA	0	180	–	–
Lens redshift	–	0.9	–	0.9
Lens shear strength	0.001 (–)	0.02 (–)	0.001	0.02
Lens shear PA	0 (–)	180 (–)	0	180
Einstein radius (arcsec)	1.2 (2)	5 (–)	1.2	5
Boost factor	=100 (40)		=1200	
Image flux$_{2B}$	>23		>23	
Image flux$_{tot}$	<19		<20	

Notes. (a) () corresponds to quantities used for Grp–Gal scale lenses, if they are different from Gal–Gal. (b) $_{2B}$ denotes the second brightest lensed image. (c) $_{tot}$ denotes the total flux integrated over all of the lensed images. (d) All fluxes are in AB mag. PA is in degrees measured east of north.

This sets the velocity dispersion of the halo hosting the galaxy, which will be later used in the model. We assume that the knee of the luminosity function of galaxies, L_* evolves such that there is a decline of 1.5 mag between $z = 1$ to 0 (Faber et al. 2007).[8]

We adopt a singular isothermal ellipsoid (SIE) model for each of our galaxies (Kormann, Schneider & Bartelmann 1994), such that the convergence is given by

$$\kappa(x, y) = \frac{b\sqrt{q}}{2}\frac{1}{\left(\theta_1^2 + q^2\theta_2^2\right)^2}. \quad (4)$$

Here, b is called the Einstein radius, and its dependence on the velocity dispersion of the SIE is given by

$$b = 4\pi \left(\frac{\sigma_v^2}{c^2}\right)\left(\frac{D_{ls}}{D_s}\right). \quad (5)$$

The SIE model results in a caustic and a pseudo-caustic on the source plane, which demarcates the regions of different image multiplicities. We make use of the parametric solutions, $r(\theta)$, for the caustics in such a model from Keeton, Mao & Witt (2000a) where θ is the polar angle. We take the maximum of the radial and tangential caustic at every polar angle in order to obtain the area of the lensing cross-section, σ_{lens}, for every galaxy,

$$\sigma_{lens} = \frac{b^2 q}{2} \int_0^{2\pi} r^2(\theta)d\theta. \quad (6)$$

We also add external shear at the centre of the potential lensing galaxy drawn randomly from a set range (see Table 1). The shear is expected to affect the lens cross-section for a small number of cases when the shear strength is high in addition to high lens ellipticity or the position angle (PA) of the shear is almost orthogonal to that of the lens ellipticity. However, the effect of shear on the lens cross-section is expected to be small for most of the cases and is ignored in the current implementation of SIMCT.

The luminosity functions of the background galaxies and quasars are determined as follows. We use the results of Faure et al. (2009)

---

[8] We anchor our L_* evolution at low redshifts using the determination of L_* in the r band by Blanton et al. (2001). To maintain consistency in magnitude systems, we have converted the CFHT MegaCAM magnitudes to SDSS magnitudes and k-corrected them to $z = 0.1$.
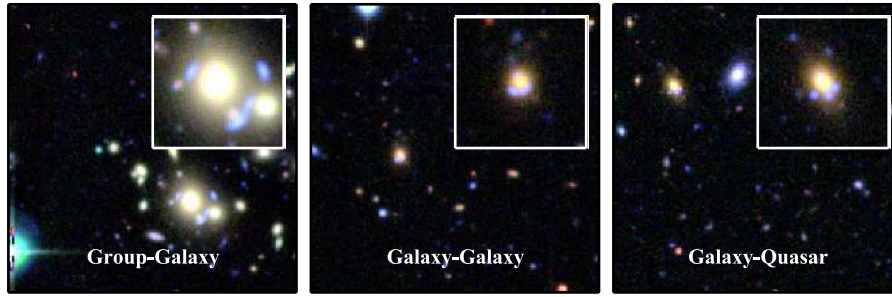
Figure 1. Examples of the three types of simulated lenses.

to specify the luminosity function of galaxies where the redshift distribution of sources is given by

$$p_s = \frac{\beta z_s^2 \exp(\frac{z_s}{z_0(m_{lim})})^\beta}{\Gamma(3/\beta) z_0^3(m_{lim})}, \tag{7}$$

where $\beta = 3/2$ and $z_0(m_{lim}) = 0.13 m_{lim} - 2.2$ and the source counts as a function of the limiting magnitude are given by

$$n_s = \int_{-\infty}^{m_{lim}} \frac{n_0 dm}{\sqrt{10^{2a(m_1-m)} + 10^{2b(m_1-m)}}}, \tag{8}$$

with parameters $a = 0.30, b = 0.56, m_1 = 20$ and $n_0 = 3 \times 10^3$ deg^{-2}.

For quasars, we assume the luminosity function prescription of Oguri & Marshall (2010) and adopt k-corrections by Richards et al. (2006).

The luminosity function is expressed as

$$\frac{d\Phi}{dM} = \frac{\Phi_*}{10^{0.4(\alpha+1)(M_{abs}-M_*)} + 10^{0.4(\beta+1)(M_{abs}-M_*)}}, \tag{9}$$

where the normalization, $\phi_* = 5.34 \times 10^{-6} h^3$ Mpc^{-3} and break magnitude, $M_* = -20.90 + 5\log h - 2.5\log f(z)$. The redshift-dependent factor in M_* is given by

$$f(z) = \frac{e^{\zeta z_s}(1 + e^{\xi z_*})}{(\sqrt{e^{\xi z_s}} + \sqrt{e^{\xi z_*}})^2}. \tag{10}$$

We adopt the best-fitting values $\zeta = 2.98$, $\xi = 4.05$, $z_* = 1.60$ (Oguri & Marshall 2010). For the faint end slope, we use $\beta = -1.45$ whereas for the bright end slope, we use $\alpha = -3.31$ when $z_s < 3$ and $\alpha = -2.58$ at higher redshifts, as prescribed by Oguri & Marshall (2010).

With the cross-section, and the luminosity functions specified, we calculate the expected number of sources behind a candidate lensing galaxy using equation (1). We need to generate a large number of simulated lenses (larger than the number of real galaxy lenses we expect to find in CFHTLS) in order to have a reasonably large and diverse training sample for thousands of SPACE WARPS volunteers. Therefore, we artificially boost the average number of sources by a factor (see Table 1), which increases the occurrence of lensing. We draw a Poisson deviate, N_{src} with a mean equal to the boosted average number of sources. If N_{src} is greater than zero, then this galaxy is flagged as a potential lensing galaxy.

Next, we determine properties of the background sources for every lens system. We follow similar procedures for both background galaxies and quasars. We draw source redshifts and luminosities from the aforementioned distributions. We note that the sources are being drawn from a much fainter magnitude range compared to the limiting magnitude of the CFHTLS imaging and thus, the magni-

fication bias[9] is naturally taken into account. The source positions with respect to the lens are drawn randomly from an area inside the caustic. When populating the sources within the caustics, the finite size of the background galaxies is expected to affect the lens cross-sections to some extent. As this factor is not critical for the purpose of our training sample, for simplicity we assume the background galaxies to be point like when computing cross-sections. We perform ray-tracing for all of the N_{src} sources using the publicly available code GRAVLENS (Keeton et al. 2000b) and choose sources that satisfy our selection criteria given below. We determine the fluxes of the lensed images and the total magnification of each of the lensed sources. We draw a source randomly for which the flux of the second brightest lensed image and the total magnification of all lensed images meet the thresholds given in Table 1.

Since we want to produce realistic looking lens systems, we simulate lenses in each of the five CFHTLS filters. The colours of the background galaxies are drawn randomly from the photometric CFHTLenS catalogue (Hildebrandt et al. 2012; Erben et al. 2013). Similarly, we use a quasar catalogue from the SDSS Data Release 9 (Pâris et al. 2012) from which colours are drawn to simulate quasar lenses. Next, we assume a Gaussian profile[10] for the galaxies. The ellipticity and the PA are drawn randomly from within the range given in Table 1. The effective radius of the galaxy is estimated from the luminosity–size relation (Bernardi et al. 2003, with a redshift scaling, to account for size evolution) given by

$$R_{eff} = 10^{0.52} \frac{L_r^{2/3}}{(1 + z_s)^2} \text{ kpc}, \tag{11}$$

where $L_r = L_s/10^{10.2}$ L$_\odot$. On the other hand, quasars are assumed to be point sources and the PSF, with which quasars are convolved, is assumed to have a Gaussian profile. The full width at half-maximum of the PSF is equated to that of the mean seeing for every filter. The mean seeing values are taken from table 4 of the official TERAPIX T0007 release explanatory document (see footnote 3).

Once all the parameters are determined for the lens and source models, we once again use GRAVLENS to generate simulated lensed images. After accounting for the shot noise in the lensed images and convolving them with the seeing in each of the filters, the simulated image is added to the real CFHTLS image centred on the galaxy chosen to act as a lens (see Fig. 1). Note that we ensure that the

[9] In a flux-limited sample from a survey, sources fainter than the flux limit end up in the sample owing to the magnification by lensing which is known as the magnification bias. This affects the source luminosity function and needs to be accounted for when comparing the true and observed luminosity functions.

[10] This was due to an oversight. We intended to use either an exponential or de Vaucouleurs' profile that will be adopted for future implementations of SIMCT.

lensed galaxies and lensed quasars are not superposed on the same 'lensing' galaxy. Similarly, these 'lensing' galaxies at the galaxy scales are ensured to be distinct from those chosen for the group scales. The framework for the group scale is described below.
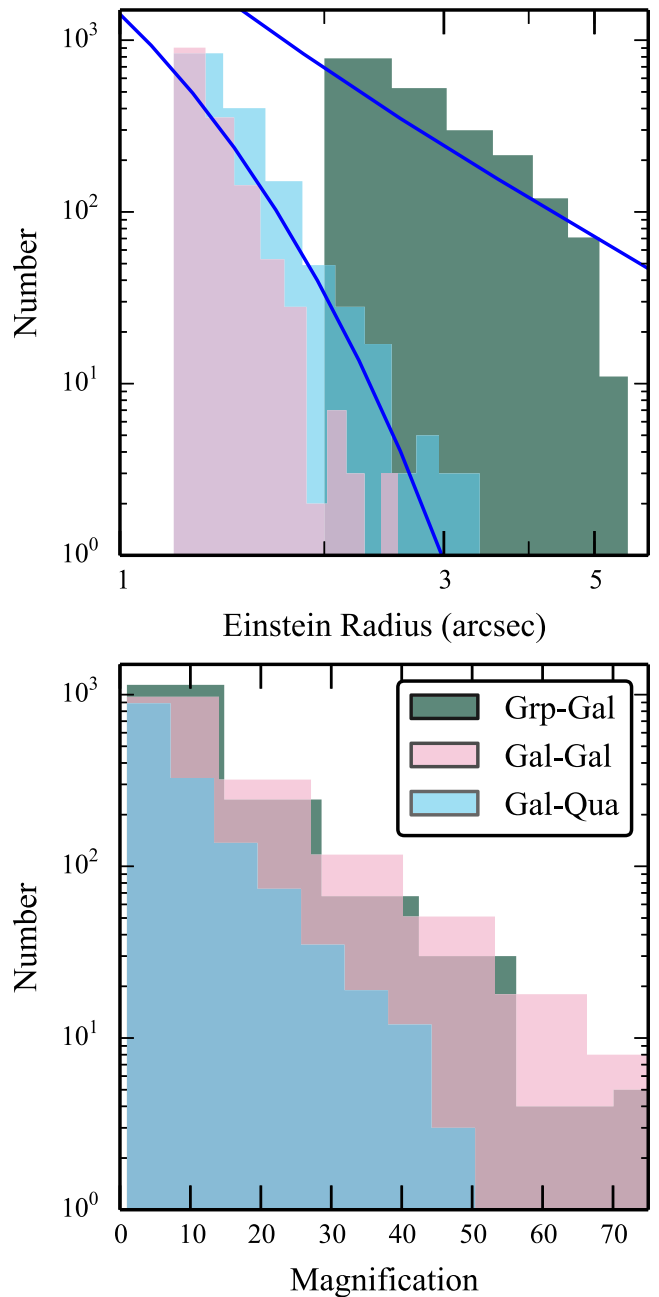
3.1.2 Group-scale lenses

At group or cluster scales, the mass distribution is more complex. The convergence in the inner regions, which are typically responsible for the multiple lensed images, arises from not only the brightest group galaxy (BGG) at the centre, but also from the dark matter component and the satellite galaxies (Oguri, Keeton & Dalal 2005; Oguri 2006). We generate a basic group catalogue based on the magnitudes and photometric redshifts available for the CFHTLS. We select all galaxies with $10^{10.8}$ M$_{\odot}$ as plausible BGGs. We select the member galaxies such that their photometric redshifts are within $\delta z = 0.01$ of the BGG and within an aperture of 250 kpc. If another BGG is found within the aperture, then the fainter BGG is removed from our list of BGGs.

We assume a constant mass-to-light ratio of $3 \times 0.7\,h\,M_*/L_*$, to convert the BGG luminosity to a stellar mass estimate. The stellar mass–halo mass relation (Behroozi, Wechsler & Conroy 2013), including random scatter, is then used to calculate the halo mass for the lens. We adopt an NFW (Navarro, Frenk & White 1997) density profile for the underlying dark matter halo. Given the halo mass, other key parameters such as the scale radius (r_s) and the density at the scale radius (ρ_s) can be determined for an NFW profile. In addition, we adopt an SIE model for the BGG and members whenever the ellipticities are available from the galaxy catalogue (else we use an isothermal sphere, SIS).

We calculate the luminosities and velocity dispersions for the BGG and each of the member galaxies following the same prescription as in Section 3.1.1. To calculate the average number of sources that get lensed by such a system, we need to calculate the lensing cross-section for each of these potential lensing groups. The complexity in the lens models makes it analytically intractable to calculate the size of the caustics.[11] Therefore, we generate the caustics numerically using GRAVLENS and then determine the area covered by the caustics. We consider only galaxies as our background source population since group- or cluster-scale quasar lenses are expected to be extremely rare in the CFHTLS. Following the same procedure as described in Section 3.1.1, we calculate the number of galaxies expected to lie behind every potential lensing group (see equation 1). As before, for each background galaxy within the lens cross-section, a redshift and an i-band magnitude is determined by drawing galaxies randomly from the respective distributions (see equations 7 and 8).

All those groups that are found to have no background galaxies within the cross-sectional area are rejected and the rest are included as potential lenses. As mentioned earlier, we artificially boost the total number of sources behind every lens but ensure that the statistical properties such as the profile of the image separation distribution are not affected (see Fig. 2). We follow the same procedure and apply thresholds to determine properties of the lensed galaxies for every lens as described for galaxy—galaxy lenses in the previous section. The thresholds are same as those used for galaxy lenses (see

---

[11] The lens mass distribution determines size and shape of the caustics. Any source located within the caustics will form multiple lensed images which is the criteria for strong lensing. To further understand caustics, see e.g. Schneider, Ehlers & Falco (1992).

Figure 2. Einstein radius and total magnification distributions for all types of lenses. The solid (blue) curves show the theoretical prediction assuming an SIS model at galaxy scales and a total (NFW+Hernquist) model at group scales taken from More et al. (2012).

Table 1) and are reported within '()', if different for group scales. The simulated lensed images are then added to the real CFHTLS images with the BGGs as the centre by following exactly the same procedure as described in the previous section (see Fig. 1).

3.2 Simulated lens sample and catalogue description

In this section, we describe some of the properties of our simulated sample for each of the three types of lens samples. We have made an attempt to generate as realistic a lens sample as possible within the requirements of the SPACE WARPS project. The statistical properties of the lens sample are expected to be similar to real lens samples.
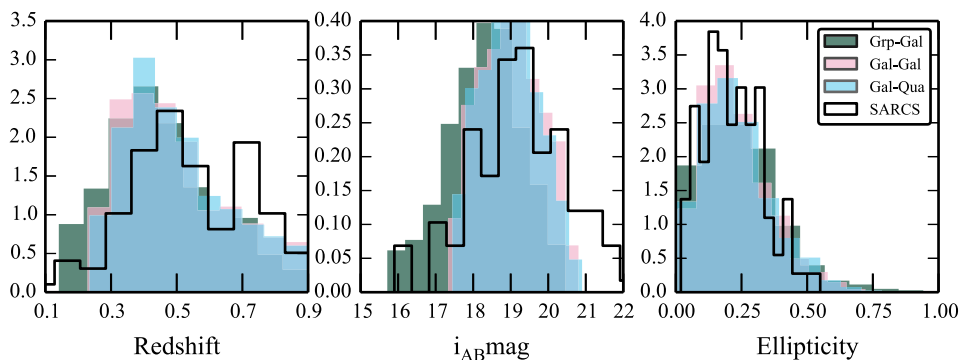
Figure 3. Distributions of properties of the 'lensing' galaxies of the simulated sample compared to the known lens sample SARCS.

In Fig. 2, we show the Einstein radius (R_E) distribution for the galaxy-scale and group-scale simulated lenses. For comparison, we give the expected distributions (blue solid curves) for an SIS (singular isothermal sphere) like density profile at galaxy scales and an NFW+Hernquist profile at group scales. The theoretical curves are taken from More et al. (2012) wherein the models are explained in detail. We note that the model we adopt at the group scale also includes SIS or SIE components for the group members unlike the theoretical prediction. The theoretical curves have arbitrary normalizations. We also show the distribution of the total magnification for all three samples.

Next, we consider the redshift, magnitude and ellipticity distributions of the 'lensing' galaxies from the simulated sample as shown in Fig. 3. For reference, we also show SARCS lenses from More et al. (2012), with arbitrary normalizations. We find that the properties of the foreground lenses in the simulated and the real lens samples are broadly similar.

We produce catalogues with lens and source properties for each of the three types of simulated lenses. The catalogues typically have lens position, redshift, magnitudes, Einstein radius, ellipticity (whenever available) and shear (for galaxy-scale lenses only). For the background sources, we provide the offset from the lens centre, redshift, magnitudes, total magnification, number of lensed images. Additionally, when possible, ellipticity and effective radius of the background galaxies have also been provided. These catalogues are available from https://github.com/anupreeta27/SIMCT and the simulated lens image sample is available from the authors on request.

3.3 Limitations of the simulated lens sample

The simulated lens sample, although realistic, is not perfect, due to the simplicity of the lensing models and our limited understanding of the uncertainties in the model parameters. Comments from citizen scientists were very helpful in order to identify some of these failures, which make up roughly 5 per cent of the simulated sample.[12] Here, we describe some of the cases or aspects in which the simulations were known to have failed to look realistic.

The parameters required by the models' various scaling relations primarily depend on the photometry of the galaxies, groups and quasars detected in the survey. For galaxy-scale lenses, the fainter or higher redshift galaxies, chosen to act as lenses, tend to

have poor photometric redshift measurements. Consequently these galaxies were occasionally assigned the wrong luminosity and velocity dispersion estimates, resulting in simulated lenses which look implausible or unrealistic. For example, the lensed images for some of the failed simulations have larger image separation than what one would expect from the luminosity and/or size of the galaxy. We roughly expect mass to follow light, so more massive galaxies typically look brighter and/or bigger.

At group scales, the magnitudes and photometric redshifts were used when defining the group membership. Therefore, errors in redshift estimates occasionally generated galaxy groups having member galaxies with unrealistically dissimilar properties. In some cases, low-redshift spiral galaxies were incorrectly assigned high redshift. Spiral galaxies are typically less massive and low-redshift spiral galaxies are unlikely to act as gravitational lenses. Hence, some such instances did not appear convincing, as the lensed images again did not have the expected configurations or separations.

We also use a single component to describe the light distribution of the background galaxies. This is clearly not the most accurate description for galaxies, especially for the irregular star-forming galaxies which comprise a significant fraction of the lensed galaxy population. Star-forming galaxies have complex structures such as star-forming knots, spiral arms, bars and discs. The simulated lensed images do not display these features. This is not problematic for most of the images taken from ground-based telescopes such as the CFHT but sometimes the profiles of the lensed arcs can appear very symmetric (along the length or width of the arc) and featureless, especially, if the images are very bright.

3.4 Duds and impostors

Citizen scientists need training not only to identify gravitational lenses, but also to reject images which either contain no lenses, or contain objects which could be mistaken for lenses. Hence, in addition to the simulated lenses, we added a sample of 'duds' and 'impostors' to the training sample. Duds are images which have been visually inspected by experts and confirmed to contain no lenses. Impostors are systems which have lens like features but are not lenses in reality, for example, spiral galaxies, star-forming galaxies, chance alignments of features arranged in a lensing configuration and stars.

We selected a sample of 450 duds for the Stage 1 classification in Space Warps and a sample of 500 impostors for the Stage 2 inspection. The sample of impostors was selected from the candidates who passed the Stage 1 of Space Warps. We note that this is the first time we have a systematically compiled sample

---

[12] This estimate is based on the number of #simfail tags from TALK, the discussion forum.

of visually inspected impostors by the SPACE WARPS volunteers and categorized by the science team. We produced an additional larger sample of a few thousand false positive detections by scanning through the low-probability images after the completion of Stage 2. All of these data products will be made available at http://spacewarps.org/#/projects/CFHTLS/. Such a sample has tremendous utility for training and testing of various lens-finding algorithms (e.g. Chan et al. 2015).

4 METHODOLOGY TO PRODUCE THE SPACE WARPS-CFHTLS LENS SAMPLE

SPACE WARPS works as a single unified system which uses the method of visual inspection to find gravitational lenses. For the first SPACE WARPS lens search, the volunteers were shown images at two stages. At Stage 1, volunteers were asked to carry out a rapid inspection to select lens candidates ranging from possible lenses to almost certain lenses. At Stage 2, volunteers were asked to inspect the candidates from Stage 1 and select only promising lens candidates. A daily snapshot of the classifications performed by volunteers was provided to the science team every night. This daily batch was analysed by the Space Warps Analysis Pipeline (SWAP). The philosophy and the details of SWAP are described in detail in Paper I. Here, we briefly summarize how it works.

Each subject (image cut-out) is assigned a prior probability of 2×10^{-4} of containing a lens system. Every volunteer is assigned an agent characterized by a 2×2 confusion matrix $\mathbf{M}$, which quantifies the volunteer's ability to correctly classify an image as containing a lens ($\mathbf{M}_{LL} = P_L$) or not containing a lens ($\mathbf{M}_{NN} = P_D$). The values of these confusion matrix elements are determined based on the performance of the volunteer on the training sample, specifically, P_L (and P_D) is determined based on the fraction of simulated lenses (and, respectively, duds) correctly classified. After every classification, the agent updates the probability of the classified subject based on the volunteer's classification and the confusion matrix, according to Bayes' theorem. The agent's confusion matrix is updated after the classification of every training image. The thresholds for the probabilities to accept or reject a subject if it contains a lens or does not contain one can be chosen in SWAP. In Stage 1, those images which cross these threshold values are 'retired', and are not subsequently shown to the volunteers. In this way, the crowd can use its time efficiently in inspecting previously unclassified subjects.

SWAP was run nightly during Stage 1 in order to retire subjects and inject new ones in to the classification stream. The subjects that passed the detection threshold at the end of Stage 1 were served again at Stage 2 for careful re-inspection. The goal of Stage 1 inspection was to maximize completeness and that of Stage 2 was to maximize purity. Each subject, after Stage 2, has a final posterior probability P. In the ideal case, all images containing lenses will have high P values and those without lenses will have low P values. In practice, we expect a small fraction of the real lenses (or non-lenses) to be assigned low (or high) P values, thereby decreasing the completeness (or purity) of the final lens sample. As this is the first lens search with SPACE WARPS, we want to find a threshold P value which will result in acceptable levels of completeness and purity of the final sample of lens candidates.

To achieve this, we selected a total of 665 subjects with $P > 0.3$ at Stage 2 which were then visually inspected by three of us (as 'lens experts', AM, AV, PJM). Each image was assigned a grade on a scale of 0–3, representing those images: 0 – unlikely to contain a lens; 1 – possibly containing a lens; 2 – probably containing a lens; and 3 – almost certainly containing a lens. The final sample

of SPACE WARPS-CFHTLS lens candidates was then produced by selecting candidates above a threshold on the averaged grade G, as described in the next section.

5 RESULTS

5.1 SPACE WARPS-CFHTLS lens sample

In this section, we describe the SPACE WARPS candidate lens sample from the CFHTLS. We find a total of 141 candidates with $G \geq 1.3$ (medium–high grade), of which 59 are new systems. This sample is further divided as follows. We have a total of 50 candidates with $1.3 \leq G < 2$ (medium grade), of which 30 are new. The quality of candidates in this category is such that at least one of the inspectors ('lens expert') thought the candidate was probably a lens (that is, a grade of 2) and a second inspector thought that it was possibly a lens (that is, a grade of 1). Among our high-grade sample ($G \geq 2$), there are a total of 91 candidates, of which 29 are new. In this category, the minimum grade by all of the inspectors was 2, suggesting that the candidates are probably or almost certainly lenses according to all three inspectors. To avoid duplication, only the newly discovered lens candidates with $G \geq 1.3$ (medium–high probability) are presented in this paper (see Section 5.2), and further information on SPACE WARPS-selected candidates that were previously identified in the literature (as described in Section 2.2) will be made available at http://spacewarps.org/#/projects/CFHTLS/.

We also find a total of 288 (and 245 new) candidates with averaged grade $0 < G < 1.3$ (low grade), which means that at least one of the inspector thought the candidate was possibly a lens, and in the best cases, all three inspectors thought the candidate was possibly a lens (that is, a grade of 1).[13] Further information on the low-probability sample such as their positions and images will be available at http://spacewarps.org/#/projects/CFHTLS/. Note that if all of the inspectors gave a grade of 0 to a candidate, then it was discarded from the sample.

In Table 2, we give overall statistics of the systems detected at Stage 1 and Stage 2. We give the total number of detections of the known lens candidates, known lenses and the new lens candidates at each stage. We also give the recovery fractions for the known samples and fraction of total detection for the new samples at each stage. Overall, the sample of new SPACE WARPS lens candidates comprises over 40 per cent of the total SPACE WARPS -CFHTLS lens candidates. We find that 90 per cent of the confirmed lenses found at Stage 1 are also recovered at Stage 2. However, ∼35 per cent of the known lenses were missed already at Stage 1: we return to the discussion of these false negatives below in Section 6.3. Nearly 45 per cent of the known lens candidates from Stage 1 are rejected at Stage 2. Such fractions are acceptable for 'candidates' as their quality grades vary from high to low.

In Fig. 4, we plot the distribution of false positives and the high-grade lens candidates, as a function of the P value assigned by SWAP at the end of Stage 2. On average, the fraction of lens candidates is indeed an increasing function of P. This shows that the SPACE WARPS generated P values for the subject are roughly correlated with the expert grades albeit with quite some scatter. We note that below $P \sim 0.75$, the fraction of false positives starts to exceed the fraction of real lens candidates. This could be a good threshold to choose to maximize the purity of the final sample.

---

[13] If grades from the inspectors were found to be discrepant by 2 or more, these were discussed and re-graded to resolve the discrepancy.

Table 2. Statistics of detections in SPACE WARPS.

| | Stage 1 | | Stage 2 | |
	KC	KL	KC	KL
Number	142	39	79	34
Fraction				
of total recovered	32 per cent	66 per cent	18 per cent	58 per cent
$P_{\mathrm{accthresh}}$	0.95	0.95	0.3	0.3
Averaged grade (G)	–	–	1.3	1.3
	Stage 2			
	NC	AC		
Number	59	141		
Fraction				
of detections	14 per cent	33 per cent		
$P_{\mathrm{accthresh}}$	0.3	0.3		
Averaged grade (G)	1.3	1.3		

Notes. KC– Known lens candidates

KL– Known lenses

NC– New lens candidates

AC– All (known and new) lens candidates

$P_{\mathrm{accthresh}}$ − systems with posterior probability P above this threshold are selected

Note. For KC and KL, percentages are with respect to the known population (i.e. 438 KC and 59 KL; see Section 2.2) whereas for NC and AC, percentages are with respect to the total sample of 429 lens candidates.
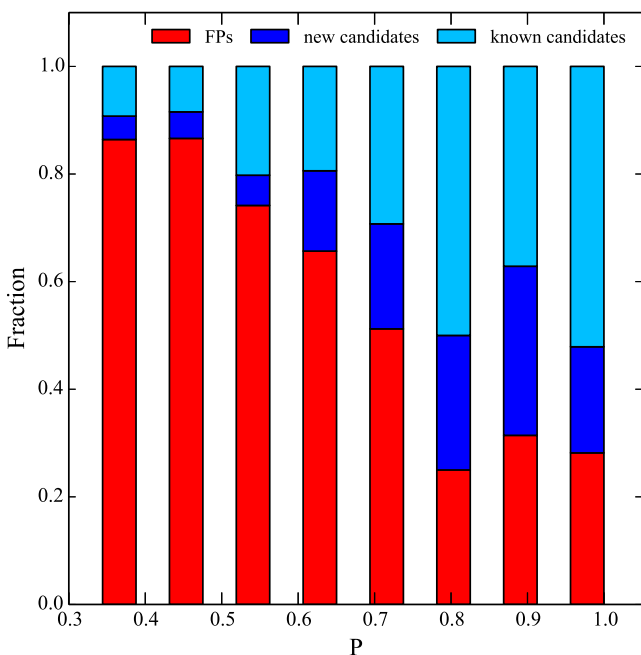
Figure 4. Distribution of different types of candidates as a function of the posterior probability P, obtained at the end of Stage 2. The types of the candidates are the false positives (FPs), the new candidates and the known candidates. The new and known candidates have higher detection rate for higher values of P, as expected.

However, choosing $P = 0.3$ gives a completeness of 92 per cent for the 'known lens' sample instead of 64 per cent for $P = 0.75$. Therefore, the new sample should also have increased completeness; expert grading then allows us to increase the purity of the sample.

5.2 New lens candidates from SPACE WARPS

We give basic information about the final sample of 59 new medium–high grade lens candidates found by SPACE WARPS in Table 3. We report the candidates with a SPACE WARPS ID and Name of the lens system. We give their positions (RA, Dec), photometric redshift (z_{phot}), i-band magnitude of the lensing galaxy, averaged grade G from the lens experts, zoo ID (identifier used in TALK), P value at Stage 2 and a visual categorization of the type of lensed images and the lensing galaxy in the 'Comments' column in Table 3. Whenever available the lens properties are taken from the CFHTLS photometric catalogue (Coupon et al. 2009); otherwise, for the lens galaxy positions, the reported values were measured manually. The visual categorization of the lens type is only suggestive and the explanation of the notations in the Comments column is given at the bottom of the table.

We show images of our new sample in Fig. 5. The panels are arranged first in the descending order of their grades, and within each grade, in ascending order of RA. As the first lens search was a blind search with no pre-selection of candidates from any algorithm, we find various types of lenses, as expected from such a search. The final sample consists of both galaxy- and group-scale lens candidates. There are detections of elongated arcs and some interesting point-like quasar lensed images. Most of them are brighter in the bluer g band, but some candidates brighter in the redder i band are also found. Since the robotic lens searches focused on the blue lensed features, they are likely to miss such interesting lens candidates. We did not find any examples of exotic lens candidates from the visually inspected $P > 0.3$ sample. There may be some more interesting candidates that were missed either at Stage 1 or Stage 2 but have been identified by volunteers in Talk. This resource is yet to be mined and is left for future work.

The new SPACE WARPS lens sample presented here illustrates some of the advantages of having citizen scientists find lenses through visual inspection. An algorithm, by definition will find objects that adhere to a selection criteria that uses either geometry or flux information from an image. On the other hand, citizen scientists can interpolate over or extrapolate beyond the basic selection criteria provided to them. For example, the lower blue arc in SW7 is split by a small red galaxy. An algorithm typically fails to detect such arcs because the arc is broken into smaller arclets which then falls below the minimum length or area allowed for an arc to be detected. Human inspectors have no problem in interpolating over the broken blue arc over the red galaxy, understanding that it is a single long arc. The system SW20 has point-like lensed images which cannot be detected by arc-finding algorithm, whereas the ring-finding algorithm may have missed this because of the atypical colour and structure of the lensing galaxy. Detection of red arcs, for example, as seen in the SW39 candidate, shows how the volunteers extrapolate on the colour parameter: the training sample contains predominantly blue arcs, because the source colours were drawn from realistic observed distributions.

The power of citizen scientists also lies in the high dynamic range that allows us to find systems which have very short (thick) to long (thin) arcs, from highly compact to low surface brightness images, from round and point-like to elongated and curved images, from blue to red, from regular to exotic kinds of lenses, while keeping the false positive rate low compared to algorithms. Discovery of this large sample of completely new candidates missed by some of these algorithms demonstrates that the SPACE WARPS system is functioning well, the self-taught citizen scientists reaching parts of discovery space that the robots did not.

Table 3. Sample of the SPACE WARPS new lens candidates.

SW ID	Name	RA (deg)	Dec (deg)	z_{phot}	m_i (mag)	R_E (arcsec)	G	ZooID	P	Comments
SW19	CFHTLS J020642.0−095157	31.675 04	−9.865 84	0.2	20.8	0.9	2.0	ASW0001ld7	0.8	A,R
SW8	CFHTLS J020648.0−065639	31.700 31	−6.944 30	0.8	20.2	1.3	2.3	ASW00099ed	0.4	A,E
SW43	CFHTLS J020810.7−040220	32.044 97	−4.038 91	1.0	20.8	1.8	1.3	ASW0001c3j	0.7	A,R
SW9	CFHTLS J020832.1−043315	32.133 96	−4.554 29	1.0	21.0	1.6	2.3	ASW0002asp	1.0	A,R
SW10	CFHTLS J020848.2−042427	32.201 10	−4.407 51	0.8	20.5	1.1	2.3	ASW0002bmc	0.9	D,D
SW11	CFHTLS J020849.8−050429	32.207 84	−5.074 94	0.8	20.6	0.9	2.3	ASW0002qtn	1.0	A,R
SW44	CFHTLS J021021.5−093415	32.589 81	−9.571 09	0.4	18.4	2.7	1.3	ASW0002k40	0.4	D,S
SW30	CFHTLS J021057.9−084450	32.741 48	−8.747 45	0.0	0.0	2.5	1.7	ASW0002p8y	0.4	A,G
SW20	CFHTLS J021221.1−105251	33.088 10	−10.881 06	0.3	17.9	1.8	2.0	ASW0002dx7	0.8	D,E/S
SW45	CFHTLS J021225.2−085211	33.105 11	−8.869 73	0.8	19.5	2.1	1.3	ASW00024id	1.0	R,R
SW46	CFHTLS J021317.6−084819	33.323 41	−8.805 49	0.5	19.8	1.3	1.3	ASW00024q6	0.4	A,R/E
SW31	CFHTLS J021514.6−092440	33.810 89	−9.411 15	0.7	19.9	2.6	1.7	ASW00021r0	0.4	A,R/G
SW32	CFHTLS J022359.8−083651	35.999 55	−8.614 39	0.0	0.0	3.1	1.7	ASW0004iye	0.4	A,E
SW12	CFHTLS J022406.1−062846	36.025 58	−6.479 63	0.4	19.6	0.9	2.3	ASW0003wsu	0.7	A,E
SW1	CFHTLS J022409.5−105807	36.039 78	−10.968 85	0.0	0.0	4.8	3.0	ASW0004dv8	1.0	A,G
SW21	CFHTLS J022533.3−053204	36.388 82	−5.534 60	0.5	19.4	3.6	2.0	ASW0004m3x	0.4	A,R/G
SW22	CFHTLS J022716.4−105602	36.818 56	−10.934 10	0.4	17.3	1.8	2.0	ASW0009ab8	0.7	A,E/G
SW33	CFHTLS J022745.2−062518	36.938 68	−6.421 83	0.6	20.5	1.2	1.7	ASW0003s0m	0.5	A,R
SW13	CFHTLS J022805.6−051733	37.023 62	−5.292 66	0.4	18.8	1.4	2.3	ASW00047ae	1.0	Q,E
SW47	CFHTLS J022843.0−063316	37.179 42	−6.554 65	0.5	19.1	1.8	1.3	ASW0003r6c	0.3	D/A,E
SW23	CFHTLS J023008.6−054038	37.535 91	−5.677 44	0.6	19.7	1.9	2.0	ASW0003r61	0.5	A,E
SW14	CFHTLS J023123.2−082535	37.846 82	−8.426 63	0.0	0.0	1.2	2.3	ASW0004xjk	0.3	A,R
SW24	CFHTLS J023315.2−042243	38.313 34	−4.378 86	0.7	19.7	1.0	2.0	ASW00050sk	0.8	A,R
SW34	CFHTLS J023453.5−093032	38.723 21	−9.508 92	0.5	19.8	0.7	1.7	ASW000051ld	0.3	A,D
SW35	CFHTLS J084833.2−044051	132.138 47	−4.680 85	0.8	20.2	0.9	1.7	ASW0004wgd	0.7	A,R
SW15	CFHTLS J084841.0−045237	132.170 84	−4.877 20	0.3	19.0	1.0	2.3	ASW0004nan	1.0	A,E
SW48	CFHTLS J090219.0−053923	135.579 47	−5.656 66	0.0	0.0	2.0	1.3	ASW0000g95	1.0	A,R/E
SW36	CFHTLS J090248.4−010232	135.702 04	−1.042 43	0.4	19.1	1.4	1.7	ASW000096t	0.6	D,E
SW25	CFHTLS J090308.2−043252	135.784 49	−4.547 89	0.0	0.0	1.3	2.0	ASW00007mq	0.6	D,D
SW49	CFHTLS J090319.4−040146	135.831 05	−4.029 71	0.0	19.8	1.2	1.3	ASW00007ls	0.5	A,R/E
SW50	CFHTLS J090333.2−005829	135.888 69	−0.974 90	0.0	0.0	2.1	1.3	ASW00008a0	1.0	A/D,E/G
SW51	CFHTLS J135724.8+561614	209.353 74	56.270 66	0.0	0.0	2.6	1.3	ASW0006e0o	0.9	D,E
SW26	CFHTLS J135755.8+571722	209.482 68	57.289 71	0.8	20.2	0.9	2.0	ASW0005ma2	0.8	A,R
SW52	CFHTLS J140027.9+541028	210.116 36	54.174 55	0.0	0.0	1.2	1.3	ASW0006a07	0.6	Q,R/E
SW16	CFHTLS J140030.2+574437	210.126 01	57.743 71	0.4	18.2	2.0	2.3	ASW0009bp2	0.6	A,E
SW2	CFHTLS J140522.2+574333	211.342 61	57.725 87	0.7	19.7	1.0	2.7	ASW000619d	0.7	A,R
SW17	CFHTLS J140622.9+520942	211.595 81	52.161 69	0.7	20.3	1.2	2.3	ASW0005rnb	0.7	A,R
SW27	CFHTLS J141432.9+534004	213.637 16	53.667 88	0.7	21.4	1.0	2.0	ASW0006jh5	0.8	A,R
SW53	CFHTLS J141518.9+513915	213.829 03	51.654 20	0.4	18.3	3.0	1.3	ASW00070vl	0.8	D,E
SW3	CFHTLS J142603.2+511421	216.513 75	51.239 35	0.0	0.0	4.4	2.7	ASW0006mea	0.7	A,G
SW54	CFHTLS J142620.8+561356	216.586 99	56.232 30	0.5	19.5	1.3	1.3	ASW0007sez	1.0	A/R,S
SW55	CFHTLS J142652.8+560001	216.720 04	56.000 44	0.0	0.0	1.5	1.3	ASW0007t5y	1.0	R,R
SW56	CFHTLS J142843.5+543713	217.181 53	54.620 36	0.4	19.7	1.3	1.3	ASW0007pga	0.6	D,D
SW4	CFHTLS J142934.2+562541	217.392 61	56.428 07	0.5	19.0	5.9	2.7	ASW0009cjs	0.8	A,G
SW28	CFHTLS J143055.9+572431	217.733 32	57.408 83	0.7	19.3	1.6	2.0	ASW0007xrs	0.9	A,R/G
SW37	CFHTLS J143100.2+564603	217.751 24	56.767 50	0.0	0.0	1.2	1.7	ASW00086xq	0.8	A,E
SW38	CFHTLS J143353.6+542310	218.473 57	54.386 24	0.8	19.8	1.8	1.7	ASW0009cp0	0.7	A,E
SW5	CFHTLS J143454.4+522850	218.727 02	52.480 80	0.6	19.4	4.4	2.7	ASW0007k4r	0.4	Q,G/R
SW6	CFHTLS J143627.9+563832	219.116 36	56.642 49	0.5	19.5	1.5	2.7	ASW0008swn	0.9	A,D
SW57	CFHTLS J143631.5+571131	219.131 55	57.192 15	0.7	20.9	1.3	1.3	ASW0008pag	0.6	D/A,R
SW58	CFHTLS J143651.6+530705	219.215 03	53.118 32	0.6	19.2	3.1	1.3	ASW0007iwp	0.7	A,E/G
SW18	CFHTLS J143658.1+533807	219.242 46	53.635 50	0.7	19.6	0.9	2.3	ASW0007hu2	0.6	D,D
SW29	CFHTLS J143838.1+572647	219.658 87	57.446 45	0.8	20.2	1.1	2.0	ASW0008qsm	0.9	A,R
SW59	CFHTLS J143950.6+544606	219.961 01	54.768 58	0.0	0.0	1.7	1.3	ASW00085cp	0.4	A,G/R
SW39	CFHTLS J220215.2+012124	330.563 48	1.356 67	0.3	17.4	4.6	1.7	ASW0005qiz	0.5	A,G
SW7	CFHTLS J220256.8+023432	330.736 91	2.575 81	0.0	0.0	6.8	2.7	ASW0007e08	0.8	A,G
SW40	CFHTLS J221306.1+014708	333.275 79	1.785 61	0.0	17.1	1.4	1.7	ASW0008wmr	0.9	A,S
SW41	CFHTLS J221519.7+005758	333.832 12	0.966 15	0.4	20.2	1.0	1.7	ASW0008xbu	0.8	A,D
SW42	CFHTLS J221716.5+015826	334.318 94	1.973 94	0.1	21.6	1.0	1.7	ASW00096rm	1.0	A/R,R

Notes. The column Comments has two type of notes. The first is about the lens image configuration where the symbols mean the following A: arc, D: double, Q: quad, R: ring. The second is a comment on the type of lens assessed visually. Note that this classification is not based on colours or spectral analysis. The symbols are E: elliptical, S: (face on) spiral, G: group-scale, D: edge-on disc, R: red star-forming galaxy.
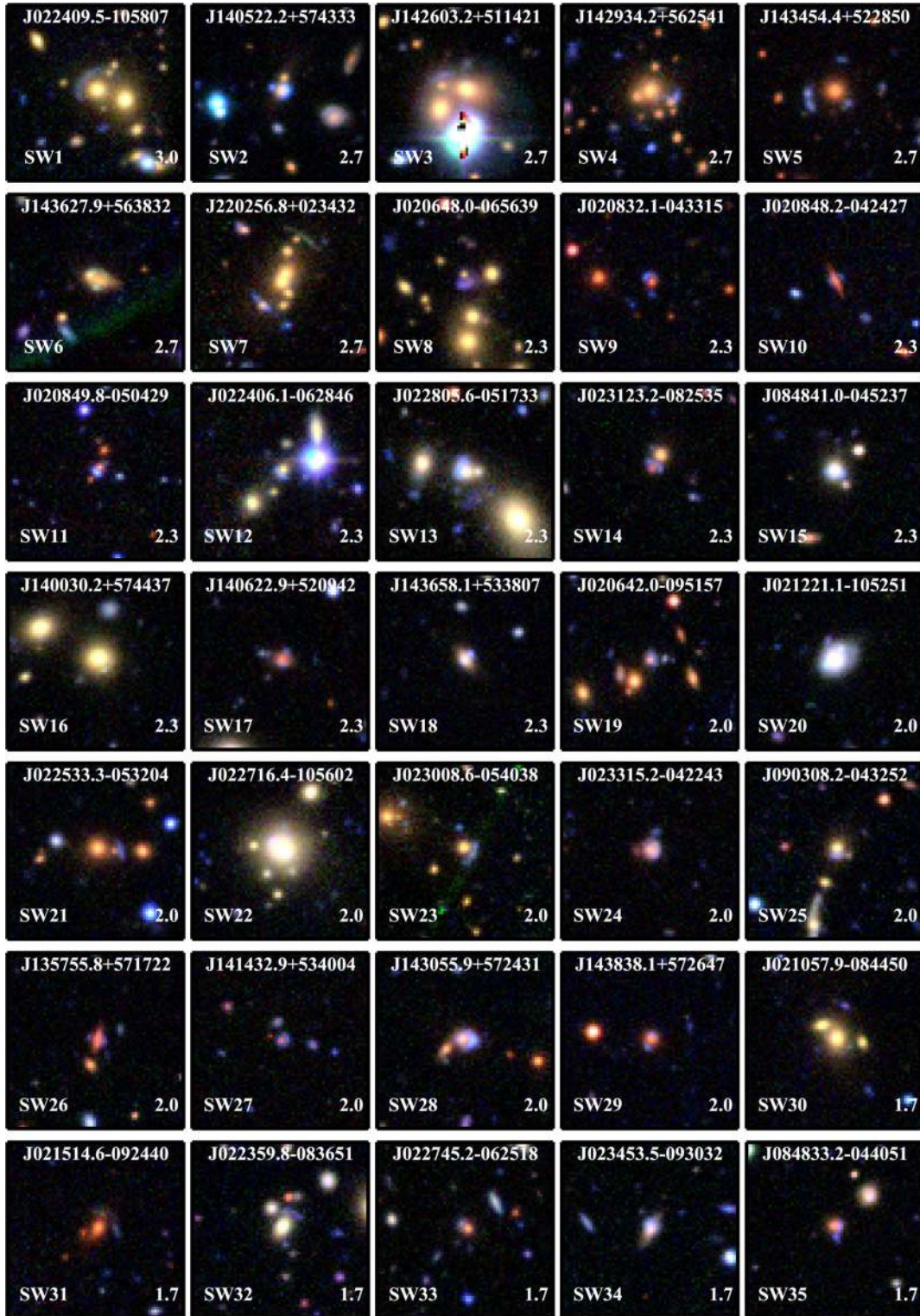
Figure 5. The new SPACE WARPS lens candidates with expert grade $G >= 1.3$. The images are 30 arcsec on the side.

Further detailed qualitative and quantitative analysis of the properties of the entire SPACE WARPS sample (new and previously identified candidates) and the mass modelling analyses for the new candidates will be presented in a subsequent SPACE WARPS paper (Verma et al., in preparation).

5.3 Measurements of properties of the lens and the lensed images

In the subsequent sections, we compare various properties of the lens candidates. Here, we describe how we extract or measure these properties, namely, the lens redshift, the Einstein radii and the total flux of the lensed images or arcs.
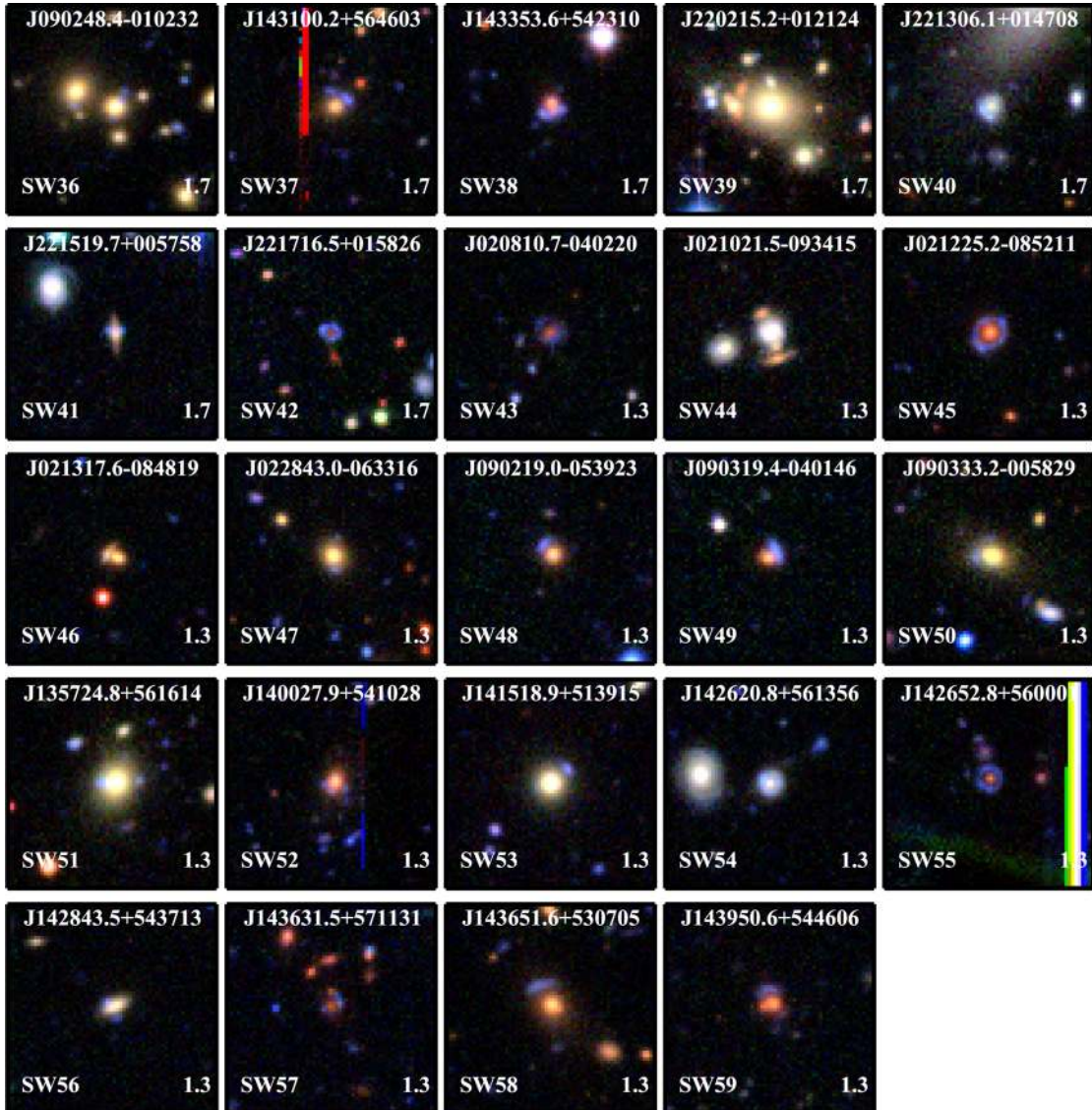
Figure 5 – *continued.*

We use the publicly available redshifts for the lens galaxy from the CFHTLS photometric catalogues (Coupon et al. 2009). The total flux of the lensed image or arc is measured in the *g* band but the adopted method is different for different samples. For the simulated sample, we multiply the magnification of the second brightest image with the source magnitude. For the RINGFINDER sample, the arcs are detected in the scaled difference image of *g* and *i* bands from which the lensing galaxy is subtracted (for details, see Gavazzi et al. 2014). Here, we use the flux of the lensed images measured by SEXTRACTOR from the scaled difference image, that is, $g - \alpha i$ and convert it to the *g*-band flux using mean colours of the foreground and background population. For the ARCFINDER and the SPACE WARPS sample, we integrate the flux in the image pixels identified by ARCFINDER or SEXTRACTOR.

The Einstein radius is also measured differently for different samples. For the galaxy-scale lenses in the simulated sample, we use the value of the input parameter of the lens model for the R_E. For group-scale lenses, since the lens model is multicomponent, we need to determine the R_E from the image positions. We use those pairs of lensed images that have the smallest and the largest angular separations. The R_E here is then half of the averaged values of these angular image separations. For the RINGFINDER sample, we use the peak position of the lensed images measured by running SEXTRACTOR on the scaled difference image. We calculate the image separation from the lens centre as a rough estimate of the R_E. For the ARCFINDER (SARCS) sample, we use the same definition as above except that the peak position is identified either by the ARCFINDER or manually. For the SPACE WARPS lens sample, the same definition is used where the peak positions are identified either with ARCFINDER or SEXTRACTOR.

5.4 Recovery of known lens samples from the CFHTLS by SPACE WARPS

We now determine the fraction of the known sample of lenses that were recovered by SPACE WARPS. Note that this sample corresponds to the RINGFINDER and ARCFINDER samples combined, as defined in Section 2.2. In Table 2, we show that ∼32 per cent of the known lens candidates, and ∼65 per cent of the known lenses were found at Stage 1. We find that 56 per cent of the known lens candidates
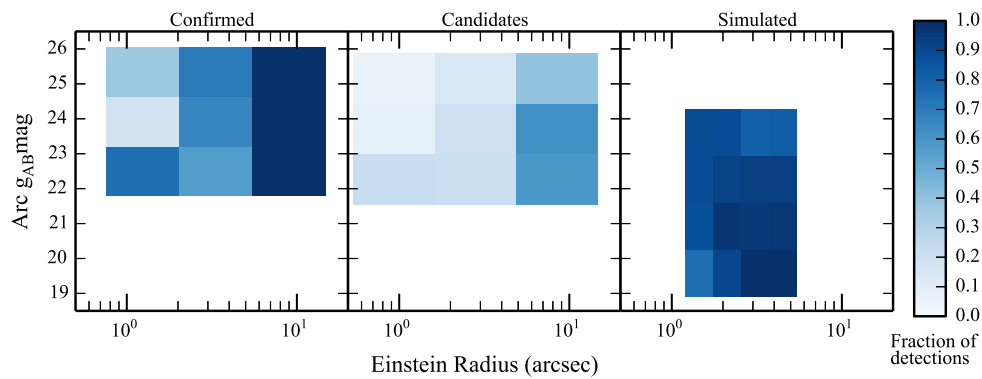
Figure 6. Fraction of lens candidates recovered by SPACE WARPS as a function of the arc magnitude (*g* band) and the Einstein radius for three lens samples, namely, the known lenses, the known lens candidates and the simulated sample.

and 87 per cent of the known lenses from Stage 1 passed our Stage 2 selection criterion of $P > 0.3$ and $G > 1.3$. The left-hand and the middle panels of Fig. 6 show the fraction of detections as a function of arc magnitude and the Einstein radius of the lens systems for the known confirmed lenses and lens candidates. As expected, we find that systems with brighter images and/or with larger R_E are detected more often in SPACE WARPS.

We find that most of the confirmed lenses and candidates that are missed by SPACE WARPS are systems from the RINGFINDER sample, with fainter arcs and smaller R_E. The main reason why RINGFINDER found such candidates is because it involves subtracting the light from the lens galaxy, making it easier to detect the lensed images during both the automated object-finding, and the visual inspection and classification phases. This approach naturally improves the detection efficiency at smaller R_E and for fainter systems. The SPACE WARPS volunteers were not shown any galaxy-subtracted images. Showing galaxy-subtracted images might be a better strategy to adopt for future lens searches at galaxy scales with SPACE WARPS. However, we note that accurate modelling and subtraction of the galaxy light profile in different bands is challenging and better techniques are being actively developed to enhance detections of lenses at small image separations. In Section 6.3 below, we further explore and discuss why some of the confirmed lenses were missed by SPACE WARPS.

5.5 Image separation distribution

The distribution of image separations (i.e. $2 R_E$) can be used to probe the average density profile of the lens population (Oguri 2006; More et al. 2012). However, the lens sample found by the ARCFINDER may have incompleteness as a function of the image separation. Thus, the lack of understanding of the selection function of the lens sample may affect the constraints on the density profile. A blind lens search done by visual inspection alone, for example, through SPACE WARPS citizen scientists may find lenses missed by the ARCFINDER search and thereby, improve completeness.

Indeed, we have found 59 new medium–high grade lens candidates that were not known before. In Fig. 7, we show the image separation distribution using all the known and new lens candidates. The different data points are the known RINGFINDER and ARCFINDER sample (green), the SPACE WARPS identified (known and new) lens sample (blue) only and the combined CFHTLS sample of RINGFINDER, ARCFINDER and the new SPACE WARPS lens sample (magenta). It is interesting to note that both the RINGFINDER +ARCFINDER and SPACE WARPS samples have very similar profiles and thus, the
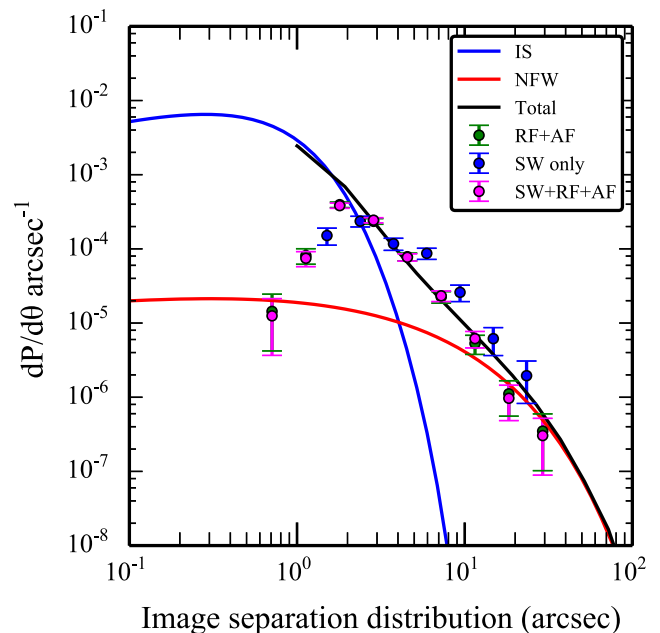
Figure 7. Image separation distribution (ISD). Comparing theoretical predictions (solid curves) with the CFHTLS known lens samples (green) and the combined sample of known and SPACE WARPS lens candidates (magenta). The sample of new and the known lens candidates discovered from SPACE WARPS alone is shown in blue. The new updated profile of the ISD (magenta) is consistent with our previous measurements and strengthens our conclusion that the average density profiles of the lenses are similar to the Total profile.

profile of the combined sample has not changed significantly. This implies that previous constraints on the image separation distribution are robust and the ARCFINDER selected sample does not suffer from significant incompleteness for medium to large R_E. This is the regime that probes density profiles of galaxy groups to clusters.

In the figure, we also show for comparison the theoretical predictions corresponding to three density profiles, namely, isothermal sphere (IS), NFW and a 'Total' profile which has NFW and Hernquist profiles combined with an adiabatically contracting model for the dark matter component (Gnedin et al. 2004). These curves are taken from More et al. (2012), which gives details of the calculation of these predictions. With the updated sample of lens candidates, we confirm our previous prediction that the mass density profiles of galaxy groups is indeed consistent with the 'Total' profile. At smaller image separations ($\lesssim 2$ arcsec), the 'Total' profile converges
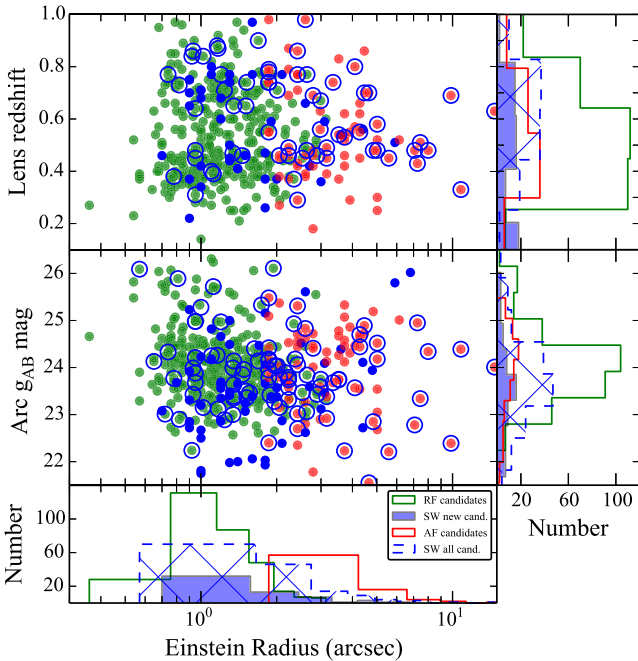
Figure 8. Comparison of the lens redshift and the arc magnitude with the Einstein radius for all of the three lens samples, namely, the RINGFINDER (green dots), SPACE WARPS (known candidates–blue circles; and new candidates only–blue dots) and ARCFINDER (red dots). All samples have broadly similar properties.

to isothermal like case and assuming these predictions are reliable, we find that the lens samples have very low completeness. This is not too surprising compared to the 40 per cent completeness expected for the RINGFINDER sample (Gavazzi et al. 2014).

6 DISCUSSION

Finding gravitational lenses is a difficult and complex task. No single method is perfect, each method has some advantages over the other. It may be the case that a single method may never be the best method for optimizing completeness and purity. Visual inspection will likely be required for pruning candidates at some stage of lens candidate selection even in the future. Therefore, we would like to understand how best we should combine the strengths of robots and humans to optimize the lens-finding method.

In this section, we first compare the lens candidates found by SPACE WARPS and the lens-finding robots and then attempt to understand why each method failed to detect lenses from the other sample.

6.1 Comparison of the RINGFINDER, SPACE WARPS and ARCFINDER samples

In Fig. 8, we show the lens redshift and the arc flux measured in *g*-band AB magnitude as a function of the Einstein radius for the RINGFINDER (green), the ARCFINDER (red) and the SPACE WARPS sample (new candidates only in blue and known candidates as blue circles). We note that the errors on the redshift measurement should not be too different across the samples since they are measured by a single method. However, the error on the total flux of the lensed images is likely to be different across the samples and the types of systematics are also different. We have not attempted to quantify these errors in this work. With that caveat, we find that the SPACE WARPS candidates sample is broadly similar to the robotically found lens candidates in terms of the flux of the lensed images and the redshift of the lensing galaxies.

The properties considered here do not show any clear differences between the types of lenses being found by each method. Other properties such as the flux of the lensing galaxies and the surface brightness of the lensed images may be useful in showing some qualitative differences but this is beyond the scope of our current analysis. A more detailed and accurate analysis is deferred to future.

In Fig. 9, we show the relative distribution of number of candidates from each sample as a function of the Einstein radius and arc magnitude. The light blue colour shows the overlap between the SPACE WARPS and the RINGFINDER samples and the purple colour shows the overlap between the SPACE WARPS candidates and the ARCFINDER samples. As noted earlier, the RINGFINDER
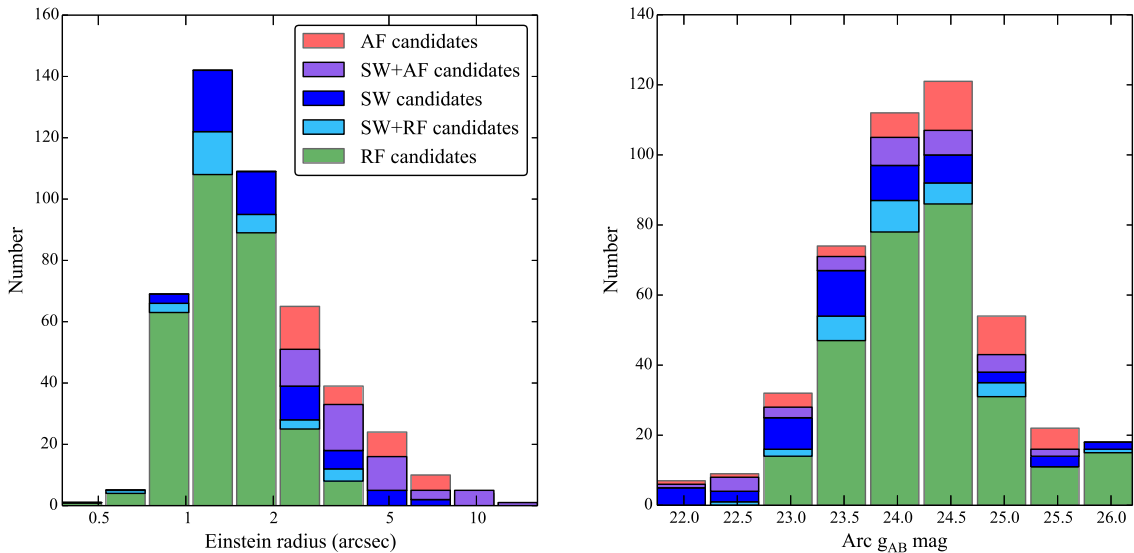
Figure 9. Candidate detections by the RINGFINDER, SPACE WARPS and the ARCFINDER as a function of the Einstein radius and *g*-band magnitude of the lensed images.

dominates the small R_E (<2 arcsec) detections although Space Warps does find a modest number of candidates in this range. At larger R_E, Space Warps sample begins to dominate and is comparable to the Arcfinder sample. As a function of the arc magnitudes, all three samples have detections at all magnitudes and median magnitudes for all samples is around $g \sim 24.5$. Relatively, the Ringfinder sample spans a narrower range compared to the Space Warps and Arcfinder sample. However, this can be verified only after understanding and accounting for the systematic uncertainties in our measurements.

6.2 Why were the new Space Warps candidates missed by the robots?

We test the Ringfinder and Arcfinder on images centred on the new Space Warps candidates to trace and understand at what stage the algorithm failed to detect them.

First, we reran Ringfinder on the new Space Warps sample. At the beginning, a galaxy catalogue is generated based on magnitude, redshift and SED (spectral energy distribution) type (see Gavazzi et al. 2014) to select galaxies which are most likely to act as lenses. We find that about 40 per cent of the new Space Warps candidates failed to meet this initial selection criteria, for example, SW1, SW14, SW20, SW23, SW27 and SW30. All of the lensing galaxies are bright enough to satisfy the magnitude criterion ($i < 22$). However, some of them have a bright companion galaxy, some of them do not look like E/S0 type galaxies and some are edge-on galaxies which could be the reason for these galaxies having failed the photometric redshift and SED type pre-selection.

In subsequent Ringfinder steps, the flux from the galaxy is subtracted from the scaled difference image to enhance the visibility of the faint blue lensed features. An object finder is then run on this image to quantify the lensed image properties.

Another $\sim$50 per cent of the Space Warps candidates could not be detected by the object finder because properties such as the image area, axis ratio, magnitude/colour and alignment with respect to the lensing galaxy were not satisfied. Some of the candidates missed at this stage are, for example, SW4, SW5, SW6, SW26, SW36, SW39 and SW46.

Next, we reran the Arcfinder on the same Space Warps sample of new candidates. The Arcfinder is directly run on the images to look for elongated arc-like objects and does not require a list of targets to begin with. Objects are identified by placing thresholds on the noise level in the images. Thus, Arcfinder detections are sensitive to changes in the noise levels.

Originally, the Arcfinder was run on a large image with an area of $\sim$19 350 × 19 350 pixels squared. For the rerun, we worked with much smaller images because this is faster. However, this alters the measured noise and hence affects the number and type of arc detections. We find that about 30 per cent of the new candidates were detected without changing any of the thresholds in the code, suggesting that these could have been detected by Arcfinder had its noise thresholds been set differently.

The Arcfinder code calculates second-order brightness moments around every pixel to decide if the distribution of flux is elongated in some direction in order to detect elongated arc-like objects. An elongation estimator is assigned to every pixel. All pixels with a value of the elongation estimator above a certain threshold are connected to form the arc feature. This is called the segmentation of the arc candidate. Subsequently, arc properties such as the area, mean flux, length and curvature are determined. We relaxed the threshold at the segmentation stage and also relaxed thresholds

mainly on the area of the arc. These new settings led to the detection of about 75 per cent of the new Space Warps candidates. We find that relaxing thresholds on other arc properties does not improve the detection rate significantly.

Typically then, the Space Warps candidates were missed from the Arcfinder sample either because (a) the arcs were fainter, (b) the flux of the arc and the galaxy were blended together (such that the Arcfinder incorrectly connected part of the galaxy to the arc), (c) the arcs were unusually short or thick, or (d) the lensed images are almost circular or point-like (Arcfinder was not designed to detect lensed quasars).

Relaxing the Arcfinder thresholds obviously increases the number of candidate arc detections but this also increases the false positive rate. For example, the number of arc candidate detections increased by a factor of $\sim$2 when we relaxed the thresholds in the rerun described above, while the number of false positives increased by a factor of $\sim$5. While the Arcfinder sample purity could be increased by cross-correlating the arc candidate positions with a putative lens galaxy catalogue, these numbers illustrate the predicament facing automated lens-finding algorithms, and the continuing benefits of visual screening.

## 6.3 False negatives: known lenses missed by Space Warps

Like any lens-finding method, the Space Warps system could potentially be failing to detect certain kinds of lenses. We find that about 35 per cent of the known sample of lenses are missed at Stage 1; about 10 per cent losses were incurred during the Stage 2 refinement (see Table 2). Below, we focus on the known lens sample at Stage 1 to understand why some of them are being missed and possibly find a way to improve the detection rate which can be adopted in the future Space Warps lens searches.

Many of the missed lenses are from the Ringfinder sample with small Einstein radii and faint lensed images (see Fig. 6). Among the confirmed lenses from Ringfinder, about 45 per cent are missed. Out of the missed sample of 15 lenses, about half of them are visually difficult to detect. The other half appear to have faint blue smudges around galaxies which should have been easier to identify. Similarly, if we consider the Arcfinder lens sample, $\sim$20 per cent are missed by Space Warps. This is a relatively small sample of $\sim$5 systems and visual inspection of them suggests that, by and large, either the lensed features are faint or they have odd properties which makes them difficult to identify correctly. For further tests, we combine the Ringfinder and Arcfinder sample.

For a lens-finding method which uses the collective skill, experience and knowledge of a group of volunteers, it may be difficult to find a single factor with certainty which causes a lens candidate to be missed. We attempt to understand whether there is indeed a single dominant factor that is resulting in the loss of these lenses, or if the lenses are being missed due to a combination of multiple reasons. Below, we consider some of the factors that could affect the efficiency of finding lenses.

6.3.1 Number of classifications

First, we check whether the number of classifications (Nclass) is significantly lower for the missed sample compared to the detected one. Most of the lenses in the known sample (including both those that were detected and those that were missed) received similar numbers of classifications to the other subjects. A few received
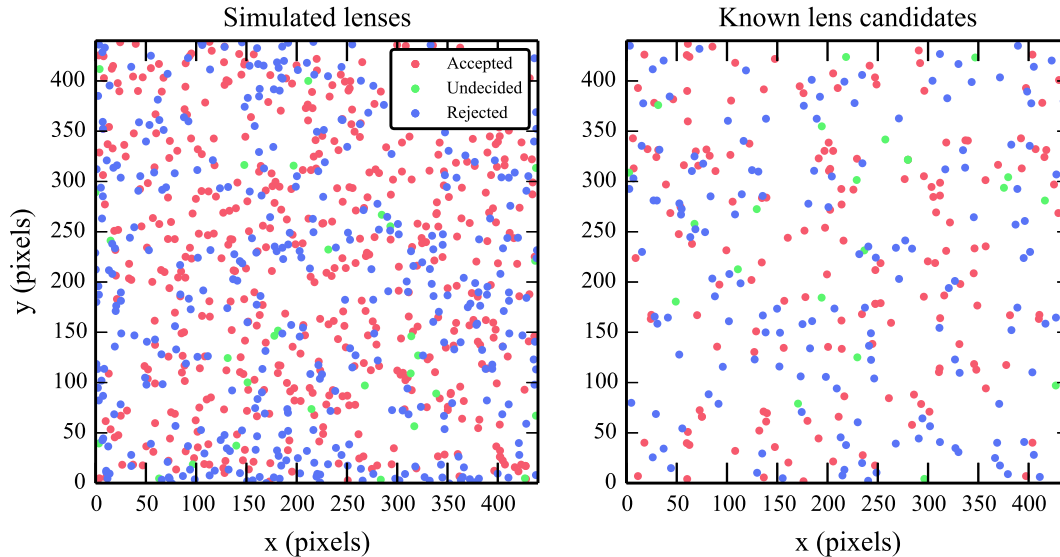
Figure 10. Completeness as a function of the positions of the lens systems. Simulated lenses (left) and real lens candidates (right) are shown. Irrespective of the status of the lenses, that is, detected, undecided or rejected, there is no strong dependency on the location of the lenses, both for the simulated and the real sample of candidates.

Nclass>20, possibly as a result of continuing to remain for a long time in the data base because there was uncertainty over whether or not to reject them. Overall, we do not find any difference in the number of classifications between the detected and the missed lenses.

6.3.2 Lens positions within the image cut-outs

The efficiency of a visual search could potentially vary in different sections of an image. Our eyes tend to focus usually at the centre of an image and lens candidates close to the borders could go undetected. Therefore, it is important to check whether SPACE WARPS could be missing some of the known lenses because they happen to be close to the borders of the image cut-outs.

From the SWAP, the image cut-outs inspected by the SPACE WARPS volunteers receive a status of detected (if $P > P_{\mathrm{accthresh}}$), rejected (if $P < P_{\mathrm{rejthresh}}$) and undecided (if $P_{\mathrm{rejthresh}} < P < P_{\mathrm{accthresh}}$). In Fig. 10, we compare the positions of lenses which are detected (red), undecided (green) and rejected (blue). The left-hand and the right-hand panels have the simulated lens sample and the known lens candidates sample, respectively. We note that the density of points do not represent the actual number of detections because, for cases with large sample size, randomly drawn subsamples are shown for the ease of visual comparison.

We do not find any strong correlation in the detection rate of lenses as a function of their positions in the image, for either the simulated or the known lens sample. Thus, the completeness of the lens sample is most likely not significantly affected by lenses located close to the image borders.

6.3.3 Classification power

Each SPACE WARPS image classification is based on the markers placed by around 10 volunteers (on average, Paper I). This number could be small enough to introduce some scatter in the system performance, arising from the variations between the small groups of volunteers inspecting each subject. In Paper I we investigated the system performance in terms of the 'Skill' of each agent; in

Appendix A we define a complementary property, the 'Power' of a classification to make a large difference in the probability P of an image containing a lens. Here, we investigate whether the distribution of classification power is systematically different between the detected and missed lenses.

We check how the posterior probability P (see Paper I for the mathematical definition) of an image or a subject to contain a lens changes as the image receives more classifications from multiple volunteers. A graphical representation of changing probabilities for increasing classifications is called a trajectory plot. In Fig. 11, we show the trajectory plots of a few examples of detected lenses (top row of panels) and missed lenses (bottom row of panels) from Stage 1 of SPACE WARPS. Every subject is assigned a prior probability $P_0 = 2 \times 10^{-4}$ (grey dashed line) and starts at the middle of the trajectory plot. The number of classifications (Nclass) for a subject increases from top to bottom (subjects move down the trajectory plots as they are classified). The P value of a subject is updated with every classification from the volunteer. If a volunteer identifies a lens candidate, the trajectory moves to the right otherwise moves to the left. A subject is accepted if it crosses the blue-dashed line marking the $P_{\mathrm{accthresh}}$ (set to 0.95 for Stage 1) on the right. It is rejected if it crosses the red-dashed line marking the $P_{\mathrm{rejthresh}}$ (set to 10^{-7} for Stage 1) on the left.

The amount by which the posterior probability P value of a subject will change depends on how well the volunteers are performing on the training sample and its current probability. Thus, for a given current probability, some volunteers will change the P by a large factor compared to others. This is evident in the trajectory plots, which show both large and small distances between consecutive points which we refer to as kicks. Comparison of the kick sizes between the detected and the missed lenses suggests that *the missed lenses do not have as many volunteers giving large kicks*. We also note that most of the large kicks seen in the trajectories of the missed lenses seem to be moving the subjects to the right. In other words, certain 'high-power' volunteers are mostly classifying them as subjects with lens candidates.

The bottom panels of Fig. 11, show the trajectories of missed lenses for the cases which are visually easier (light green) and more
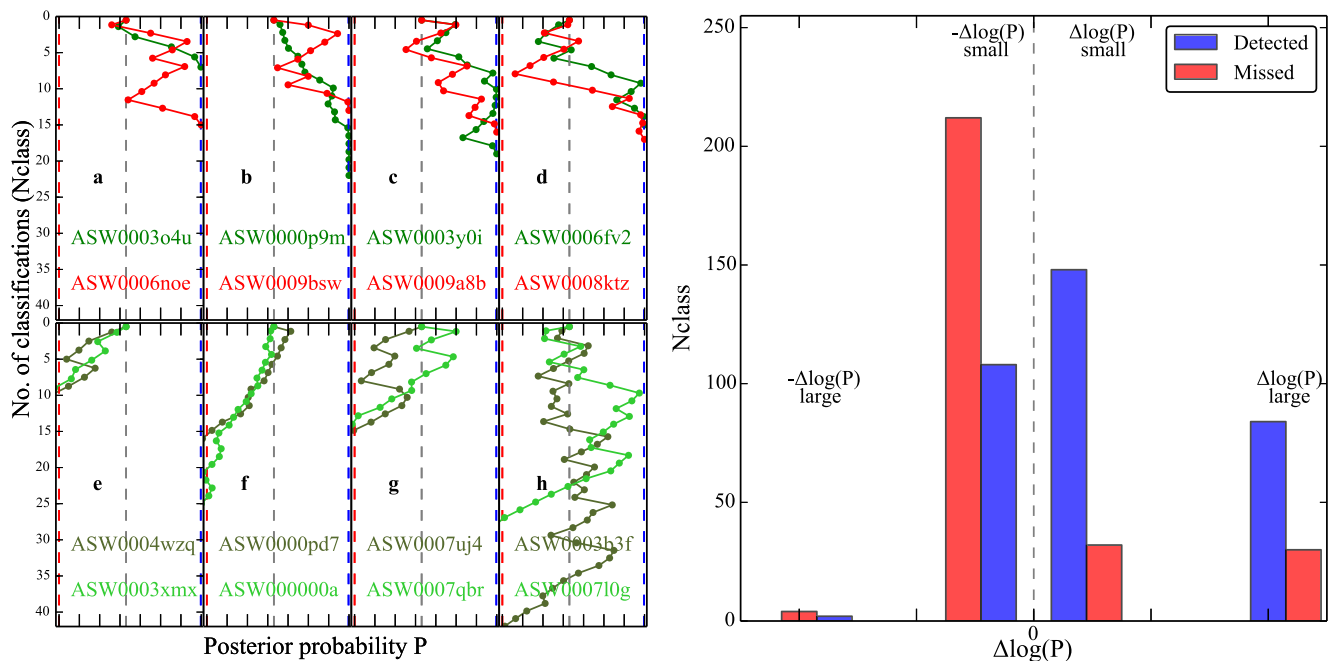
Figure 11. Left: examples of Stage 1 trajectories of some known lenses. Upper panels show detected lenses, lower panels show missed lenses (false negatives). The green trajectories for the detected sample are counterparts of the light and dark green trajectories of the missed sample except that the kicks are positive for the detected sample. The red trajectories of the detected sample demonstrate that sufficient number of positive large kicks can lead to the detection of the lenses. Within the missed sample, the light and dark green correspond to visually easier and more difficult to identify systems, respectively. Right: histogram of classification power received by known lenses. Among the sample of missed lenses, most classifications are from the low-power volunteers identifying these images incorrectly (i.e. $-\Delta \log(P)$) which overshadows the small number of correct identifications (i.e. $\Delta \log(P)$) coming from both the less- and the high-power volunteers combined. For the detected lens sample, most classifications are correct identifications coming from both low- and high-power volunteers.

difficult (dark green) to identify. In spite of some mild qualitative differences, both set of trajectories have very similar behaviour. The trajectories in panel (e) are typical of this sample in terms of Nclass and the dominance of small negative kicks. Panel (f) represents a small fraction of this sample where the kicks are only small and negative. The panel (g) shows how some lenses receive a bunch of large positive kicks which are led to rejection by still mostly small negative kicks. Finally, panel (h) shows those cases of lenses which received almost sufficient number of large positive kicks to be detected but ended up being rejected.

The detected lenses shown with green trajectories, in the upper panels of Fig. 11, can be thought of as counterparts of the trajectories of the missed lenses in the corresponding bottom panels except that their classification power is different. Most detected lenses are similar to the case in panel (a) that are detected within a few classifications coming from large positive kicks. Panel (b) represents a few odd cases which are dominated mainly by small positive kicks. Panel (c) shows a lens getting more classifications, but not reaching the detection or rejection thresholds because of the tug between positive and negative kicks mostly from experienced volunteers. Panel (d) represents two extreme cases when the images are on the verge of being rejected but are saved thanks to a series of large positive kicks. The red trajectories are some more examples of randomly selected cases which demonstrate how having sufficient number of large positive kicks allows lenses to be detected in spite of several small negative kicks.

For a quantitative comparison of the large and small kicks for the entire samples of detected and missed known lenses, we show a plot of histogram on the right of Fig. 11. Qualitatively, there are four types of volunteers making classifications: those causing large,

positive kicks (correct classifications by high-power volunteers), those causing small, positive kicks (correct classifications by low-power volunteers), those causing small, negative kicks (incorrect classifications by low-power volunteers) and those causing large, negative kicks (incorrect classifications by high-power volunteers). The four histograms in the figure correspond to these four types of volunteers for each sample (that is, detected or missed). In this plot, the kick size is defined as small if $\Delta \log P (= \log P_{\text{current}} - \log P_{\text{previous}}) < \Delta \log P_{\text{cut}}$ (chosen as 1.2) and is large if greater than $\Delta \log P_{\text{cut}}$.

Some of the key inferences are as follows. (i) The ratio of positive kicks to negative kicks for the detected sample is higher than in the missed sample, suggesting that the fraction of volunteers making positive kicks is higher for the detected sample. (ii) The number of classifications received by the missed sample is dominated by small negative kicks. In contrast, for the detected sample there are comparable contributions from all three types – small positive kicks, large positive kicks and small negative kicks. (iii) The number of classifications providing large negative kicks is lower for both the detected and the missed samples. This is consistent with our expectation that high-power volunteers should not be making incorrect classifications. We conclude that one of the major factors in the SPACE WARPS system missing the known lenses is a lack of high-power classifications.

As a demonstration, we reran SWAP for Stage 1 using only classifications that produced large kicks, with $|\Delta \log P| > 1.2$. This obviously meant reducing the total number of classifications per subject by a large fraction. As a result, we also needed to change the $P_{\text{accthresh}}$. Choosing this to be 0.1, we found that about a third of the lenses that had previously been missed were now detected, while

all the previously detected lenses were again detected. The remaining missed lenses simply do not have enough classifications from volunteers producing large positive kicks. This experiment shows that it may be possible to increase the SPACE WARPS completeness by preferentially showing certain rejected systems – those that had never received a high-power classification – to volunteers capable of making such classifications.

Changing the rejection and acceptance thresholds will likely decrease the purity along with improved completeness. This will need to be further quantified before detailed recommendations can be made. However, dynamically assigning certain subjects to volunteers according to various measures of their skill seems like a fruitful line of investigation when seeking to improve the system performance in future.

7 SUMMARY AND CONCLUSIONS

We report the discovery of gravitational lens candidates from the first SPACE WARPS lens search. In this search, volunteers were shown $g - r - i$ colour images of random regions of the sky taken by the CFHT Legacy Survey. The aim of this blind lens search was to find lenses that had been missed by previous searches done on the CFHTLS with lens-finding algorithms.

The search was carried out in two stages. In Stage 1, volunteers inspected ∼430 000 images, and selected a smaller sample of ∼3000 images as having interesting lens candidates. In Stage 2, after a careful second inspection of the candidates from Stage 1, a purer sample of ∼500 candidates was obtained. In a final step, these images were inspected by three of us (AM, AV and PJM) to produce a sample of candidates with grades ranging from possibly a lens (1) to almost certainly a lens (3). In this paper, we presented this new SPACE WARPS sample and compared it with the previously known samples from two robotic searches from the CFHTLS, namely, RINGFINDER and ARCFINDER.

Our conclusions are as follows.

(i) SPACE WARPS works well as a discovery engine for gravitational lenses through citizen science. While a targeted visual search may be more efficient, we show that the blind search works reasonably well too.

(ii) We use a sample of simulated lenses, duds and impostors tailored to the CFHTLS data to train the volunteers and calibrate their performance. The volunteers not only perform well on the training sample (see Paper I) but also find lenses that are fainter, more compact or redder than covered by the training sample demonstrating their adaptability in this task.

(iii) We present a sample of 29 new gravitational lens candidates, and an additional 30 medium grade systems. These 59 candidates received averaged grade $G \geq 1.3$ from three experts following the scale where 1 means possibly, 2 means probably, and 3 means almost certainly, a lens. In addition, among the $G \geq 1.3$ sample, we rediscovered 82 lens candidates from various samples published in the literature.

(iv) Compared to the sample of RINGFINDER and ARCFINDER robotically detected lens candidates, the SPACE WARPS sample finds lens systems with statistically similar properties, including the range of lens redshifts, lensed image total magnitudes, and Einstein radii. However, having only displayed images without the lens galaxy light subtracted, the SPACE WARPS sample does not contain many of the RINGFINDER-identified-lensed images with subarcsecond R_E, just because the flux of the typically faint lensed images is obscured by the flux from bright lensing galaxies.

(v) Qualitatively, SPACE WARPS seems to have found lens systems with different types of lensing galaxies, for example, elliptical, spiral (face on and edge on) and small red galaxies unlike those found from robotic searches. Similarly, the lensed images too have diverse properties such as different colours, morphologies and sizes which are again typically missed by any given algorithm.

(vi) Based on the known sample of lenses and lens candidates, we find that we lose a small fraction of them during Stage 2 refinement. It is more important to improve the lens detection sensitivity at the initial Stage 1 classification step. About 35 per cent of the known lenses (20 in total) were missed at Stage 1. Two-thirds of these missed lenses were found to be galaxy-scale RINGFINDER systems, with faint arcs blended with the bright lens galaxies.

(vii) It is possible to improve the SPACE WARPS completeness by changing the strategy of when and who is shown an image: only using high-power classifications recovers 40 per cent of the missed lenses.

The discovery of many new lens candidates through the first SPACE WARPS lens search has demonstrated that the citizen scientists have successfully taught themselves to identify lenses within a short span of time. They have found lens candidates that the algorithms failed to discover. Upcoming and planned wide-field imaging surveys such as the DES, HSC, *Euclid* and LSST will produce formidable amounts of data. Blind lens searches as described here will be impractical with these very large surveys. However, it should be possible to conduct a blind search on a subarea of a large survey in order to assess the performance of either the algorithms or the volunteers, the results of which can be extrapolated to the entire survey. As demonstrated in this paper, any one approach for finding lenses from the entire survey data may not be sufficiently complete and pure. Thus, combining robotic methods for pre-selection with the citizen science approach for visual screening might be a good strategy for finding lenses in these large imaging surveys.

(CNRS) of France, and the University of Hawaii. This work is based in part on data products produced at TERAPIX available at the Canadian Astronomy Data Centre as part of the CFHTLS, a collaborative project of NRC and CNRS.

REFERENCES

Alard C., 2006, preprint (astro-ph/0606757)
Barnabè M., Czoske O., Koopmans L. V. E., Treu T., Bolton A. S., Gavazzi R., 2009, MNRAS, 399, 21
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Benjamin J. et al., 2007, MNRAS, 381, 702
Bernardi M. et al., 2003, AJ, 125, 1866
Blanton M. R. et al., 2001, AJ, 121, 2358
Brault F., Gavazzi R., 2015, A&A, 577, A85
Cabanac R. A. et al., 2007, A&A, 461, 813
Cardamone C. et al., 2009, MNRAS, 399, 1191
Chan J. H. H., Suyu S. H., Chiueh T., More A., Marshall P. J., Coupon J., Oguri M., Price P., 2015, ApJ, 807, 138
Collett T. E., Auger M. W., 2014, MNRAS, 443, 969
Collett T. E., Auger M. W., Belokurov V., Marshall P. J., Hall A. C., 2012, MNRAS, 424, 2864
Coupon J. et al., 2009, A&A, 500, 981
Elyiv A., Melnyk O., Finet F., Pospieszalska-Surdej A., Chiappetti L., Pierre M., Sadibekova T., Surdej J., 2013, MNRAS, 434, 3305
Erben T. et al., 2009, A&A, 493, 1197
Erben T. et al., 2013, MNRAS, 433, 2545
Faber S. M. et al., 2007, ApJ, 665, 265
Faure C. et al., 2009, ApJ, 695, 1233
Gavazzi R., Marshall P. J., Treu T., Sonnenfeld A., 2014, ApJ, 785, 144
Gnedin O. Y., Kravtsov A. V., Klypin A. A., Nagai D., 2004, ApJ, 616, 16
Gwyn S. D. J., 2012, AJ, 143, 38
Hildebrandt H. et al., 2012, MNRAS, 421, 2355
Jaskot A. E., Oey M. S., 2013, ApJ, 766, 91
Keel W. C. et al., 2012, AJ, 144, 66
Keeton C. R., Mao S., Witt H. J., 2000a, ApJ, 537, 697
Keeton C. R., Christlein D., Zabludoff A. I., 2000b, ApJ, 545, 129
Koopmans L. V. E., Treu T., Bolton A. S., Burles S., Moustakas L. A., 2006, ApJ, 649, 599
Kormann R., Schneider P., Bartelmann M., 1994, A&A, 284, 285
Küng R. et al., 2015, MNRAS, 447, 2170
Leier D., Ferreras I., Saha P., Falco E. E., 2011, ApJ, 740, 97
Lenzen F., Schindler S., Scherzer O., 2004, A&A, 416, 391
Limousin M. et al., 2008, A&A, 489, 23
Lintott C. J. et al., 2008, MNRAS, 389, 1179
Lintott C. J. et al., 2009, MNRAS, 399, 129
Lupton R., Blanton M. R., Fekete G., Hogg D. W., O'Mullane W., Szalay A., Wherry N., 2004, PASP, 116, 133
Marshall P. et al., 2015, preprint (arXiv:1504.06148) (Paper I)
Maturi M., Mizera S., Seidel G., 2014, A&A, 567, A111
More A., Cabanac R., More S., Alard C., Limousin M., Kneib J.-P., Gavazzi R., Motta V., 2012, ApJ, 749, 38
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Newman A. B., Treu T., Ellis R. S., Sand D. J., 2013, ApJ, 765, 25
Oguri M., 2006, MNRAS, 367, 1241
Oguri M., Marshall P. J., 2010, MNRAS, 405, 2579
Oguri M., Keeton C. R., Dalal N., 2005, MNRAS, 364, 1451
Oguri M., Bayliss M. B., Dahle H., Sharon K., Gladders M. D., Natarajan P., Hennawi J. F., Koester B. P., 2012, MNRAS, 420, 3213
Pâris I. et al., 2012, A&A, 548, A66
Parker L. C., Hudson M. J., Carlberg R. G., Hoekstra H., 2005, ApJ, 634, 806
Richards G. T. et al., 2006, AJ, 131, 2766
Schneider P., Ehlers J., Falco E. E., 1992, Gravitational Lenses. Springer-Verlag, Berlin, Heidelberg, New York
Seidel G., Bartelmann M., 2007, A&A, 472, 341
Sereno M., Paraficz D., 2014, MNRAS, 437, 600
Sonnenfeld A., Gavazzi R., Suyu S. H., Treu T., Marshall P. J., 2013a, ApJ, 777, 97
Sonnenfeld A., Treu T., Gavazzi R., Suyu S. H., Marshall P. J., Auger M. W., Nipoti C., 2013b, ApJ, 777, 98
Sonnenfeld A., Treu T., Marshall P. J., Suyu S. H., Gavazzi R., Auger M. W., Nipoti C., 2015, ApJ, 800, 94
Suyu S. H., Halkola A., 2010, A&A, 524, A94
Sygnet J. F., Tu H., Fort B., Gavazzi R., 2010, A&A, 517, A25
Thanjavur K. G., 2009, PhD thesis, Univ. Victoria, Canada
Whitaker K. E., Rigby J. R., Brammer G. B., Gladders M. D., Sharon K., Teng S. H., Wuyts E., 2014, ApJ, 790, 143
Zheng W. et al., 2012, Nature, 489, 406
Zitrin A., Broadhurst T., 2009, ApJ, 703, L132
Zitrin A., Broadhurst T., Barkana R., Rephaeli Y., Benítez N., 2011, MNRAS, 410, 1939

APPENDIX A: LENS DETECTION POWER

In Paper I, we defined the 'Skill' of an agent as being given by the expectation value of the information gain per classification. This quantity is a non-linear function of both the P_L, the probability of correctly identifying a lens as a lens and P_D, the probability of correctly identifying a dud as a dud. This means that one can get the same value of Skill for different combinations of P_L and P_D (see the
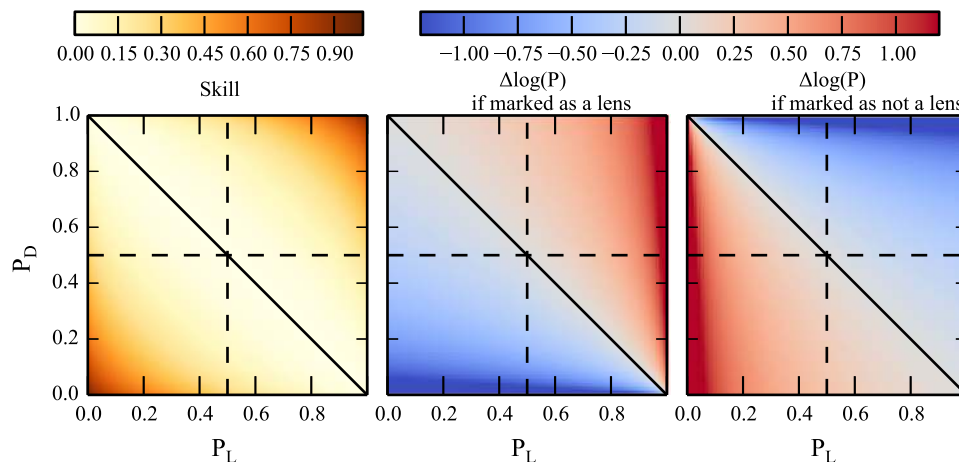
Figure A1. Skill of the volunteers and $\Delta \log P$ given a lens or not a lens in an image as a function of P_L and P_D. These quantities indicate the ability of the volunteers but do not have a simple linear relation with P_L and P_D.

left-hand panel of Fig. A1). The skill reflects the all-round ability of a classifier to contribute information.

As described in Paper I, the posterior probability P of a subject is determined by the P_L and P_D of all the volunteers who clicked on the subject, via Bayes' Theorem. Each agent will apply a 'kick' of a different size to the subject probability, $\Delta \log P$, which can be either positive (if the classifier thinks the subject contains a lens) or negative (if the classifier thinks the subject does not contain a lens). For instance, given a subject containing a lens, a volunteer with high P_L implies a large positive kick irrespective of the value of P_D, as shown in the middle panel of Fig. A1. However, large positive kicks are still possible for a volunteer located in the upper triangle with different combinations of (P_L, P_D) suggesting that the kick is not a simple function of (P_L, P_D).

The kicks appear as steps on the subject's trajectory plot. This kick magnitude gives a useful measure of an agent's 'Power' to move images closer to detection. Note that a volunteer who is very good at rejecting duds, but not so good at identifying lenses, may have a high Skill but a low Power (since they may fail to detect many of the interesting lenses): Power provides a more precise quantification of a classifier's ability to detect lenses (compared to rejecting non-lenses).

This paper has been typeset from a TEX/LATEX file prepared by the author.