Mon. Not. R. Astron. Soc. **000**, 1–12 (2013) Printed 30 September 2014 (MN LATEX style file v2.2)

SPACE WARPS: II. Did we beat the lens finding robots in CFHTLS?

Anupreeta More,[1⋆] Aprajita Verma,[2] Phil Marshall,[2,3] Amit Kapadia,[4]
Michael Parrish,[4] Chris Snyder,[4] Julianne Wilcox, Elisabeth Baeten,
Christine Macmillan, Claude Cornen, Surhud More,[1] Raphael Gavazzi,[6] Chris Lintott,[1]
Robert Simpson,[1] David Miller,[3] Arfon Smith,[3] Edward Paget,[3]
Prasenjit Saha,[4] Rafael Kueng,[4] Kelly Borden,[3] Tom Collett,
Thomas Jennings, Matthias Tecza,[1] Layne Wright and possibly others

[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] *Adler Planetarium, Chicago, IL, USA*
[5] *Department of Physics, University of Zurich, Switzerland*
[6] *Institut dAstrophysique de Paris, UMR7095 CNRS Université Pierre et Marie Curie, 98bis bd Arago, 75014 Paris, France*

to be submitted to MNRAS

ABSTRACT

The CFHT Legacy Survey has been searched for strong lenses with semi-automated algorithms both at galaxy and groups scales. With the aim of improving these lens finding robots, we carry out a blind lens search in the complete CFHTLS WIDE survey with SPACE WARPS. We describe the training sample used for both training the citizen scientists that participated in SPACE WARPS and calibrating their performance. We generate realistic looking simulated samples of lenses both at galaxy and group-scales as part of this training sample. We present 80 new strong gravitational lens candidates discovered from the SPACE WARPS-CFHTLS search out of which 40 candidates are promising. Furthermore, we compile a sample of false positives which can be used for robot testing, first of its kind for the lensing community. The simulated sample is also used to make a comparison between human performance against the robots. We find that the XX per cent of the known lens sample is recovered and the new lens sample has XX completeness with respect to the simulated sample. We make available the following data products: simulated sample, simulation code and the false positives sample at XXX.

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 they have in astrophysics and cosmology. Strong lenses are routinely used to probe the dark matter distribution from galaxy (ref) to cluster scales (ref), to study distant young galaxies by using the lensing magnification as a natural telescope (ref), to test the cosmological model by constraining cosmological parameters such as the Hubble constant (ref) and dark energy (ref) and many more. Even though strong lenses are rare, since a foreground massive object needs to be sufficiently aligned with a distant background source to produce multiple images, systematic lens searches have led to discovery of over 500 lenses till date (XXX add mld url).

Rarity of lenses implies searching for them is a painstaking task. Efficient automated methods are thus imperative

⋆ anupreeta.more@ipmu.jp

to finding a reasonably complete and pure sample of strong lenses.

Since the inception of the first citizen science project, Galaxy Zoo, to classify galaxy morphologies (**?**), several astronomy and non-astronomy projects have been launched by the Zooniverse leading to many interesting project. For example, PlanetHunters has discovered XXX transiting exoplanets (**?**), Supernova XXX add ref and XXX add ref

In Paper I, we describe Space Warps, an online system that enables crowd-sourced detection of gravitational lenses. In this paper (referred to as Paper II), we describe our first lens search with Space Warps in the CFHTLS data.

This paper is organised as follows. In Section **??** we give a brief overview of the Space Warps system, focusing on the aspects most relevant to the interpretation of the results of this first lens search. In Section 2 we introduce the CFHTLS imaging data and the known lens samples from the CFHTLS. In Section **??**, we explain the training sample generated to aid the Space Warps users in the lens search. In Section 5, we present the new lens candidates and our findings. We discuss the implications of our results for future lens searches in Section 6 and draw conclusions in Section 7.

2 DATA

2.1 The CFHT Legacy Survey

The Canada-France-Hawaii Telescope Legacy Survey (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″. The CFHTLS WIDE covers a total area of 171 deg^2 on the sky and it 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[1]. 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 great 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 lenses. We use the data from the final T0007 release taken from the Terapix website[2] 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 DES, KIDS and HSC and planned surveys such as the LSST. Searching for lenses with Space Warps in the CFHTLS will teach us important lessons and help prepare us for these larger imaging surveys.

[1] These numbers are estimated from http://terapix.iap.fr/cplt/table_syn_T0006.html
[2] http://terapix.iap.fr/cplt/T0006-doc.pdf

2.2 Existing CFHTLS lens samples

The CFHTLS data has been searched for lenses using semi-automated algorithms, primarily, in the g band where the predominantly faint blue source galaxies are bright compared to the predominantly red deflector galaxies. Here, we briefly mention the lens samples which were known to the authors prior to the lens search with Space Warps.

At galaxy-scales, we focus on two primary lens searches. The RingFinder (**?**) was used for finding compact rings or arcs around centers of isolated and massive early-type galaxies. By subtracting the PSF-matched i-band images from the g-band images, the algorithm looks for excess flux in the bluer g-band. An object detector measures the properties of these residual blue features, and candidates meeting length-width ratio and tangential alignment criteria are then visually inspected to form the final sample. **?** first selected some 638,000 targets as either photometrically-classified early type galaxies, or objects selected to have red centers and blue outer parts, from the T06 CFHTLS data release catalogs. 14370 were found to show detectable blue residuals, and 2524 were visually inspected, having passed the automatic feature selection process. This led to a sample of 42 good quality (`q_flag` = 3) and 288 medium quality (`q_flag` = 2) lens candidates. In addition to this well defined sample, **?** reported a further 71 serendipitously detected lens candidates. From this sample of "RingFinder candidates," the SL2S team found, during their follow-up campaign, 39 confirmed lenses (and 17 promising candidates). We use this sample of 39 "confirmed RingFinder lenses" in our completeness analysis.

The second galaxy-scale lens search was to find edge-on galaxy lenses in the CFHTLS (Sygnet et al. 2010) by selecting the galaxy's profile from the output of Sextractor in the i band and with low inclination angle. This sample has about 3 promising and a total of 18 lens candidates.

On the other hand, the arcfinder More et al. (2012) was used for finding blue arc-like features in the complete CFHTLS data withouth any pre-selection on the type of the lensing object. The search was carried out in the g-band which is the most efficient wavelength to find typical lensing galaxies. This sample, called the SARCS, has 55 promising and a total of 127 lens candidates and consists of both galaxy and groups/cluster scale lens candidates. Arc finding is better suited for lensed images or arcs with larger image separations i.e. more massive systems like groups and clusters. Thus, the SARCS sample has mostly groups/cluster-scale lenses and a few galaxy-scale lenses.

For the purposes of transparency and to help a little with their training, the volunteers participating in Space Warps-CFHTLS lens search were made aware of these known lens samples. Images containing the systems from the above samples were labelled as "Known Lens Candidates" in the Talk forum, where volunteers have the opportunity to discuss their findings with other volunteers and the science team.

XXX refer Paper I ?

2.3 Image Presentation

Preparation of data: divide survey into overlapping tiles.

Presentation of images. Uniform scales, to build in-

tuition and avoid rescales due to bright objects. Arcsinh stretch, to bring out low SB features. Approximately optimized, how? Examples of images.

We carry out a blind search over the CFHTLS fields where we, mainly, use the g-r-i color information to look for signs of lensing. We extract contiguous cutouts of size $81.84''$ (440 pixels). The neighbouring cutouts have an overlapping region of $10''$(54 pixels). If a possible lens candidate is too close to the edge of a cutout, this overlap allows the inspector to get a clearer view of the same candidate in a neighbouring cutout. We note that since the images are shown randomly, a given inspector may not necessarily come across the neighbouring cutout unless the inspector classifies a lot of images. However, this is not a problem since we our user base is extremely large and we get multiple classifications of the same image.

3 TRAINING SAMPLE: SIMULATED LENSES

State the importance

3.1 Methodology

We create two main types of simulated lens sample a) galaxy-scale lenses and b) group or cluster-scale lenses. The galaxy-scale lenses are further divided into two types based on the nature of the background sources, namely, galaxies and quasars. Below, we describe how each type of the lens sample was generated.

3.1.1 Galaxy-scale lenses

The $N_{\rm src}$ behind a lens is then calculated by doing the following integral,

$$N_{\rm src} = N_{\rm src}(> L_s, z_s) \int_{z_l}^{\infty} \sigma_{\rm lens}(\sigma, z_l, z_s, q) D_s^2 (1+z_s)^2 \frac{{\rm d}\chi}{{\rm d}z_s} {\rm d}z_s \quad (1)$$

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

where $\Phi(L_s, z_s)$ is the source luminosity function per unit comoving volume, q is the projected axis ratio of the lens ellipticity, χ and D_s are the comoving and angular diameter distances to the source,

We use the elliptical galaxy (LRG) catalog from the CFHTLS XXX to select all the foreground galaxies (e.g. $z < 1$) that are potential lenses for the simulated sample. We exclude all those galaxies whose positions match with the lensing galaxies from the known CFHTLS SL2S lens samples More et al. (2012) within 2 arcsec (XXX check).

First, we calculate the luminosity and velocity dispersion of each potential lensing galaxy using the CFHT Megacam g and r band magnitudes along with the photometric redshift (z_l) from the LRG catalog. The Megacam magnitudes are converted to SDSS magnitudes[3] and are further k-corrected to redshift $z = 0.1$ (Frei & Gunn 1994). We assume that the evolution of galaxy luminosities is similar to that determined by (Faber et al. 2007), that is, a decline of 1.5 in the m_{r*} from redshift $z = 1$ to $z = 0$ (see Eq. 4).

$$\frac{L}{L_*} = 10.0^{-0.4\ (m_{r\rm SDSS} - m_{r*})} \quad (3)$$

where m_{r*} is

$$m_{r*} = -20.44 + 1.5\ (z_l - 0.1)\,. \quad (4)$$

We use the $L - \sigma$ relation from (**?**) to get the velocity dispersion as given in Eq. 5.

$$\sigma = 142 \left(\frac{L}{L_*}\right)^{1/3} \quad (5)$$

Next, in order to decide whether a galaxy is likely to act as a strong lens, we calculate the lens cross-section ($\sigma_{\rm lens}$) and the number of sources ($N_{\rm src}$) that are in the background. Following (**?**), the lens cross-section is calculated analytically for an isothermal model and is given by

$$\sigma_{\rm lens} = b_I^2 \int_0^{2\pi} 0.5 r^2(\theta) d\theta \quad (6)$$

where b_I is

$$b_I = b_{\rm SIS} \epsilon_3 / sin^{-1}(\epsilon_3)\,, \quad (7)$$

the eccentricity (ϵ_3) is

$$\epsilon_3 = (1 - q_3^2)^{1/2} \quad (8)$$

and the projected axis ratio is given by

$$q_k = \sqrt{q_3^2 {\rm sin}^2 i_e + {\rm cos}^2 i_e}\,. \quad (9)$$

In the above equations, q_3 is the 3d axis ratio of the ellipsoid and i_e is the inclination angle. Also, $b_{\rm SIS} = 4\pi \frac{c^2}{\sigma^2} \frac{D_s}{D_{ls} D_l}$ and is referred to as the Einstein radius where D_s,D_l and D_{ls} are angular diameter distances to the source, the lens and between the lens and source, respectively.

Next, if the foreground galaxy can act as a lens and has at least one source in the background, then we determine a redshift (z_s) and i-band magnitude of the background source(s). We assume two types of background sources namely, galaxies and quasars. For each source, the redshift and magnitude are generated by drawing randomly from the following redshift and luminosity distributions. For galaxies, we assume the redshift distribution is

$$p_s = \frac{\beta z_s^2 {\rm exp}(\frac{z_s}{z_0(m_{\rm lim})})^\beta}{\Gamma(3/\beta) z_0^3(m_{\rm lim})} \quad (10)$$

where $\beta = 3/2$ and $z_0(m_{\rm lim}) = 0.13 m_{\rm lim} - 2.2$ and the luminosity function is

$$n_s = \int_{-\infty}^{m_{\rm lim}} \frac{n_0 {\rm d}m}{\sqrt{10^{2a(m_1 - m)} + 10^{2b(m_1 - m)}}} \quad (11)$$

with parameters $a = 0.30$, $b = 0.56$, $m_1 = 20$ and $n_0 = 3 \times 10^3\ deg^{-2}$ as given in (Faure et al. 2009) and references therein. For quasars, we calculate the luminosity function

[3] http://www3.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/megapipe/docs/filters.html

Figure 1. Examples of the three types of simulated lenses.

by following the prescription of (**?**) and use k-corrections by (Richards et al. 2006).

The luminosity function is expressed as

$$\frac{\mathrm{d}\Phi}{\mathrm{d}M} = \frac{\Phi_*}{10^{0.4(\alpha+1)(M_{\rm abs}-M_*)} + 10^{0.4(\beta+1)(M_{\rm abs}-M_*)}} \quad (12)$$

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

$$f(z) = \frac{e^{\zeta z_s}(1+e^{\xi z_*})}{(\sqrt{e^{\xi z_s}} + \sqrt{e^{\xi z_*}})^2}\,. \quad (13)$$

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. We note that when calculating $N_{\rm src}$, the source number density is artificially boosted by a factor (see Table 1) to increase the occurrence of simulated lenses. This helps in creating a large enough sample to carry out various performance tests.

Next, we determine properties of the background source for every lens. We follow similar procedures for both background galaxies and quasars. For simplicity, we simulate a single background source behind every lens. In order to select one background source from the $N_{\rm src}$ per lens, we do ray-tracing for all of the $N_{\rm src}$ sources with GRAVLENS (Keeton et al. 2000) and choose sources that satisfy criteria as given below. 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 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 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) given by

$$R_{\rm eff} = 10^{0.52}\frac{L_r^{2/3}}{(1+z_s)^2} \quad (14)$$

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 [4].

Once all the parameters are determined for the lens and source models, GRAVLENS is used to simulate 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 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 brightest group galaxy (BGG) at the center alone does not cause strong lensing. We need to account for the extra convergence arising from the dark matter component as well as satellite galaxies, at least, in the inner regions which are typically responsible for the multiple lensed images (**?**Oguri 2006). Owing to the lack of an appropriate group catalog for our purposes, we create a basic group catalog based on the magnitudes and photometric redshifts available for the CFHTLS. We select all galaxies with $10^{10.8} M_\odot$ as the 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.

We adopt an isothermal ellipsoid for the BGG and members whenever the ellipticities are available else we use an isothermal sphere. On the other hand, we adopt an NFW profile for the underlying dark matter halo. Assuming a constant mass-to-light ratio of 3×0.7 h M_*/L_*, we use the BGG luminosity to estimate the stellar mass. The stellar mass−halo mass relation (**?**), including random scatter, is then used to calculate the halo mass for the lens. Given the halo mass, other key parameters such as the scale radius (r_s)

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

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

Name	gal		qso	
	min	max	min	max
Source Redshift	1.0	4.0	1.0	5.9
Source Magnitude	21.0	25.5	21.0	25.5
boost factor	100 (40†)		1200	
Second brightest image	23		23	
Total magnification	19		20	
Lens shear strength	0.001	0.02	0.001	0.02
Lens shear pa	0	180	0	180
Source ellipticity	0.1	0.6		
Source PA	0	180		

† – corresponds to the factor used for Groups scale lenses.

and the density at the scale radius (ρ_s) can be determined for an NFW profile.

As described in Section 3.1.1, we calculate the luminosity and velocity dispersion for the BGG and each of the member galaxies. Next, we calculate the lens cross-section for each potential lensing group. The complexity in the lens models makes it analytically intractable to calculate the size of the caustics[5]. Hence, we use GRAVLENS to 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 found 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 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. 10-11).

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 images are added to 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.

The Figure 2 shows 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 and magnitude distributions of the lensing galaxies in the left and right panels of Figure 3 respectively. Furthermore, we overplot the distributions of respective properties of the SARCS lenses from (More et al. 2012) for comparison with arbitrary normalizations. We note qualitative similarities between the simulated and the real lens samples.

Similarly, we show the ellipticity and the position angle of the simulated lens galaxy population extracted from the T0007 release of the CFHTLS catalogs in the left and right panels of Figure 3, respectively. As before, the dashed-dotted (blue) curves show the same distribution for the SARCS lens population with arbitrary normalization (More et al. 2012).

We produce catalogs with lens and source properties for each of the three types of lenses. These catalogs are available XXX. 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, ellipticity and effective radius when the background sources are galaxies.

4 TRAINING SAMPLE: DUDS AND FALSE POSITIVES

A good training sample consists of a representative set of objects that one wants to find and another set of objects which appear to be from the former set but are of different origin in reality and which one can learn to discard efficiently. Indeed, we wanted to have a good training sample for the Space Warps users so that they can correctly identify the true lens candidates. 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 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 (see XXX) and a sample of XXX false positives for the Stage II inspection (see XXX). 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 users and categorized by the science team. Such a sample is tremendously helpful for training and understanding performances of various lens finding algorithms (refer james, sherry's paper ? XXX) **TBA data products; make the FP sample available**

[5] 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.

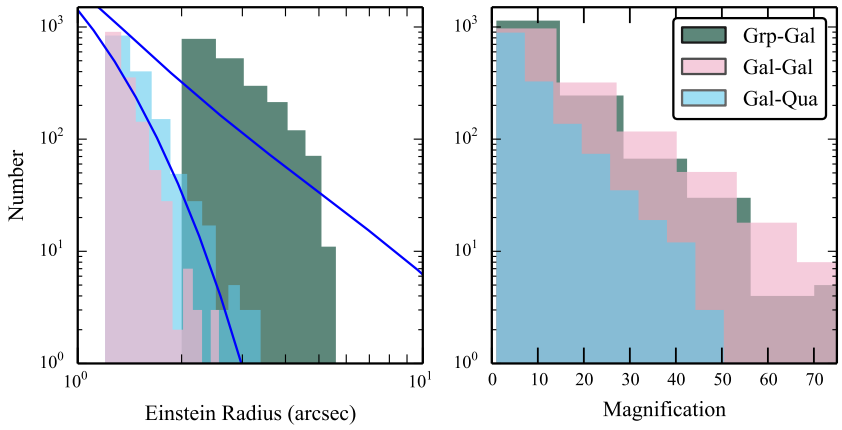

Figure 2. Einstein radius distribution for all types of lenses. The dashed-dotted (blue) curves show the theoretical prediction assuming an SIS model at galaxy-scales and a total (NFW+Hernquist) model at groups-scales taken from (More et al. 2012).

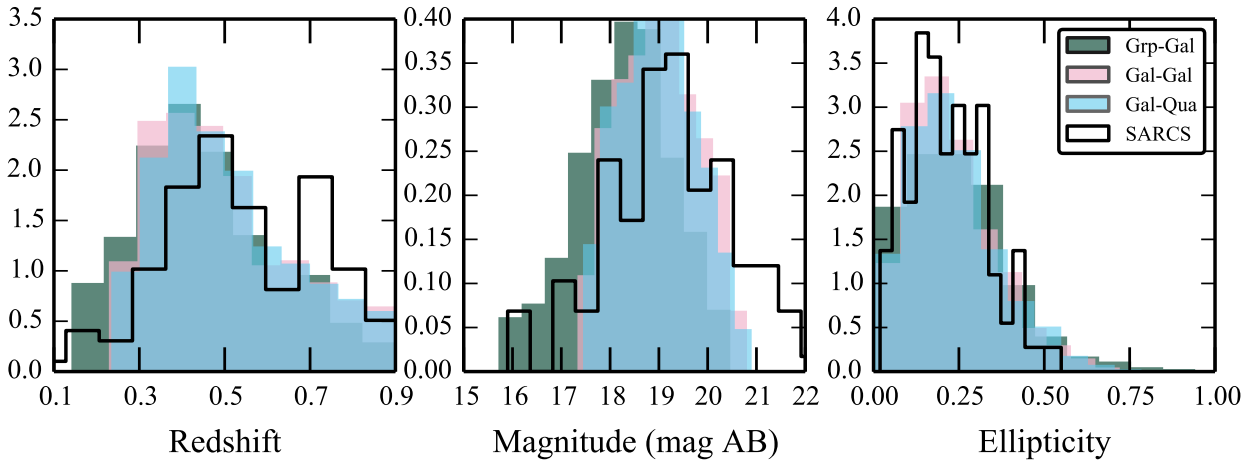

Figure 3. Distributions of properties of the lensing galaxies of the simulated sample compared to the known lens sample SARCS XXXX check?

5 RESULTS

Table 2: Thresholds used in the selection of the simulated lenses.

SW ID	Name	RA (deg)	Dec (deg)	$z_{\rm phot}$	m_i (mag)	$R_{\rm A}$ (")	R_{avg}	ZooID	P	Comments
SW1	CFHTLS J020338-051901	30.9097	-5.3171	0.4	18.8	1.6	2.0	ASW0001sqw	0.7	D,E
SW2	CFHTLS J020341-074722	30.9223	-7.7897	0.5	18.8	1.6	2.0	ASW000993q	1.0	A,E
SW3	CFHTLS J020457-110309	31.2392	-11.0526	0.7	20.0	1.6	2.3	ASW00099b9	1.0	A,R
SW4	CFHTLS J020642-095157	31.6751	-9.8659	–	–	0.9	2.0	ASW0001ld7	1.0	A,R
SW5	CFHTLS J020810-040220	32.0450	-4.0389	–	–	1.8	1.3	ASW0001c3j	1.0	A,R
SW6	CFHTLS J020832-043315	32.1339	-4.5544	–	–	1.6	2.3	ASW0002asp	1.0	A,R
SW7	CFHTLS J020848-042427	32.2011	-4.4076	–	–	1.1	2.7	ASW0002bmc	1.0	D,D
SW8	CFHTLS J020849-050429	32.2078	-5.0748	–	–	0.9	2.3	ASW0002qtn	1.0	A,R
SW9	CFHTLS J021021-093415	32.5898	-9.5711	0.4	18.4	2.7	1.3	ASW0002k40	0.5	D,S
SW10	CFHTLS J021057-084450	32.7415	-8.7474	–	–	2.5	1.7	ASW0002p8y	0.4	A,G
SW11	CFHTLS J021221-105251	33.0881	-10.8811	0.3	17.9	1.8	2.0	ASW0002dx7	1.0	D,E/S
SW12	CFHTLS J021225-085211	33.1051	-8.8697	0.8	19.5	2.1	1.7	ASW00024id	1.0	R,R
SW13	CFHTLS J021230-074727	33.1264	-7.7909	0.9	20.3	1.7	1.7	ASW0001ze0	0.5	A,R/G
SW14	CFHTLS J021317-084819	33.3234	-8.8055	0.5	19.8	1.3	1.3	ASW00024q6	0.6	A,R/E
SW15	CFHTLS J021514-092440	33.8109	-9.4111	0.7	19.9	2.6	1.3	ASW00021r0	0.5	A,R/G
SW16	CFHTLS J022016-102446	35.0688	-10.4129	0.6	19.4	1.5	2.3	ASW0009a4w	0.7	D,E
SW17	CFHTLS J022359-083651	35.9995	-8.6144	–	–	3.1	1.3	ASW0004iye	0.7	A,E
SW18	CFHTLS J022406-062846	36.0256	-6.4796	0.4	19.6	0.9	2.3	ASW0003wsu	1.0	A,E
SW19	CFHTLS J022409-105808	36.0398	-10.9689	–	–	4.8	3.0	ASW0004dv8	1.0	A,G
SW20	CFHTLS J022533-053204	36.3888	-5.5346	0.5	19.4	3.6	2.0	ASW0004m3x	0.7	A,R/G
SW21	CFHTLS J022716-105602	36.8186	-10.9341	0.4	17.3	1.8	2.0	ASW0009ab8	0.8	A,E/G
SW22	CFHTLS J022745-062518	36.9387	-6.4218	0.6	20.5	1.2	1.7	ASW0003s0m	0.7	A,R
SW23	CFHTLS J022805-051733	37.0236	-5.2927	0.4	18.8	1.4	2.3	ASW0009ans	1.0	Q,E
SW24	CFHTLS J022817-080242	37.0727	-8.0452	0.5	19.6	0.9	2.0	ASW00042d9	1.0	A,E/R
SW25	CFHTLS J022843-063316	37.1794	-6.5547	0.5	19.1	1.8	1.3	ASW0003r6c	0.4	D/A,E
SW26	CFHTLS J023008-054038	37.5359	-5.6774	0.6	19.7	1.9	2.3	ASW0003r61	0.6	A,E
SW27	CFHTLS J023010-110409	37.5420	-11.0694	0.9	20.6	1.2	1.7	ASW0004fgb	0.7	A,R
SW28	CFHTLS J023051-082423	37.7141	-8.4064	–	–	0.8	2.3	ASW000412m	0.6	A,E
SW29	CFHTLS J023123-082535	37.8468	-8.4266	–	–	1.2	2.3	ASW0004xjk	0.4	A,R
SW30	CFHTLS J023315-042243	38.3133	-4.3789	0.7	19.7	1.0	2.0	ASW00050sk	1.0	A,R
SW31	CFHTLS J023325-053104	38.3547	-5.5178	0.5	18.8	1.3	1.7	ASW0005ire	1.0	Q,M
SW32	CFHTLS J023453-093032	38.7232	-9.5089	0.5	19.8	0.7	1.7	ASW00051ld	0.5	A,D
SW33	CFHTLS J084833-044051	132.1385	-4.6809	0.7	20.2	0.9	1.7	ASW0004wgd	0.9	A,R
SW34	CFHTLS J084841-045237	132.1708	-4.8772	0.3	19.0	1.0	2.3	ASW0004nan	1.0	A,E
SW35	CFHTLS J084941-051650	132.4216	-5.2808	0.4	19.1	1.5	1.3	ASW0004nh3	0.6	A,E/S
SW36	CFHTLS J085135-052232	132.8980	-5.3756	–	–	1.9	1.3	ASW0004n2x	0.7	D,R
SW37	CFHTLS J085317-020312	133.3233	-2.0535	0.7	20.6	1.2	1.3	ASW0000vtc	0.7	A,R
SW38	CFHTLS J090218-053924	135.5790	-5.6567	–	–	2.0	1.3	ASW0000g95	1.0	A,R/E
SW39	CFHTLS J090248-010232	135.7020	-1.0424	0.4	19.1	1.4	1.7	ASW000096t	0.8	D,E
SW40	CFHTLS J090308-043252	135.7840	-4.5479	–	–	1.2	2.0	ASW00007mq	0.8	A,E
SW41	CFHTLS J090319-040146	135.8311	-4.0297	–	19.8	1.2	1.7	ASW00007ls	0.6	A,R/E
SW42	CFHTLS J090333-005829	135.8890	-0.9749	–	–	2.1	1.7	ASW00008a0	1.0	A/D,E/G
SW43	CFHTLS J135724+561614	209.3540	56.2707	–	–	2.6	1.3	ASW0006e0o	1.0	D,E
SW44	CFHTLS J135755+571722	209.4827	57.2897	0.8	20.2	1.3	2.0	ASW0005ma2	1.0	D,D
SW45	CFHTLS J140027+541028	210.1160	54.1745	–	–	1.2	1.3	ASW0006a07	0.8	Q,R/E
SW46	CFHTLS J140030+574437	210.1260	57.7437	0.4	18.2	2.0	2.3	ASW0009bp2	0.9	A,E
SW47	CFHTLS J140425+520506	211.1062	52.0850	0.4	18.9	1.4	1.3	ASW0005o0w	0.9	D,E
SW48	CFHTLS J140522+574333	211.3426	57.7259	0.7	19.7	1.0	2.7	ASW000619d	1.0	A,R
SW49	CFHTLS J140622+520942	211.5958	52.1617	0.7	20.3	1.2	2.3	ASW0005rnb	1.0	A,R
SW50	CFHTLS J140845+514913	212.1907	51.8205	0.7	19.7	1.5	2.7	ASW0005o38	1.0	A,E
SW51	CFHTLS J141056+533225	212.7364	53.5405	0.6	19.4	1.7	2.0	ASW0006kjx	0.6	A,R/G
SW52	CFHTLS J141432+534004	213.6372	53.6679	0.7	21.4	0.9	2.0	ASW0006jh5	1.0	A,R
SW53	CFHTLS J141448+545548	213.7010	54.9301	0.4	18.3	2.1	1.3	ASW0006zc9	0.4	A,G
SW54	CFHTLS J141518+513915	213.8290	51.6542	0.4	18.3	3.0	1.7	ASW00070vl	1.0	D,E
SW55	CFHTLS J141927+533919	214.8627	53.6554	0.7	20.5	2.3	1.7	ASW0006yrw	0.5	A,R/G
SW56	CFHTLS J142432+550019	216.1354	55.0055	0.5	19.5	1.1	2.0	ASW0007vx2	1.0	A,E
SW57	CFHTLS J142603+511421	216.5140	51.2393	–	–	4.4	2.7	ASW0006mea	1.0	A,G
SW58	CFHTLS J142620+561356	216.5870	56.2323	0.4	19.5	1.3	1.3	ASW0007sez	1.0	A/R,S

SW ID	Name	RA (deg)	Dec (deg)	$z_{\rm phot}$	m_i (mag)	$R_{\rm A}$ (")	R$_{avg}$	ZooID	P	Comments
SW59	CFHTLS J142652+560001	216.7200	56.0004	–	–	1.5	1.7	ASW0007t5y	1.0	R,R
SW60	CFHTLS J142843+543713	217.1815	54.6204	0.4	19.7	1.3	1.3	ASW0007pga	0.9	D,D
SW61	CFHTLS J142934+562541	217.3926	56.4281	0.5	19.0	5.9	2.7	ASW0009cjs	1.0	A,G
SW62	CFHTLS J143055+572431	217.7333	57.4088	0.7	19.3	1.0	2.0	ASW0007wfj	1.0	A,R
SW63	CFHTLS J143100+564603	217.7510	56.7675	–	–	1.8	1.7	ASW00086xq	1.0	A,E
SW64	CFHTLS J143353+542310	218.4736	54.3862	0.8	19.8	1.6	2.0	ASW0009cox	0.9	A,R/G
SW65	CFHTLS J143454+522850	218.7270	52.4808	0.6	19.4	4.4	2.3	ASW0007k4r	0.4	Q,G/R
SW66	CFHTLS J143627+563832	219.1164	56.6425	0.5	19.5	1.5	2.3	ASW0008swn	1.0	A,D
SW67	CFHTLS J143631+571131	219.1315	57.1922	0.7	20.9	1.3	1.3	ASW0008pag	0.8	D/A,R
SW68	CFHTLS J143651+530705	219.2150	53.1183	0.6	19.2	3.1	1.3	ASW0007h27	1.0	A,E/G
SW69	CFHTLS J143658+533807	219.2425	53.6355	0.7	19.6	0.9	2.7	ASW0007hu2	0.9	D,D
SW70	CFHTLS J143838+572647	219.6589	57.4464	0.8	20.2	1.1	2.0	ASW0008qsm	1.0	A,R
SW71	CFHTLS J143950+544606	219.9610	54.7686	–	–	1.7	1.3	ASW00085cp	0.6	A,G/R
SW72	CFHTLS J220215+012124	330.5635	1.3567	0.3	17.4	4.6	1.7	ASW0005qiz	0.5	rA,G
SW73	CFHTLS J220256+023432	330.7370	2.5758	–	–	6.8	2.3	ASW0007e08	1.0	A,G/C
SW74	CFHTLS J220722+013610	331.8441	1.6030	0.8	20.3	1.0	2.0	ASW0001uuz	0.6	A,R
SW75	CFHTLS J221101+003401	332.7582	0.5671	0.9	20.0	1.1	1.3	ASW00090l5	1.0	A,E/R
SW76	CFHTLS J221306+014708	333.2758	1.7856	–	17.1	1.4	1.7	ASW0008wmr	1.0	A,S
SW77	CFHTLS J221513+010240	333.8060	1.0445	–	–	0.8	1.7	ASW0008dxh	0.6	A,R/G
SW78	CFHTLS J221519+005758	333.8321	0.9661	0.4	20.2	1.0	1.7	ASW0008xbu	1.0	A,D
SW79	CFHTLS J221716+015826	334.3189	1.9739	0.1	21.6	1.0	1.7	ASW00096rm	1.0	A/R,R
SW80	CFHTLS J222007-002505	335.0307	-0.4182	0.3	21.4	1.2	2.0	ASW00094fq	1.0	Q,R/G

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.

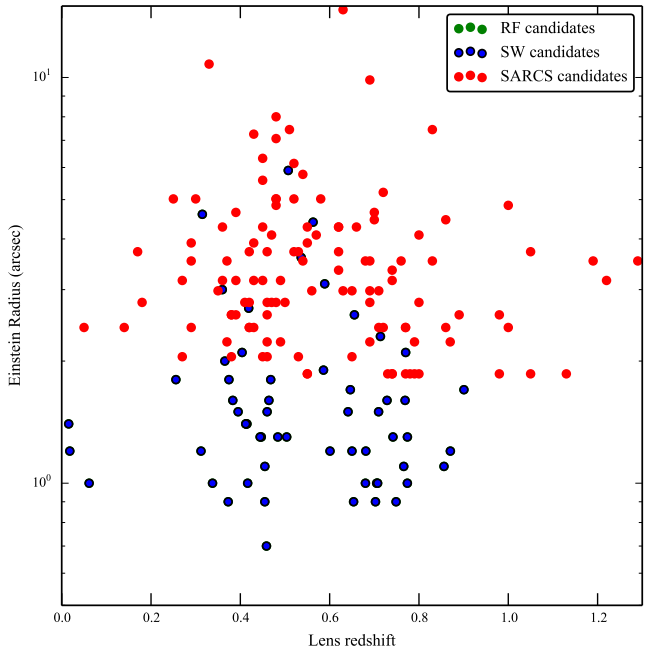

Figure 6. Comparison of Lens redshift and Arc radius for all three lens samples, namely, from Space Warps, SARCS and those from the RingFinder.

5.1 New lens candidates from Space Warps

Describe the newly found lenses and explain for a few cases why they were missed

5.2 Recovery of known CFHTLS lenses with Space Warps

Describe which lenses were recovered and explain any cases missed by citizen scientists

5.3 CFHTLS completeness and purity

Rejection rate. Completeness and purity at P ¿ retirement, P ¿ 95%, and as function of probability P.

Summarize performance at some fiducial threshold: eg P = 95%.

5.4 Stage 1 Classification

5.5 Stage 2 Classification

5.6 Comparison with "Expert" Classification

6 DISCUSSION

Differences with robots: what types of lenses are found at Space Warps?

Selection function. Missing system.

Further work.

Describe the importance of this kind of study - being able to quantify the completeness of the lens samples - state how this will be used in a future Space Warps paper

TBD: add section on offline analysis - extra lenses found etc; expand FP section; mention EGS moustakas and other lens samples; comment on any lensed quasars, on any exotic lenses; check if some candidates were detected because they were hidden underneath the sims ie. from the D11;

We looked at the locations of simulated and real lenses from our data those were missed by Space Warps compared to the locations of the lenses that got detected. The real lens sample consists of a total of 383 candidates which have $P_l > 2.e-5$ and received a rank of $R_{avg} > 0$ from the expert. We do not find any obvious dependency in the rate of detections as a function of the position of a given lens for both simulated and real lens sample. Thus, the completeness of the lens sample is not affected by whether a lens is located close to the border or well within the center (see Figure 7).

6.1 Limitations and Caveats of the training sample

The simulated sample has certain limitations due to lack of our understanding of various phenonmena in the Universe and due to uncertainties in various parameters of our model. Here, we describe some of the cases or aspects in which the simulations are known to have failed or seem unrealistic.

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 measurements. This causes relatively larger uncertainties in its luminosity and velocity dispersion and leads to simulated lenses which look implausible. For example, due to a larger uncertainty in the velocity dispersion of the lensing galaxy, the lensed images may have larger image separation than what is expected given the visual priors from the galaxy.

At group-scales, the photometric and redshift estimates are used when defining 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 are incorrectly assigned high redshift. Spiral galaxies are typically less massive and low redshift spiral galaxies are unlikely to act as gravitational lenses. Hence, the resulting simulated lenses are not convincing.

We use 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.

7 SUMMARY AND CONCLUSIONS

In this paper, we describe the framework and procedure used to generate simulated lens sample for the blind lens search in the CFHTLS survey in collaboration with Space Warps. The aim of this lens search is to find lenses that have been missed by lens finding algorithms. The simulated lens sample is used for training the citizen scientists, calibrating their performance and rejecting unlikely lenses from the sample

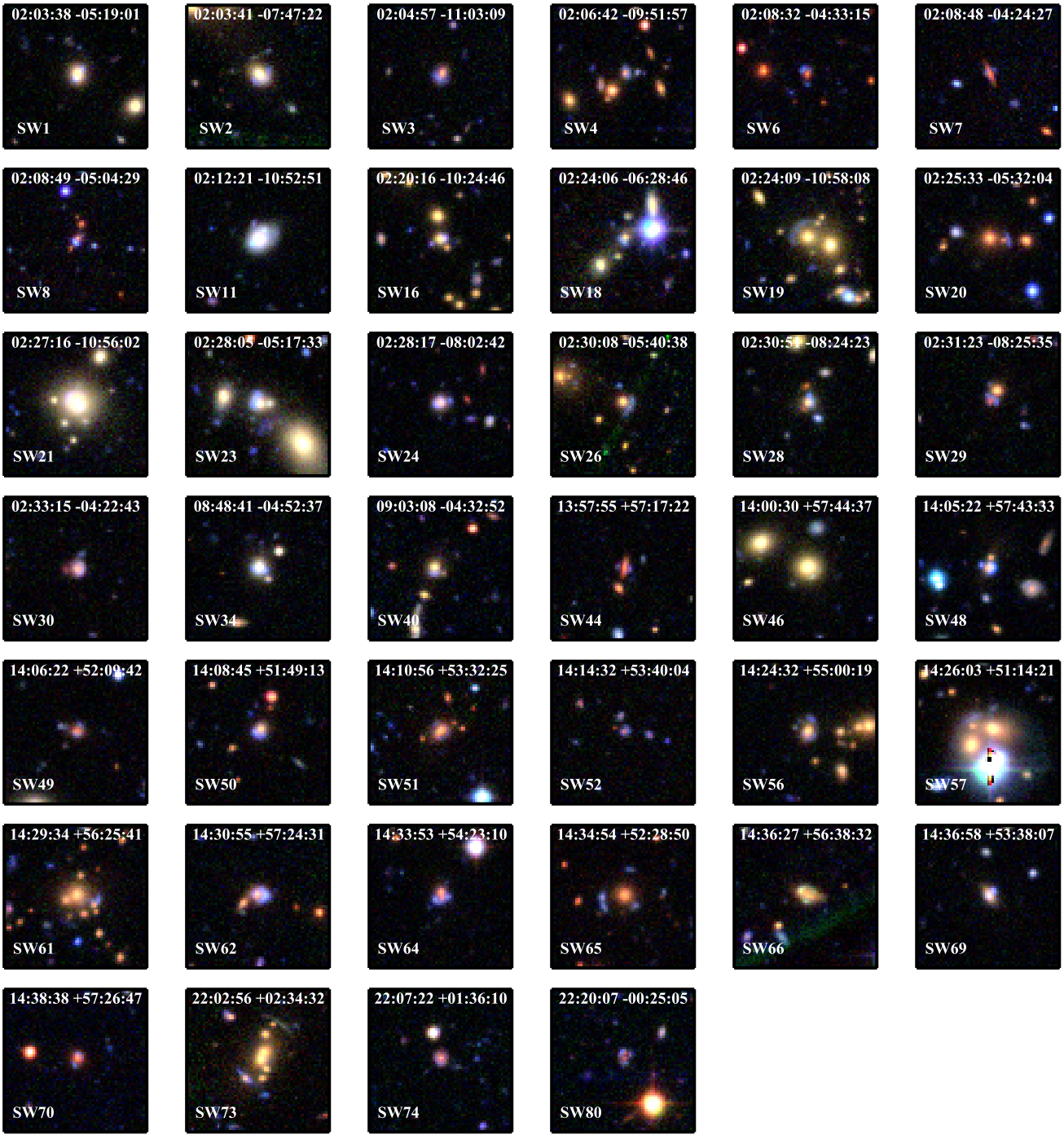

Figure 4. Sample of high probability lens candidates ($R_{avg} >= 2.0$).

based on the classifications of citizen scientists. As a result, the simulated lenses need to consist of realistic looking lenses.

We use the photometric and redshift catalogues for the foreground galaxies and additionally, color catalogues for background galaxies and quasars. We use scaling relations to relate light properties to properties such as mass and size to generate lens models. We further account for the instrumental effects such as seeing and noise before creating the final lensed images of model background sources. We add the lensed images on top of the real galaxies and groups in the CFHTLS data in all filters.

We draw the following conclusions:

• Crowd-sourced gravitational lens detection works, as shown in by comparing with real lenses in CFHTLS: Real (robotically-detected and expert-confirmed) lenses are recovered/missed at similar/comparable/different rates C%

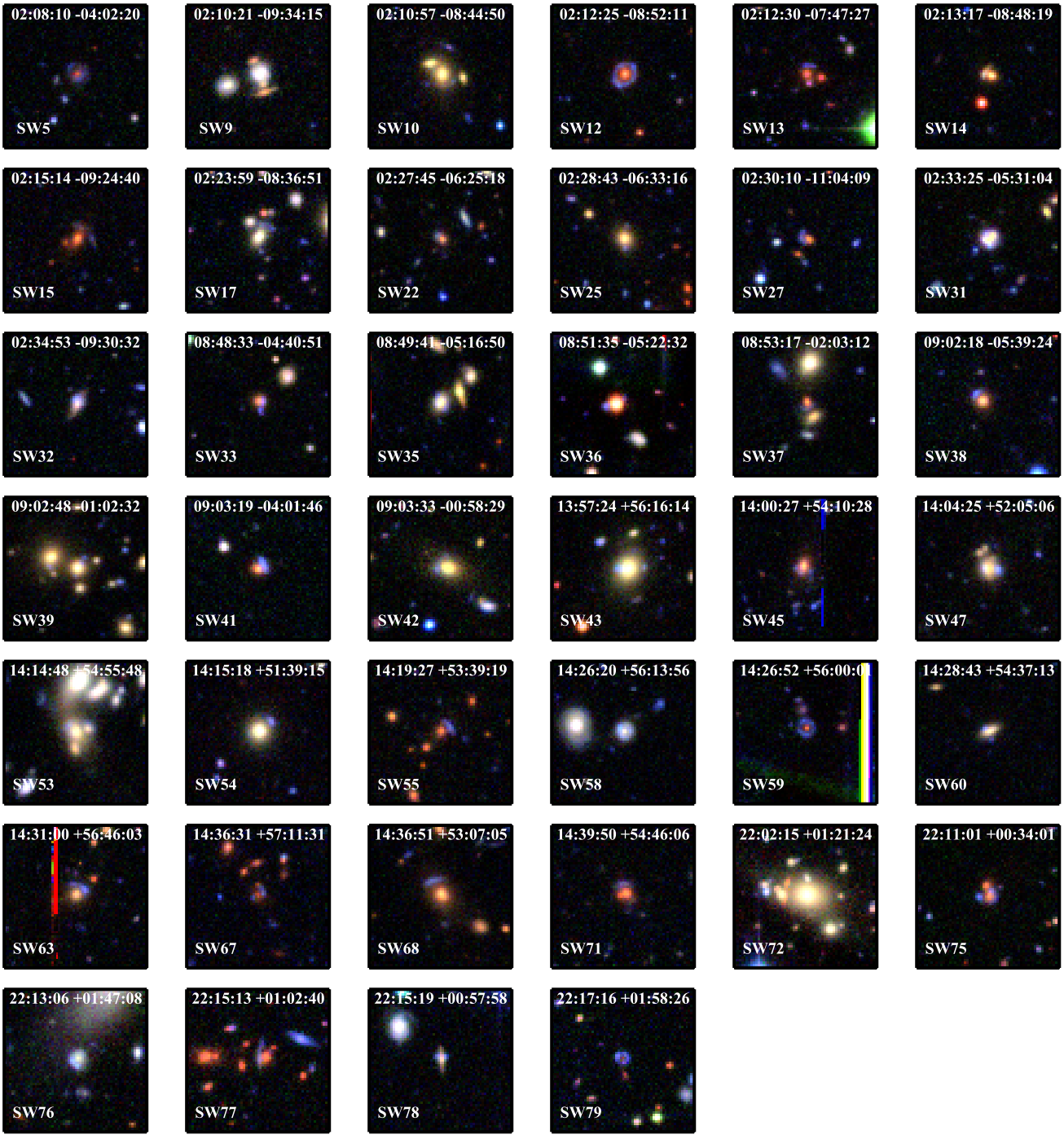

Figure 5. Sample of medium probability lens candidates ($1.0 < R_{avg} < 2.0$).

and P% - this is a partial test of supervised vs unsupervised learning

- We found a sample of new gravitational lens candidates. An expert-graded sample of 74 with 44 promising and 30 low probability candiates.

The SW lenses are different from the robotically (RingFinder and ArcFinder) detected lenses, in the following ways. XXXXX

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 citi-

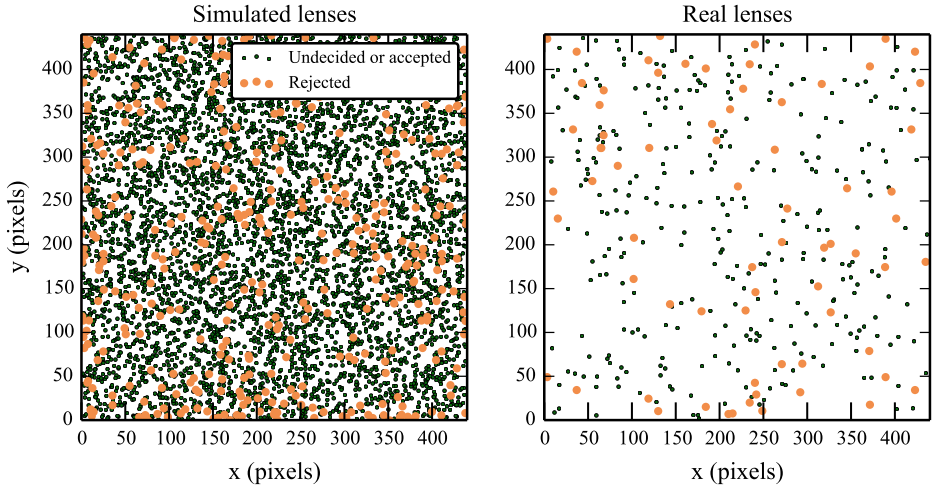

Figure 7. Completeness as a function of location of lenses. Simulated lenses (left) and real lens candidate (right) are shown. **: TBD - Add a third color for P> Pthresh which will be defn of "accepted"** Irrespective of whether the lenses are detected or rejected and whether they are simulated or real, there is no obvious dependency on where the lenses are located.

zen 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. This work was supported by World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan.

REFERENCES

Bernardi, M., et al. 2003, AJ, 125, 1866
Erben, T., et al. 2013, MNRAS, 433, 2545
Faber, S. M., et al. 2007, ApJ, 665, 265
Faure, C., et al. 2009, ApJ, 695, 1233
Frei, Z., & Gunn, J. E. 1994, AJ, 108, 1476
Hildebrandt, H., et al. 2012, MNRAS, 421, 2355
Keeton, C. R., Christlein, D., & Zabludoff, A. I. 2000, ApJ, 545, 129
More, A., Cabanac, R., More, S., Alard, C., Limousin, M., Kneib, J.-P., Gavazzi, R., & Motta, V. 2012, ApJ, 749, 38
Oguri, M. 2006, MNRAS, 367, 1241
Oguri, M., & Marshall, P. J. 2010, MNRAS, 405, 2579
Pâris, I., et al. 2012, A&A, 548, A66
Richards, G. T., et al. 2006, AJ, 131, 2766
Sygnet, J. F., Tu, H., Fort, B., & Gavazzi, R. 2010, A&A, 517, A25

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