An Extended Catalog of Galaxy–Galaxy Strong Gravitational Lenses Discovered in DES Using Convolutional Neural Networks

C. Jacobs[1,2], T. Collett[3], K. Glazebrook[1,2], E. Buckley-Geer[4], H. T. Diehl[4], H. Lin[4], C. McCarthy[5], A. K. Qin[5], C. Odden[4,6], M. Caso Escudero[4,6], P. Dial[4,6], V. J. Yung[4,6], S. Gaitsch[4], A. Pellico[4], K. A. Lindgren[4], T. M. C. Abbott[7], J. Annis[4], S. Avila[8], D. Brooks[9], D. L. Burke[10,11], A. Carnero Rosell[12,13], M. Carrasco Kind[14,15], J. Carretero[16], L. N. da Costa[13,17], J. De Vicente[12], P. Fosalba[18,19], J. Frieman[4,20], J. García-Bellido[8], E. Gaztanaga[18,19], D. A. Goldstein[21], D. Gruen[10,11,22], R. A. Gruendl[14,15], J. Gschwend[13,17], D. L. Hollowood[23], K. Honscheid[24,25], B. Hoyle[26,27], D. J. James[28], E. Krause[29], N. Kuropatkin[4], O. Lahav[9], M. Lima[13,30], M. A. G. Maia[13,17], J. L. Marshall[31], R. Miquel[16,32], A. A. Plazas[33], A. Roodman[10,11], E. Sanchez[12], V. Scarpine[4], S. Serrano[18,19], I. Sevilla-Noarbe[12], M. Smith[34], F. Sobreira[13,35], E. Suchyta[36], M. E. C. Swanson[15], G. Tarle[37], V. Vikram[38], A. R. Walker[7], and Y. Zhang[4]

(DES Collaboration)

[1] Centre for Astrophysics and Supercomputing, Swinburne University of Technology, P.O. Box 218, Hawthorn, VIC 3122, Australia
[2] ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Swinburne University of Technology, Hawthorn, VIC 3122, Australia
[3] Institute of Cosmology & Gravitation, University of Portsmouth, Portsmouth, po1 3fx, UK
[4] Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510, USA
[5] School of Software and Electrical Engineering, Swinburne University of Technology, P.O. Box 218, Hawthorn, VIC 3122, Australia
[6] Phillips Academy, Andover, MA 01810, USA
[7] Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, Casilla 603, La Serena, Chile
[8] Instituto de Fisica Teorica UAM/CSIC, Universidad Autonoma de Madrid, E-28049 Madrid, Spain
[9] Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK
[10] Kavli Institute for Particle Astrophysics & Cosmology, P.O. Box 2450, Stanford University, Stanford, CA 94305, USA
[11] SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA
[12] Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain
[13] Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ—20921-400, Brazil
[14] Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA
[15] National Center for Supercomputing Applications, 1205 West Clark Street, Urbana, IL 61801, USA
[16] Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, E-08193 Bellaterra (Barcelona) Spain
[17] Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ—20921-400, Brazil
[18] Institut d'Estudis Espacials de Catalunya (IEEC), E-08034 Barcelona, Spain
[19] Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain
[20] Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA
[21] California Institute of Technology, 1200 East California Boulevard, MC 249-17, Pasadena, CA 91125, USA
[22] Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA
[23] Santa Cruz Institute for Particle Physics, Santa Cruz, CA 95064, USA
[24] Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA
[25] Department of Physics, The Ohio State University, Columbus, OH 43210, USA
[26] Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, D-85748 Garching, Germany
[27] Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, D-81679 München, Germany
[28] Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA
[29] Department of Astronomy/Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721-0065, USA
[30] Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, São Paulo, SP, 05314-970, Brazil
[31] George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, and Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA
[32] Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain
[33] Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, NJ 08544, USA
[34] School of Physics and Astronomy, University of Southampton, Southampton, SO17 1BJ, UK
[35] Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, 13083-859, Campinas, SP, Brazil
[36] Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
[37] Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA
[38] Argonne National Laboratory, 9700 South Cass Avenue, Lemont, IL 60439, USA

Abstract

We search Dark Energy Survey (DES) Year 3 imaging for galaxy–galaxy strong gravitational lenses using convolutional neural networks, extending previous work with new training sets and covering a wider range of redshifts and colors. We train two neural networks using images of simulated lenses, then use them to score postage-stamp images of 7.9 million sources from DES chosen to have plausible lens colors based on simulations. We examine 1175 of the highest-scored candidates and identify 152 probable or definite lenses. Examining an additional 20,000 images with lower scores, we identify a further 247 probable or definite candidates. After including 86 candidates discovered in earlier searches using neural networks and 26 candidates discovered through visual inspection of blue-near-red objects in the DES catalog, we present a catalog of 511 lens candidates.

Key words: gravitational lensing: strong – methods: data analysis – methods: statistical – surveys

Supporting material: extended figure, machine-readable tables

1. Introduction

Gravitational lensing is a phenomenon arising from the relativistic curvature of spacetime around massive objects (Einstein 1936; Zwicky 1937, see Treu 2010 for an overview). When strong gravitational lensing occurs, we sometimes observe multiple magnified images of distant sources that lie behind the lensing mass. When the lens is a massive galaxy, group, or cluster, strong lensing can be detectable across cosmological distances. Lensing observables, such as the Einstein radius, are sensitive to the mass of the lens as well as to cosmological parameters, lending strong lensing analyses to many applications across astrophysics and cosmology.

One of strong lensing's applications is as a precise probe of lens mass and dark and baryonic, out to redshift 1 and beyond. Early-type galaxies (ETGs) contain much of the local universe's stellar mass (Renzini 2006) and are the majority of known galaxy lenses due to their high surface mass densities. By measuring the evolution of the total-mass density slopes of ETGs (i.e., constraining the exponent γ, where $\rho(r) \propto r^{\gamma}$), we can test the two-phase model of galaxy assembly predicted by theorists. Simulations predict that at early times, gas-rich assembly from filaments and gas-rich mergers lead to in situ star formation, concentrating baryons in galaxy centers and steepening the density profile. At later times, mass assembly is dominated by dry minor mergers, depositing mass on the outskirts of galaxies and thus increasing the size while decreasing γ (Wellons et al. 2015; Bellstedt et al. 2018). Observations have so far failed to confirm this prediction, with a weak steepening over time of γ observed instead (Sonnenfeld et al. 2013; Remus et al. 2017). At nonlocal redshifts, galaxy-scale strong lensing remains the only feasible method for measuring these density slopes. However, the current lens sample is not large enough to conclusively resolve this tension between simulations and the existing observations. More galaxy-scale strong lenses are needed and at higher redshifts.

The statistics of strong lenses may also prove important in ruling in or out particular models of dark matter. Strong lensing can produce bright arcs and, in some cases, near-perfect Einstein rings. These rings and arcs are perturbed if they intersect the substructure within the lens's dark matter halo, producing detectable "kinks" in the ring. The Sloan Digital Sky Survey (SDSS) strong lens system, SDSS J120602.09 +514229.5 described in Vegetti et al. (2010), contains a visible dwarf that lies on the Einstein ring and introduces a visible distortion in the ring; the same effect will be detectable for dark subhaloes. Exploiting this effect using a large sample of bright arcs and rings will allow us to constrain the subhalo mass function (Koopmans 2005; Vegetti & Koopmans 2009), following a technique demonstrated by Vegetti et al. (2014) on 11 strong lensing systems that resulted in a single detection. Li et al. (2016) used simulations to calculate that as few as 100 bright arcs, with sufficient image resolution to detect subhaloes down to $10^7 \, h^{-1} \, M_\odot$ (which is consistent with future observations) would tightly constrain the subhalo cutoff mass. Such an analysis may confirm the ΛCDM paradigm by providing direct evidence for the low-mass subhaloes predicted by theory; conversely, a detection of low-mass subhaloes could provide strong evidence for a warm dark matter candidate, such as a keV-mass sterile neutrino. Despali et al. (2018) also demonstrated a method to constrain the subhalo mass function using lensing and line-of-sight substructure.

Double-source plane lenses, where two strongly lensed sources at different redshifts are detectable, can function as unique cosmological probes. The ratio of the Einstein radii of the two lenses, β, is independent of the Hubble parameter but is sensitive to the dark energy equation of state, w, and to both Ω_M and Ω_k. Collett & Auger (2014) used a model of the double-source plane lens SDSSJ0946+1006 to constrain w with 30% greater precision than *Planck* alone. Only a few examples of such lenses have been discovered so far (Gavazzi et al. 2008; Tanaka et al. 2016; Diehl et al. 2017).

Artificial neural networks (ANNs) are the key machine-learning technique that underpins recent advances in so-called "deep learning." An overview of ANNs and deep learning can be found in Schmidhuber (2015). Convolutional neural networks (CNNs; LeCun et al. 1989) are a type of ANN optimized for problems involving image data. For standard computer vision tasks, such as image classification and object detection, CNNs have proven highly successful and now routinely exceed human performance. CNNs have already found many successful applications in astronomy—for instance, galaxy morphology classification (Dieleman et al. 2015; Dai & Tong 2018), star-galaxy separation (Kim & Brunner 2017; Cabayol et al. 2019), or identifying quasars from spectra (Busca & Balland 2018).

Here, we are concerned with finding and exploiting strong lenses on the galaxy and group scales. Several hundred examples of galaxy–galaxy strong lenses are currently known.[39] Simulations, such as Collett (2015), predict that several thousand lenses should be discoverable in current-generation surveys, such as the Dark Energy Survey (DES; Dark Energy Survey Collaboration et al. 2016), Kilo Degree Survey (KiDS; de Jong et al. 2015), and Subaru Hyper-Suprime Cam (Miyazaki et al. 2018). Although, in decades past most strong lenses were discovered serendipitously or through visual inspection of an entire survey, the scale of modern surveys means a more targeted approach is required. Previous strategies included searching catalogs for multiple blue sources near red ETGs (Diehl et al. 2017), modeling all sources as strong lenses and testing for goodness of fit (Marshall et al. 2009; Chan et al. 2015), or recruiting citizen scientists to examine images in quantity (Marshall et al. 2016; More et al. 2016). Recently, many efforts have employed modern computer vision and machine-learning techniques. Neural nets have been shown to be effective at distinguishing between simulated lenses and nonlenses (Avestruz et al. 2017; Hezaveh et al. 2017; Lanusse et al. 2018). When applied to imaging surveys, Jacobs et al. (2017, hereafter Paper I) used CNNs to recover several hundred known lenses and 17 new candidates in an hour of inspection time, and Petrillo et al. (2019) and Petrillo et al. (2017) used neural networks to discover over 300 candidate lenses in KiDS.

In Jacobs et al. (2019, hereafter Paper II), we presented 84 candidate lenses at redshifts of 0.8 and above discovered in the DES Year 3 imaging using CNNs.

In the present paper, we present the results of a wider search of the DES images. We apply the technique developed in Papers I and II to the DES Year 3 coadd images (Abbott et al. 2018; Morganson et al. 2018),[40] using newly trained networks

---

Table 1
Strong Lens Candidates from Visual Inspection of the Neural Network-selected Sources, Sorted by Grade

Candidate	Object ID	R.A.	Decl.	Grade	Photoz	Imag	Notes
DESJ0002-3507	139741252	0.59845	−35.12122	2.3	0.51	19.2	⋯
DESJ0003-3348	139823797	0.81825	−33.80120	2.7	0.68	19.3	⋯
DESJ0006-4429	142775105	1.68592	−44.49735	2.0	0.53	18.6	a[1]
DESJ0007-4434	142779522	1.87201	−44.57949	3.0	0.52	18.1	a[1]
DESJ0010-4315	182452355	2.62678	−43.25413	2.3	0.84	19.9	d
DESJ0011-4614	182003535	2.97135	−46.23942	3.0	0.60	18.5	a[1,2] d
DESJ0013+0040	179698697	3.29016	0.66767	2.3	0.75	20.0	⋯

Note. Redshifts are BPZ (http://www.stsci.edu/dcoe/BPZ/) photometric redshifts except where indicted with ∧ where spectroscopic redshifts are substituted. Imag is the *i*-band magnitude from the DES Year 3 A1 Catalog aperture photometry. Notes: (a) previously known (74 candidates); (b) discovered through the BNR and rich cluster search described in Section 2.5; (c) found in both searches; (d) found in previous CNN-based search. References. (1) Diehl et al. (2017), (2) Bleem et al. (2015b), (3) Furlanetto et al. (2013), (4) Stark et al. (2013), (5) More et al. (2012), (6) More et al. (2016), (7) Cabanac et al. (2007), (8) Gavazzi et al. (2014), (9) Hammer (1991), (10) Postman et al. (2012), (11) Bayliss (2012), (12) Lin et al. (2017), (13) Bayliss et al. (2016), (14) Menanteau et al. (2010), (15) Bleem et al. (2015a), (16) Abell et al. (1989), (17) Nord et al. (2016), (18) Buckley-Geer et al. (2011), (19) Sonnenfeld et al. (2018), (20) Kostrzewa-Rutkowska et al. (2014), (21) Bettinelli et al. (2016).

(This table is available in its entirety in machine-readable form.)

and searching for lenses from a wider range of redshifts, morphologies, and colors. In Section 2, we outline the method used to train the networks and score candidates. In Section 3, we present the results of the search and discuss the likely completeness of the search. In Section 4, we offer concluding remarks.

2. Method

The lenses in the catalog presented in this paper were discovered using the methodology presented in Papers I and II. Here, we summarize the method and describe refinements made since the earlier searches. The catalog presented in this work includes candidates discovered in searches using variations on the color cuts, network architectures, and simulation parameters employed in previous works. The search described in Paper II used simulations to target strong lenses in a constrained redshift range; here, we refine the simulations and expand the search, targeting discoverable lenses across all redshifts, morphologies, and colors, aiming for a larger and purer candidate set. We note variations from the earlier search in the text and in Table 1 where appropriate. The method described below describes the parameters of the latest and most comprehensive search.

2.1. Creating Simulated Lenses

In order to train a CNN to distinguish between lenses and nonlenses, we require a training set of labeled examples. To train a network with tens of millions of trainable weights of the size required for reliably processing image data, we require a training set of the order of tens or hundreds of thousands of labeled examples (e.g., Krizhevsky et al. 2012). Since this exceeds the number of known lenses by orders of magnitude, we must instead use simulations to create training sets.

To generate simulations, we use the LENSPOP code described in Collett (2015). LENSPOP generates a population of synthetic elliptical galaxies as lenses, with singular isothermal ellipsoid mass profiles, using masses, ellipticities, and redshifts drawn from realistic distributions. The lenses are simulated with an elliptical De Vaucouleurs profile, and lensed sources are modeled as exponential disks with properties drawn from the Cosmic Evolution Survey (COSMOS) sample (Ilbert et al. 2009). Lens images are created using the GRAVLENS ray-tracing code

(Keeton 2001), with simulated seeing and shot noise appropriate for DES imaging (see Paper II). The simulations produced by LENSPOP follow a realistic distribution of lensing parameters such as an Einstein radius and magnification; for the purposes of training the networks, we want clear, bright examples of strong lensing, so we both discard undetectable lenses and make simulated sources brighter by one magnitude. The thresholds used for detectability are: signal to noise >20, magnification >5, and an Einstein radius >2″. Simulated lens and source images are combined with random patches of DES imaging to create postage stamps with realistic sky noise, foreground objects, artifacts, and other contaminants.

In addition to the simulations described above and in Paper II, we create a further set of simulated lenses using the red-sequence Matched-filter Galaxy Catalog (redMaGiC) catalog (Rozo et al. 2016) of luminous red galaxies (LRGs). For each of the 88,307 galaxies in the catalog, we use the supplied photometric redshift of the galaxy and a nominal velocity dispersion value, which was calculated using the Hyde & Bernardi (2009) fundamental plane of SDSS ellipticals to convert between the rest frame *r*-band absolute magnitude and velocity dispersion (assuming a ΛCDM cosmology, $h = 0.7$). We used the redMaGiC photometric redshifts and assumed a 10 Gyr old passive spectral energy distribution (SED) to convert the observed *i*-band magnitude into the rest frame *r*-band absolute magnitude. We then sample three simulated sources at different positions in the source plane and produce images via ray-tracing for each of the lensed sources. These are then combined with the actual DES imaging for the galaxy to create simulated lens images.

Figure 1 depicts simulations, with and without synthetic lenses, used for training.

2.2. Training Neural Networks

We use the simulations to create training sets for the neural networks. For the positive examples, we use the simulated lenses described in Section 2.1. For the negative examples, we use a combination of simulated nonlenses and real sources from the field. For the former, we use simulations without any flux from a lensed source added, depicting only the simulated ETG. For the second, we take postage stamps of galaxies randomly selected from the target search catalog (Section 2.3). Since strong lenses are rare—as few as 1 in 100,000 galaxies—a
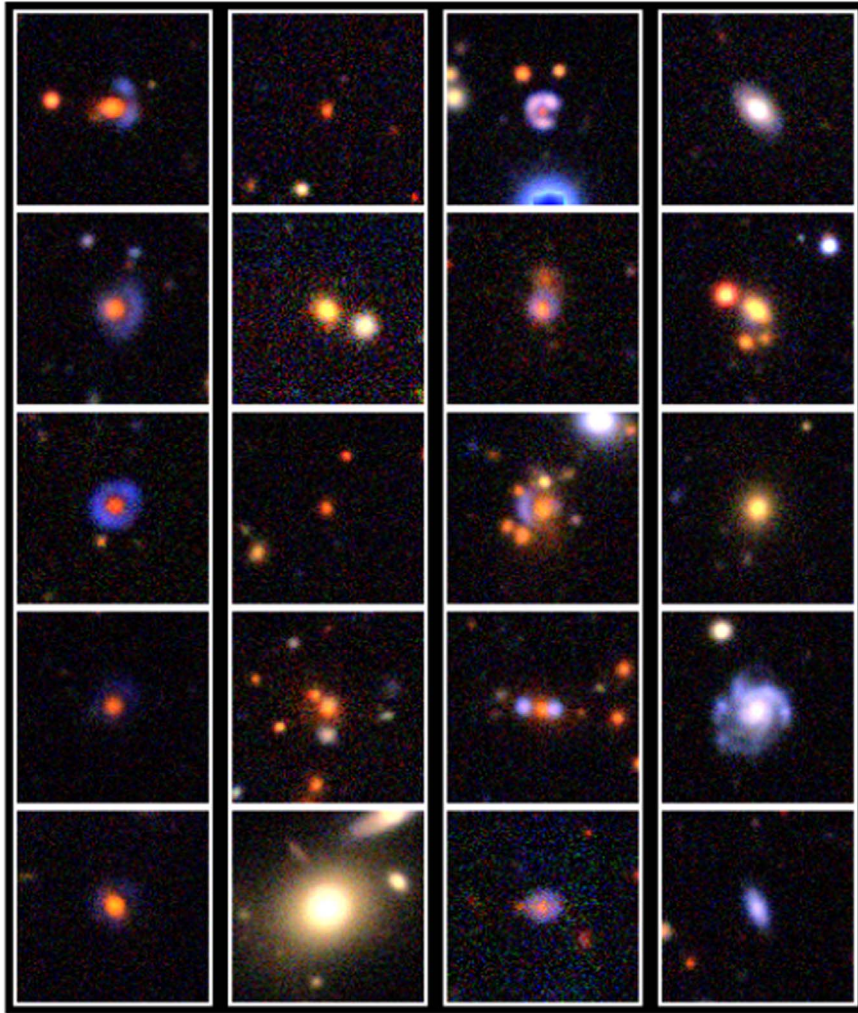
Figure 1. Images used in training the neural networks. Left column: simulated lenses and lensed sources. Second column from left: simulated ETGs without a lensed source. Third column from left: redMaGiC galaxies and simulated lensed sources. Right column: field galaxies, used as negative examples.

random sample of sources from the survey catalog is unlikely to contain any contamination with strong lenses. (We expect to find a few hundred lenses in our search catalog of 7.9 million galaxies.)

We use two types of negative examples for the following reasons. First, we use simulated ETGs with and without a lensed source. This way, the network learns that the presence of a lensed background source is the significant feature and that the presence of the elliptical lens is not in itself indicative of lensing. Even if the simulated ellipticals are unrealistic in some way, such as color, the networks should learn to ignore them as they are present in both positive and negative examples and are, therefore, not discriminatory to arrive at the correct class. The second type of training set, which incorporates real galaxies as negative examples, exposes the networks to spiral galaxies, mergers, and other sources that represent nonlenses and will be present in the images to be tested. In this way, the network learns that an elliptical galaxy with lensing is the target and that potentially confusing objects, such as spiral arms and tidal tails, are to be ignored.

We normalize the training data so that in each band, the mean value of the supplied image data is zero and the standard deviation is one; this aids in quicker convergence of the neural network training process.

We create training sets of up to 200,000 images, consisting of equal numbers of positive and negative examples. We construct a neural network with the following architecture: four convolutional layers, with kernel sizes of 11, 5, 3, and 3 and ReLU[41] activations; 2×2 max pooling[42] after each convolutional layer; and two fully connected layers of 1024 neurons each, with an output layer of two neurons. The network architecture of CN1 and CN2 is described in detail in Table 2.

The process of training, which employs the backpropagation algorithm, is described in Papers I and II and LeCun et al. (1989). Briefly, for each training example, the algorithm determines a correction to each of the weights in the network that would decrease a loss function, L, where $L = 0$ if all classifications are correct and increases as accuracy decreases. With each iteration, each of the weights of the network are updated with the mean optimal adjustment calculated over a batch of 128 training images. By this process, the network learns key features of the images and converges on higher classification accuracy. During training, we measure the loss

---

[41] Rectified linear unit: $f(x) = \max(0, x)$.
[42] This reduces the spatial extent of the input, such that for each 2×2 pixel area in the input, the output is a single value, which is the maximum of the four values.

Table 2
Output of the Keras Model Summary for the CNNS Used in This Lens Search

Layer (Type)	Output Shape	Param Count
conv2d_13 (Conv2D)	(None, 96, 50, 50)	34944
max_pooling2d_10 (MaxPooling)	(None, 96, 24, 24)	0
conv2d_14 (Conv2D)	(None, 128, 24, 24)	307328
activation_19 (Activation)	(None, 128, 24, 24)	0
max_pooling2d_11 (MaxPooling)	(None, 128, 11, 11)	0
conv2d_15 (Conv2D)	(None, 256, 11, 11)	295168
activation_20 (Activation)	(None, 256, 11, 11)	0
conv2d_16 (Conv2D)	(None, 256, 11, 11)	590080
dropout_13 (Dropout)	(None, 256, 11, 11)	0
activation_21 (Activation)	(None, 256, 11, 11)	0
max_pooling2d_12 (MaxPooling)	(None, 256, 5, 5)	0
dropout_14 (Dropout)	(None, 256, 5, 5)	0
flatten_4 (Flatten)	(None, 6400)	0
dense_10 (Dense)	(None, 1024)	6554624
activation_22 (Activation)	(None, 1024)	0
dropout_15 (Dropout)	(None, 1024)	0
dense_11 (Dense)	(None, 1024)	1049600
activation_23 (Activation)	(None, 1024)	0
dropout_16 (Dropout)	(None, 1024)	0
dense_12 (Dense)	(None, 2)	2050
activation_24 (Activation)	(None, 2)	0
Total params:	8833,794	
Trainable params:	8833,794	
Nontrainable params:	0	

Table 3
A Summary of Training Sets Used to Train Neural Networks to Search DES Imaging

Network	Positive Examples	Negative Examples	Training Set Size
RM1	redMaGiC sims	redMaGiC galaxies	160,000
RM2	redMaGiC sims	catalog galaxies	200,000
CN1	LensPop sims	LensPop sims	200,000
CN2	LensPop sims	catalog galaxies	200,000
TS1	LensPop high-z sims	LensPop sims	250,000
TS2	LensPop high-z sims	Real galaxies	150,000

Note. redMaGiC sims use real galaxies for the simulated deflector; LensPop simulates both deflector and lensed source. Networks TS1 and TS2 are described in Paper II.

and accuracy on a validation set, consisting of images not shown to the network during the training steps. This allows us to determine whether the improvements are a due to an overfitting to the training set or will generalize well to new examples.

We train until the validation loss (the loss on the validation set) does not improve by more than 10^{-4} over six epochs (where an epoch is a run through the entire training set).

In addition to the networks described in Paper II, we train four new CNNs using these training sets, as described in Table 3.

2.3. Selecting a Catalog to Search

Even a highly accurate classifier will produce false positives, especially if it is likely to see irregular objects at classification time that were not represented in training, which will be the case for some proportion of sources in any imaging survey. We can minimize false positives by minimizing the number of
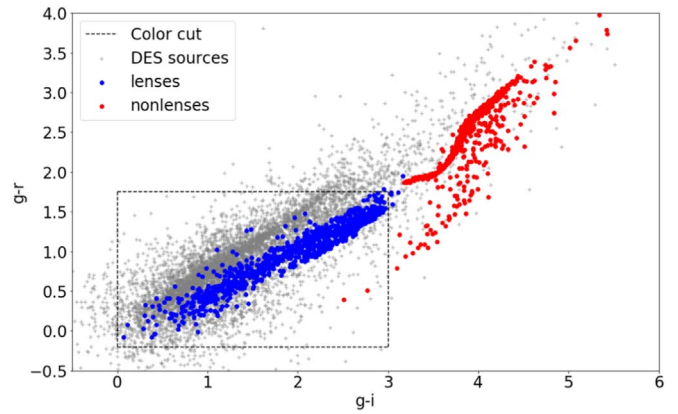

Figure 2. Integrated colors of simulated ETGs (blue), simulated strong lenses (red), and a sample of 5000 sources selected at random from the DES catalog (gray). The color cuts used to assemble our search catalog are shown in dashed lines. Note that some simulated galaxies lie at the edge of the DES g-band magnitude limit, resulting in large magnitude errors and a consequent scatter along the diagonal.

objects we classify, if we can do so without discarding any true positives. We restrict our search to only objects that have the colors of plausible lenses. Although the color of likely lensing ETGs is well known, the color of combined lens and lensed source systems in the aperture photometry of the survey catalog is not known a priori. We again turn to simulations as a guide. Figure 2 depicts the combined $g - r$ and $g - i$ colors of a sample of 10,000 simulated lenses.

In Paper II, we searched 1.4 million sources with colors $2 < g - i < 5$ and $0.6 < g - r < 3$, or $1.8 < g - i < 2$ and $0.8 < g - r < 1.2$, based on simulated lenses at redshifts of 0.8 and above. Here, we search for lenses across all redshifts. Using cuts of $0 < g - i < 3$ and $-0.2 < g - r < 1.75$, which encompasses 98.7% of our simulations, we assemble a catalog of 7.9 million sources from the 300 million objects in the DES catalog for scoring by the neural networks.

2.4. Scoring and Examining Candidates

We score 100×100 pixel postage-stamp images in *griz* bands[43] with CNNs trained using the two training sets, RM1 and RM2, described in Section 2.1. Each network produces a score in the interval (0, 1) for each image. We choose thresholds for the CNN scores, producing lists of candidates with scores greater than those thresholds. Adjusting this parameter produces candidate sets of varying sizes, with different (and unknown) purities and completenesses. The thresholds are initially chosen to produce a candidate set of a few thousand, which is a convenient size to inspect. We then determine the purity of the sample, lower the threshold, and inspect further candidates with lower scores. We repeat this process until the purity has dropped to the point where diminishing returns (less than one quality lens candidate per thousand inspections) make further inspection unworkable.

We inspect RGB images of these candidates using the software LensRater[44] developed for that purpose, which displays PNG images made with *gri* imaging using three different scaling parameters. We assign each source a grade from 0–3, where 0 = not a lens, 1 = "possibly a lens," 2 = "probably a lens,"

---

[43] Previous searches used simulations and survey data in *gri* bands only.
[44] https://github.com/coljac/lensrater

and 3 = "definitely a lens." The grades used throughout the paper represent the mean grade assigned by authors C.J., T.C., E.B.G., and K.G.

2.5. BNR and Rich Cluster Search

In addition to candidates discovered with the CNN search, for completeness in our catalog, we include 26 candidates discovered through two visual searches. A search was performed on 53,000 candidates selected from the DES Year 3 A1 catalog using a methodology similar to that described in Diehl et al. (2017), which was extended to an extra 3500 square degrees of sky (covering the entire DES footprint in line with the CNN-based search). Then, excluding sources examined in Diehl et al. (2017), blue-near-red (BNR) sources were selected from the DES Year 3 A1 catalog as follows:

1. select LRGs with a i-band magnitude <22, with a redshift of $0.22 < z < 0.70$;
2. count blue-colored sources ($-1 \leqslant g - r < 1$, $-1 \leqslant r - i < 1$) within $10''$;
3. and examine sources where two blue sources were found near LRGs brighter than 21st mag in r, and three or more were found near LRGs brighter than 22nd mag.

We also examined 759 sources from the red-sequence Matched-filter Probabilistic Percolation (redMaPPer) galaxy cluster catalog (Rykoff et al. 2014) that matched with high-flux sources detected by the *Chandra X-ray Observatory* (Weisskopf et al. 1996).

After visual inspection and systematic grading according to the prescription above, 40 candidates were graded as "likely" or "definite" lenses, of which 14 were also discovered in the CNN search; 13 are previously known; 5 were both rediscovered by the CNN search and previously known; and 26 were new. These candidates are indicated in Table 1.

3. Results and Discussion

3.1. Training the Network and Scoring Sources

The networks described in Section 2.2 required approximately five hours each to train on an NVidia Tesla P90 GPU. Training converged on accuracies of between 99% and 99.9% on the validation sets (images not used in the training process), as shown in Figure 3. A typical receiver operating characteristic (ROC) curve depicting the trade-off between false positives and false negatives on our training sets can be found in Paper II.

Including the overheads of loading images into memory, scoring the 8 million images in our catalog took a total of 20 hr using a Tesla P90 GPU; scoring an individual object required of the order of 1 μs. Scores assigned by the networks are shown in Table 4.

3.2. Selecting Candidates

We choose candidates to examine by selecting a CNN score threshold and examining the candidates with a score greater than this number. We seek a candidate set that is as complete as possible in detectable lenses included but is of a tractable size and has as high a purity (lowest fraction of false positives) as possible. If the networks are working, i.e., the score adds significant information, then we should find fewer good candidates at lower score values. We first examine a smaller candidate set, with a high score threshold, and grade the candidates; then, we examine
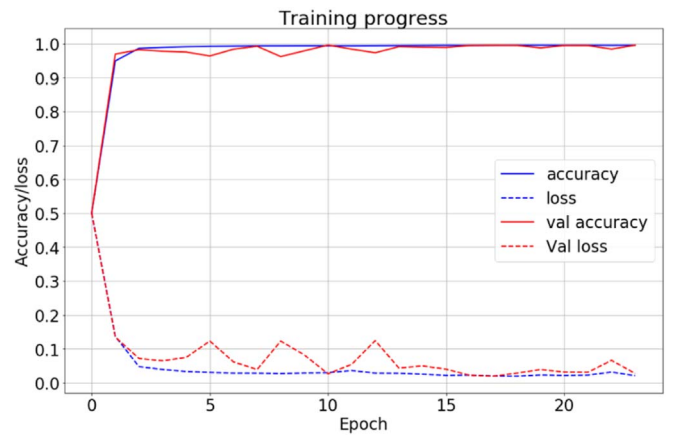
Figure 3. Training a neural network on simulated lenses and nonlenses. Blue dashed line: the loss value, optimized by the training process, decreasing over time. Red dashed line: the loss evaluated on a validation set not used for training. Blue solid line: accuracy in classification on the training set. Red solid line: accuracy measured on the validation set.

Table 4
How the CNNS Scored the Catalog of 7.9 Million DES Sources

Network	Scores = 0	Scores >0.5	Scores = 1
RM1	6248566	64525	3708
RM2	3295227	1840383	330069
CN1	7869097	2799	1000
CN2	2090100	1868791	228776

Note. The number of sources scored 1.0 by both RM1 and RM2 was 1175 but by CN1 and CN2 was 164.

further candidates with lower scores, until diminishing returns suggest further searching is not feasible. The scores, subsequent candidate set sizes, and results are summarized in Table 4.

After examining 1175 images with scores of 1.0 by both RM1 and RM2, we grade 152 with a grade of $> =2$ and a further $148 > =1$. We then test for diminishing returns by examining further, lower-scored sources. We lower the thresholds and examine further candidates. Using a score threshold of >0.5 for both networks, we examine a further 15,172 images and grade them as $247 > =2$, $401 > =1$.

Finally, we include candidates from other searches. This includes networks and catalogs prepared for the search in Paper II and accounts for approximately 20,000 further image inspections. We identify a further 86 candidates with a grade of $> =2$ and 188 with a grade of $> =1$ not identified in the other inspections. Including 26 candidates from the rich cluster and BNR searches, we assemble a total catalog of 511 "probable" and "definite" lenses. These candidates are presented in Table 1. Postage-stamp images of these candidates are presented in Figure 4, which also depicts their CNN scores and human grades. The 742 candidates with grades $> =1$ are presented in Table 5 for reference in future lens searches.

3.3. Purity and Completeness of the Candidate Catalog

Previous lens searches have uncovered of the order of a few hundred potential lenses in DES. Diehl et al. (2017) conducted a search of DES science verification (SV) and Year 1 (Y1) imaging and identified 374 candidate strong lens systems, approximately half of which were graded as "probable" or "definite" lenses. The candidates were selected using the survey
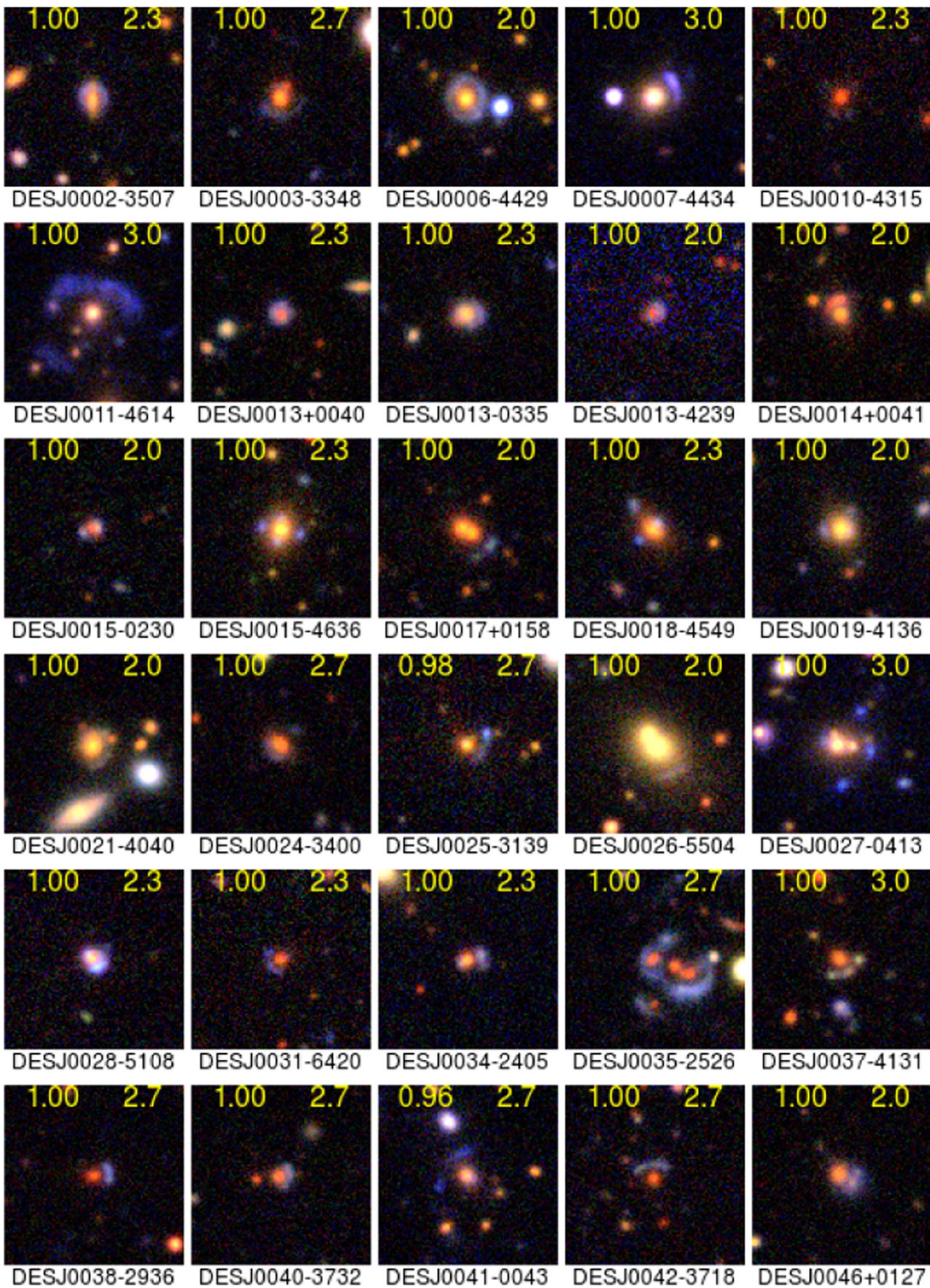
Figure 4. Candidate lenses found in DES using CNNs. In yellow, left: best CNN score; right: human grade. (An extended version of this figure is available.)

catalog, searching for BNR objects and examining a known catalog of ETGs. Assembling this catalog involved inspection of approximately 400,000 cutout images. Nord et al. (2016) searched DES SV and Y1 data for group and cluster-scale strong lenses, identifying 99 candidates of which 21 were confirmed spectroscopically.

Table 5
Possible Strong Lens Systems Selected by CNN but with Grades
$0 <$ Grade < 2, Indicating Possible, but Not Probable or Definite, Lensing

Object ID	R.A.	Decl.	Notes
140003287	00 02 29.66	−52 29 19.90	
141268015	00 02 41.13	02 48 52.85	
142078524	00 03 38.47	−51 55 57.40	
142345819	00 04 28.14	−38 44 08.95	
182976851	00 06 10.34	−55 07 50.09	
182217410	00 06 25.41	−54 24 33.77	
182434686	00 08 35.18	−39 22 38.50	

Note. The object ID is the identifier from the DES Year 3 A1 coadd object tables. Where the source has previously been flagged as a potential lens, this is noted. References. (1) Diehl et al. (2017), (2) Petrillo et al. (2019), (3) Sonnenfeld et al. (2018), (4) Wong et al. (2018), (5) Sonnenfeld et al. (2013).

(This table is available in its entirety in machine-readable form.)

Figure 5. Colors of the CNN lens candidates (red), non-CNN lens candidates (blue), and source catalog (gray).

On simulated data, the CNNs are able to achieve accuracies above 99% for a score threshold of 0.5—in other words, we consider each candidate with a score above this value to be a lens. At this rate, we would expect up to 80,000 false positives from our search catalog in the best case where the networks were as accurate on real data as on simulations.

Collett's (2015) simulations suggest ∼1300 lenses should be findable in DES imaging, using detectability criteria of a signal to noise in $g > 20$, an Einstein radius >2 times seeing, and a magnification of at least 3. However, blind tests on simulated images of lenses (see Paper II) indicate that these detectability criteria may be overly optimistic by a factor of up to five at estimating what an astronomer can confidently pick from RGB images. Ideally, we could also examine lens-subtracted images; however, this requires accurate point-spread function (PSF) modeling, which is complicated for the DES coadd imaging. Poor PSFs lead to ringlike artifacts in the resultant imaging, making the discovery of genuine Einstein rings difficult.

LENSPOP suggests that ∼1300 lenses should be detectable; if only 1/5 of these can be confirmed by human experts, we expect that ∼300 lenses should be detectable with confidence. We, therefore, conclude that our sample is mostly complete for the survey imaging. As noted below in Section 3.5, 41 previously identified high-quality candidates were not recovered in the CNN search. This indicates that improvements to the networks would yield further candidates.

If subsequent improvements can be made in visualization and grading of DES images, it is that possible the detectability threshold could be lowered and new lenses would become findable.

3.4. Suitability of the Search Catalog

The area of $g - i$, $g - r$ color space encompassed by our search catalog contained 98.5% of our simulated lenses. Figure 5 depicts the location of our candidate lenses in this space. The density of candidates with $g - i < 1$ is low, indicating that the search is likely complete at the blue end of our sample. At the red end, the density of candidates with $g - i > 2.5$ also diminishes. Only 27.5% of our search catalog lies below this value, but 79% of our candidates are in this region. If we had restricted our search to sources bluer than $g - i$ of 2.5, we would have recovered 3/4 of our best
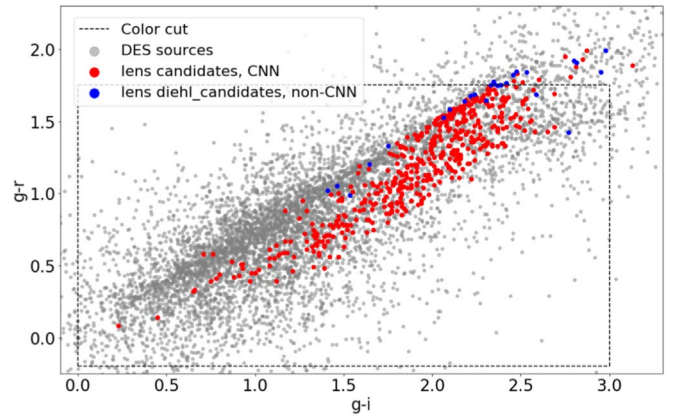
candidates but tested only a quarter as many (∼2 million) sources. This would have yielded a purer sample for human inspection, at the cost of some completeness. Of the Diehl et al. (2017) candidates, 16 high-quality candidates lay outside (redward) of our color cuts by up to 0.25 mag, which is an indication that future searches would benefit from relaxing the criteria.

More candidates lie in the redder end ($g - r > 1$), suggesting that searching further into the red area may be worthwhile. However, fewer sources overall lie in this region of the color space (∼5% of the DES catalog are redder than our cuts), and so we expect diminishing returns to be evident in this area as well.

When we examine the location in this space of candidates with grades of 1 compared to those with a grade of 3, no particular trend is apparent. More ambiguous candidates in our sample have colors similar to higher-quality ones.

3.5. Comparison with BNR and Rich Cluster Search

The BNR search, using the methodology described in Section 2.5 and Diehl et al. (2017), was able to discover several high-quality lens candidates but was less efficient than the CNN search. Visual inspection of over 50,000 sources yielded 40 probable or definite lenses: a rate of 1 in 1250; the CNN-based search required the inspection of approximately 30,000 candidates and yielded 485 probable or definite lenses: a rate of 1 in 62 (as high as one in five in the purest sample).

Of the 26 lenses discovered only in the BNR search, 4 are of galaxy–galaxy scale; the remainder are groups and clusters. Since our training set did not simulate group- and cluster-scale lenses, we do not expect the CNN to discover these lenses, many of which have Einstein radii larger than the postage-stamp images scored with the CNNs.

The BNR search methodology was also employed in the Diehl et al. (2017) bright arcs survey. Forty-one high-quality candidates from that search that satisfied our color cuts were not recovered in this CNN search (they received scores below the thresholds we used by at least one of our networks). Conversely, of our 485 CNN-selected candidates, 48 were found by that search. Identifying those candidates involved visual inspection of over 400,000 images.
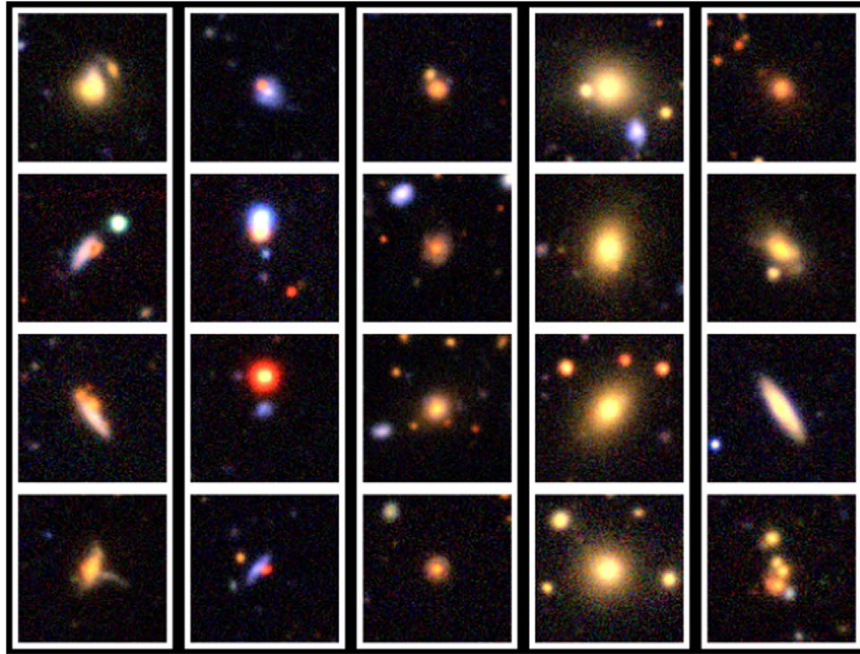
Figure 6. False positives scored as definite lenses (score = 1) by a CNN. Left: false arcs. Second from left: BNR objects. Middle: low signal to noise. Second from right: bright ETGs. Right: no clear pattern.

3.6. False Positives

The purest candidate sample inspected yielded 255 quality candidates from a sample of 1175, or about one in five. Figure 6 depicts some examples of false positives, which fall into four loose categories.

1. False arcs: arc-like features in the image appear to have confused the networks (5%).
2. Blue near red: a chance alignment of blue and red sources may have confused the network (24%).
3. Low signal to noise: could be a lens but the image is not deep enough to be sure (10%).
4. Unknown: no clear reason (61%; however, 33% are bright ETGs that could have lensed sources obscured by the lens in the RGB imaging inspected by us).

The rate of one in five represents a significant improvement from previous searches, but the fact that a human astronomer would reject the majority of the CNN-selected candidates clearly implies there is further room for improvement of the method.

3.7. Improvements for Future Searches

Our networks produced a candidate set of 1175 candidates with a purity of ∼13%, defined as the proportion of probable or definite lenses; this figure is greater than 20% if we include possible lenses. Although this represents high accuracy given the number and variety of sources scored by the networks, with a false positive rate of 1 in 8000, it suggests that the network could be improved to be more aggressive in rejecting certain types of candidates that a human would classify as unlikely. Retraining networks with highly scored false positives classified by human inspectors may drop the false positive rate without significantly impacting the false negative rate. Reducing the false positive rate would also make wider

searches—for instance the entire survey catalog—more feasible.

The use of transfer learning (Bengio 2012; Vilalta 2018), where a network trained on one particular problem domain or training set can be applied to a different problem domain with minimal need for retraining, could assist future searches. The use of transfer learning for a network with an understanding of galaxy morphology was demonstrated on SDSS and DES data in Sánchez et al. (2019). Retraining networks trained to find lenses in another survey or with larger training sets of known lenses could improve networks used in future searches of DES or other surveys.

An improvement in the quality of simulations used for training is also likely to result in improved accuracy. The use of the redMaGiC simulations resulted in a noticeable improvement in the quality of candidates, indicating that there was some property of the simulations that the networks relied on too heavily in scoring. A greater diversity in synthetic stellar populations, redshifts, and morphology may lead to an improvement in the completeness. Our simulations use a PSF drawn from a distribution consistent with DES Year 1 SV data, which may not be optimal for the Y3 coadd imaging we searched. This may bias the networks to images with seeing closer to the simulated distribution. This could also be tested with more varied simulations.

Finally, if the quality of the simulations is not the limiting factor in performance, then deeper networks may also lead to an improvement. In theory, larger (more trainable weights) or deeper (more layers) networks would have the ability to extract more relevant information from the training sets but can also prove more difficult to train. Further work exploring this balance is warranted.

4. Conclusion

In this paper, we present a catalog of 485 strong gravitational lens candidates discovered in the DES Year 3 A1 coadd images

using CNNs. We used simulated lenses to determine color cuts, which we applied to the 300 million sources in the DES survey catalog, yielding 7.9 million sources to search. We scored each image with several neural networks trained with simulated lenses and real galaxies, and combined the scores to produce small sub-sets of our search catalog that were examined by human inspectors and graded for quality. Examining one set of 1175 images (0.01% of the search catalog) resulted in 152 high-quality (probable or definite) lenses. By experimenting with networks of different architectures and training sets, a further approximately 20,000 images were inspected, bringing the total catalog of high-quality candidates to 399. To this, we add 86 candidates found in previous CNN searches and 26 new candidates discovered in other visual searches examining rich clusters and BNR sources. The 511 candidates we discovered in DES with a grade of $>=2$ ("probably a lens") are presented in Table 1.

For reference by future lens searches, in Appendix B we include 742 lenses with grades <2, i.e., possible lenses. If a significant proportion of these are confirmed in future, it may indicate that the CNNs are able to distinguish lensing features that are difficult for a human astronomer to confidently identify.

Future searches will seek to improve the purity of the samples by retraining networks using discovered lenses and false positives graded by human experts including citizen scientist volunteers.

ORCID iDs

C. Jacobs https://orcid.org/0000-0003-4239-4055
T. Collett https://orcid.org/0000-0001-5564-3140
K. Glazebrook https://orcid.org/0000-0002-3254-9044
J. Annis https://orcid.org/0000-0002-0609-3987
D. L. Burke https://orcid.org/0000-0003-1866-1950
M. Carrasco Kind https://orcid.org/0000-0002-4802-3194
D. A. Goldstein https://orcid.org/0000-0003-3461-8661
D. Gruen https://orcid.org/0000-0003-3270-7644
R. A. Gruendl https://orcid.org/0000-0002-4588-6517
D. J. James https://orcid.org/0000-0001-5160-4486
R. Miquel https://orcid.org/0000-0002-6610-4836
A. A. Plazas https://orcid.org/0000-0002-2598-0514
A. Roodman https://orcid.org/0000-0001-5326-3486
E. Sanchez https://orcid.org/0000-0002-9646-8198
M. Smith https://orcid.org/0000-0002-3321-1432
F. Sobreira https://orcid.org/0000-0002-7822-0658
M. E. C. Swanson https://orcid.org/0000-0002-1488-8552
G. Tarle https://orcid.org/0000-0003-1704-0781
A. R. Walker https://orcid.org/0000-0002-7123-8943

References

Abbott, T. M. C., Abdalla, F. B., Allam, S., et al. 2018, ApJS, 239, 18
Abell, G. O., Corwin, H. G., Jr., & Olowin, R. P. 1989, ApJS, 70, 1
Avestruz, C., Li, N., Lightman, M., Collett, T. E., & Luo, W. 2017, arXiv:1704.02322
Bayliss, M. B. 2012, ApJ, 744, 156
Bayliss, M. B., Ruel, J., & Stubbs, C. W. 2016, ApJS, 227, 3
Bellstedt, S., Forbes, D. A., Romanowsky, A. J., et al. 2018, MNRAS, 476, 4543
Bengio, Y. 2012, in Proc. ICML Workshop on Unsupervised and Transfer Learning (New York: ACM), 17, http://dl.acm.org/citation.cfm?id=3045796.3045800
Bettinelli, M., Simioni, M., Aparicio, A., et al. 2016, MNRAS, 461, L67
Bleem, L. E., Stalder, B., Brodwin, M., et al. 2015a, ApJS, 216, 20
Bleem, L. E., Stalder, B., de Haan, T., et al. 2015b, ApJS, 216, 27
Buckley-Geer, E. J., Lin, H., Drabek, E. R., et al. 2011, ApJ, 742, 48
Busca, N., & Balland, C. 2018, arXiv:1808.09955
Cabanac, R. A., Alard, C., Dantel-Fort, M., et al. 2007, A&A, 461, 813
Cabayol, L., Sevilla-Noarbe, I., Fernández, E., et al. 2019, MNRAS, 483, 529
Chan, J. H. H., Suyu, S. H., Chiueh, T., et al. 2015, ApJ, 807, 138
Collett, T. E. 2015, ApJ, 811, 20
Collett, T. E., & Auger, M. W. 2014, MNRAS, 443, 969
Dai, J.-M., & Tong, J. 2018, arXiv:1807.10406
Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270
de Jong, J. T. A., Verdoes-Kleijn, G. A., Boxhoorn, D. R., et al. 2015, A&A, 582, A62
Despali, G., Vegetti, S., White, S. D. M., Giocoli, C., & Bosch, V. D. C. F. 2018, MNRAS, 475, 5424
Diehl, H. T., Buckley-Geer, E. J., Lindgren, K. A., et al. 2017, ApJS, 232, 15
Dieleman, S., Willett, K. W., & Dambre, J. 2015, MNRAS, 450, 1441
Einstein, A. 1936, Sci, 84, 506
Furlanetto, C., Santiago, B. X., Makler, M., et al. 2013, MNRAS, 432, 73
Gavazzi, R., Marshall, P. J., Treu, T., & Sonnenfeld, A. 2014, ApJ, 785, 144
Gavazzi, R., Treu, T., Koopmans, L. V. E., et al. 2008, ApJ, 677, 1046
Hammer, F. 1991, ApJ, 383, 66
Hezaveh, Y. D., Levasseur, L. P., & Marshall, P. J. 2017, Natur, 548, 555
Hyde, J. B., & Bernardi, M. 2009, MNRAS, 396, 1171
Ilbert, O., Capak, P., Salvato, M., et al. 2009, ApJ, 690, 1236
Jacobs, C., Collett, T., Glazebrook, K., et al. 2019, MNRAS, 484, 5330
Jacobs, C., Glazebrook, K., Collett, T., More, A., & McCarthy, C. 2017, MNRAS, 471, 167
Keeton, C. R. 2001, arXiv:astro-ph/0102340

Kim, E. J., & Brunner, R. J. 2017, MNRAS, 464, 4463
Koopmans, L. V. E. 2005, MNRAS, 363, 1136
Kostrzewa-Rutkowska, Z., Wyrzykowski, Ł., Auger, M. W., Collett, T. E., & Belokurov, V. 2014, MNRAS, 441, 3238
Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012, in Advances in Neural Information Processing Systems, Vol. 25, ed. F. Pereira (red Hook, NY: Curran Associates, Inc.), 1097
Lanusse, F., Ma, Q., Li, N., et al. 2018, MNRAS, 473, 3895
LeCun, Y., Boser, B., Denker, J. S., et al. 1989, Neural Comput., 1, 541
Li, R., Frenk, C. S., Cole, S., et al. 2016, MNRAS, 460, 363
Lin, H., Buckley-Geer, E., Agnello, A., et al. 2017, ApJL, 838, L15
Marshall, P. J., Hogg, D. W., Moustakas, L. A., et al. 2009, ApJ, 694, 924
Marshall, P. J., Verma, A., More, A., et al. 2016, MNRAS, 455, 1171
Menanteau, F., Hughes, J. P., Barrientos, L. F., et al. 2010, ApJS, 191, 340
Miyazaki, S., Oguri, M., Hamana, T., et al. 2018, PASJ, 70, 27
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
Morganson, E., Gruendl, R. A., Menanteau, F., et al. 2018, PASP, 130, 074501
Nord, B., Buckley-Geer, E., Lin, H., et al. 2016, ApJ, 827, 51
Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2017, MNRAS, 472, 1129
Petrillo, C. E., Tortora, C., Vernardos, G., et al. 2019, MNRAS, 484, 3879
Postman, M., Coe, D., Benítez, N., et al. 2012, ApJS, 199, 25
Remus, R.-S., Dolag, K., Naab, T., et al. 2017, MNRAS, 464, 3742
Renzini, A. 2006, ARA&A, 44, 141
Rozo, E., Rykoff, E. S., Abate, A., et al. 2016, MNRAS, 461, 1431
Rykoff, E. S., Rozo, E., Busha, M. T., et al. 2014, ApJ, 785, 104
Sánchez, H. D., Huertas-Company, M., Bernardi, M., et al. 2019, MNRAS, 484, 93
Schmidhuber, J. 2015, NN, 61, 85
Sonnenfeld, A., Treu, T., Gavazzi, R., et al. 2013, ApJ, 777, 98
Sonnenfeld, A., Treu, T., Gavazzi, R., et al. 2018, PASJ, 70, S29
Stark, D. P., Auger, M., Belokurov, V., et al. 2013, MNRAS, 436, 1040
Tanaka, M., Wong, K. C., More, A., et al. 2016, ApJL, 826, L19
Treu, T. 2010, ARA&A, 48, 87
Vegetti, S., Czoske, O., & Koopmans, L. V. E. 2010, MNRAS, 407, 225
Vegetti, S., & Koopmans, L. V. E. 2009, MNRAS, 400, 1583
Vegetti, S., Koopmans, L. V. E., Auger, M. W., Treu, T., & Bolton, A. S. 2014, MNRAS, 442, 2017
Vilalta, R. 2018, arXiv:1812.10403
Weisskopf, M. C., O'dell, S. L., & van Speybroeck, L. P. 1996, Proc. SPIE, 2805, 2
Wellons, S., Torrey, P., Ma, C.-P., et al. 2015, MNRAS, 449, 361
Wong, K. C., Sonnenfeld, A., Chan, J. H. H., et al. 2018, ApJ, 867, 107
Zwicky, F. 1937, PhRv, 51, 290