A&A 653, L6 (2021)
https://doi.org/10.1051/0004-6361/202141758
© R. Cañameras et al. 2021

Astronomy & Astrophysics

LETTER TO THE EDITOR

HOLISMOKES

VI. New galaxy-scale strong lens candidates from the HSC-SSP imaging survey*

R. Cañameras[1], S. Schuldt[1,2], Y. Shu[1,3], S. H. Suyu[1,2,4], S. Taubenberger[1], T. Meinhardt[5], L. Leal-Taixé[5], D. C.-Y. Chao[4], K. T. Inoue[6], A. T. Jaelani[7], and A. More[8,9]

[1] Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany
e-mail: rcanameras@mpa-garching.mpg.de
[2] Technische Universität München, Physik Department, James-Franck Str. 1, 85741 Garching, Germany
[3] Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing, 44780 Bochum, Germany
[4] Institute of Astronomy and Astrophysics, Academia Sinica, 11F of ASMAB, No.1, Section 4, Roosevelt Road, Taipei 10617, Taiwan
[5] Technical University of Munich, Department of Informatics, Boltzmann-Str. 3, 85748 Garching, Germany
[6] Faculty of Science and Engineering, Kindai University, Higashi-Osaka 577-8502, Japan
[7] Astronomy Research Group and Bosscha Observatory, FMIPA, Institut Teknologi Bandung, Jl. Ganesha 10, Bandung 40132, Indonesia
[8] Kavli Institute for the Physics and Mathematics of the Universe (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan
[9] The Inter-University Centre for Astronomy and Astrophysics (IUCAA), Post Bag 4, Ganeshkhind, Pune 411007, India

Received 9 July 2021 / Accepted 3 August 2021

ABSTRACT

We have carried out a systematic search for galaxy-scale strong lenses in multiband imaging from the Hyper Suprime-Cam (HSC) survey. Our automated pipeline, based on realistic strong-lens simulations, deep neural network classification, and visual inspection, is aimed at efficiently selecting systems with wide image separations (Einstein radii $\theta_E \sim 1.0$–$3.0''$), intermediate redshift lenses ($z \sim 0.4$–0.7), and bright arcs for galaxy evolution and cosmology. We classified *gri* images of all 62.5 million galaxies in HSC Wide with *i*-band Kron radius $\geq 0.8''$ to avoid strict preselections and to prepare for the upcoming era of deep, wide-scale imaging surveys with Euclid and Rubin Observatory. We obtained 206 newly-discovered candidates classified as definite or probable lenses with either spatially-resolved multiple images or extended, distorted arcs. In addition, we found 88 high-quality candidates that were assigned lower confidence in previous HSC searches, and we recovered 173 known systems in the literature. These results demonstrate that, aided by limited human input, deep learning pipelines with false positive rates as low as $\simeq 0.01\%$ can be very powerful tools for identifying the rare strong lenses from large catalogs, and can also largely extend the samples found by traditional algorithms. We provide a ranked list of candidates for future spectroscopic confirmation.

Key words. gravitational lensing: strong – methods: data analysis

1. Introduction

Strong gravitational lensing systems are very powerful tools for probing galaxy evolution and cosmology. They provide constraints to the level of a few percent on the total mass of the foreground galaxies or galaxy clusters producing the light deflections (e.g., Bolton et al. 2008; Shu et al. 2017; Caminha et al. 2019). This leads to unique diagnostics on the dark matter mass distributions, and to test galaxy evolution models and the flat Lambda cold dark matter (ΛCDM) cosmological model. Moreover, strongly lensed time-variable sources with observed time delays between multiple images provide independent and competitive measurements of the Hubble constant H_0 (e.g., Refsdal 1964; Wong et al. 2020). However, identifying statistical samples of strong lenses remains a major challenge.

Convolutional neural networks (CNNs; LeCun et al. 1998) have proven extremely efficient for pattern recognition tasks and have given a strong impetus to image analysis and processing. Recent studies largely demonstrate the ability of supervised CNNs to identify the rare gravitational lenses among large datasets (e.g., Jacobs et al. 2017, 2019; Petrillo et al. 2019; Huang et al. 2021), extending previous automated algorithms (e.g., Gavazzi et al. 2014; Joseph et al. 2014) generally with better classification performance (Metcalf et al. 2019). In Cañameras et al. (2020, hereafter C20), we show that realistic simulations and careful selection of negative examples are crucial for successfully conducting a systematic search over 30 000 deg^2 with PanSTARRS multiband imaging.

We develop here new supervised neural networks for automated selection of galaxy-scale strong lenses in large-scale

* Full Table 1 is only available at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/653/L6

Open Access article, published by EDP Sciences, under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Open Access funding provided by Max Planck Society.

multiband surveys, using the Hyper-Suprime Cam Subaru Strategic Program (HSC-SSP; Aihara et al. 2018) for testing on images approaching the expected depth and quality of the Rubin Observatory Legacy Survey of Space and Time (LSST) final stacks (see Ivezić et al. 2019). Previous non-machine learning identification of galaxy-, group-, and cluster-scale lenses from the Survey of Gravitationally lensed Objects in HSC Imaging (SuGOHI; Sonnenfeld et al. 2018, 2019, 2020; Wong et al. 2018; Chan et al. 2020; Jaelani et al. 2020, 2021, hereafter S18, S20, W18 for SuGOHI-g) offers an independent observational set to test our classification completeness. In this Letter we validate our deep learning pipeline and present new high-confidence galaxy-scale lens candidates with wide image separations from the Wide layer of the HSC survey, which is only ≃1 mag shallower than LSST ten-year stacks, and has sufficient sky coverage for lens searches. In Sect. 2 we describe the neural network and the datasets used for training, validating, and testing the network, and for searching for new lenses. The classification procedure is presented in Sect. 3, and the results are discussed in Sect. 4.

2. Methodology

We conducted our search on *gri* cutouts from HSC Wide public data release 2 (PDR2) covering nearly 800 deg^2 (Aihara et al. 2019) in all three bands, out of the final 1400 deg^2. In PDR2, about 300 deg^2 reach the nominal 5σ point-source sensitivities of 26.8, 26.4, and 26.2 mag respectively in g, r, and i. We focused on systems with luminous red galaxies (LRGs) acting as lenses and with image separations $\gtrsim$0.75″, larger than the median seeing FWHMs in g, r, and i bands. Such systems are ideal for constraining the lens mass-density profiles and for finding strongly lensed supernovae for cosmography and stellar physics, which are major goals of our ongoing Highly Optimized Lensing Investigations of Supernovae, Microlensing Objects, and Kinematics of Ellipticals and Spirals (HOLISMOKES; Suyu et al. 2020). To go beyond previous studies that relied on strict catalog preselections, we demonstrate that a dedicated neural network trained on a carefully constructed dataset can automatically and efficiently identify such lenses over an extended galaxy sample. We focused on the 62.5 million galaxies observed in *gri* bands in PDR2, without flagged artifacts, and with i-band Kron radius $\geq$0.8″. We used 12″ × 12″ cutouts, sufficient for galaxy-scale lenses, downloaded from the data archive server (Bosch et al. 2018). The design of the dataset and choice of network architecture resulted from thorough tests of classification completeness and purity, using the test set described in Sect. 2.3. The performance of our different networks will be compared in a future paper (Cañameras et al., in prep.).

2.1. Constructing the ground truth dataset

Supervised machine learning classification depends strongly on the construction of the ground truth data used by the network to learn the morphological features relevant to each class. We trained and validated our binary classification network with a balanced set of 40 000 positive and 40 000 negative examples (Fig. 1) obtained from random sky positions to limit biases from small-scale seeing and depth variations. The GAMA09H field was excluded and reserved for a future comparison study of various lens search pipelines (More et al., in prep.).

As positive examples, we produced realistic galaxy-scale lens simulations by painting lensed arcs on HSC *gri* images of LRGs. This approach accounts for the quality of HSC imaging and for the presence of artifacts and neighboring galaxies.

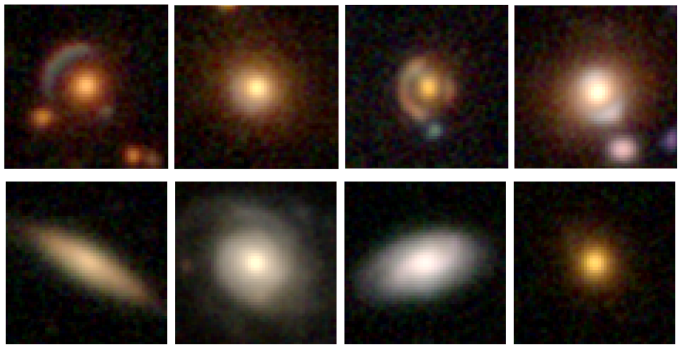

Fig. 1. Examples of positives (mock lenses, *top row*) and negatives (spirals, LRGs, and random nonlenses, *bottom row*) in the ground truth dataset for training and validation. Each cutout is 12″ × 12″.

We followed the procedure described in Schuldt et al. (2021a) and C20, by modeling the lens mass distributions with Singular Isothermal Ellipsoids (SIE) using LRG redshifts and velocity dispersions from SDSS, and inferring axis ratios and position angles from the light profiles. Unlike in C20 we included external shear, and we chose lens–source pairs to produce a uniform Einstein radius distribution in the range 0.75″−2.5″, increasing the number of wide separations and of fainter ($z > 0.7$) lens galaxies to help recover these configurations. As background sources, we used high S/N galaxies with spectroscopic redshifts from the *Hubble* Ultra Deep Field (Inami et al. 2017), applying color corrections that match HST filter passbands to HSC, and a common flux boost to the three bands. Sources were lensed with GLEE (Suyu & Halkola 2010; Suyu et al. 2012), convolved with the PSF model at the location of the lens from the HSC archive, and coadded with the lens HSC cutout. Mocks that have lensed images with $\mu \geq 5$, $S/N > 5$, and that are brighter than the lens at the position of peak lensed image emissions are accepted by the pipeline. We used similar fractions of quadruply- and doubly-imaged systems.

As negative examples, we selected a sample of spirals, isolated LRGs, and random galaxies with $r_{\rm Kron} < 23$ mag in similar proportions, and a few compact galaxy groups. We obtained spirals with Kron radius <2″ from the catalog of Tadaki et al. (2020, also from HSC Wide) in order to boost the fraction of examples mimicking lensed arcs. Isolated LRGs helped the network to learn that lensed arcs are the relevant features, and groups were selected from Wen et al. (2012). Other types frequently misclassified as lenses (e.g., rings, mergers) were more difficult to include due to limited morphological classifications available in the HSC footprint (e.g., Willett et al. 2013).

2.2. Training the neural network

Building upon the success of CNNs, deeper architectures have been developed to optimize performance such as image classification accuracies. In particular, the residual learning concept (ResNet; He et al. 2016a) enables one to increase the network depth and performance without requiring drastic computing resources. Such ResNets have obtained excellent results on the ImageNet Large Scale Visual Recognition Challenge 2015 (He et al. 2016a). They resemble deep CNNs with multiple building blocks (preactivated bottleneck residual units in He et al. 2016b), and shortcut connections between these blocks that make the convolutional layers learn residual functions with respect to the previous layer, and help avoid vanishing gradients

during optimization. In the recent past, Lanusse et al. (2018) have developed ResNet architectures for lens finding on LSST-like simulations, and obtained better results than classical CNNs on the strong lens finding challenge (Metcalf et al. 2019). Subsequent studies confirm that such ResNets can efficiently select lenses on real survey data (e.g., Li et al. 2020; Huang et al. 2021).

We used a ResNet adapted from the ResNet-18 architecture (He et al. 2016a) which provides a good trade-off between performance and total training time for binary classification in ground-based imaging. The network has a total of 18 layers with eight blocks comprising two convolutional layers with batch normalization and nonlinear ReLU activations. We added a fully connected layer of 16 neurons before the last single-neuron layer with sigmoid activation that outputs a score p.

As data augmentation to prevent overfitting and improve generalization, the image centroids were randomly shifted between −5 and +5 pixels, negative pixels were clipped to zero, and square root stretch was applied to boost low-luminosity features. Other techniques such as image normalization did not improve the performance and were thus not used. The dataset was split into 80% for training and 20% for validation and, after randomly initializing weights, the ResNet was trained over 100 epochs using mini-batch stochastic gradient descent with 128 images per batch, a learning rate of 0.0006, a weight decay of 0.001, and a momentum fixed to 0.9. We used early stopping by saving the final network at epoch 21 that corresponds to the minimal binary cross-entropy loss in the validation set without overfitting.

2.3. Testing the performance

We used HSC Wide PDR2 images to design a test set closely representative of the overall search sample. First, the completeness was measured on SuGOHI galaxy-scale lenses that are spectroscopically confirmed or have grades A or B (S18; W18; S20). We visually rejected a few lenses with large image separations $\gtrsim 4''$ suggesting major perturbation from the lens environment as we do not intend to recover such configurations. Out of the 220 SuGOHI systems remaining, 202 match our Kron radius $\geq 0.8''$ threshold and were kept as test lenses. Second, the expected rate of false positives was automatically measured with a set of nonlens galaxies representative of our overall search sample, including observational artifacts and various types of interlopers. We collected nonlenses in the COSMOS field (Scoville et al. 2007), excluding flagged HSC cutouts and sources with Kron radius lower than $0.8''$, as described above. We excluded all 130 strong lenses and lens candidates previously listed in the MasterLens database[1], or in Faure et al. (2008), Pourrahmani et al. (2018), and SuGOHI, assuming that the unparalleled coverage of COSMOS guarantees a nearly complete lens selection. We then classified *gri* images of the 91 000 remaining nonlenses.

As shown in the Receiver Operating Characteristic (ROC) curve (Fig. 2), our ResNet reaches extremely low false positive rates (FPRs) at least a factor of 10 lower than classical CNNs (C20). By drastically limiting the number of contaminants, this network saves significant human inspection time which makes it very promising for rapid lens finding in any deep, wide-scale imaging survey. We adopted a ResNet score threshold $p > 0.1$ for lens selection to maintain FPR $\lesssim 0.01\%$ with completeness >50% in SuGOHI (Fig. 2). A comprehensive discussion on classification accuracies as a function of galaxy properties will be presented in a future paper, together with our other networks.

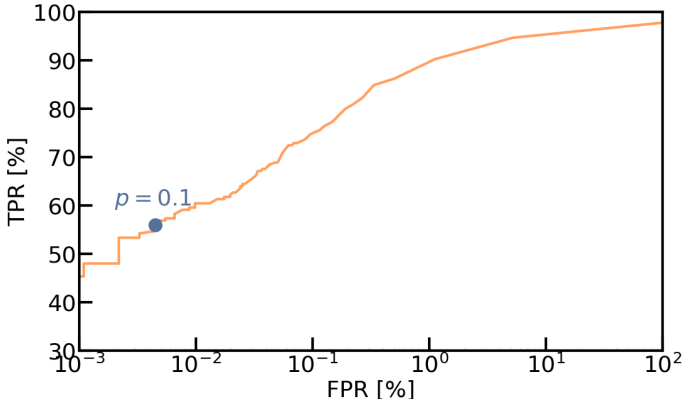

Fig. 2. Receiver Operating Characteristic curve for our ResNet using an independent test set from HSC Wide PDR2 survey data. Different threshold scores of the ResNet trace out the orange curve, and the threshold score of $p = 0.1$ is indicated by the blue dot. The true positive rate (TPR) corresponds to the number of SuGOHI galaxy-scale test lenses correctly classified over the 202 test lenses. The false positive rate (FPR) is measured using random nonlenses in the COSMOS field with Kron radius larger than $0.8''$ and is defined as the number of nonlenses identified as lenses over the total number of nonlenses.

Using 6000 COSMOS nonlenses with $r < 22$ and the 202 SuGOHI test lenses, we tested the stability of ResNet scores through few-pixel translations, $k \times \pi/2$ rotations, and flipping of the *gri* images. Applying 100 random transformations and computing the output distribution of p showed that predictions with mean $\mu_p < 0.1$ and >0.9 are systematically stable, with a scatter $\sigma_p < 0.05$. Galaxies with $p < 0.1$ discarded before visual inspection therefore have robust ResNet predictions. Scores with $\mu_p = 0.2$–0.8 have higher scatter $\sigma_p \simeq 0.05$–0.35.

3. The classification procedure

The trained ResNet was applied to the *gri* cutouts of all 62.5 million galaxies with Kron radius larger than $0.8''$ in order to estimate their score p. A few hundred cutouts with residual sky background due to imperfect subtraction or to nearby saturated stars were assigned high scores, and we automatically excluded these cutouts with SExtractor (Bertin & Arnouts 1996). This resulted in 9651 neural network candidates with $p > 0.1$, 0.015% of the input sample, including 114/202 (56%) galaxy-scale test lenses from SuGOHI. We qualitatively observe that the misclassified test lenses (see Fig. 3) tend to have either compact and fainter lens galaxies, lensed sources with redder colors, stronger blending with lens light, or lower source-to-lens flux ratios. Each of these configurations is less represented in our simulations. Our ResNet also recovers 102 group- and cluster-scale lens candidates, although it is not optimized for these systems.

The sample with $p > 0.1$ contains a large number of false positives, and we conducted a visual inspection stage to collect a final list of high-confidence lens candidates. Five authors (R. C., S. S., Y. S., S. H. S., and S. T.) inspected three-color images displayed with different scaling and contrasts, and assigned grades following explicit criteria described in S18 and C20. In short, grade 3 corresponds to unambiguous lenses with resolved multiple images, grade 2 corresponds to probable lenses with

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

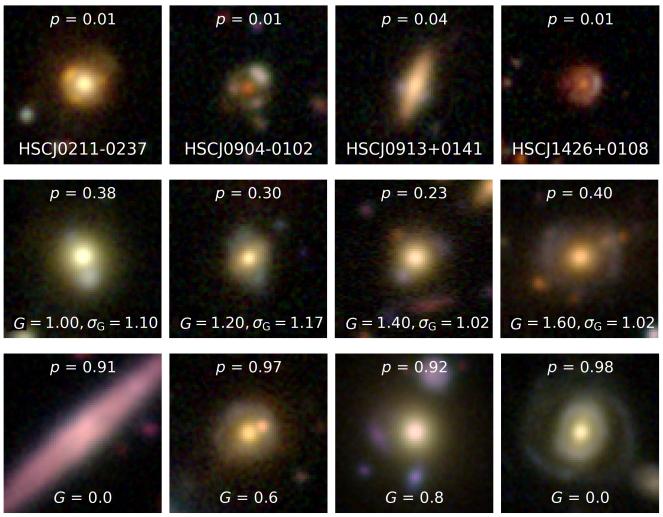

Fig. 3. Illustration of the network and visual classification stages. *Top row*: HSC three-color *gri* images of some SuGOHI galaxy-scale test lenses missed by the neural network, with scores $p < 0.1$. *Second row*: subset of ResNet candidates with elevated dispersion σ_G among our visual inspection grades. *Third row*: examples of interlopers with $p > 0.1$ and low average grades G, showing dust lanes or arc-like features around LRGs.

extended and distorted arcs but no obvious counter-image, grade 1 corresponds to possible lenses such as LRGs with a single, weakly distorted companion, and grade 0 includes obvious interlopers such as spirals, mergers, and galaxy groups. Blind tests using 70 galaxies with $p > 0.1$ and 30 SuGOHI lenses led to comparable average grades per classifier, 30% of cases with zero dispersion among our grades, and systematic recovery of known lenses, which illustrates the benefit of averaging individual grades and validates our approach.

While the network scores are not calibrated as probabilities, the fraction of contaminants clearly increases for lower scores. For this reason, the top 2092 candidates with $p > 0.2$ were directly graded by the five authors, while author R. C. excluded obvious interlopers from the 7559 candidates with $0.1 < p < 0.2$ and forwarded the 739 objects with grades ≥1 for inspection by the other authors. After this first iteration, we reinspected the 332 candidates with dispersion ≥0.75 among our five grades. The final grades have averages of 0.36–0.62 and dispersions of 0.80–0.85 per classifier and were not normalized. Cases with dispersed visual grades often show ambiguous blue arcs that could either be lensed arcs from background galaxies (without clear counter-images), spiral arms, or tidal features. The number of grade 3 lens candidates from each classifier spans a broad range between 69 and 204. Moreover, our inspection recovers 101/114 SuGOHI test lenses with $p > 0.1$.

Most contaminants turn out to be underrepresented in our training set and include edge-on spirals, spirals with diffuse or unresolved arms discarded from the "S/Z" classification of Tadaki et al. (2020), lenticular galaxies, and LRGs with dust lanes or with faint and unlensed companions. In the future, morphological classifications with unsupervised machine learning (e.g., Martin et al. 2020) or crowdsourcing will offer interesting avenues for collecting large samples of these galaxy types in the HSC Wide footprint, aiding the selection of negative examples for supervised lens searches. Image artifacts are already well represented in the training set and were better excluded.

4. Results and discussion

We used the average visual grades G among the five examiners to rank our final sample. We compiled a total of 88 grade A ($G \geq 2.5$) and 379 grade B ($1.5 \leq G < 2.5$) that have convincing lensing features, corresponding to ≃5% of network recommendations and to ≃0.6 candidate per deg^2 (close to expectations from simulations by Collett 2015). Our findings are summarized in Table 1. The purity, defined as the fraction of grades $G \geq 1.5$ among ResNet recommendations, decreases rapidly when lowering the threshold p. We estimate that 38% of the highest scores $0.9 < p < 1.0$ have $G \geq 1.5$, decreasing to 20% for $0.6 < p < 0.7$, 7% for $0.2 < p < 0.3$, and 3% for the lowest interval $0.1 < p < 0.2$.

To find duplicates, this list was cross-matched with our extended compilation of strong gravitational lenses previously published as confirmed systems or as candidates with confidence levels equivalent to our grades A and B (see C20). Given the dataset overlap, we extended our cross-match to the full SuGOHI database including grades A, B, and C. We also checked the SIMBAD Database[2] and the *Hubble* Source Catalog[3]. A total of 21/88 grade A and 185/379 grade B lens candidates are newly discovered, and our inspection increases confidence for 4/88 grade A and 84/379 grade B that were assigned grade C in SuGOHI. A subset of these 294 new high-quality candidates best suited for spectroscopic follow-up is shown in Fig. 4. References of lenses in the literature we recovered are listed in Table 1. We analyzed SDSS DR16 spectra available for a subset of candidates, and systematically found signatures of LRGs at intermediate redshift, but no robust confirmation of background lensed sources as they mostly fall outside the 2″ SDSS fibers.

Our independent selection has moderate overlap with galaxy-scale lenses in SuGOHI (similar to the comparison in KiDS/GAMA from Knabel et al. 2020). On the one hand, by relying on YattaLens, an algorithm combining lens light subtraction, arc-finding, and lens modeling (S18), SuGOHI could be more efficient at finding lenses with blended components than our analysis of brute *gri* cutouts. On the other hand, our approach classifies a large catalog from PDR2, while SuGOHI have either focused on spectroscopically confirmed LRGs (S18); (W18), or have used S17A release that covers a 35% smaller area with full-color full-depth imaging than PDR2 (S20). Our newly discovered candidates exhibit both extended arcs and simple double or quad configurations and are not drastically different from those in SuGOHI. Lensed sources mostly have blue colors, and our visual inspection tends to preferentially retain brighter sources. While the vast majority of lenses are isolated LRGs, small compact groups also contribute in a few cases. Quantitative properties from lens modeling will be presented in a forthcoming paper.

This analysis paves the way for limiting human inspection in future lens searches not only with LSST, but also with Euclid and Roman. Unsupervised machine learning has not yet reached the performance of supervised CNNs for lens search, but the results are promising, especially for identifying peculiar lens configurations that could be omitted in human-assisted training sets (Cheng et al. 2020). In the future, combining the two approaches could therefore help increase completeness and purity.

[2] http://simbad.u-strasbg.fr/simbad/sim-fcoo
[3] https://catalogs.mast.stsci.edu/hsc/

Table 1. High-confidence galaxy-scale lens candidates with ResNet scores >0.1, and average grades ≥1.5 from visual inspection.

Name	RA	Dec	p	G	σ_G	#$G = 3$	$g_{\rm Kron}$	$r_{\rm Kron}$	$i_{\rm Kron}$	z	Notes
HSCJ0102+0158	15.6597	1.98247	0.12	3.0	0.0	5	27.61	21.62	20.58	0.8669$^{(*)}$	(i)
HSCJ0157−0330	29.3812	−3.51603	0.16	3.0	0.0	5	21.9	21.02	20.14	0.6212$^{(*)}$	(a)
HSCJ0200−0344	30.1981	−3.73763	0.6	3.0	0.0	5	22.23	21.17	20.04	0.72	
HSCJ0232−0323	38.2078	−3.39058	0.19	3.0	0.0	5	20.89	19.37	18.62	0.42	(h) (e) (i)
HSCJ0236−0332	39.1554	−3.53893	1.0	3.0	0.0	5	20.48	19.09	18.54	0.2695$^{(*)}$	(a) (e)
HSCJ0238−0545	39.5741	−5.76545	0.72	3.0	0.0	5	20.75	19.96	19.18	0.5993$^{(*)}$	(a) (c)
HSCJ0850+0039	132.6942	0.65146	0.61	3.0	0.0	5	24.66	22.08	21.25	1.00	(d)
HSCJ0904−0059	136.0331	−0.99807	0.69	3.0	0.0	5	22.42	21.23	20.26	0.56	(k)

Notes. The full table is available at the CDS. Some systems in this table are also found in separate lens searches and cross-references will be added to the corresponding publications (Shu et al., in prep.; Jaelani et al., in prep.). Columns are: source name; right ascension; declination; output score from the ResNet; average visual grades from five authors; dispersion in the grades; number of classifiers assigning the highest grade of 3; *g*-, *r*-, and *i*-band Kron magnitudes from PDR2; CNN photometric redshift estimates from Schuldt et al. (2021b) or spectroscopic redshifts marked as $^{(*)}$ where available; references for systems previously published, either as spectroscopically-confirmed lenses or as grade A or B candidates.
References. (a) Sonnenfeld et al. (2018), (b) Wong et al. (2018), (c) Chan et al. (2020), (d) Sonnenfeld et al. (2020), (e) Jaelani et al. (2020), (f) Cañameras et al. (2020), (g) Huang et al. (2020), (h) Stark et al. (2013), (i) Jacobs et al. (2019), (j) Li et al. (2020), (k) More et al. (2012), (l) Petrillo et al. (2019), (m) Brownstein et al. (2012), (n) Gavazzi et al. (2014), (o) Diehl et al. (2017), (p) More et al. (2016), (q) Shu et al. (2016), (r) Jacobs et al. (2017), (s) Ratnatunga et al. (1995), (t) Tanaka et al. (2016), and (u) More et al. (2017). Candidates marked with a † have a lower grade C in SuGOHI.

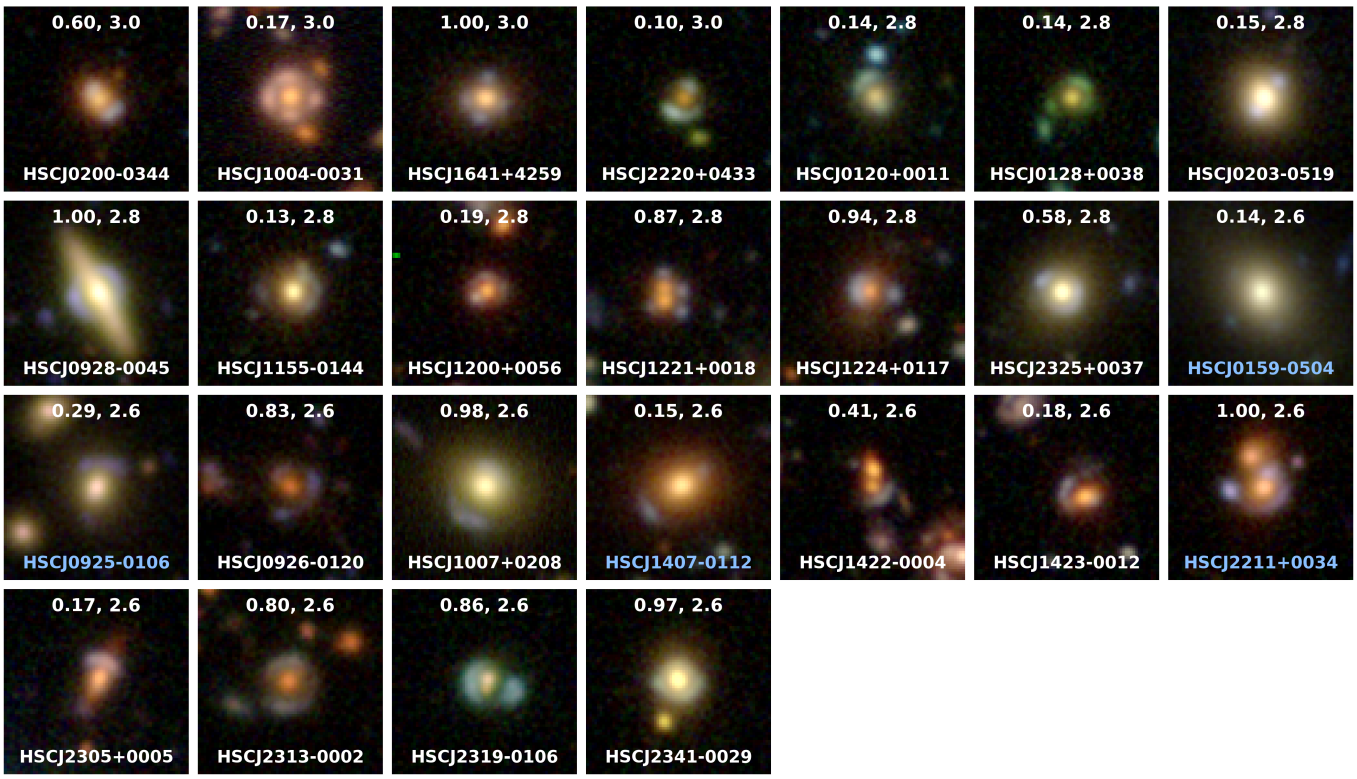

Fig. 4. Postage stamps ($12'' \times 12''$) of grade A lens candidates we have discovered in the HSC Wide survey, using *gri* multiband imaging. At the top of each panel we list the ResNet scores p, and the average grades G from visual inspection of scores $p > 0.1$. Grade A corresponds to $G \geq 2.5$. Candidates with white labels are newly discovered as they are not part of our compilation of previously confirmed strong lenses and grade A or B lens candidates in the literature. Those marked in light blue are listed as grade C in SuGOHI and obtained higher confidence in our classification.

Acknowledgements. We would like to thank the referee for useful comments that helped improve the paper. We thank D. Sluse for useful feedback about this work. RC, SS and SHS thank the Max Planck Society for support through the Max Planck Research Group for SHS. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (LENSNOVA: grant agreement No 771776). YS acknowledges support from the Max Planck Society and the Alexander von Humboldt Foundation in the framework of the Max Planck– Humboldt Research Award endowed by the Federal Ministry of Education and Research. Also supported by Riset ITB 2021 (ATJ). This paper is based on data collected at the Subaru Telescope and retrieved from the HSC data archive system, which is operated by Subaru Telescope and Astronomy Data Center at National Astronomical Observatory of Japan. The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute

for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University. This paper makes use of software developed for the LSST. We thank the LSST Project for making their code available as free software at http://dm.lsst.org.

References

Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, PASJ, 70, S4
Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, PASJ, 106, 103
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Bolton, A. S., Burles, S., Koopmans, L. V. E., et al. 2008, ApJ, 682, 964
Bosch, J., Armstrong, R., Bickerton, S., et al. 2018, PASJ, 70, S5
Brownstein, J. R., Bolton, A. S., Schlegel, D. J., et al. 2012, ApJ, 744, 41
Cañameras, R., Schuldt, S., Suyu, S. H., et al. 2020, A&A, 644, A163
Caminha, G. B., Rosati, P., Grillo, C., et al. 2019, A&A, 632, A36
Chan, J. H. H., Suyu, S. H., Sonnenfeld, A., et al. 2020, A&A, 636, A87
Cheng, T.-Y., Li, N., Conselice, C. J., et al. 2020, MNRAS, 494, 3750
Collett, T. E. 2015, ApJ, 811, 20
Diehl, H. T., Buckley-Geer, E. J., Lindgren, K. A., et al. 2017, ApJs, 232, 15
Faure, C., Kneib, J.-P., Covone, G., et al. 2008, ApJS, 176, 19
Gavazzi, R., Marshall, P. J., Treu, T., & Sonnenfeld, A. 2014, ApJ, 785, 144
He, K., Zhang, X., Ren, S., Sun, J., et al. 2016a, in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770
He, K., Zhang, X., Ren, S., Sun, J., et al. 2016b, ArXiv e-prints [arXiv:1603.05027]
Huang, X., Storfer, C., Ravi, V., et al. 2020, ApJ, 894, 78
Huang, X., Storfer, C., Gu, A., et al. 2021, ApJ, 909, 27
Inami, H., Bacon, R., Brinchmann, J., et al. 2017, A&A, 608, A2
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Jacobs, C., Glazebrook, K., Collett, T., More, A., & McCarthy, C. 2017, MNRAS, 471, 167
Jacobs, C., Collett, T., Glazebrook, K., et al. 2019, ApJS, 243, 17
Jaelani, A. T., More, A., Oguri, M., et al. 2020, MNRAS, 495, 1291
Jaelani, A. T., Rusu, C. E., Kayo, I., et al. 2021, MNRAS, 502, 1487
Joseph, R., Courbin, F., Metcalf, R. B., et al. 2014, A&A, 566, A63
Knabel, S., Steele, R. L., Holwerda, B. W., et al. 2020, AJ, 160, 223
Lanusse, F., Ma, Q., Li, N., et al. 2018, MNRAS, 473, 3895
LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, in Proceedings of the IEEE, 2278
Li, R., Napolitano, N. R., Tortora, C., et al. 2020, ApJ, 899, 30
Martin, G., Kaviraj, S., Hocking, A., Read, S. C., & Geach, J. E. 2020, MNRAS, 491, 1408
Metcalf, R. B., Meneghetti, M., Avestruz, C., et al. 2019, A&A, 625, A119
More, A., Cabanac, R., More, S., et al. 2012, ApJ, 749, 38
More, A., Verma, A., Marshall, P. J., et al. 2016, MNRAS, 455, 1191
More, A., Suyu, S. H., Oguri, M., More, S., & Lee, C.-H. 2017, ApJ, 835, L25
Petrillo, C. E., Tortora, C., Vernardos, G., et al. 2019, MNRAS, 484, 3879
Pourrahmani, M., Nayyeri, H., & Cooray, A. 2018, ApJ, 856, 68
Ratnatunga, K. U., Ostrander, E. J., Griffiths, R. E., & Im, M. 1995, ApJ, 453, L5
Refsdal, S. 1964, MNRAS, 128, 307
Schuldt, S., Suyu, S. H., Meinhardt, T., et al. 2021a, A&A, 646, A126
Schuldt, S., Suyu, S. H., Cañameras, R., et al. 2021b, A&A, 651, A55
Scoville, N., Abraham, R. G., Aussel, H., et al. 2007, ApJS, 172, 38
Shu, Y., Bolton, A. S., Kochanek, C. S., et al. 2016, ApJ, 824, 86
Shu, Y., Brownstein, J. R., Bolton, A. S., et al. 2017, ApJ, 851, 48
Sonnenfeld, A., Chan, J. H. H., Shu, Y., et al. 2018, PASJ, 70, S29
Sonnenfeld, A., Jaelani, A. T., Chan, J., et al. 2019, A&A, 630, A71
Sonnenfeld, A., Verma, A., More, A., et al. 2020, A&A, 642, A148
Stark, D. P., Auger, M., Belokurov, V., et al. 2013, MNRAS, 436, 1040
Suyu, S. H., & Halkola, A. 2010, A&A, 524, A94
Suyu, S. H., Hensel, S. W., McKean, J. P., et al. 2012, ApJ, 750, 10
Suyu, S. H., Huber, S., Cañameras, R., et al. 2020, A&A, 644, A162
Tadaki, K.-I., Iye, M., Fukumoto, H., et al. 2020, MNRAS, 496, 4276
Tanaka, M., Wong, K. C., More, A., et al. 2016, ApJ, 826, L19
Wen, Z. L., Han, J. L., & Liu, F. S. 2012, ApJS, 199, 34
Willett, K. W., Lintott, C. J., Bamford, S. P., et al. 2013, MNRAS, 435, 2835
Wong, K. C., Sonnenfeld, A., Chan, J. H. H., et al. 2018, ApJ, 867, 107
Wong, K. C., Suyu, S. H., Chen, G. C. F., et al. 2020, MNRAS, 498, 1420

Appendix A: Complete list of candidates

Fig. A.1. HSC three-color *gri* postage stamps of our grade B lens candidates with $1.5 \leq G < 2.5$. The same format as in Fig. 4 is used.

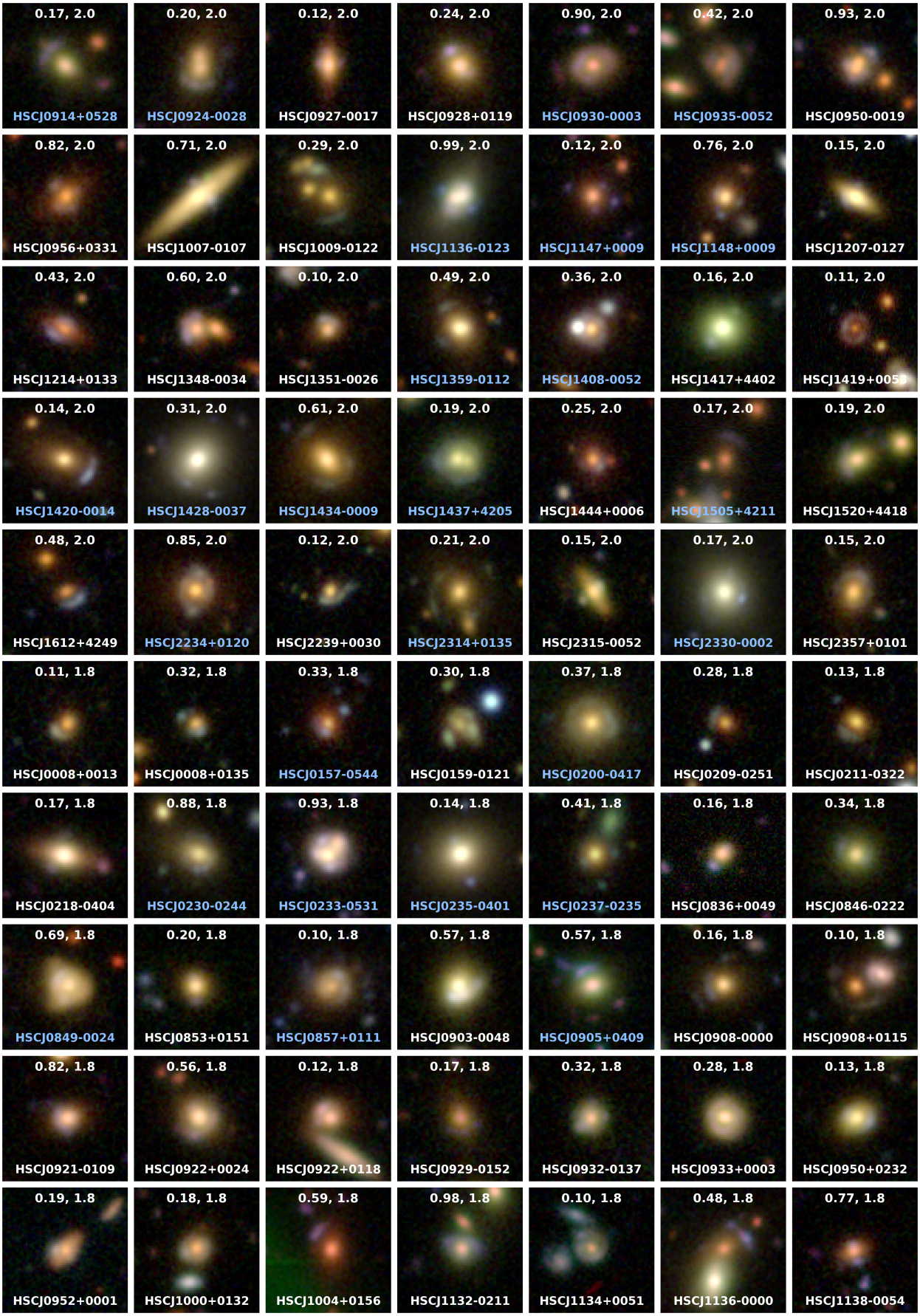

Fig. A.1. continued.

Fig. A.1. continued.

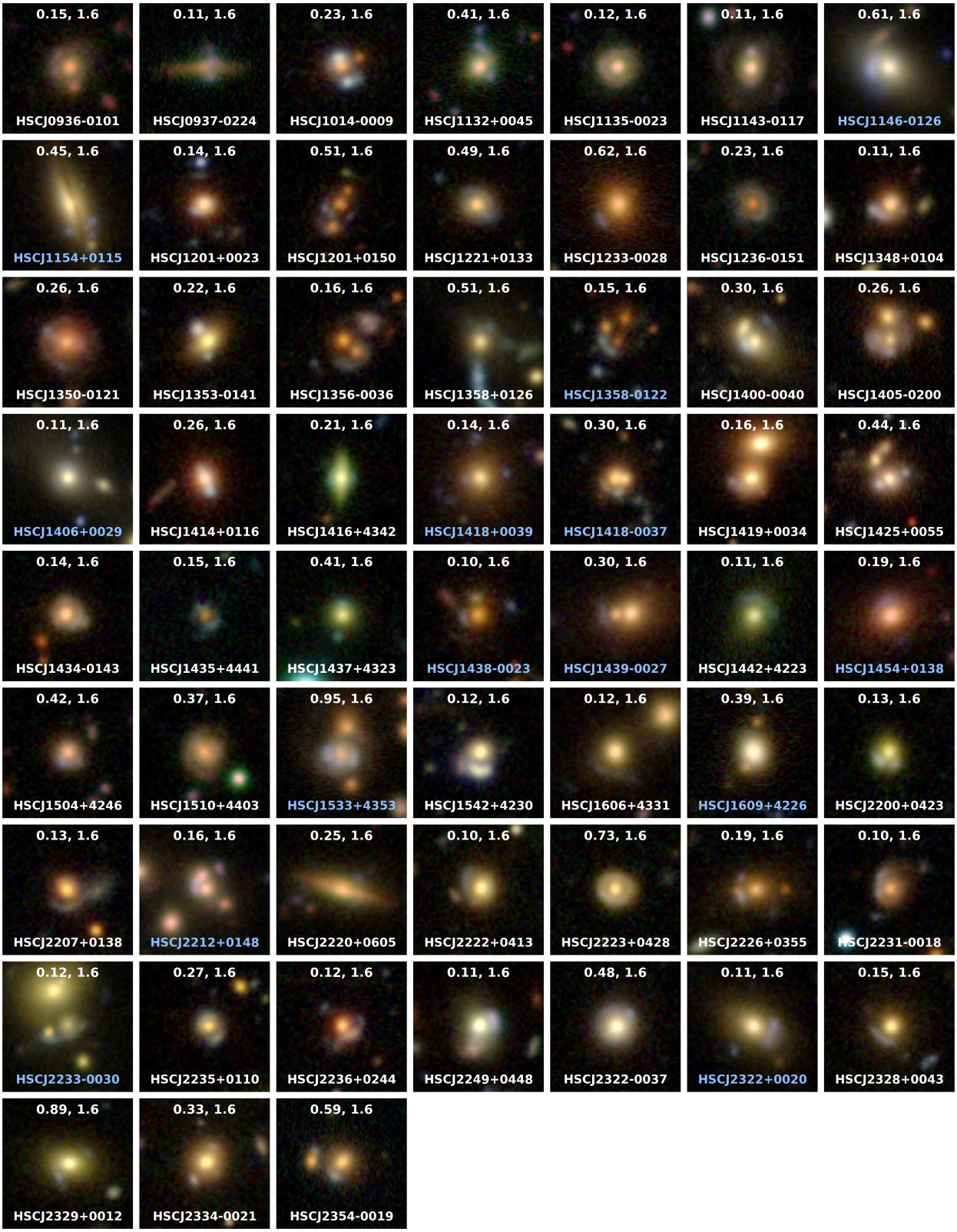

Fig. A.1. continued.