Mon. Not. R. Astron. Soc. **000**, 1–22 (2013) Printed 10 April 2015 (MN LATEX style file v2.2)

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

Anupreeta More,[1⋆] Aprajita Verma,[2] Phil 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] Chris Davis,[3] Raphael Gavazzi,[6] Chris Lintott,[2] Robert Simpson,[2] David Miller,[4] Arfon Smith,[4] Edward Paget,[4] Prasenjit Saha,[7] Rafael Kueng,[7] Tom Collett,[8] Matthias Tecza[2]

[1] *Kavli IPMU (WPI), University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa 277-8583, Japan*
[2] *Dept. 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, USA*
[6] *Institut dAstrophysique de Paris, UMR7095 CNRS Université Pierre et Marie Curie, 98bis bd Arago, 75014 Paris, France*
[7] *Department of Physics, University of Zurich, Switzerland*
[8] *Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth P01 3FX, UK*

to be submitted to MNRAS

ABSTRACT

We report the discovery of 28 promising and a total of 58 new lens candidates from the CFHT 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 lenses missed by lens finding robots (the RINGFINDER on galaxy scales and ARCFINDER on group/cluster scales), which have been previously used to mine the CFHTLS for lenses. We compare some properties of lens samples detected by these algorithms to the SPACE WARPS sample and found that they are broadly similar. The image separation distribution calculated from the SPACE WARPS discovered sample shows that our previous constraints on the average density profile of the lens population are robust. SPACE WARPS recovers about 60% of the known sample and the new candidates show a richer variety compared to the lenses found by the two robots. We find that analyzing only those classifications which are performed by the high power volunteers, SPACE WARPS can achieve a detection rate of up to 75% of the known lens sample. Training and calibration of the performance of citizen scientists is crucial for the success of SPACE WARPS. We also present the SIMCT pipeline, used for generating a sample of realistic simulated lensed images in the CFHTLS, and a sample of duds and false positives used in the training. Such a training sample has a legacy value for testing future lens finding algorithms. We make our training sample publicly available.

Key words: gravitational lensing – methods: statistical – methods: citizen science

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

⋆ anupreeta.more@ipmu.jp

matter distribution from galaxy (e.g. Koopmans et al. 2006; Barnabè et al. 2009; Sonnenfeld et al. 2015) to groups-cluster scales (e.g. Limousin et al. 2008; Zitrin et al. 2011; Oguri et al. 2012; More 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 dark energy (e.g. Suyu & Halkola 2010; Collett et al. 2012; Collett & Auger 2014; Sereno & Paraficz 2014) and many more. Strong lenses are rare since 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 till date[1].

The search for gravitational lenses is a needle-in-a-haystack problem. Efficient automated methods are imperative to find a reasonably complete and pure sample of strong lenses. Several lens finding algorithms have been developed so far (e.g. Lenzen et al. 2004; Alard 2006; Seidel & Bartelmann 2007; More et al. 2012; Gavazzi et al. 2014) but they can not 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, colors, 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 artifacts 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 the problems of high rate of false positives. To minimize 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 find a sample of plausible lens candidates.

Recognising patterns amidst an array of noisy pixelated images is one of the strengths of human brains. Humans are also capable of dealing with multi-tiered complex web of questions before arriving at a conclusion 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. Therefore, participation of a large community of volunteers to help with the visual identification of lenses is extremely beneficial for the lensing community. In particular, as we enter the era of large area imaging surveys (spanning thousands of square degrees), it is a perfect time to tap into the potential of citizen science.

Galaxy Zoo was the first citizen science project in astronomy, and addressed the problem of how to classify large numbers of galaxies by their morphology (Lintott et al. 2008). From the early results based on morphological classifications of galaxies, 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 harness the potential of citizen scientists. Since then 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 image configurations. With these significant challenges at hand, we designed the Space Warps project to enable the discovery of lenses through citizen science. In a companion paper, Paper I (Marshall et al., in prep.), we describe the design of Space Warps and how the entire system functions as a discovery service.

In this paper (Paper II hereafter), we describe our first lens search using data from the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS). This paper is organised 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 false positives, 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 4. In Section 5, we briefly describe how the classifications of images from the volunteers are turned into a catalog of plausible candidates (for further details, see Paper I). In Section 6, we present the new lens candidates from Space Warps and compare it to the lens samples produced by a few robotic searches from the CFHTLS in the past. Next, we discuss what kind of lenses are detected or missed by the algorithms and Space Warps in Section 7. The summary and conclusions are given in Section 8.

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″(Gwyn 2012). The CFHTLS WIDE covers a total non-overlapping area of 150 deg^2 on the sky and consists of four fields W1, W2, W3 and W4. The field W1 has the largest sky coverage of 63.65 deg^2. The fields W2 and W4 have similar sky coverages of 20.32 deg^2 and 20.02 deg^2, respectively[2]. The field W3 has a sky coverage of 42.87 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 thus, 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″. 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

[1] http://masterlens.astro.utah.edu

[2] These numbers are estimated from http://terapix.iap.fr/cplt/table_syn_T0006.html

lenses. We use the stacked images from the final T0007 release taken from the Terapix website[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 (KiDS) 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.

In the thesis dissertation of **?**, 9 promising and 2 low probability candidates are reported which were discovered serendipitiously. 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 galaxy lenses in the CFHTLS WIDE. They identified galaxies, using Sextractor, which had $18 < i < 21$ and inclination angle < 25 deg. 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. The algorithm 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 bonafide lens sample from the CFHTLS WIDE. Also, the total number of lens candidates is 108 from the CFHTLS WIDE alone.

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 W1 field. The goal was to find instances of lensed quasars. Their sample consists of a total of 18 candidates out of which 3 candidates are found to be promising.

The RingFinder (Gavazzi et al. 2014) was used for finding compact rings or arcs around centers of isolated and massive early-type galaxies. The RingFinder subtracts the 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 centers and blue outer parts, from the T0006 CFHTLS data release catalogs. A total of 14370 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 are good quality (`q_flag` = 3) and 288 are 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 be detected by earlier version 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.

The work by Maturi et al. (2014) used the arc finding code of Seidel & Bartelmann (2007) and color properties of typical arcs to optimize arc detection. This new approach was tested on the CFHTLS-Archive-Research Survey (CARS, Erben et al. 2009) which covers an area of 37 sq. deg only and this entire data was also visually inspected by the authors to estimate 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 are 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 has been run on the entire CFHTLS imaging. Thus, these can be considered as 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.

[3] http://terapix.iap.fr/cplt/T0006-doc.pdf

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

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

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 cutouts of images selected from the survey region. We briefly describe the image presentation here for completeness but more information can be found in Paper I. We use the g, r and i-band imaging from CFHTLS which are most useful for visual identification of lenses. We made color composite images using the publicly-available code, HumVI[6] following the prescription of Lupton et al. (2004). The color scales are chosen to maximize the contrast between faint extended objects. These parameters are 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 extract contiguous cutouts of size $81.84''$ (440 pixels) with an overlapping region of $10''$(54 pixels) between the neighbouring cutouts. This results in a catalog of $\sim$ 430 000 cutouts for the entire CFHTLS WIDE region. The size of the individual cutout is determined by optimising factors such as the typical angular scales of gravitational lenses, the number of objects seen in a single cutout and the total number of image cutouts in the survey. If a lens candidate happens to be too close to the edge of a cutout, then the overlap between neighbouring cutout allows a volunteer to get a clearer view of the same candidate in at least one of the cutouts. We note that since the images are shown randomly, a volunteer may not necessarily come across the neighbouring cutout unless the volunteer classifies 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 cutout.

3 TRAINING SAMPLE: SIMULATED LENSES

The simulated lenses are important to train citizen scientists who are novice to the task of finding lenses but they are also crucial for analyzing the classifications performed by the citizen scientists (more details can be found in Paper I but see Section 5 for a brief summary). In this section, we describe the framework used for generating the simulated lens sample and give details of the sample itself along with some known limitations.

3.1 Methodology

For the purpose of generating simulated lens systems, we divide them in to 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.

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

3.1.1 Galaxy-scale lenses

We begin by considering all elliptical galaxies at $z < 1$ in our parent CFHTLS catalog 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 (XXX check).

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

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

where

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

Here, $\Phi(L', z_s)$ denotes the source luminosity function per unit comoving volume, $\sigma_{\rm lens}$ 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 LRG catalog. We use the $L - \sigma$ scaling relation from the bright sample of (Parker et al. 2005) to set the velocity dispersion of the halo hosting the galaxy which will be later used in the model

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

We assume that the knee of the luminosity function of galaxies, L_* evolves such that there is a decline of 1.5 magnitudes between $z = 1$ to $z = 0$ (Faber et al. 2007).[7]

We assume a singular isothermal ellipsoid model for each of our galaxies (Kormann et al. 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 demarcate the regions of different image multiplicities. Parametric solutions, $r(\theta)$ (where θ is the polar angle), for the caustics in such a model were given by Keeton et al. (2000b). We use these parametric solutions to obtain the area of the lensing cross-section, σ, for every galaxy,

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

[7] 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.

At every polar angle we use the maximum of the radial and tangential caustic for the purpose of calculating the cross-section.

We use the results of Faure et al. (2009) to specify the source luminosity function of galaxies. We assume that the redshift distribution of sources is given by

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

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

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

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

$$\frac{{\rm d}\Phi}{{\rm d}M} = \frac{\Phi_*}{10^{0.4(\alpha+1)(M_{\rm abs}-M_*)} + 10^{0.4(\beta+1)(M_{\rm abs}-M_*)}} \quad (9)$$

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

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

We adopt the best-fit 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). respectively.

With the cross-section, and the luminosity functions specified, we calculate the expected number of sources behind a candidate 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 use it as a training sample as well as for calibrating the performance of our citizen scientists (see Paper I). 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 with a mean equal to the boosted average number of sources. If the Poisson deviate is greater than zero, then we mark this galaxy to act as a lens galaxy. We then draw the properties of the lensed source based on the redshift and the luminosity distribution of the sources. The source positions with respect to the lens are drawn randomly from an area inside the caustic.

Next, we determine properties of the background source for every lens. We follow similar procedures for both background galaxies and quasars. We perform ray-tracing for all of the $N_{\rm src}$ sources using the publicly available code GRAVLENS (Keeton et al. 2000a) and choose sources that satisfy our selection criteria. We determine fluxes of the lensed images and the total magnification of each of the lensed source. We draw a random source for which the flux of the second brightest lensed image and the total magnification of all lensed images are above 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 colors of the background source galaxies are drawn randomly from the photometric CFHTLenS catalog (Hildebrandt et al. 2012; Erben et al. 2013). Similarly, we use a quasar catalog from the SDSS Data Release 9 (Pâris et al. 2012) from which colors are drawn to simulate quasar lenses. Next, we assume deVaucoleur's profile to account for the size and shape of the galaxies. The ellipticity and the position angle (PA) are drawn randomly between 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 some size evolution) given by

$$R_{\rm eff} = 10^{0.52} \frac{L_r^{2/3}}{(1+z_s)^2} \quad (11)$$

where $L_r = L_s/10^{10.2}$. On the other hand, quasars are assumed to follow a Gaussian profile where the σ is equated to that of the median seeing for every filter. The median seeing values are taken from Table 4 of the official Terapix T0007 release explanatory document [8].

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 median seeing in each of the filters, the simulated image is added to the real CFHTLS image centered on the lensing galaxy. Note that we ensure that the lensed galaxies and lensed quasars do not have the same lensing galaxy in the foreground. Similarly, the lensing galaxies from the galaxy-scale

[8] http://terapix.iap.fr/cplt/T0007/doc/T0007-doc.pdf

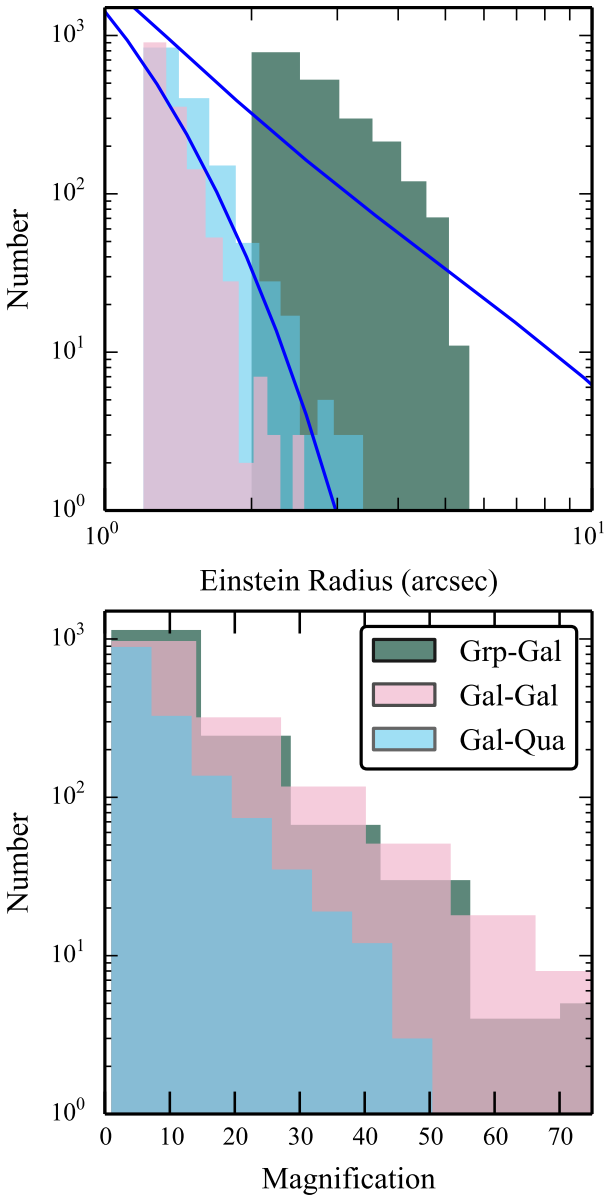


Figure 2. Einstein radius distribution for all types of lenses. The dashed-dotted (blue) curves show the theoretical prediction assuming an SIS model at galaxy-scales and a total (NFW+Hernquist) model at groups-scales taken from (More et al. 2012).

lenses are distinct from the central galaxies of groups-scale lenses which are described in the following section.

3.1.2 Groups-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 center, but also has contribution from the dark matter component as well as the satellite galaxies (Oguri et al. 2005; Oguri 2006). We generate a basic group catalog based on

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

Name	Gal-Gal (Grp-Gal)		Gal-Qua	
	min	max	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	

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

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 with the aperture, then the fainter BGG is removed from our list of BGGs.

We assume a constant mass-to-light ratio of 3×0.7 h M_*/L_*, to convert the BGG luminosity to a stellar mass estimate. The stellar mass−halo mass relation (Behroozi et al. 2013), including random scatter, is then used to calculate the halo mass for the lens. We adopt an NFW 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 isothermal ellipsoid model for the BGG and members whenever the ellipticities are available from the galaxy catalog (else we use an isothermal sphere).

We calculate the luminosity and velocity dispersion 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[9]. 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 not expected to be extremely rare in the CFHTLS (check XXX). 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 Eq. 1). As before, for each

[9] 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 XXX.

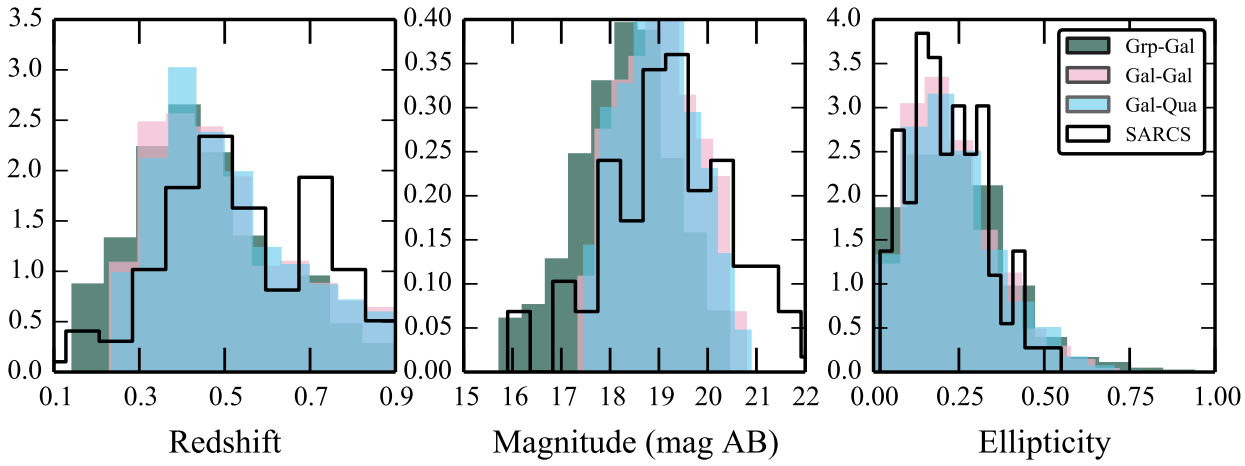


Figure 3. Distributions of properties of the lensing galaxies of the simulated sample compared to the known lens sample SARCS.

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 Eqs. 7-8).

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 (check XXX) the statistical properties such as the profile of the image separation distribution are not affected (see Figure 2). We follow the same procedure and apply the same thresholds to determine properties of the lensed galaxies for every lens as are described for galaxy-galaxy lenses in the previous section. The simulated lensed images are then added on top of the real CFHTLS images with the BGGs as the center.

3.2 Simulated Lens Sample and Catalog Description

In this section, we describe some of the properties of our simulated sample for each of the three types of lens samples.

Figure 2 shows the Einstein radius distribution for the galaxy-scale (dashed for background quasars and dotted for background galaxies) and groups-scale simulated lenses. For comparison, we show the expected distributions (blue dashed-dotted curves) for an SIS-like density profile at galaxy-scales and an NFW+Hernquist profile at groups-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 groups-scale also includes SIS or SIE components for the group members unlike the theoretical prediction. The theoretical curves have arbitrary normalizations.

We show the redshift, magnitude and ellipticity distributions of the lensing galaxies in the three panels of Figure 3, respectively. We also show the distributions of respective properties of the SARCS lenses from (More et al. 2012) for comparison with arbitrary normalizations. We note the qualitative similarities between the simulated and the real lens samples.

We produce catalogs with lens and source properties for each of the three types of lenses. These catalogs are available from https://github.com/anupreeta27/SIMCT. The catalogs 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 center, redshift, magnitudes, total magnification, number of lensed images. Additionally, when possible, ellipticity and effective radius of the background galaxies have also been provided.

3.3 Limitations of the simulated 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 parameters for the models. Comments from citizen scientists were very helpful in order to sort some of these failures. Here, we describe some of the cases or aspects in which the simulations were known to have failed (or appeared unrealistic). This typically happens for a small fraction of the simulated images (roughly 5 percent of the cases)[10].

The parameters required by various scaling relations and the models primarily depend on the photometry of the galaxies, groups and quasars detected in the survey. For galaxy-scale lenses, the lensing galaxies at higher redshifts or which are fainter have poor photometric redshift measurements. Consequently these galaxies are assigned a wrong luminosity and velocity dispersion estimates, which sometimes result in simulated lenses which look implausible. For example, the lensed images for some of the failed simulations have larger image separation than expected given the visual priors from the galaxy.

At group-scales, the photometric and redshift estimates are used when define the group membership. Therefore, errors in redshift estimates generate galaxy groups with BGG or member galaxies with dissimilar properties. In some cases, low redshift spiral galaxies have been incorrectly assigned high redshift. Spiral galaxies are typically less massive and

[10] Based on the hashtag simfail in the discussion forum TALK.

low redshift spiral galaxies are unlikely to act as gravitational lenses. Hence, some of the simulated lenses were not as convincing.

We also use a single Sersic component to describe the light profiles of background galaxies. This is clearly not the most accurate description for galaxies, especially, star-forming galaxies which form a significant fraction of the lensed galaxy population. Star-forming galaxies have complex structures such as star forming knots, spiral arms, bars and disks. The simulated lensed images do not display these features.

4 TRAINING SAMPLE: DUDS AND FALSE POSITIVES

Citizen scientists not only need training to identify gravitational lenses, but also to reject images which either contain impostors or have no lenses. Hence, in addition to the simulated lenses, we added a sample of duds and false positives to the training sample. Duds are images which have been visually inspected by experts and confirmed to contain no lenses. False positives (FPs) are systems which look like lenses but are not, for example, spiral galaxies, starforming galaxies, chance alignments of features arranged in a lensing configuration and stars.

We selected a sample of 450 duds for the Stage I classification in Space Warps and a sample of 500 false positives for the Stage II inspection. The sample of false positives was selected from the candidates which passed the Stage I of Space Warps. We note that this is the first time, we have a systematically compiled sample of visually inspected false positives by the Space Warps volunteers and categorized by the science team. The data products are available at http://spacewarps.org/#/projects/CFHTLS/. Such a sample is tremendously helpful for training and understanding performances of various lens finding algorithms (Chan et al. 2014).

5 METHODOLOGY TO PRODUCE THE Space Warps-CFHTLS LENS SAMPLE

The Space Warps works as a single unified neural network which uses the method of visual inspection to find gravitational lenses. The volunteers are shown images at two stages. At Stage I, volunteers are requested to undertake a rapid inspection to select lens candiates ranging from possible lenses to almost certain lenses. At Stage II, volunteers carefully inspect the candidates from Stage I and select only promising lens candidates. A daily snapshot of classifications performed by volunteers are provided to the science team every night. This daily batch is analyzed 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 (or image cutout) is assigned a prior probability of 2×10^{-4} to contain a lens system. Every volunteer or classifier is assigned an agent characterised by a 2×2 confusion matrix, which quantifies the volunteer's ability to correctly classify an image to contain (P_L) or not contain a lens system (P_D). The values of the confusion matrix

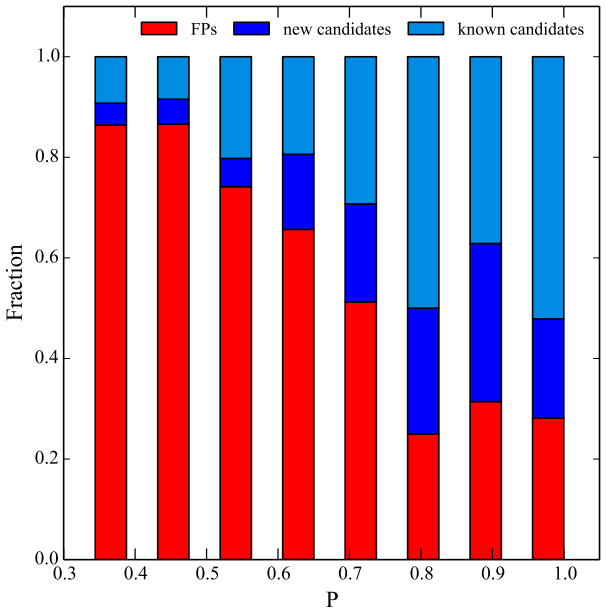


Figure 4. Distribution of different types of candidates as a function of the posterior probability P. The types of the candidates are the FPs, the new candidates and the known candidates. The new and known candidates have higher detection rate for higher values of P, as expected.

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 duds) correctly classified. After every classification, the agent updates the prior probability of the classified subject based on the classification and the volunteer's confusion matrix, according to the Bayes theorem. Note that the volunteer's confusion matrix is updated after the classification of every training image. Threshold values for the probabilities to accept (or reject) a subject if it contains (or doesn not contain) a lens can be chosen in SWAP. Those images which cross these threshold values are retired and are not subsequently shown to the volunteers, so that volunteers can use their time efficiently to classify previously unclassified subjects.

SWAP was run nightly during Stage I in order to retire subjects and inject new ones in to the classification stream. The subjects that passed the acceptance threshold at the end of Stage I are served again at Stage II for careful reinspection. Each subject, after Stage II, has an associated 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 and thus, decreasing the completeness (or purity) of the final lens sample. As this is the first candidate 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.

We decided to inspect the top few hundreds of images from a catalog of descending P values as a reasonable sample size for visual inspection. We find a total of 665 subjects with P$>$ 0.3 in the catalog. These images were then visually

inspected by three lens experts (AM, AV, PJM). The images are assigned grades on a scale of 0 to 3, thus dividing the subjects into candidates that almost certainly do not contain a lens, possibly a lens, probably a lens and almost certainly a lens, respectively. The final sample of SPACE WARPS-CFHTLS lens candidates (see Section 6.1) is produced by selecting candidates above some threshold on the averaged grades.

6 RESULTS

6.1 SPACE WARPS-CFHTLS Lens Sample

In this section, we describe the SPACE WARPS-CFHTLS lens sample (henceforth, called the SPACE WARPS lens sample). We find a total of XXX (and XXX new) candidates with averaged grade 0 <G< 1.3 (low probability sample) which means at least one of the inspector gave a grade of 1 and the maximum grade given by all three inspectors was 1[11]. Further information on the low probability sample such as their positions and images are available at http://spacewarps.org/#/projects/CFHTLS/.

There are a total of XXX candidates with 1.3 ⩽G< 2 (medium probability sample) out of which 30 are new. These grades suggest that, most of the times, at least one of the inspectors gave a grade of 2 and the maximum grade by all of the inspectors was below 2. Among our high probability sample (G⩾2), there are a total of XXX candidates out of which 28 are new which means at least two of the inspectors gave a grade of 2 or higher. To avoid duplication, further information on SPACE WARPS candidates that were previously identified in the literature is available at http://spacewarps.org/#/projects/CFHTLS/ whereas the new lens candidate sample of medium-high probability is presented in this paper (see Section 6.2).

In Table 2, we give overall statistics of the systems detected at Stage I and Stage II. We give the total number of detections of the known lens candidates, known lenses and the new lens candidates at each stage. We also show their fractions with respect to the total number of lens candidates found at these stages. We note that most (XXX check) of the confirmed lenses found at Stage I are also recovered at Stage II whereas a small fraction of the known lens candidates are rejected at Stage II. Also, the sample of new lens candidates increases the sample of CFHTLS lens candidates by over 50%.

In Figure 4, we plot the distribution of false positives and the high probability lens candidates, categorized by the lens experts, as a function of the P value assigned by SWAP at the end of Stage II. 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 from P∼0.75 onwards to lower P values, 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. However, the choice of P=0.3 as the threshold gives a more acceptable completeness of XXX for our sample which combined with expert grading allows us to achieve further purity of our sample.

[11] 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 I		Stage II	
	KC	KL	KC	KL
Number	128	34	107	34
Percentage of recovery	29	58	25	58
P_{thresh}	0.95	0.95	0.3	0.3

	Stage II	
	NC	AC
Number	58	168
Percentage of detection	12	34
P_{thresh}	0.3	0.3

KC– Known lens candidates
KL– Known lenses
NC– New lens candidates
AC– All (known and new) lens candidates
P_{thresh} – systems with P above this threshold are selected
Note: For KC and KL, Percentages are with respect to the known population whereas for NC and AC, percentages are with respect to the total population of lens candidates.

6.2 New lens candidates from SPACE WARPS

We give basic information about the final sample of 58 new lens candidates (medium-high probability) 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 II and a visual categorization of 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 catalog (Coupon et al. 2009) otherwise the reported lens galaxy positions are 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 Figure 5 which is arranged first in the decreasing order of their grades and then increasing order of their positions. As the first lens search was a blind search with no preselection 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 groups-scale lens candidates. There are detections of elongated arcs and some interesting point-like quasar lensed images. Most of them are brighter in bluer g band but some candidates brighter in the redder i band are also found. Since robotic lens searches often look for blue colored lensed features, they are very much 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 interesting candidates

lurking in TALK but these will have to be sifted through (**is there someone actively looking into this whose work should be referred here ?**).

The new Space Warps lens sample presented here shows 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 do amazing amount of interpolation or extrapolation over 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 typically. Human brains have no problem in interpolating over the broken blue arc over the red galaxy and understand that it is a single long arc. The system SW19 has point-like lensed images which cannot be detected by arc finding algorithm whereas the ring finding algorithm may miss this because of atypical color and structure of the galaxy.

The power of citizen scientists lies in the high dynamic range that allows us to detect systems with very from small-to-large 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 comparatively low. 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, Paper III.

6.3 Measurements of properties of the lens and the lensed images

In the subsequent sections, we make comparison of various properties of the lens candidates. In this section, we describe how we extract these properties, namely, the lens redshift, the Einstein radii and the total flux of the lensed images or arcs.

We use the publicly available redshifts for the lens galaxy from the CFHTLS photometric catalogs (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 between g and i bands where the lensing galaxy is subtracted (Gavazzi et al. 2014, for details, see). 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 colors 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 ($R_{\rm E}$) is also measured differently for different samples. For the galaxy-scale lenses in the simulated sample, we use the value of the input lens model parameter for the $R_{\rm E}$. For groups-scale lenses, since the lens model is multi-component, we need to determine the $R_{\rm E}$ from the image positions. We use those pairs of lensed images that have the smallest and the largest angular separations. The $R_{\rm E}$ 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 center as an estimate of the $R_{\rm E}$. 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.

6.4 Recovery of known lens samples from the CFHTLS by Space Warps

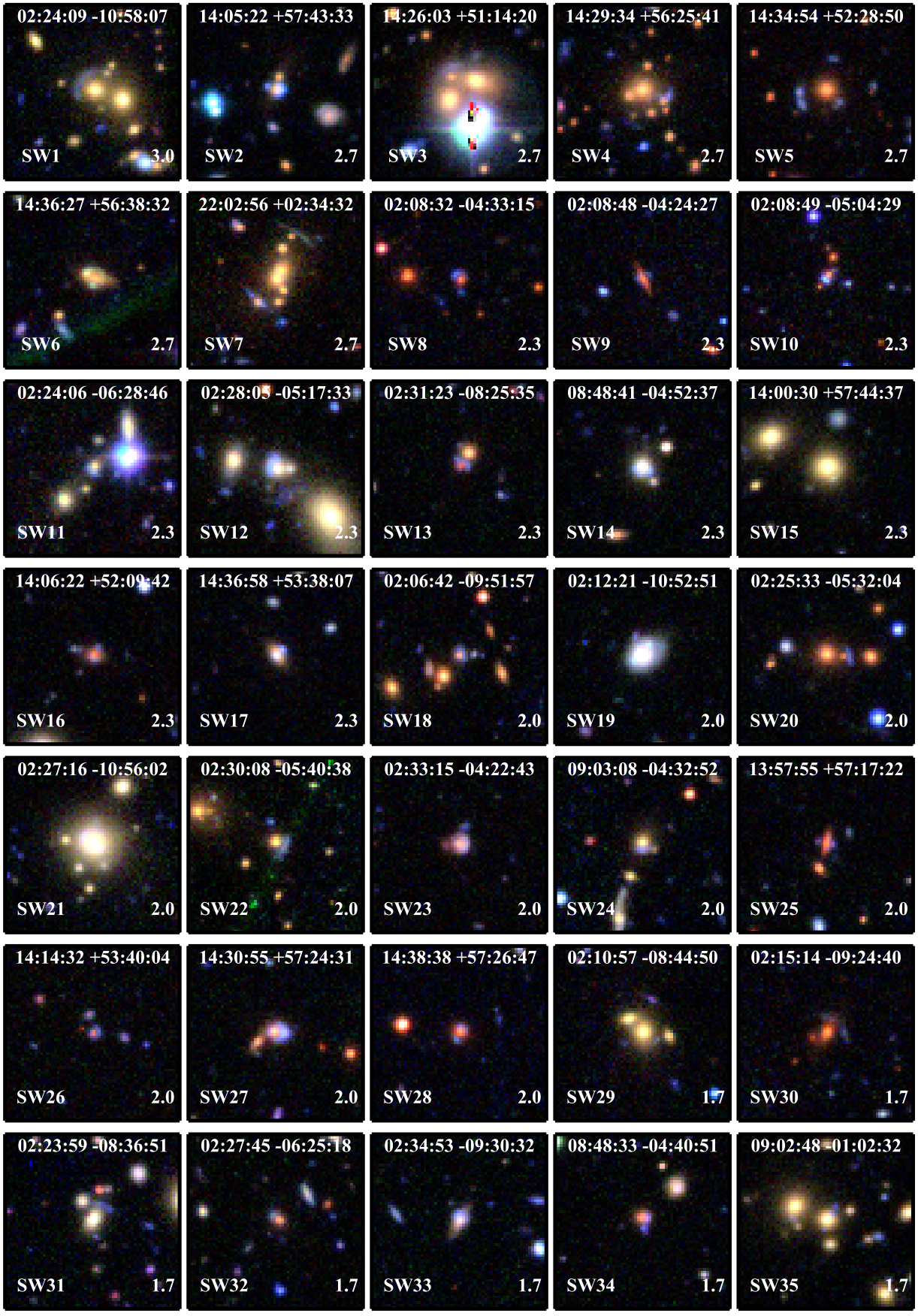
We determine what fraction of the known sample of lenses are recovered by Space Warps. Note that this sample corresponds to the RingFinder and ArcFinder samples combined, as mentioned in Section 2.2. In Table 2, we show that $\sim$ 25%-30% of the known lens candidates and $\sim$ 60% of the known lenses are found at Stage I and Stage II in Space Warps. The left and the middle panels of Figure 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 Einstein radii 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 with fainter arcs and smaller Einstein radii and they come from the RingFinder sample. The main reason why RingFinder found such candidates is because their team used lensing galaxy-subtracted images to detect the presence of the lensed images both during the automated object-finding phase and during the visual inspection and classification of their candidates. This approach naturally improves the detection efficiency at smaller Einstein radii and for fainter systems. The Space Warps volunteers were not shown any galaxy-subtracted images. However, showing galaxy subtracted images might be a better strategy to adopt for future lens searches at galaxy-scales with Space Warps. In the Discussion Section 7.3, we further explore and discuss why some of the confirmed lenses are missed by Space Warps.

6.5 Image separation distribution

The distribution of image separations (i.e. twice the Einstein radius) 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 find several new lens candidates that were not known before. In Figure 7, we show the image


02:24:09 -10:58:07
SW1 3.0
14:05:22 +57:43:33
SW2 2.7
14:26:03 +51:14:20
SW3 2.7
14:29:34 +56:25:41
SW4 2.7
14:34:54 +52:28:50
SW5 2.7
14:36:27 +56:38:32
SW6 2.7
22:02:56 +02:34:32
SW7 2.7
02:08:32 -04:33:15
SW8 2.3
02:08:48 -04:24:27
SW9 2.3
02:08:49 -05:04:29
SW10 2.3
02:24:06 -06:28:46
SW11 2.3
02:28:05 -05:17:33
SW12 2.3
02:31:23 -08:25:35
SW13 2.3
08:48:41 -04:52:37
SW14 2.3
14:00:30 +57:44:37
SW15 2.3
14:06:22 +52:09:42
SW16 2.3
14:36:58 +53:38:07
SW17 2.3
02:06:42 -09:51:57
SW18 2.0
02:12:21 -10:52:51
SW19 2.0
02:25:33 -05:32:04
SW20 2.0
02:27:16 -10:56:02
SW21 2.0
02:30:08 -05:40:38
SW22 2.0
02:33:15 -04:22:43
SW23 2.0
09:03:08 -04:32:52
SW24 2.0
13:57:55 +57:17:22
SW25 2.0
14:14:32 +53:40:04
SW26 2.0
14:30:55 +57:24:31
SW27 2.0
14:38:38 +57:26:47
SW28 2.0
02:10:57 -08:44:50
SW29 1.7
02:15:14 -09:24:40
SW30 1.7
02:23:59 -08:36:51
SW31 1.7
02:27:45 -06:25:18
SW32 1.7
02:34:53 -09:30:32
SW33 1.7
08:48:33 -04:40:51
SW34 1.7
09:02:48 -01:02:32
SW35 1.7

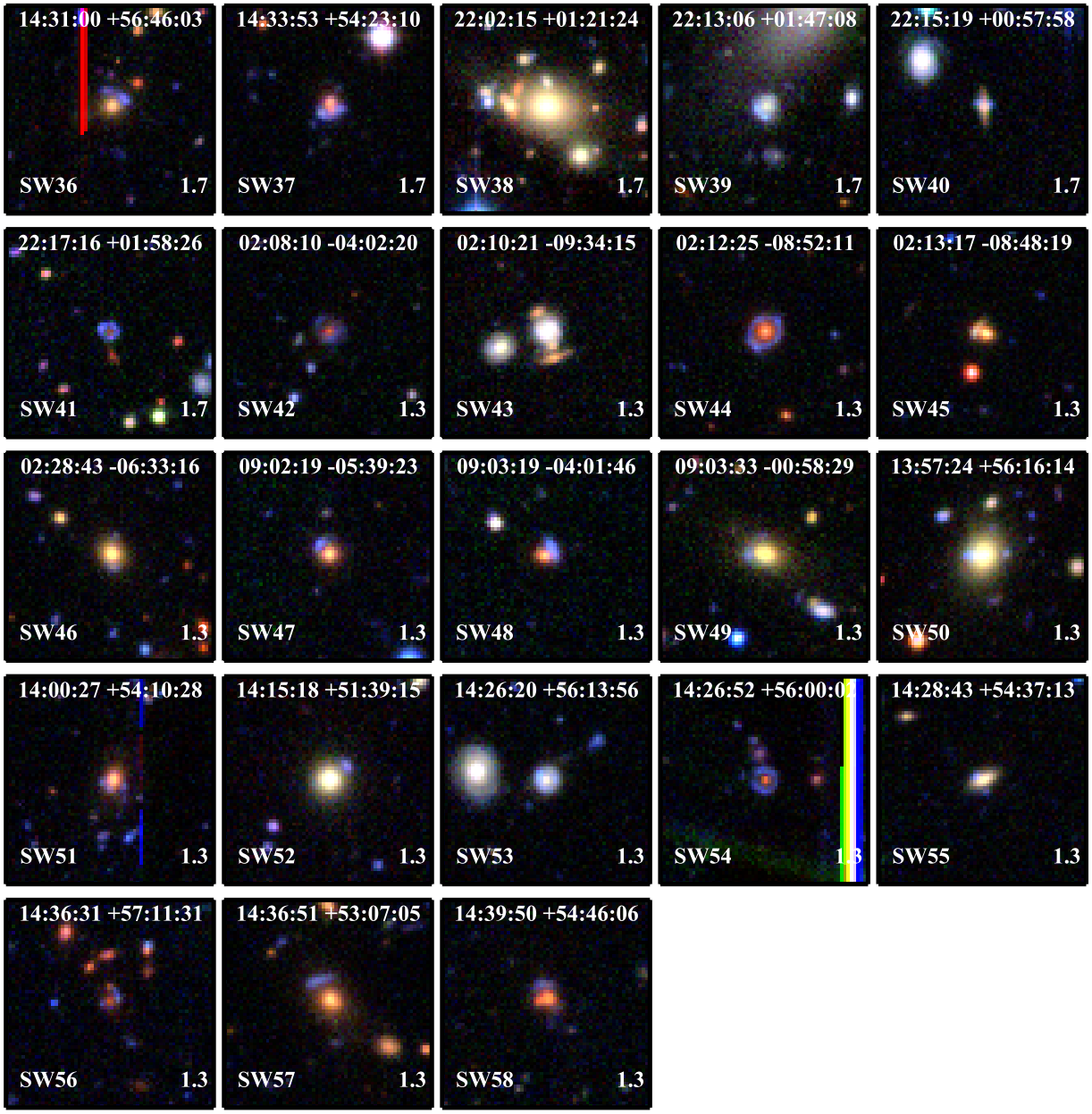


Figure 5. The new Space Warps lens candidates with expert grade G>=1.3. The images are 30″on the side.

separation distribution for the known sample (green), for the Space Warps identified (known and new) lens sample (blue) only and combined known with the new Space Warps lens sample (magenta). It is interesting to note that both these samples have very similar profiles and thus, the profile of the combined sample has not changed much. 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 Einstein radii. 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 (Navarro et al. 1997) and Total profile which has NFW and Hernquist profiles combined with an adiabatically contracting model for 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 similar to the Total profile.

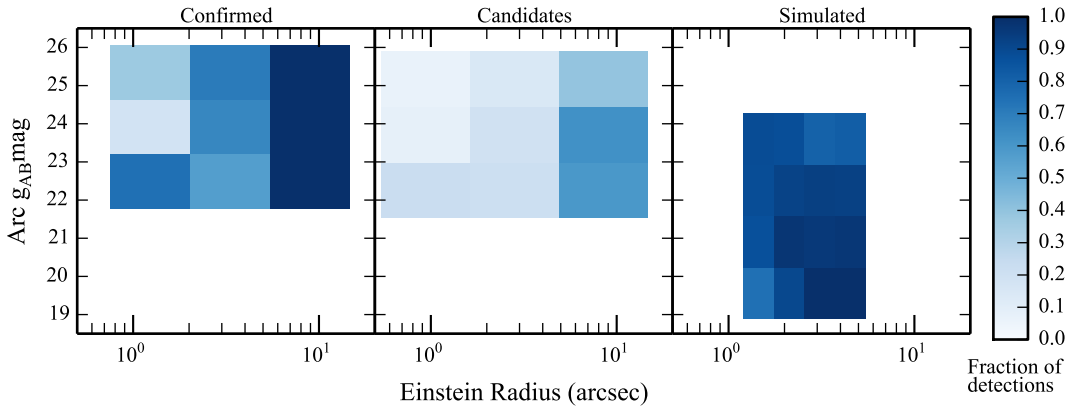


Figure 6. Fraction of lens candidates recovered by SPACE WARPS as a function of the arc magnitude (g band) and the Einstein radius for three lens samples, namely, the known lenses, the known lens candidates and the simulated sample.

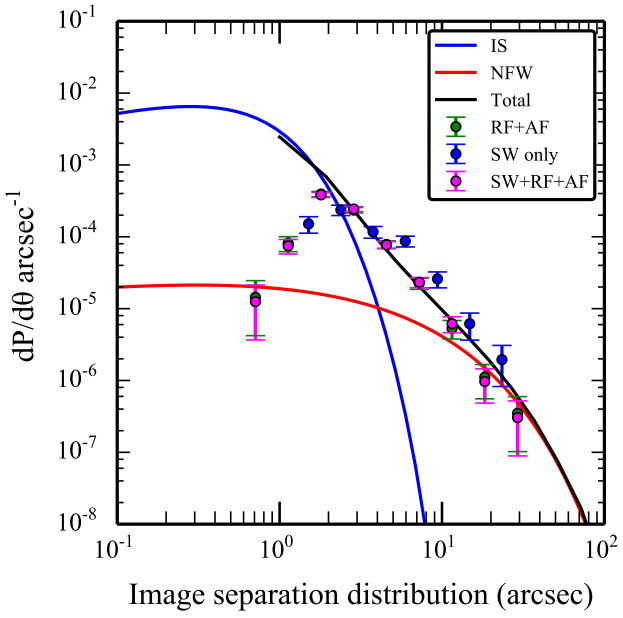


Figure 7. Image separation distribution. Comparing theoretical predictions (solid curves) with the CFHTLS known lens samples (green) and the combined sample of known and SPACE WARPS lens candidates (magenta). The sample of new and the known lens candidates discovered from SPACE WARPS alone is shown in 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.

7 DISCUSSION

Finding gravitational lenses is a difficult and complex task. No one 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 optimising 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 attemp to understand why each method failed to detect lenses from the other sample.

7.1 Comparison of the RINGFINDER, SPACE WARPS and ARCFINDER samples

Here, we compare the properties of the lens candidates from each sample to understand if there are any clear differences between them. We consider the lens redshift, the total flux of the lensed images and the Einstein radii of the systems for this purpose.

In Figure 8, we show the lens redshift and the arc flux measured in g AB mag 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 black 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 are 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 or the redshift of the lensing galaxies.

Since SPACE WARPS is visual based search, it is difficult to detect many lens candidates with small R_E in the presence of a bright lensing galaxy. While this is true for algorithms too, the RINGFINDER team adopted a different method to circumvent this problem whose goal was to find galaxy-scale lenses which have small R_E. The RINGFINDER sample is generated by working with lensing galaxy −subtracted images and this led to increased detections of small image separation candidates. The SPACE WARPS lens search did not show any galaxy subtracted images and hence, this results in a quantitative difference between the SPACE WARPS and the RINGFINDER sample as a function of R_E.

The ARCFINDER, on the other hand, is capable of find-

ing smaller image separation lenses but a lower limit on R_E was imposed by the ArcFinder team to simplify the selection function of their sample. Thus, we can compare the Space Warps and ArcFinder samples at about $R_E > 2''$. There is a good overlap between the two samples given that most of these are candidates and could turn out be not lenses. We again caution the readers that the uncertainties on the measurements are not quantified and may change the results to some extent.

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 the future.

Lastly, in Figure 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 color shows the overlap between the Space Warps and the RingFinder samples and the purple color shows the overlap between the Space Warps candidates and the ArcFinder samples. As noted earlier, the RingFinder dominates the small $R_E(< 2'')$ range although Space Warps does find 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 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.

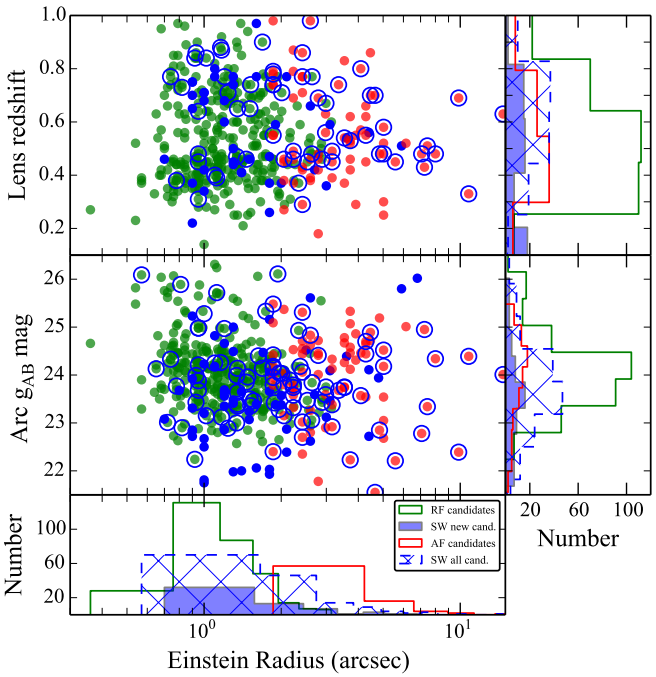


Figure 8. Comparison of the lens redshift and the arc magnitude with the Einstein radius for all of the three lens samples, namely, the RingFinder (green dots), Space Warps (known candidates−black circles and new candidates only−blue dots) and ArcFinder (red dots). Alls samples have broadly similar properties.

7.2 Why Space Warps candidates were missed by lens finding robots?

We rerun the RingFinder and ArcFinder on images centered on the new Space Warps candidates to trace and understand at what stage the algorithm failed to detect them.

First, we rerun RingFinder on the new Space Warps sample. At the beginning, a galaxy catalog is generated based on magnitude, redshift and SED type (Gavazzi et al. 2014, see) to select galaxies which are most likely to act as lenses. We find that about 40% of the new Space Warps candidates fail to meet this initial selection criteria, for example, SW1, SW13, SW19, SW22, SW26 and SW29. All of the lensing galaxies are bright enough to satisfy the $i < 22$ criterion. 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.

In the following 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 run on this image to quantify the lensed image properties. About 50% of the Space Warps candidates could not be detected by the object finder because properties such as the image area, axis ratio, magnitude/color and alignment with respect to the lensing galaxy are not satisfied. Some of the candidates missed at this stage are, for example, SW4, SW5, SW6, SW25, SW35, SW38 and SW45.

Next, we rerun the ArcFinder on the same Space Warps sample of new candidates. The ArcFinder is directly run on 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 $\sim 19350 \times 19350$ pixels^2. For the rerun, we work with much smaller images because this is faster but this alters the measured noise and hence, affects the number and type of arc detections. We find that about 30% of the new candidates are detected without changing any of the thresholds in the code because of the change in the noise level.

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 stimator above a certain threshold are connected to form the arc feature. This is building of the arc candidate. Subsequently, arc properties such as the area, mean flux, length and curvature are determined. We relax the threshold at the building stage and also relax thresholds mainly on the area of the arc which leads to detection of about 75% of the new Space Warps candidates. We find that relaxing thresholds on other arc properties do not improve the detection rate significantly and thus are left unchanged.

Typically, the candidates are missed from the ArcFinder sample if a) they are brighter in the redder band

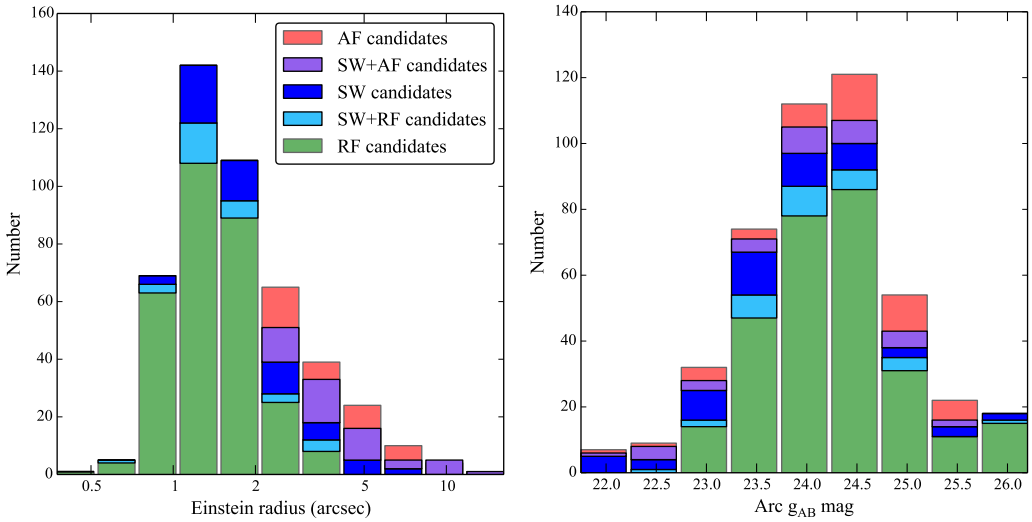


Figure 9. Candidate detections by the RingFinder, Space Warps and the ArcFinder as a function of the Einstein radius and g band magnitude of the lensed images.

rather than the bluer g band used for arc detection because then the arc feature will be at the noise level and will fail detection b) the flux of the arc and the galaxy are blended in the g band such that the ArcFinder mistakenly connects part of the galaxy to the pixels belonging to the arc candidate which then results in a candidate with no morphological characteristics of an arc or c) if the candidate is not elongated which means most of the lensed quasars with circular looking lensed images will be missed.

Relaxing the thresholds obviously increases the total number of candidate detections which includes a large sample of false positives. For example, the number of arc candidate detections increase by a factor of ∼2 when we relaxed the thresholds in the reruns described above whereas the number of false positives increase by a factor of ∼5. Thus, it is not recommended to relax these thresholds alone. A better alternative is to relax the thresholds to increase the completeness and cross-correlate the arc candidate positions with a galaxy catalog to discard those candidates which are not close to a putative lensing galaxy within typical radius in order to reduce the false positives.

7.3 False negatives: known lenses missed by Space Warps

Like any lens finding method, the Space Warps system can fail to detect certain kinds of lenses. We find that about 40% of the known sample of lenses are missed both at Stage I and II (see Table 2). In fact, most of the known lenses from Stage I are discovered at Stage II suggesting that most lenses are lost right at Stage I. Below, we focus on the known lens sample at Stage I 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 Figure 6). Among the confirmed lenses from the RingFinder, about 50% are missed. Out of the missed sample of 18 lenses, about half of them are visually difficult to detect and 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, ∼23% are missed by Space Warps. This is a relatively small sample of ∼ 5 systems and visual inspection 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 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 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.

7.3.1 Number of classifications

First, we check if the number of classifications (Nclass) is significantly lower for the missed sample compared to the detected one. Surprisingly, most of the lenses in the known sample have few classifications (Nclass<10) which includes both the detected and missed lenses. Many of the remaining lenses have between 10 to 20 classifications. And, only a few subjects have Nclass>20. They continue to remain for long in the database because these candidates are possibly difficult to identify. Therefore, the distribution of Nclass can not be the reason for losing the lenses from the missed sample.

7.3.2 Blind lens search

Efficiency of a visual search can vary in different sections of an image. Our eyes tend to focus usually at the center of an image and lens candidates close to the borders could go undetected. Therefore, it is essential to test and understand

if Space Warps is missing some of the known lenses because they happen to be close to the borders of the image cutouts.

From the SWAP, the image cutouts inspected by the Space Warps volunteers receive a status of detected (if $P>P_{accthresh}$), rejected (if $P<P_{rejthresh}$) and undecided (if $P_{rejthresh} < P < P_{accthresh}$). In Figure 10, we compare the positions of lenses which are detected (red), undecided (green) and rejected (blue). The left and the right 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 some cases, randomly selected subsamples are shown for the ease of visual comparison.

We do not find any strong visual correlation in the detection rate of lenses as a function of their positions in the image for both the simulated and the known lens sample. Thus, the completeness of the lens sample is most likely not significantly affected by whether a lens is located close to the border or well within the center.

7.3.3 Volunteer profile

Here, we investigate if the power[12] of volunteers is systematically different between the sample of 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. In Figure 11, we show the trajectory plots of a few examples of detected lenses (top left panels) and missed lenses (bottom left panels) by Space Warps at Stage I. The number of classifications (Nclass) for a subject increase from top to bottom. Also, every subject is assigned a prior probability P0= 2×10^{-4} (grey dashed line) and starts at the middle of the trajectory plot. 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 ($P_{accthresh} = 0.95$) on the right and is rejected, if it crosses the red-dashed line marking the $P_{rejthresh} = 10^{-7}$ on the left.

By how much the P value will change depends on how well the volunteers are performing on the training sample. Thus, high power volunteers will change the P by a large factor compared to the low power volunteers. This is evident in the trajectory plots as large and small distances in between the consequent 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, the high power volunteers are mostly classifying them as subjects with lens candidates. Further qualitative comparison and discussion of the trajectory plots is given in Appendix A.

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 Figure 11. Qualitatively, there are four types of volunteers making classifications - those causing positive large kicks (correct classifications by high power volunteers, those causing positive small kicks (correct classifications by low power volunteers), those causing negative small kicks (incorrect classifications by low power volunteers) and those causing negative large 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). The kick size is small if Δ log(P)=log ($P_{current}$)−log($P_{previous}$) < Δlog(P)$_{cut}$ (chosen to be 1.2) and is large if greather than Δlog(P)$_{cut}$.

Some of the key differences are i) the ratio of positive large kicks to positive small kicks for the detected sample is higher than the missed sample suggesting that the fraction of high power volunteers is large for the detected sample ii) in terms of Nclass, the missed sample is dominated by volunteers causing negative small kicks whereas for the detected sample there is a comparable contribution from three types of volunteers - positive small and large kicks and negative small kicks. iii) the Nclass from volunteers with negative large kicks are lower both for the detected and the missed samples. this is consistent with our expectation that high power volunteers should not be making incorrect classifications.Therefore, lack of high power volunteers classifying the missed lenses seems to be one of the major factors.

We rerun SWAP for Stage I where we use classifications from volunteers who produce Δ log(P)> 1.2 only. This obviously means reducing the total Nclass per subject by a large fraction. As a result, we also need to change the $P_{accthresh}$ which is chosen to be 0.1 and we find that about a third of the missed lenses are detected while all the previously detected lenses remain detected too. The others simply do not have enough classifications from volunteers producing large positive kicks. Thus, it may be possible to detect more promising candidates if we preferentially show the rejected systems to a few high power volunteers. We understand that changing the rejection and acceptance thresholds will likely increase the rate of FPs along with improved completeness but most probably the increase in FPs will not be too significant and that it will still be worth using the new strategy.

8 SUMMARY AND CONCLUSIONS

We report the discovery of new gravitational lens candidates from the the first lens search through Space Warps. In this search, volunteers are shown $g-r-i$ color images of random regions of the sky taken by the CFHT Legacy Survey. The aim of this blind lens search is to find lenses that have been missed by previous searches done on the CFHTLS with lens finding algorithms. In this search, we use a training sample to train the volunteers and calibrate their performance which helps us in more efficient pruning of the lens candidates. The training sample has simulated lenses, duds and FPs. In this paper, we describe the details of the training sample generated and tuned for the CFHTLS data. Volunteers receive feedback messages when they click or fail to click on the training images during live classification of real images which helps them remain focused and improve their performance over time. More details about the design of

[12] see the Appendix B for the meaning of this term and how it is different from the "Skill" defined in paper I

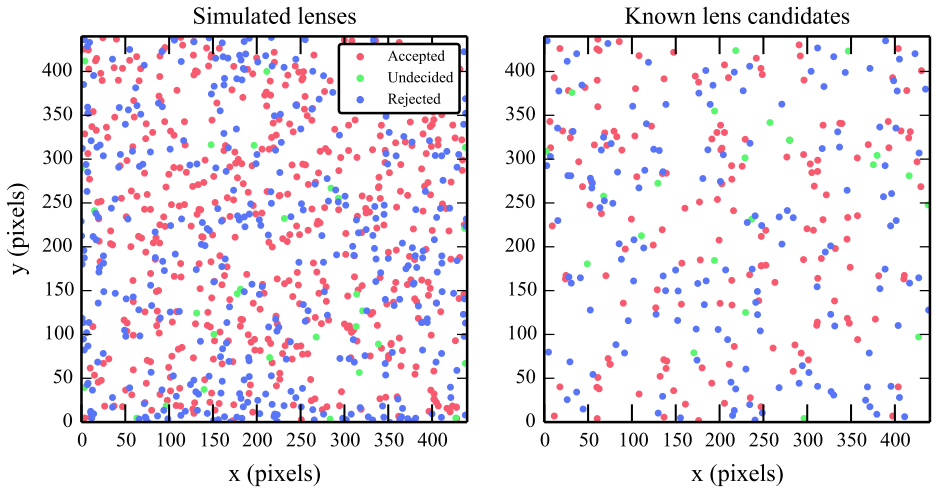


Figure 10. Completeness as a function of the positions of the lens systems. Simulated lenses (left) and real lens candidates (right) are shown. Irrespective of the status of the lenses, that is, detected, undecided or rejected, there is no strong dependency on the location of the lenses, both for the simulated and the real sample of candidates.

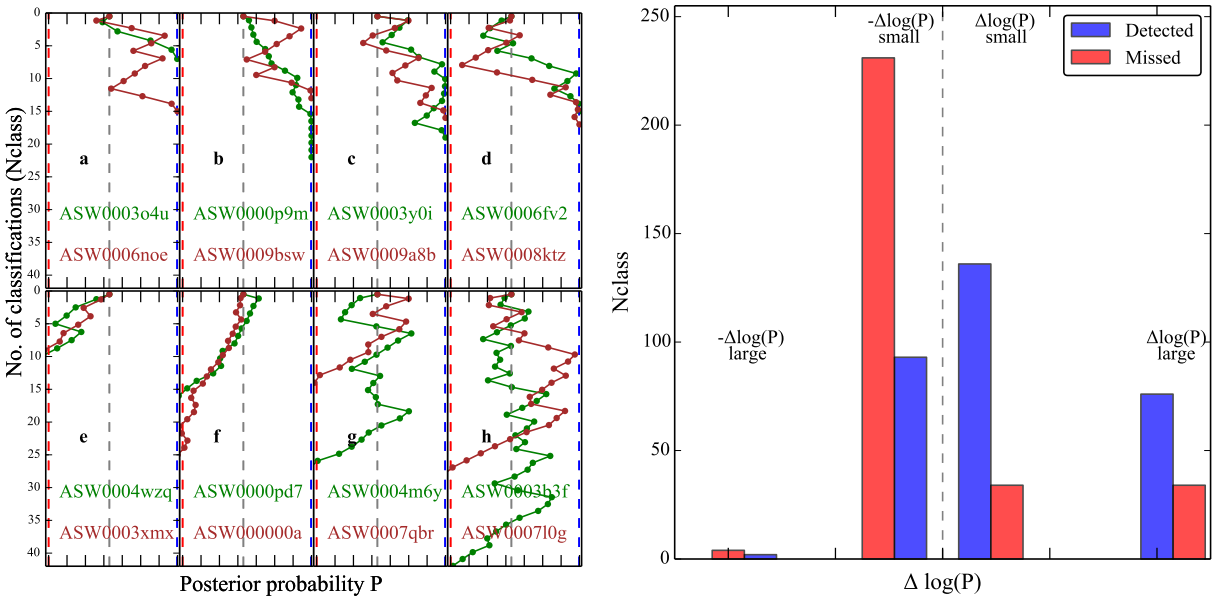


Figure 11. Examples of trajectories of some known lenses (top-left: detected and bottom-left: missed) at Stage I and histogram of the known lenses for different types of volunteers (right). Among the sample of missed lenses, most classifications are from the low power volunteers identifying these images incorrectly (i.e. $-\Delta$ logP) which overshadows the small number of correct identifications (i.e. Δ logP) 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.

Space Warps system and how it works are given in the companion paper, Paper I.

The blind lens search in the CFHTLS is done in two stages: Stage I and Stage II. In Stage I, volunteers inspected $\sim$ 430 000 subjects and selected a smaller sample of $\sim$ 3000 subjects with interesting lens candidates. In Stage II, after a careful reinspection of the candidates from Stage I, a smaller pool of more pure sample of $\sim$ 500 candidates is obtained. In the next step, the images are graded by three lens experts (AM, AV and PJM) producing candidates that range from mildly interesting to certainly a lens system.

In this paper (Paper II), we present the new Space Warps sample and compare this with the previously known samples from two robotic searches from the CFHTLS namely, the RingFinder and ArcFinder. And, further details about the new sample such as any follow up confirmation and investigations including mass modelling analysis will be presented in a subsequent Space Warps paper (Paper III).

Below, we give some of our conclusions from the first Space Warps lens search and the comparison with robotic searches:

• 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.

• We present a sample of 58 **new** gravitational lens candidates out of which 28 are more promising systems. These candidates received averaged grades, G$\geqslant$1.3 from three experts where 1-mildly interesting, 2-probably, a lens and 3-certainly, a lens. Another sample of over 100 lens candidates is recovered from previously known samples of lens candidates.

• Compared to the sample of robotically (the RingFinder and ArcFinder) detected lens candidates, the Space Warps sample finds lens systems with similar range in the lens redshift and properties of the lensed images such as the total magnitude and the Einstein radii. However, the Space Warps search being a purely visual one cannot find some of the RingFinder-identified-lensed images with sub-arcsecond Einstein radii because the flux of the typically faint lensed images is obscured by the flux from bright lensing galaxies. The RingFinder finds these candidates because it works with lensing galaxy-subtracted images.

• Qualitatively, the Space Warps sample finds 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 different colors, morphologies and sizes which are again typically missed by any given algorithm.

• Based on the known sample of lenses and lens candidates, we find that we lose a very small fraction of them as we go from Stage I to Stage II. In other words, it is more important to improve the lens detection efficiency right at Stage I to increase the overall completeness of the final lens sample

• From the sample of known lenses, Space Warps missed about 40% at Stage I. Among the missed sample of about 20 lenses, roughly half of them are visually difficult to identify because most of them are faint with small image separations which are identified through the RingFinder

• If we use the classifications from more skilled volunteers alone at Stage I, then we recover over 30% of the lenses from the missed sample. Thus, it is possible to improve the completeness of lenses by changing the strategy of when and who is shown an image.

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

SW ID	Name	RA (deg)	Dec (deg)	$z_{\rm phot}$	m_i (mag)	$R_{\rm A}$ (")	G	ZooID	P	Comments
SW1	CFHTLS J022409-105807	36.0398	-10.9688	0.0	0.0	4.8	3.0	ASW0004dv8	1.0	A,G
SW2	CFHTLS J140522+574333	211.3426	57.7259	0.7	19.7	1.0	2.7	ASW000619d	0.7	A,R
SW3	CFHTLS J142603+511420	216.5149	51.2390	0.0	0.0	4.4	2.7	ASW0006mea	0.7	A,G
SW4	CFHTLS J142934+562541	217.3926	56.4281	0.5	19.0	5.9	2.7	ASW0009cjs	0.8	A,G
SW5	CFHTLS J143454+522850	218.7270	52.4808	0.6	19.4	4.4	2.7	ASW0007k4r	0.4	Q,G/R
SW6	CFHTLS J143627+563832	219.1164	56.6425	0.5	19.4	1.5	2.7	ASW0008swn	0.9	A,D
SW7	CFHTLS J220256+023432	330.7369	2.5758	0.0	0.0	6.8	2.7	ASW0007e08	0.8	A,G/C
SW8	CFHTLS J020832-043315	32.1340	-4.5543	1.0	21.0	1.6	2.3	ASW0002asp	1.0	A,R
SW9	CFHTLS J020848-042427	32.2011	-4.4075	0.8	20.5	1.1	2.3	ASW0002bmc	0.9	D,D
SW10	CFHTLS J020849-050429	32.2078	-5.0749	0.8	20.6	0.9	2.3	ASW0002qtn	1.0	A,R
SW11	CFHTLS J022406-062846	36.0256	-6.4796	0.4	19.6	0.9	2.3	ASW0003wsu	0.7	A,E
SW12	CFHTLS J022805-051733	37.0236	-5.2927	0.4	18.8	1.4	2.3	ASW0009ans	1.0	Q,E
SW13	CFHTLS J023123-082535	37.8468	-8.4266	0.0	0.0	1.2	2.3	ASW0004xjk	0.3	A,R
SW14	CFHTLS J084841-045237	132.1708	-4.8772	0.3	19.0	1.0	2.3	ASW0004nan	1.0	A,E
SW15	CFHTLS J140030+574437	210.1260	57.7437	0.4	18.2	2.0	2.3	ASW0009bp2	0.6	A,E
SW16	CFHTLS J140622+520942	211.5958	52.1617	0.7	20.3	1.2	2.3	ASW0005rnb	0.7	A,R
SW17	CFHTLS J143658+533807	219.2425	53.6355	0.7	19.6	0.9	2.3	ASW0007hu2	0.6	D,D
SW18	CFHTLS J020642-095157	31.6750	-9.8658	0.2	20.8	0.9	2.0	ASW0001ld7	0.8	A,R
SW19	CFHTLS J021221-105251	33.0881	-10.8811	0.3	17.9	1.8	2.0	ASW0002dx7	0.8	D,E/S
SW20	CFHTLS J022533-053204	36.3888	-5.5346	0.5	19.4	3.6	2.0	ASW0004m3x	0.4	A,R/G
SW21	CFHTLS J022716-105602	36.8186	-10.9341	0.4	17.2	1.8	2.0	ASW0009ab8	0.7	A,E/G
SW22	CFHTLS J023008-054038	37.5359	-5.6774	0.6	19.7	1.9	2.0	ASW0003r61	0.5	A,E
SW23	CFHTLS J023315-042243	38.3133	-4.3789	0.7	19.7	1.0	2.0	ASW00050sk	0.8	A,R
SW24	CFHTLS J090308-043252	135.7844	-4.5479	0.0	0.0	1.2	2.0	ASW00007mq	0.6	A,E
SW25	CFHTLS J135755+571722	209.4827	57.2897	0.8	20.2	1.3	2.0	ASW0005ma2	0.8	D,D
SW26	CFHTLS J141432+534004	213.6372	53.6679	0.7	21.4	0.9	2.0	ASW0006jh5	0.8	A,R
SW27	CFHTLS J143055+572431	217.7333	57.4088	0.7	19.3	1.0	2.0	ASW0007wfj	0.9	A,R
SW28	CFHTLS J143838+572647	219.6589	57.4464	0.8	20.2	1.1	2.0	ASW0008qsm	0.9	A,R
SW29	CFHTLS J021057-084450	32.7414	-8.7474	0.0	0.0	2.5	1.7	ASW0002p8y	0.4	A,G
SW30	CFHTLS J021514-092440	33.8109	-9.4111	0.7	19.9	2.6	1.7	ASW00021r0	0.4	A,R/G
SW31	CFHTLS J022359-083651	35.9995	-8.6143	0.0	0.0	3.1	1.7	ASW0004iye	0.4	A,E
SW32	CFHTLS J022745-062518	36.9387	-6.4218	0.6	20.5	1.2	1.7	ASW0003s0m	0.5	A,R
SW33	CFHTLS J023453-093032	38.7232	-9.5089	0.5	19.8	0.7	1.7	ASW00051ld	0.3	A,D
SW34	CFHTLS J084833-044051	132.1385	-4.6809	0.8	20.2	0.9	1.7	ASW0004wgd	0.7	A,R
SW35	CFHTLS J090248-010232	135.7020	-1.0424	0.4	19.1	1.4	1.7	ASW000096t	0.6	D,E
SW36	CFHTLS J143100+564603	217.7511	56.7675	0.0	0.0	1.8	1.7	ASW00086xq	0.8	A,E
SW37	CFHTLS J143353+542310	218.4736	54.3862	0.8	19.8	1.6	1.7	ASW0009cox	0.6	A,R/G
SW38	CFHTLS J220215+012124	330.5635	1.3567	0.3	17.4	4.6	1.7	ASW0005qiz	0.5	rA,G
SW39	CFHTLS J221306+014708	333.2758	1.7856	0.0	17.1	1.4	1.7	ASW0008wmr	0.9	A,S
SW40	CFHTLS J221519+005758	333.8321	0.9661	0.4	20.2	1.0	1.7	ASW0008xbu	0.8	A,D
SW41	CFHTLS J221716+015826	334.3189	1.9739	0.1	21.6	1.0	1.7	ASW00096rm	1.0	A/R,R
SW42	CFHTLS J020810-040220	32.0450	-4.0389	1.0	20.8	1.8	1.3	ASW0001c3j	0.7	A,R
SW43	CFHTLS J021021-093415	32.5898	-9.5711	0.4	18.4	2.7	1.3	ASW0002k40	0.4	D,S
SW44	CFHTLS J021225-085211	33.1051	-8.8697	0.8	19.5	2.1	1.3	ASW00024id	1.0	R,R
SW45	CFHTLS J021317-084819	33.3234	-8.8055	0.5	19.8	1.3	1.3	ASW00024q6	0.4	A,R/E
SW46	CFHTLS J022843-063316	37.1794	-6.5547	0.5	19.1	1.8	1.3	ASW0003r6c	0.3	D/A,E
SW47	CFHTLS J090219-053923	135.5794	-5.6566	0.0	0.0	2.0	1.3	ASW0000g95	1.0	A,R/E
SW48	CFHTLS J090319-040146	135.8311	-4.0297	0.0	19.8	1.2	1.3	ASW00007ls	0.5	A,R/E
SW49	CFHTLS J090333-005829	135.8886	-0.9749	0.0	0.0	2.1	1.3	ASW00008a0	1.0	A/D,E/G
SW50	CFHTLS J135724+561614	209.3536	56.2707	0.0	0.0	2.6	1.3	ASW0006e0o	0.9	D,E
SW51	CFHTLS J140027+541028	210.1164	54.1746	0.0	0.0	1.2	1.3	ASW0006a07	0.6	Q,R/E
SW52	CFHTLS J141518+513915	213.8290	51.6542	0.4	18.3	3.0	1.3	ASW00070vl	0.8	D,E
SW53	CFHTLS J142620+561356	216.5870	56.2323	0.5	19.5	1.3	1.3	ASW0007sez	0.8	A/R,S
SW54	CFHTLS J142652+560002	216.7201	56.0006	0.0	0.0	1.5	1.3	ASW0007t5y	1.0	R,R
SW55	CFHTLS J142843+543713	217.1815	54.6204	0.4	19.7	1.3	1.3	ASW0007pga	0.6	D,D
SW56	CFHTLS J143631+571131	219.1315	57.1922	0.7	20.9	1.3	1.3	ASW0008pag	0.6	D/A,R
SW57	CFHTLS J143651+530705	219.2150	53.1183	0.6	19.2	3.1	1.3	ASW0007h27	1.0	A,E/G
SW58	CFHTLS J143950+544606	219.9609	54.7686	0.0	0.0	1.7	1.3	ASW00085cp	0.4	A,G/R

SW ID	Name	RA (deg)	Dec (deg)	$z_{\rm phot}$	m_i (mag)	$R_{\rm A}$ (")	G	ZooID	P	Comments

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 colors or spectral analysis. The symbols are E: Elliptical, S: (face on) Spiral, G: Group-scale, D: Edge on disk, R: Red starforming galaxy. This galaxy falls within the masked region as per the catalog from which the magnitudes and the redshift are extracted.

ACKNOWLEDGEMENTS

We thank all XXXmembers of the Space Warps community for their contributions to the project so far. A complete list of collaborators is given at... In particular we would like to recognise the efforts of XXX, YYY etc in moderating the discussion.

We are also grateful to Brooke Simmons, David Hogg, XXX and YYY for many useful conversations about citizen science and gravitational lens detection. PJM was given support by the Royal Society, in the form of a research fellowship, and by the U.S. Department of Energy under contract number DE-AC02-76SF00515. AV acknowledges support from the Leverhulme Trust in the form of a research fellowship. The work of AM and SM was supported by World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan. The work of AM was also supported in part by National Science Foundation Grant No. PHYS-1066293 and the hospitality of the Aspen Center for Physics. The Space Warps project is open source. The web app was developed at https://github.com/Zooniverse/Lens-Zoo while the SWAP analysis software was developed at https://github.com/drphilmarshall/SpaceWarps. This work is based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/IRFU, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This research used the facilities of the Canadian Astronomy Data Centre operated by the National Research Council of Canada with the support of the Canadian Space Agency. CFHTLenS data processing was made possible thanks to significant computing support from the NSERC Research Tools and Instruments grant program.

APPENDIX A: QUALITATIVE COMPARISON OF DETECTED AND MISSED KNOWN LENSES

The bottom panels show the missed or rejected lenses. The green (and red) trajectories show visually easier (and more difficult) to identify lenses. 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 short negative kicks. The panel f represents a small fraction of this sample where the kics are only short and negative. The panel g shows how some lenses receive a bunch of large positive kicks which are led to rejection by still mostly negative short 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 top panels show the detected lenses. The cases shown with green trajectories in each panel can be thought of counterparts of the trajectories of the missed lenses in the corresponding bottom panel except that the type of volunteers are different. Most detected lenses are similar to the case in panel a which are detected within a few classifications but coming from large positive kicks. Panel b represents a few odd cases which are dominated mainly by short positive kicks. Panel c shows a lens getting more classifications because of the tug between positive and negative kicks whereas the panel d represents the extreme cases when the subjects 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 short negative kicks.

APPENDIX B: 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 $P_{\rm L}$, the probability of correctly identifying a lens as a lens and $P_{\rm D}$, 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 $P_{\rm L}$ and $P_{\rm D}$ (see the left panel of Figure B1). The skill reflects the all-round ability of a classifier to contribute information.

As described in the appendix of Paper I, the posterior probability P of a subject is determined by the $P_{\rm L}$ and $P_{\rm D}$ 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 (not) containing a lens, a volunteer with high ($P_{\rm D}$) $P_{\rm L}$ implies a large (negative) positive kick irrespective the value of ($P_{\rm L}$) $P_{\rm D}$ as shown in the (right most) middle panel of Figure B1. However, large (negative) positive kicks are still possible for a volunteer located in the (lower) upper triangle with different combinations of ($P_{\rm L}$,$P_{\rm D}$) suggesting that the kick is not a simple function of ($P_{\rm L}$,$P_{\rm D}$).

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 subjects 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).

REFERENCES

Alard, C. 2006, ArXiv Astrophysics e-prints
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
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. 2014, ArXiv e-prints

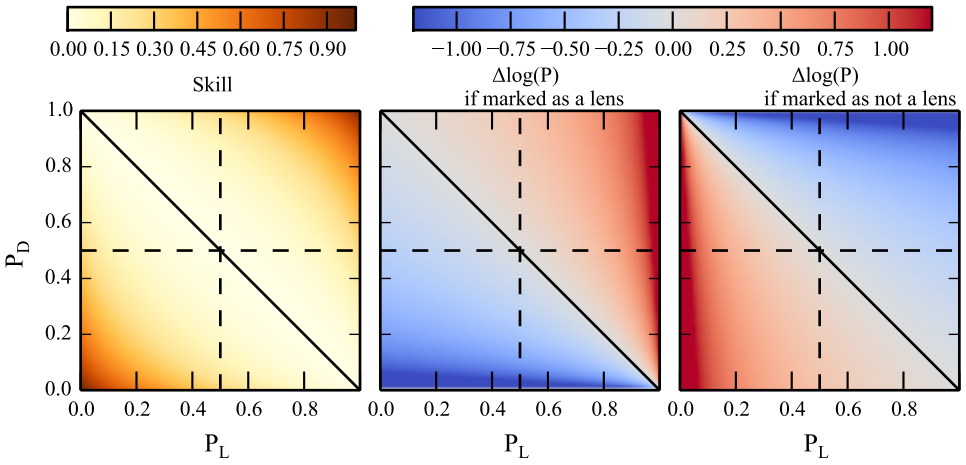


Figure B1. Skill of the volunteers, $\Delta \log(P)$ given a lens or not a lens in an image as a function of $P_{\rm L}$ and $P_{\rm D}$. Each of these quantities represent ability of the Volunteers but do not have a simple linear relation with $P_{\rm L}$ and $P_{\rm D}$. Question from AM: PJM, should we just use power everywhere instead of the $\Delta \log(P)$? This figure and this section still needs to be polished.

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
—. 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., Christlein, D., & Zabludoff, A. I. 2000a, ApJ, 545, 129
Keeton, C. R., Mao, S., & Witt, H. J. 2000b, ApJ, 537, 697
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
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
—. 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
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., Bayliss, M. B., Dahle, H., Sharon, K., Gladders, M. D., Natarajan, P., Hennawi, J. F., & Koester, B. P. 2012, MNRAS, 420, 3213
Oguri, M., Keeton, C. R., & Dalal, N. 2005, MNRAS, 364, 1451
Oguri, M., & Marshall, P. J. 2010, MNRAS, 405, 2579
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
Seidel, G., & Bartelmann, M. 2007, A&A, 472, 341
Sereno, M., & Paraficz, D. 2014, MNRAS, 437, 600
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
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, ApJL, 703, L132
Zitrin, A., Broadhurst, T., Barkana, R., Rephaeli, Y., & Benítez, N. 2011, MNRAS, 410, 1939

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