Lawrence Berkeley National Laboratory

Title

New Strong Gravitational Lenses from the DESI Legacy Imaging Surveys Data Release 9

Permalink

https://escholarship.org/uc/item/3k47h3s0

Journal

The Astrophysical Journal Supplement Series, 274(1)

ISSN

0067-0049

Authors

Storfer, C

Huang, X

Gu, A

et al.

Publication Date

2024-09-01

DOI

10.3847/1538-4365/ad527e

Copyright Information

Peer reviewed

New Strong Gravitational Lenses from the DESI Legacy Imaging Surveys Data Release 9

C. Storfer[1,2] , X. Huang[3] , A. Gu[4,5] , W. Sheu[2,6] , S. Banka[5], A. Dey[7] , J. Inchausti Reyes[3] , A. Jain[5], K. J. Kwon[8] ,
D. Lang[9,10] , V. Lee[4], A. Meisner[7] , J. Moustakas[11] , A. D. Myers[12], S. Tabares-Tarquinio[2,3], E. F. Schlafly[13] , and
D. J. Schlegel[2]

[1] Institute for Astronomy, University of Hawaii, Honolulu, HI 96822-1897, USA; cstorfer@hawaii.edu
[2] Physics Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
[3] Department of Physics & Astronomy, University of San Francisco, San Francisco, CA 94117-1080, USA; xhuang22@usfca.edu
[4] Department of Physics, University of California, Berkeley, Berkeley, CA 94720, USA
[5] Department of Electrical Engineering & Computer Sciences, University of California Berkeley, Berkeley, CA 94720, USA
[6] Department of Physics & Astronomy, University of California Los Angeles, 430 Portola Plaza, Los Angeles, CA 90095, USA
[7] NSF's National Optical-Infrared Astronomy Research Laboratory, 950 N. Cherry Avenue, Tucson, AZ 85719, USA
[8] Department of Physics, University of California Santa Barbara, Santa Barbara, CA 93106, USA
[9] Perimeter Institute for Theoretical Physics, Waterloo, ON N2L 2Y5, Canada
[10] Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[11] Department of Physics and Astronomy, Siena College, 515 Loudon Road, Loudonville, NY 12211, USA
[12] Department of Physics & Astronomy, University of Wyoming, 1000 E. University, Dept 3905, Laramie, WY 82071, USA
[13] Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550-9234, USA

Abstract

We have conducted a search for strong gravitational lensing systems in the Dark Energy Spectroscopic Instrument (DESI) Legacy Imaging Surveys Data Release 9. This is the third paper in a series. These surveys together cover $\sim$19,000 deg^2 visible from the Northern Hemisphere, reaching a z-band AB magnitude of $\sim$22.5. We use a deep residual neural network, trained on a compilation of known lensing systems and high-grade candidates as well as nonlenses in the same footprint. After applying our trained neural network to the survey data, we visually inspect and rank images with probabilities above a threshold which has been chosen to balance precision and recall. We have found 1895 lens candidates, of which 1512 are identified for the first time. Combining the discoveries from this work with those from Papers I (335) and II (1210), we have discovered a total of 3057 new candidates in the Legacy Surveys.

Unified Astronomy Thesaurus concepts: Strong gravitational lensing (1643); Cosmology (343); Convolutional neural networks (1938); Galaxy masses (607); Surveys (1671); Sky surveys (1464); Elliptical galaxies (456)

1. Introduction

Strong gravitational lensing systems are a powerful tool for cosmology. They have been used to study how dark matter is distributed in galaxies and clusters (e.g., Kochanek 1991; Koopmans & Treu 2002; Bolton et al. 2006, 2008; Koopmans et al. 2006; Huang et al. 2009; Grillo et al. 2015; Shu et al. 2017), and are uniquely suited to probe substructure in cluster and galaxy scale lenses, as well as line-of-sight low-mass halos and test the predictions of the cold dark matter model beyond the local Universe (e.g., Vegetti & Koopmans 2009; Hezaveh et al. 2016; Wagner-Carena et al. 2023; Fagin et al. 2024). Multiply lensed supernovae (SNe) are ideal for measuring time delays and the Hubble constant H_0 because of their well-characterized light curves, and in the case of Type Ia SNe, with the added benefit of standardizable luminosity (e.g., Refsdal 1964; Oguri & Marshall 2010; Treu 2010), provided microlensing can be accurately characterized (Yahalomi et al. 2017). Furthermore, lens models can be constructed after the SNe have faded (Ding et al. 2021). In the last decade, strongly lensed SNe, both core collapse (Rodney et al. 2016; Chen et al. 2022) and Type Ia (e.g., Pierel et al. 2023, 2024; Pascale et al. 2024) have been discovered. Very recently, Sheu et al. (2023) conducted a retrospective search for strongly lensed SNe in the Dark Energy Spectroscopic Instrument (DESI) Legacy Imaging

Surveys (Dey et al. 2019), and found seven promising candidates. Time-delay H_0 measurements from multiply imaged SNe (e.g., Pierel & Rodney 2019; Suyu et al. 2020; Huber et al. 2021), combined with measurements from distance ladders (e.g., Freedman et al. 2020; Riess et al. 2021) and lensed quasars (e.g., Birrer et al. 2020; Millon et al. 2020; Wong et al. 2020), can be an important test of the tension between H_0 measured locally and the value inferred from the cosmic microwave background (Planck Collaboration et al. 2020). In addition, magnified (but not multiply lensed) Type Ia SNe were identified (Nordin et al. 2014; Patel et al. 2014; Rubin et al. 2018) and used to test the lens models. Finally strong-lensing systems can be used to constrain the properties of dark energy (e.g., Linder 2016; Sharma & Linder 2022).

The introduction of neural networks to identify gravitational lens candidates in imaging surveys has been transformative (e.g., Jacobs et al. 2017, 2019a, 2019b; Metcalf et al. 2019; Cañameras et al. 2020, 2021). In our recent work, we discovered over 1500 new strong lenses (Huang et al. 2020, 2021, Papers I and II, respectively) in the DESI Legacy Imaging Surveys (Dey et al. 2019) by using residual neural networks trained on observed images.

In this paper, we present results from our third search for strong lenses in the Legacy Surveys, using Data Release 9 (DR9). We provide an overview of the observations in Section 2. In Section 3, we describe the construction of the training sample and our neural network model. New lens candidates are presented in Section 4. We discuss our discoveries in Section 5 and conclude in Section 6.
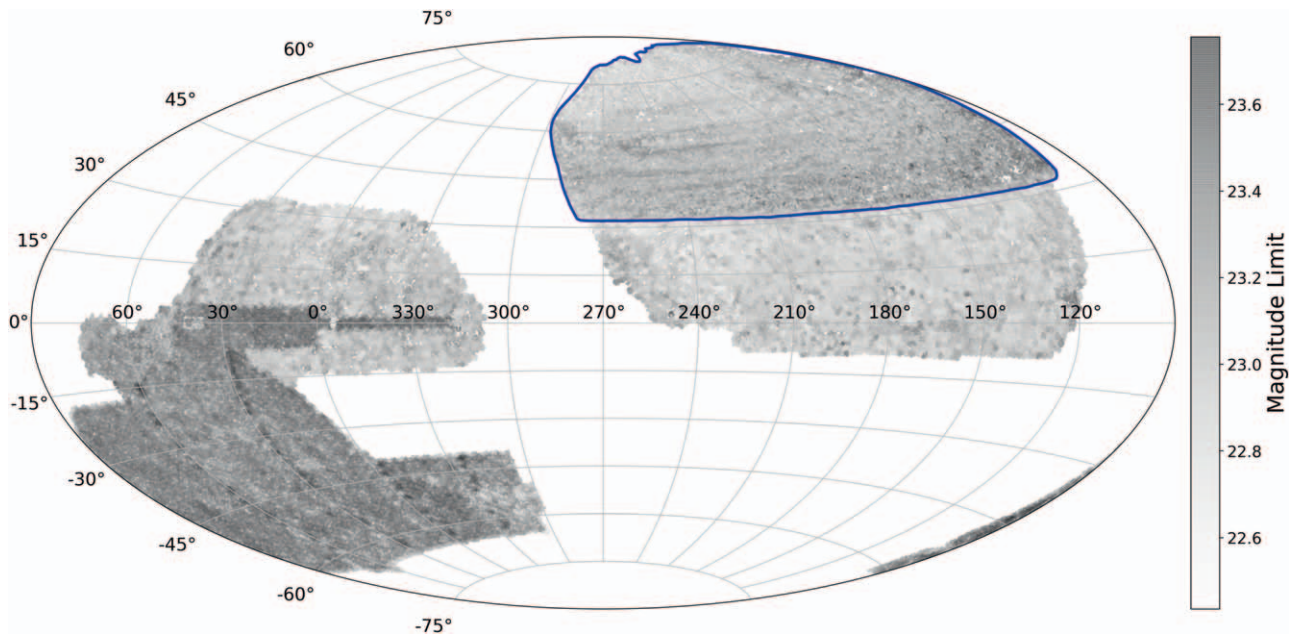
Figure 1. The DESI Legacy Imaging Surveys DR9 footprint in an equal-area Aitoff projection in equatorial coordinates. The blue border approximately outlines the MzLS/BASS region and the rest is DECaLS. The gray shading indicates the z-band 5σ magnitude limit for a galaxy ($0.''45$ round exponential profile) at randomly sampled points across the entirety of the DESI Legacy Surveys footprint. The contiguous region with a magnitude limit of $\sim$23.6 represents the original DES footprint. Note that, prior to DR9, the DES region of DECaLS at $\delta \lesssim -32°$ was not entirely reprocessed by the Legacy Surveys imaging team. The footprint of the Legacy Surveys DR9 now covers $\sim$19,000 deg^2.

Table 1
Object Counts

The Tractor Type	SER	DEV	REX	EXP	Total by Region
BASS/MzLS	3,095,371	2,942,176	2,806,991	1,963,786	10,808,324
DECaLS	15,257,135	7,223,053	7,569,493	4,404,247	34,453,928
Total by Type	18,352,506	10,165,229	10,376,484	6,368,033	45,262,252

2. Observations

The Legacy Surveys are composed of three surveys: the Dark Energy Camera Legacy Survey (DECaLS), the Beijing-Arizona Sky Survey (BASS), and the Mayall z-band Legacy Survey (MzLS). DECaLS is observed by the Dark Energy Camera (DECam; Flaugher et al. 2015) on the 4 m Blanco telescope, which covers $\sim$9000 deg^2 of the sky in the range of $-18° \lesssim \delta \lesssim +32°$. BASS/MzLS are observed in the g and r bands by the 90Prime camera (Williams et al. 2004) on the Bok 2.3 m telescope and in the z band by the Mosaic3 camera (Dey et al. 2016) on the 4 m Mayall telescope. Together BASS/MzLS cover the same $\sim$5000 deg^2 of the northern subregion of the Legacy Surveys. DR9 contains additional DECam data reprocessed from the Dark Energy Survey (DES; Dark Energy Survey Collaboration et al. 2016) for $\delta \lesssim -18°$. This provides an additional $\sim$5000 deg^2, resulting in a total footprint of $\sim$19,000 deg^2. The DECam surveys will hereafter be referred to in their entirety as DECaLS, within which we distinguish DES and non-DES regions. The Legacy Surveys image with a median 5σ point-spread function (PSF) depth of 22.5 AB mag in z band (22.6 AB mag for MzLS). The FWHMs for the delivered images are $1.''29$ (g), $1.''18$ (r), and $1.''11$ (z) for DECaLS; $1.''61$ (g) and $1.''47$ (r) for BASS; and $1.''01$ (z) for MzLS.

The Tractor (Lang et al. 2016) is a forward modeling algorithm that performs probabilistic astronomical source detection and typing and constructs the source catalogs for the Legacy Surveys. Source extraction is done on pixel-level data, taking as input the individual images from multiple exposures in multiple bands, with different seeing in each. _The Tractor_ treats the fitting process as a χ^2 minimization problem. A detected source is retained if the initial penalized χ^2 is improved by 25.[14] _The Tractor_ models detected sources as the better of either a point source (PSF) or round exponential (REX) galaxy. A detected source can be further classified as a de Vaucouleurs (DEV; Sérsic index $n = 4$) or an exponential (EXP; $n = 1$) profile over REX/PSF if such a fit improves the χ^2 by 9. The classification becomes a Sérsic profile (SER) with an improvement in the χ^2 by another nine over DEV/EXP. Earlier data releases included a composite (COMP) profile, which is no longer fit for in DR9.

Figure 1 shows a depth map of z-band observations in the Legacy Surveys DR9 by plotting the depth of objects typed as SER with $z < 20.0$ mag. SER is the most common galaxy type in this magnitude regime. Table 1 shows the total counts for each galaxy type (SER, DEV, REX, and EXP) with $z < 20.0$ mag for both the BASS/MzLS and DECaLS regions.

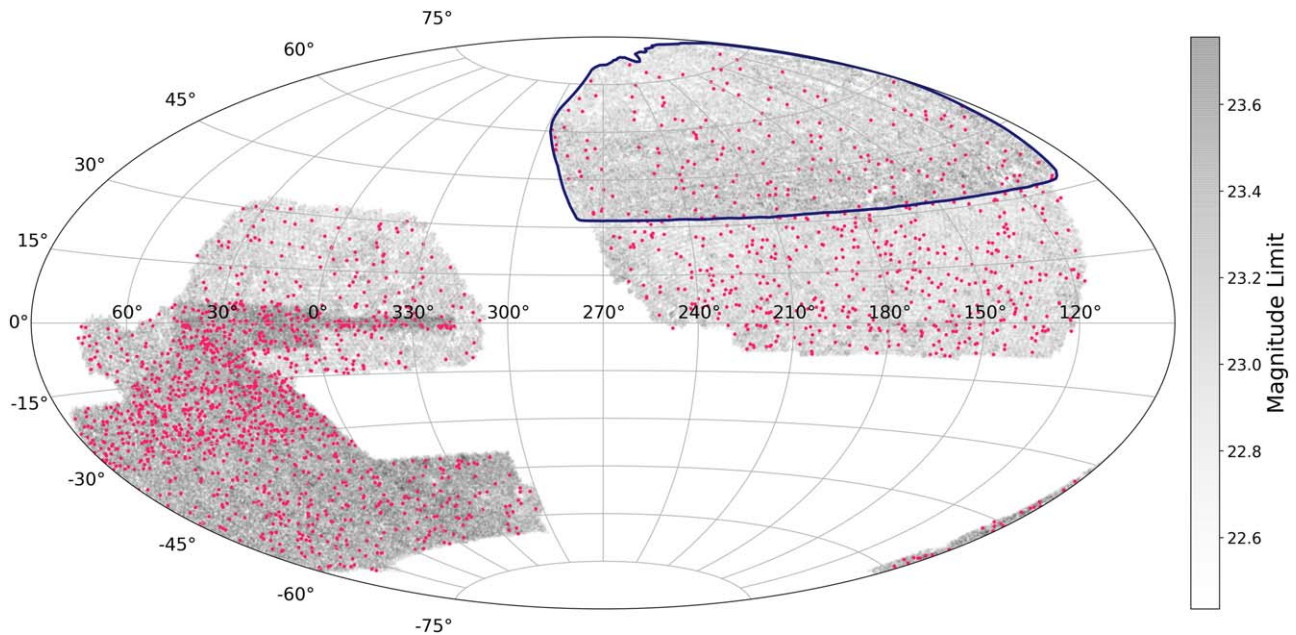[14] For more details, see https://www.legacysurvey.org/dr9/description/.

Figure 2. The 1961 lenses included in the training sample over the depth map of the Legacy Surveys DR9 shown in Figure 1.

3. The Training Sample and Neural Network Model

Convolutional neural networks (CNNs) have been very successful in identifying strong gravitational lenses in simulated and real imaging data. As in Papers I and II, we continue to use observed images of both lenses and nonlenses for training. In what follows, we describe our training sample (Section 3.1) and neural network model (Section 3.2).

3.1. Training Sample

Our training sample consists of 1961 lenses, 1400 of which come from Papers I and II with the rest selected from More et al. (2012, 2016), Carrasco et al. (2017), Diehl et al. (2017), Jacobs et al. (2017, 2019a, 2019b), Pourrahmani et al. (2018), Sonnenfeld et al. (2018, 2020), Wong et al. (2018), Petrillo et al. (2019), Jaelani et al. (2020), Cañameras et al. (2020), and Talbot et al. (2021).

Some of the lens candidates used in the training sample were found by searches conducted by other groups in various imaging surveys, using different search methods and criteria. To select the lenses for the training sample, we inspect the ∼7000 known lenses and candidates reported in the literature. We apply a uniform set of selection criteria (similar to those in Papers I and II) based solely on the Legacy Surveys images for these systems, regardless of where they were originally discovered. The locations of the selected systems are shown in Figure 2.

As in our previous searches, we apply cuts on the type (SER, DEV, or REX), magnitude ($z < 20.0$ mag), and depth (⩾3 passes in g, r, and z bands) of the objects included in the training sample. The z-band magnitude cut was first tested in Paper I where we found that the inclusion of objects fainter than 20 mag results in diminishing returns for discovering lens candidates. In our experience this is the faintest limit that is practical for human inspection.

These same cuts are applied to images in the training sample and for deployment with the exception of images centered on EXP objects, which are included in deployment, but not in the

training sample. To date there are relatively few lens galaxies typed as EXP, none of which pass our selection criteria to be included in the training sample. However, low surface brightness, extended lensed source images are sometimes typed as EXP.

For the selection of nonlenses, for each type (SER, DEV, and REX) and region (DECaLS and BASS/MzLS), we bin by the number of passes in z band. We then select nonlenses randomly, but keep the proportionality to the lenses the same in each bin (∼33:1). This is to prevent potential bias by the neural net based on the number of passes (see Paper II). We cross-reference this selection with the ∼7000 known lenses and candidates found in the literature to remove potential contaminants. In an effort to remove possible additional contamination, we deploy the trained ResNets from Paper II to this selection and inspect the images that receive a probability greater than 0.1 (the same threshold used in Paper II). In the inspection process we remove ∼100 potential contaminants from this selection.[15] This results in a total of 64,584 nonlenses in the training sample.

3.2. The Neural Network Model

The "shielded" residual network (ResNet) model, as described in Paper II, is an improvement upon the neural network architecture presented in Lanusse et al. (2018, hereafter L18). The "shielding" layers (Szegedy et al. 2014) perform 1×1 convolutions that reduce the dimensionality of the output from each ResNet block. This modification has resulted in improved performance and the training time has been shortened by 17% due to the reduction in trainable parameters (see Paper II). During the training, the ResNet attempts to minimize the cross-entropy loss function:

$$\mathcal{L}_{CE} = -\sum_{i=1}^{N} y_i \log \hat{y}_i + (1 - y_i)\log(1 - \hat{y}_i), \quad (1)$$

---

[15] The contaminants we remove are comparable to or better than Grade D (see Section 5).
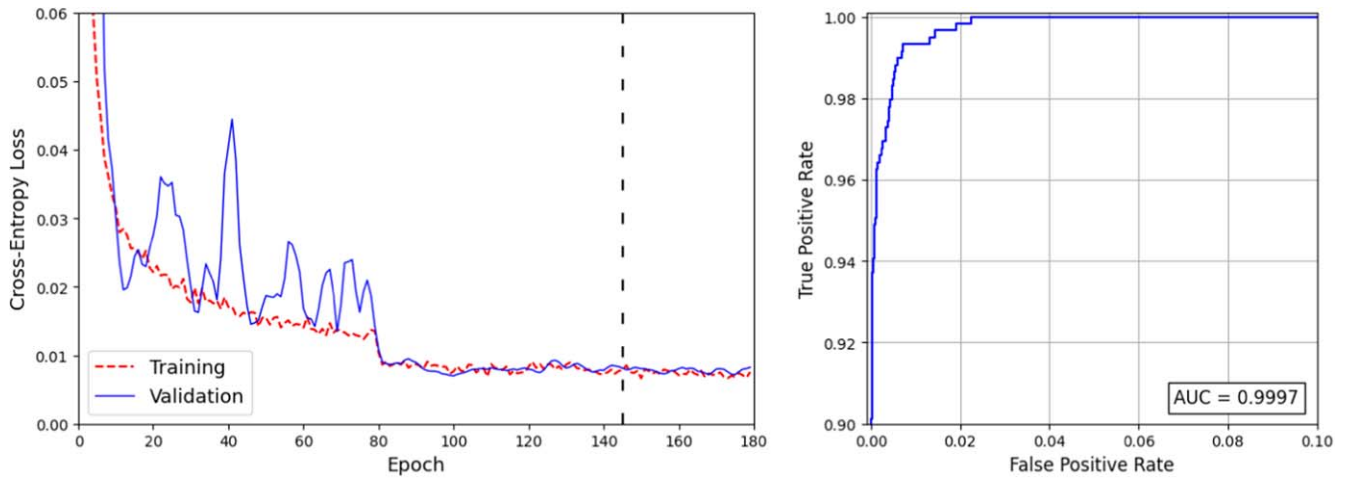
Figure 3. Left: the cross-entropy loss for the training and validation sets for the full 180 epochs of training. The vertical dashed black line at epoch 145 indicates the epoch of the trained model that achieved the highest area under the ROC curve (AUC). We used the trained model at this epoch in the deployment. Right: the ROC curve for the validation set using the best-performing model (epoch 145).

Table 2
Strong Lens Candidates

Grade	A	B	C	Total by Type
SER	68 (64, 4)	309 (296, 13)	650 (630, 20)	1027 (990, 37)
DEV	37 (32, 5)	113 (108, 5)	302 (283, 19)	452 (423, 29)
REX	6 (6, 0)	65 (64, 1)	224 (219, 5)	295 (289, 6)
EXP	4 (4, 0)	39 (34, 5)	78 (74, 4)	121 (112, 9)
Total by Grade	115 (106, 9)	526 (504, 22)	1254 (1206, 48)	1895 (1814, 81)

Note. The numbers shown in the parentheses correspond to the totals in DECaLS and BASS/MzLS (separated by a comma), respectively.

where y_i is label for the ith image (1 for lens and 0 for nonlens), $\hat{y}_i \in [0, 1]$ is the model-predicted probability, and N is the number of images in one training step (the same as the batch size, given below). In the training process, the "shielded" model outperformed the L18 model and thus we decide to exclusively use it in this work. We use a 7:3 split of the training sample to create the training and validation sets, respectively. We use an image cutout size of 101 pixel × 101 pixel (which translates to ∼26″ × 26″), a batch size of 128, an initial learning rate of 5×10^{-4}, a decay rate of 1/5, and a decay epoch at 80. The model is trained on Google Colab[16] using a GPU (NVIDIA Tesla P100). We trained our ResNet model for a total of 145 epochs (approximately 5 hr). Training beyond this point results in marginal gain in model performance (see Figure 3). In addition to the cross-entropy loss, we further assess the performance of of our trained model by using the receiver operating characteristic (ROC). The ROC curve shows the true positive rate (TPR) versus the false positive rate (FPR), where P(ositive) indicates a lens and N(egative), a nonlens. With the definitions true positive (TP) = correctly identified as a lens, false positive (FP) = incorrectly identified as a lens, true negative (TN) = correctly rejected, and false negative (FN) = incorrectly rejected:

$$\text{TPR} = \frac{\text{TP}}{\text{P}} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

and:

$$\text{FPR} = \frac{\text{FP}}{\text{N}} = \frac{\text{FP}}{\text{FP} + \text{TN}}.$$

Random classifications will result in a diagonal line in this space with an AUC equal to 0.5, while for a perfect classifier, AUC = 1. The ROC curve for our best-performing model is also shown in Figure 3 for the validation set, with an AUC of 0.9997, an improvement over the training results from Paper II.

4. Results

In this section we present the lens candidates we have found in the Legacy Surveys. In Section 4.1, we present lens candidates found in DR9 (this work). In Section 4.2 we provide a summary of all lens candidates discovered by our group in the Legacy Surveys DR7, 8, and 9.

4.1. Lens Candidates in Data Release 9

In this section we present the strong lens candidates discovered exclusively in Legacy Surveys DR9 from this work. To determine the probability threshold for human inspection, we consult the precision–recall curve (PRC), where precision = TP/(TP + FP) and recall = TP/(TP + FN), which is the same as TPR (Section 3).[17]

---

[16] https://colab.research.google.com/

[17] As in Paper II, we recognize the redundancy in terminology. This results from fairly standard usage (e.g., recall or TPR depending on the context) and in part from the difference in terminology used in the fields of machine learning and astrophysics ("recall" or "completeness," "precision" or "purity").
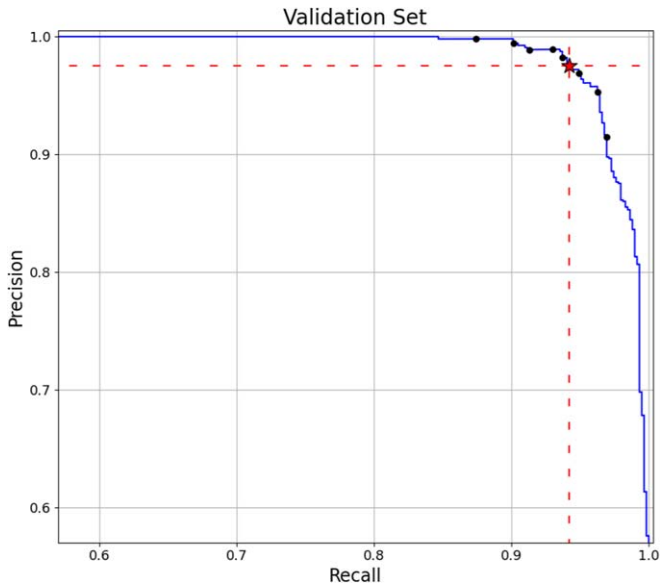
Figure 4. The PRC for the validation set from the fully trained neural network. The black dots from right to left on the curve indicate probability thresholds from 0.1 to 0.9 with a step of 0.1. The red star indicates the selected probability threshold of 0.4 with red dashed lines showing the precision and recall at that point.

Figure 4 shows the PRC for our trained ResNet model for the validation set, with marked probability thresholds. The distribution of the probability thresholds on our PRC are tightly clustered near the "shoulder" of the curve, indicating a high level of performance. A perfect PRC would pass through (1, 1) where these thresholds would converge. We consult the PRC to determine the probability threshold for inspection based on a reasonable trade-off between precision (purity) and recall (completeness). We select a probability threshold of 0.4, with the number of cutout images above this threshold ∼47,000. This is a reasonable number to inspect, comparable to those in Papers I and II. These cutouts with probability ⩾ 0.4 will be referred to as the neural network "recommendations."

For deployment, even though our ResNet model is trained on DEV, SER, and REX, we expand our search to include galaxies with EXP profiles (as mentioned in Section 3.1). Thus we deploy the trained model on 45 million cutouts with ⩾3 passes in g, r, and z bands and centered on all non-PSF objects with z-band mag ⩽ 20.0. Subsequent inspection of the resulting recommendations was conducted considering the following criteria (same as in Papers I and II): small blue galaxy/galaxies (red galaxies are rare but certainly acceptable) next to a red galaxy/galaxies at the center that satisfy the following criteria:

1. are typically 1″–5″ away,
2. have a low surface brightness,
3. curve toward the red galaxy/galaxies,
4. have counter/multiple images with similar colors (especially in Einstein-cross-like configuration), and
5. are elongated (including semi- or nearly full rings).

Typically, most candidates do not have all these characteristics. In general, the greater the number of these characteristics an image has, the higher they are ranked by humans.

After an initial round of inspections, coauthors C.S. and X.H. independently examined all preliminary candidates. The grades are assigned as follows.

Table 3
Purity for Our ResNet Model

The Tractor Type	SER	DEV	REX	EXP	Purity by Region
DECaLS	23	21	25	38	24
BASS/MzLS	29	34	128	51	40
Purity by Type	23	22	27	39	25

Note. Purities are shown as the number of ResNet recommendations inspected in order to find a lens candidate.

1. *Grade A*. We have a high level of confidence of these candidates. Many of them have one or more prominent arcs, usually blue. The rest have one or more clear arclets, sometimes arranged in multiple-image configurations with similar colors (again, typically blue). However, there are clear cases with red arcs.
2. *Grade B*. They have similar characteristics as the Grade As. Grade B giant arcs tend to be fainter than those for Grade A. Likewise, the putative arclets tend to be smaller and/or fainter, or isolated (without counterimages).
3. *Grade C*. They generally have features that are even fainter and/or smaller than what is typical for Grade B candidates, but that are nevertheless suggestive of lensed arclets. Counterimages are often not present or hardly discernible. In a number of cases, the angular scales of the candidate systems are comparable to or only slightly larger than the seeing. Therefore, for some of these candidates, to attain a higher level of certainty, higher spatial resolution, deeper data, or spectroscopic observations would be required.

We have found 1895 candidates, with 115 Grade As, 526 Bs, and 1254 Cs (Table 2). For the DEV and SER type galaxies in DECaLS we achieve a purity of ∼1 candidate in 23 ResNet recommendations. Though it is difficult to do a direct comparison given the change in *The Tractor* categorization, this seems to indicate an improvement over the purity of ∼1 in 31 reported in Paper II for DEV and COMP in DECaLS. Certainly, the overall purity of ∼1 in 25 recommendations is a clear improvement over Paper II (one in ∼40). The purity by type and region is shown in Table 3.

Of these, 125 were previously known lenses or candidates. Additionally, 258 were found in recent publications, the majority of which came from Stein et al. (2022; see Section 5). The rest are from Cañameras et al. (2021), Li et al. (2021), Savary et al. (2022), Wong et al. (2022), Rojas et al. (2022), and O'Donnell et al. (2022).[18] We therefore report 1512 new lens candidates, with 53 Grade As, 384 Bs, and 1075 Cs. We highlight in Figure 5 two examples each for five categories of new strong lens candidates that we have discovered. The positions of all new candidates on the sky are shown in Figure 6.

4.2. Lens Candidates in Data Releases 7, 8, and 9

From our three searches in the Legacy Surveys, we have found a grand total of 3057 new candidates (Paper I: 335 in DR7 among DEV and COMP in DECaLS only; Paper II: 1210 in

---

[18] We show the citations for the lenses that were also reported by these publications on our project website (https://sites.google.com/usfca.edu/neuralens/publications/lens-candidates-storfer-et-al-2022) as "identified in recent papers."
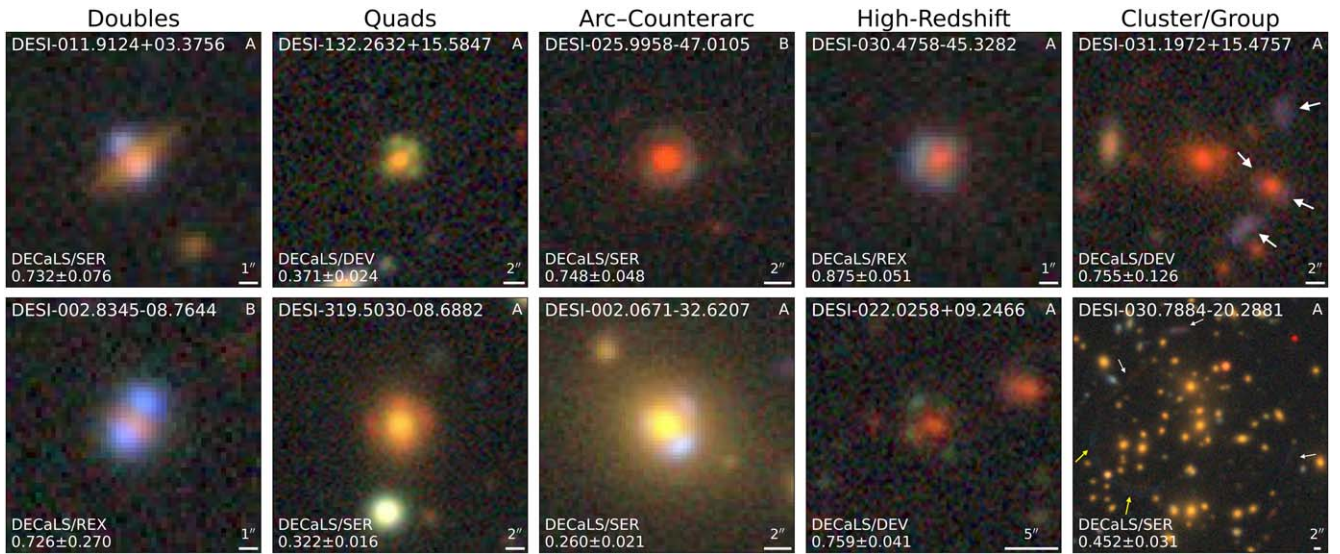
Figure 5. Ten of the 1512 new strong lens candidates discovered in this work. The naming convention is R.A. and decl. in decimal format. For all images, north is up, and east to the left. The top right corner of each image indicates the grade; bottom left corner, the region/*The Tractor* type, and the photometric redshift of the putative lens (z_d). First column: doubly lensed systems. Second column: quadruply lensed systems. Third column: arc–counterarc systems, each of which has a large arc with a smaller counterarc on the opposite side of the lens. Fourth column: high-redshift lensing systems, each of which has quadruple images. The image locations of DESI-030.4758-45.3282 suggests a fold configuration where DESI-022.0258+09.2466 shows a cross or cusp configuration (see Gu et al. 2022). These eight systems all have a single galaxy as the lens. Fifth column: group/cluster lensing systems. Putative arcs are indicated with white arrows. For DESI-031.1951+15.4749, the large lensed arc due to the galaxy group potential is also lensed by a group member, forming two small arcs to its left and right (for another example, see Huang et al. 2009). This would enable a lensing-mass measurement of the substructure. Note that this system also has a high z_d of 0.755. DESI-030.7884-20.2881 is a cluster lensing system with a faint but spectacularly large blue arc to the lower left (marked by two yellow arrows to indicate its extent). There are hints of even fainter arcs around the cluster center. High-resolution imaging will almost certainly reveal more arcs.
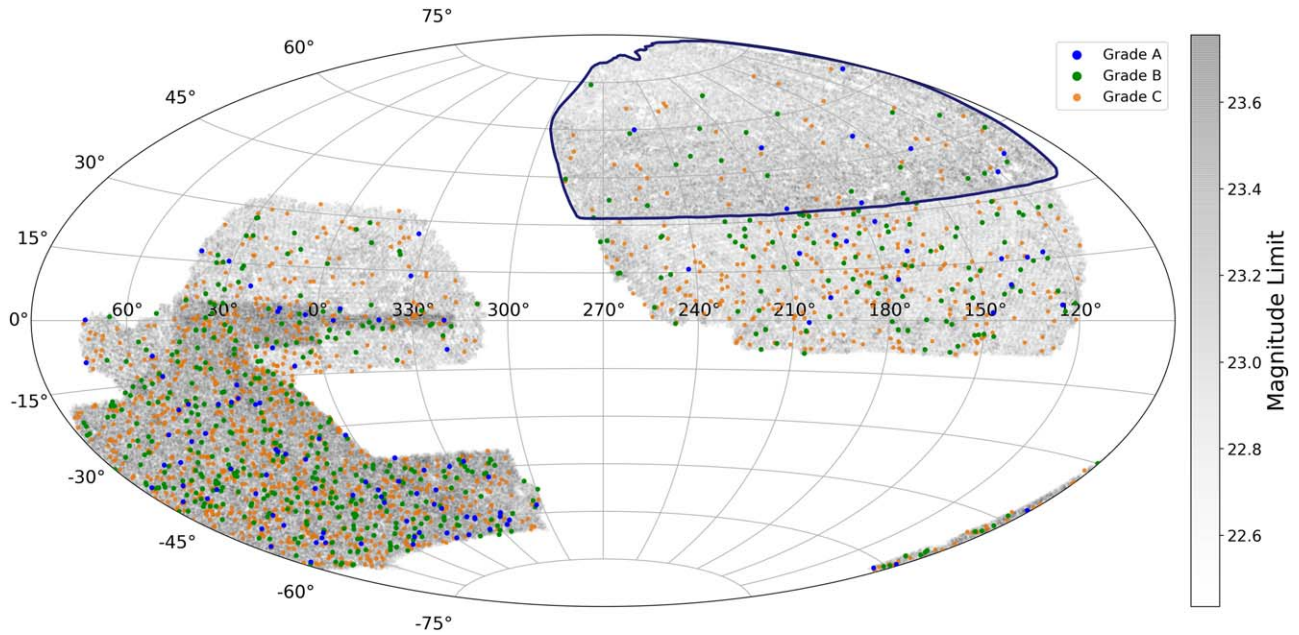
Figure 6. The 1512 new candidate lensing systems discovered in this work by grades over the depth map of the Legacy Surveys DR9 shown in Figure 1.

DR8; and this work: 1512 in DR9). The entire catalog of these lenses can be found on our project website.[19] The positions of all candidates on the sky are shown in Figure 7.

5. Discussion

We have trained and deployed on the highly heterogeneous imaging data within the Legacy Surveys, with nonuniform depth and varying image quality across BASS/MzLS and DECaLS, which includes the DES and non-DES regions. We show that despite this challenge, our neural network can be trained to perform well with a carefully constructed training sample. This training sample is larger by a factor of 3 than that in Paper II, with known lenses and lens candidates from both BASS/MzLS and DECaLS, and nonlenses that are selected proportionally in each depth bin for each region. We achieve an overall ResNet purity of one lens candidate in 25 neural net recommendations, a clear improvement over Paper II (one in
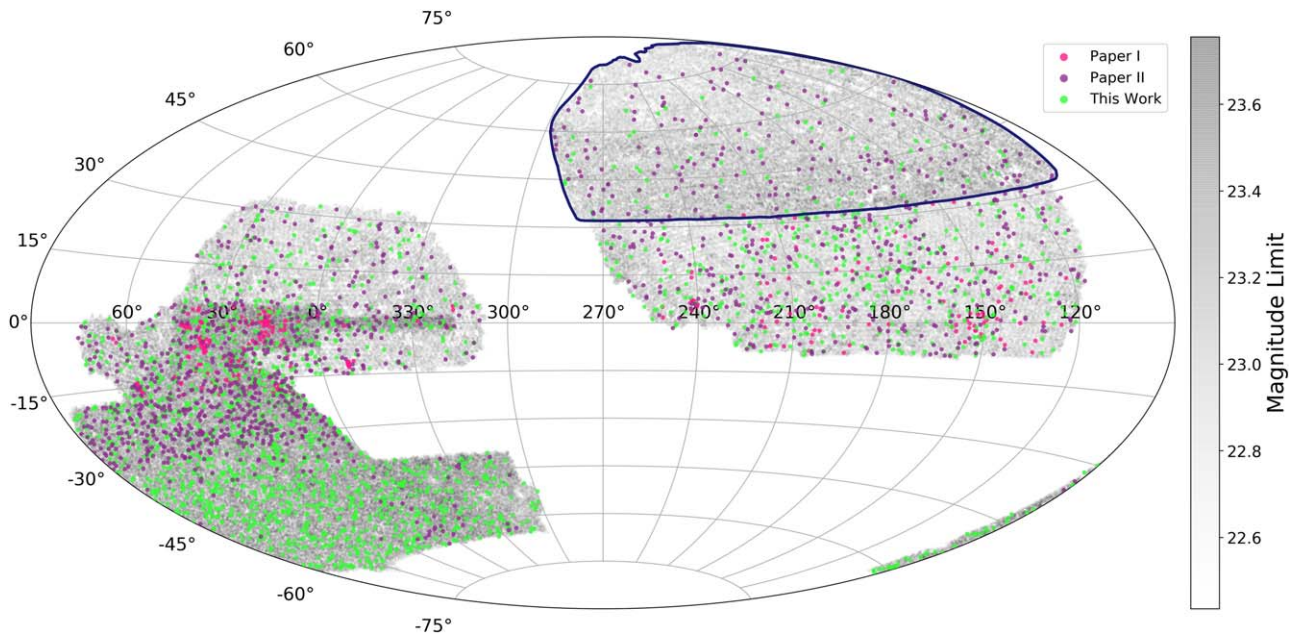
Figure 7. All 3057 candidate lensing systems discovered in the Legacy Surveys reported in Papers I (pink), II (purple), and this work (green), over the depth map of the Legacy Surveys DR9 shown in Figure 1.
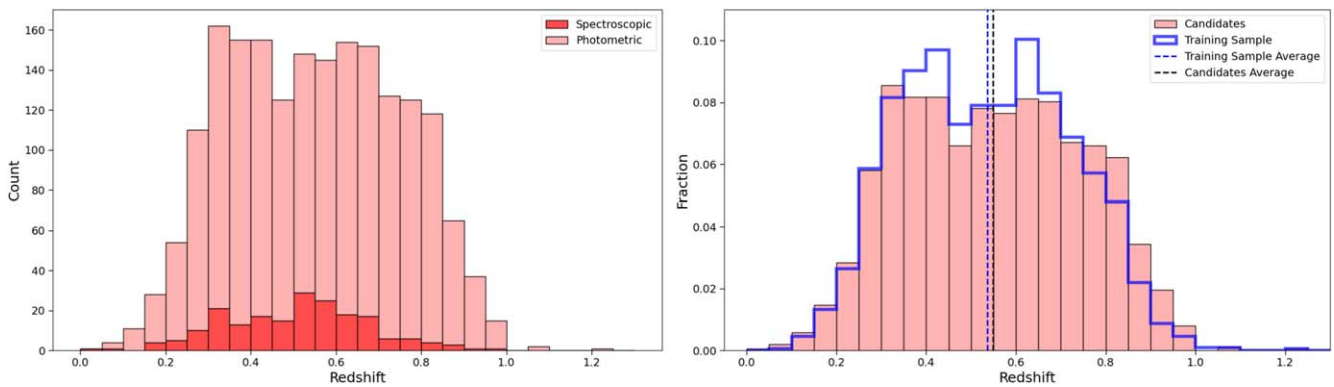
Figure 8. Left: redshift distributions of the candidates found in Legacy Surveys DR9 from SDSS Data Release 17 (spectroscopic; dark red) and Zhou et al. (2021, photometric; light red). Right: redshift distribution of the candidates and the lenses included in the training sample. Dashed vertical lines indicate the average redshift of the training sample and candidates.

~40) while reporting more candidates than Paper II using the same selection criteria. This is competitive with, if not better than, similar searches carried out in other surveys while reporting more candidates than such searches despite the data set we deploy on being the most heterogeneous.

The DECaLS region of the Legacy Surveys has been mined for lenses several times before (Diehl et al. 2017; Jacobs et al. 2019a, 2019b; Rojas et al. 2022; O'Donnell et al. 2022; Papers I and II). It is also worth keeping in mind that the Legacy Surveys overlap significantly with the Sloan Digital Sky Surveys (SDSS) and Pan-STARRS, which contain the entirety of the MzLS/BASS region. Both of these surveys have been mined for strong lenses (e.g., Sonnenfeld et al. 2013; Cañameras et al. 2020; Talbot et al. 2021). Lens searches also have been carried out in the Hyper-Suprime Cam Subaru Strategic Program, Canada–France–Hawaii Telescope Legacy Survey, and the Kilo-Degree Survey, all of which are contained within the Legacy Surveys (e.g., Jacobs et al. 2017; Cañameras et al. 2021; Li et al. 2021). Finally, the MzLS/BASS region also significantly overlaps with the Ultraviolet Near Infrared Optical Northern Survey, which has been mined for lenses as

well (Savary et al. 2022). And yet, we have succeeded in finding a large number of high-quality new lenses.

The photometric redshift distribution of our candidates is shown in the left panel of Figure 8. It is largely similar to the redshift distribution of the lenses in the training sample. The average redshift of the new candidates is higher by ~0.02 over those in the training sample. While this difference is not significant, the right panel of Figure 8 does possibly indicate our model's ability to continue to discover lenses at redshifts higher than those included in the training sample (see bins 0.7–1.0).

Stein et al. (2022, hereafter S22) also searched for strong lenses in the DESI Legacy Surveys DR9. Their approach took two steps. First they performed a form of self-supervised learning, by applying a CNN encoder on real observed images to minimize contrastive loss in a representation space. Then, after projecting image cutouts onto this representation space, they employed three search strategies: similarity search, linear binary classification trained on labeled data, and retraining ("fine-tuning") the self-supervised CNN encoder by incorporating the linear classifier. As with this work, they searched for
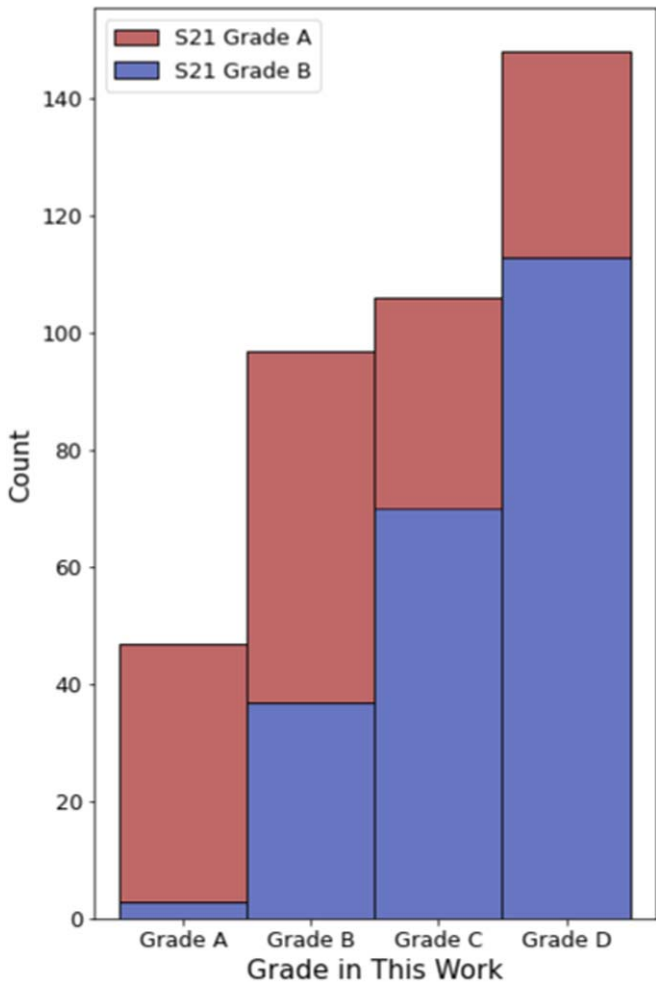
Figure 9. The stacked histogram of the distribution of S22 grades shows a comparison with our grades for the overlapping candidates. Note that grades for the lens candidates reported in this work were assigned prior to S22 being posted on arXiv.

lens candidates in all non-PSF objects (typed by *The Tractor*) with $z < 20.0$ mag. They further expanded the search to $z \geqslant 20.0$ mag, adding $\sim$17% more candidates. They reported a total of 1192 strong lens candidates.

The candidates reported in S22 were only assigned Grades A and B. A total of 398 (25.7%) candidates from S22 have been identified in our search (limited to $z < 20.0$ mag), which represent approximately 40% of their $z < 20.0$ mag candidates. This includes 148 S22 candidates receiving a grade of D[20] from our visual inspection. Figure 9 shows the distribution of the S22 grades relative to the grades in this work for the overlapping candidates. Figure 10 shows a two-way comparison.

For the overlapping candidates: from the red text in Figure 10, 25.8% of the S22 Grade A systems are given an A or B grade in this work, and a further 8.9% and 8.7% of their systems receive a C or D grade from our inspection, respectively. For the S22 Grade B systems, 9.6% are given an A or B grade in our work, 8.9% a C grade, and 14.3% a D grade. In total, 43.3% and 28.3% of the S22 Grade A and B systems, respectively, are assigned a Grade D or better from

---

[20] This grade is assigned to systems that have some evidence of strong lensing but which is even less convincing than that of Grade C.

our visual inspection. Going in the other direction, from the blue text in Figure 10, 40.9%, 18.4%, 8.5%, and 8.0% of our Grade A, B, C, and D candidates are identified in S22 (i.e., assigned an A or B grade), respectively. The agreement is the strongest for the mutual Grade As.

Figure 11 shows examples of overlapping candidates reported in S22 and this work. Grades assigned by S22 (A and B) are the same in each row, and grades from this work are the same in each column (A, B, C, and D). While grading criteria may differ between search teams, high-quality candidates (Grade A and B) are often shared, as shown in the left half of Figure 11.

We also provide comparisons for a few example systems with the largest discrepancy in grade: those assigned Grade A or B in S22 that were given a grade of C and D in this work.

DESI-070.4050-50.4559 (first row, third column) has a faint possible arclet to the NE of the putative lens. There are two other faint objects in the cutout (SE and NW of the putative lens). What we identify as the arclet has a notably different color from (is bluer than) these two faint objects. Therefore, we consider it unlikely that these objects and the arclet are from the same source. While it is possible that the two objects to the SE and NW belong to another source (i.e., doubly imaged) the object to the NW is too faint for us have a high level of confidence in this scenario. There may also be a very faint object to the SW. However, in our view, it is too faint to be considered in the grading of this system. We therefore assign a grade of C to this system.

DESI-153.6301+24.4854 (second row, third column) also shows what appears to be an arclet. While the surface brightness and color are consistent with the expectation of a lensed arc, there is no apparent counterimage and the shape of the arc does not conform to the typical morphology (there appears to be a "kink" in the putative arc, possibly indicating two galaxies in close proximity on the sky). It is also possible that the shape of the arc may be due to source structure. Given the limitations of this ground-based image and the absence of a clearly identifiable counterimage, we assign a grade of C.

DESI-066.4365-57.6926 (first row, fourth column) is a type of image that we have seen in our searches a number of times. If it is lensing, this would be a perfect Einstein ring. Since such perfect alignment is very rare even among strong-lensing systems, we are generally skeptical. Lensed arcs typically have low surface brightness, whereas this putative ring appears quite bright. The structure seen in the putative ring is, in our experience, consistent with a ring galaxy (e.g., Timmis & Shamir 2017). Furthermore, if this system is indeed a strong lens, the Einstein radius, $\theta_E \sim 2\farcs8$. Using a lens redshift of $z_d \sim 0.4$ (photometric redshift; Zhou et al. 2021), and source redshift of $z_s \sim 0.8$ (assumed to be $2 \times z_d$; e.g., Sharma & Linder 2022), we can then estimate the mass enclosed within the Einstein radius as well as the velocity dispersion (σ_v) of the lens. Assuming an singular isothermal sphere profile, we find the enclosed mass to be $\sim 2 \times 10^{12}\ M_\odot$ and $\sigma_v \sim 470$ km s^{-1}. Compared with known lenses in the SLACS program with similar r-band magnitude and lens and source redshifts, these values are extraordinarily large. Finally, the galaxy does not appear to be a part of a group. We therefore assign a grade of D.

For DESI-035.4194-46.8752 (second row, fourth column), we consider the putative arc to be more likely a spiral galaxy (perhaps with a high inclination angle), with what appears to be

This Work / S21	A (115)		B (526)		C (1254)		ABC (1895)	D (1865)		ABC+D
A (404)		10.9%		14.9%		8.9%	34.7%		8.7%	43.3%
	38.3%	44	11.4%	60	2.9%	36	140	1.9%	35	175
B (788)		0.4%		4.7%		8.9%	14.0%		14.3%	28.3%
	2.6%	3	7.0%	37	5.6%	70	110	6.1%	113	223
AB (1192)	40.9%	47	18.4%	97	8.5%	106	250	8.0%	148	398

Figure 10. Comparison of overlapping candidates discovered in S22 and our work. The number of candidates from S22 for each of the two grades are shown in red. Also in red are the overlapping systems by grade as a percentage of the number for the respective grade in S22. Conversely, the number of candidates from our search are shown in blue. Also in blue are the overlapping systems by grade as a percentage of the number in the respective grade in our work. For example, 44 are assigned Grade A by this work and S22, which is 10.9% of the total number of Grade As (404) in S22 and 38.3% of the Grade As reported in this work.
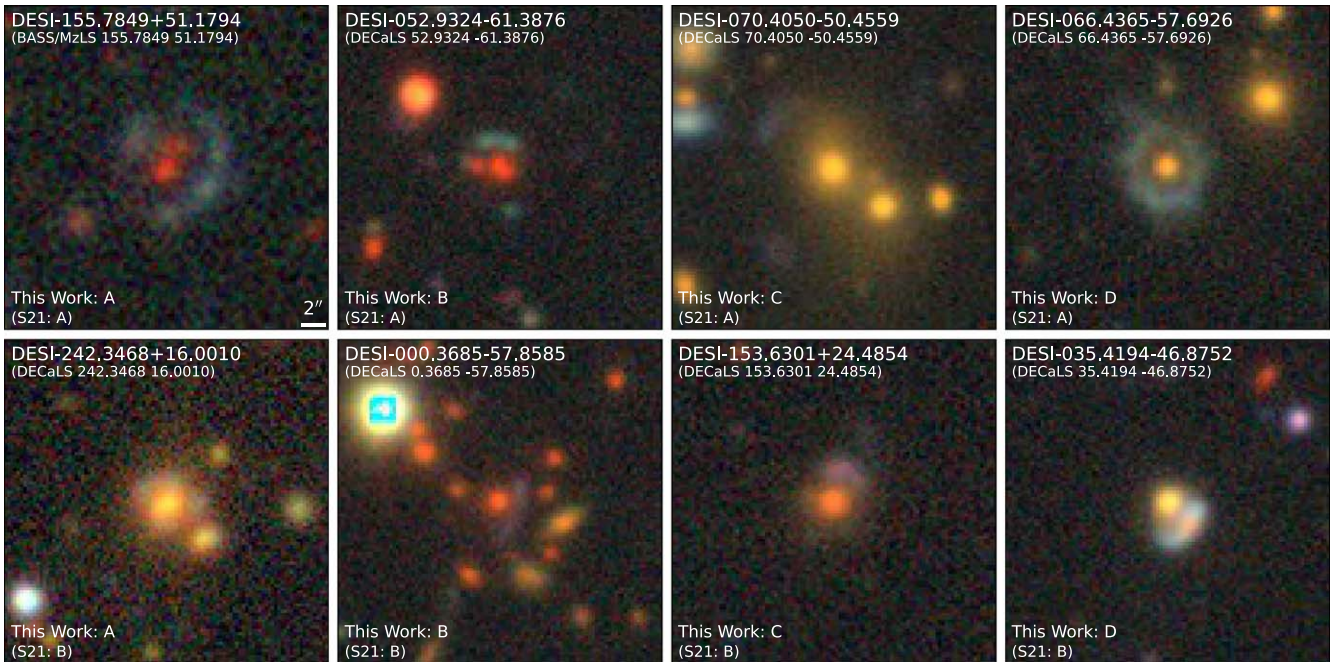
Figure 11. A comparison of a sample of systems reported in both S22 and this work. For all images, north is up, and east to the left. The scale bar shown in the first image applies to all images in this figure. We show an example system for all permutations of grade assignment given Grades A and B from S22 and Grades A, B, C, and D in this work. The name for each system from this work is shown in the top left, with the name from S22 in parentheses below that. The bottom left shows the grade from this work and below that, the S22 grade in parentheses.

a reddish core and blue spiral arms. In this and other select ambiguous cases, photometric redshifts (Zhou et al. 2021) can be helpful. The photometric redshift of the putative arc at 0.268 ± 0.047 relative to that of the putative lens (the orange galaxy at the center of the image) at 0.248 ± 0.038 is consistent with this not being a lensing system. While photometric redshifts of elliptical galaxies are often reliable, this is not always the case with blue lensed sources. The photometric redshift of the putative arc in DESI-035.4194-46.8752 is centered on the orange core and is likely more reliable. Hence we assign a grade of D.

In all four cases, we deem the possibility of lensing to be present (just not high enough to be assigned above Grade C), high-resolution images and/or spectroscopic observations are needed to be conclusive.

Finally, the technique utilized by S22 is more computationally expensive (64 GPU hours, NVIDIA V100, with an additional 288 hr for their "fine-tuned" approach) than our supervised approach (5 GPU hours, NVIDIA Tesla P100). On the other hand, the S22 approach has the advantage of facilitating searches for other kinds of objects (not just strong-lensing systems) downstream. Given future improvements on GPUs is almost a certainty, this is clearly is a promising strategy.

The discoverability of new lenses, by humans alone or with machine learning assistance, of course depends on observation conditions (e.g., depth, seeing, and filter selection) and the criteria for the deployment sample. As highlighted by the brief comparison between S22 and this work, even for the same survey, different well-designed machine-learning-based searches can lead to different discovery sets. In general, this likely is dependent on the machine learning technique employed (e.g., neural network architecture, supervised versus unsupervised), training sample (e.g., selection criteria, size, and whether it consists of simulations, observed images, or an amalgamation), and the final human inspection process itself (Rojas et al. 2023). These can lead to overlapping and complementary discoveries. A more detailed comparison between different searches and their respective results would likely be beneficial to future lens searches. This, however, is beyond the scope of this paper.

6. Conclusions

We have carried out a search for strong gravitational lensing systems in the DESI Legacy Surveys DR9. This is the third paper in a series on lens searches in the DESI Legacy Surveys, following Paper I (Huang et al. 2020, DR7 DECaLS, DEV, and COMP only) and Paper II (Huang et al. 2021, DR8, DEV, COMP, and REX). We use a customized deep residual neural network (Lanusse et al. 2018; Huang et al. 2021), trained on observed lenses and nonlenses. We apply our trained neural network to ∼45 million non-PSF (SER, DEV, REX, and EXP) cutout images with at least three passes in each of the grz bands and a z-band magnitude cut of <20.0 for the galaxy at the center of each image. We hold a high standard in grading these candidate systems. We have found 115 Grade A, 526 Grade B, and 1254 Grade C candidates, for a total of 1895. Of these, 125 were previously known systems, and 258 were also reported in recent publications, the majority of which came from S22. We therefore report 1512 new lens candidates with a grade breakdown of 51 Grade A, 383 Grade B, and 1079 Grade C. Combining all three searches, we have found a grand total of 3057 strong lens candidates. Grade D systems are not counted as candidates in this paper, but we have included them on our project website (URL provided in Section 4.2).

Along with S22, this is the largest deployment to search for strong lenses in the most heterogeneous data set. Even more notably, in all parts of the footprint of our search, multiple lens searches were performed before, by our group or other groups. By using an expanded, statistically representative training sample, we are able to find a large number of high-quality new lens candidates, with high purity. We have presented a brief comparison with S22 in Section 5. Their results and ours show that a detailed comparison will likely benefit future lens searches in large data sets.

This work, together with other searches, have clearly demonstrated that machine learning approaches are highly effective in discovering large numbers of high-quality strong-lensing candidates. This will likely continue to be the case in future surveys such as the Vera C. Rubin Observatory Legacy Survey of Space and Time, Euclid, and the Nancy Grace Roman Space Telescope, and the discovery of strong lenses will continue to accelerate. There has also been significant development on strong lens modeling, for example, the availability of open-source, widely used lens modeling packages (e.g., Birrer & Amara 2018; Nightingale et al. 2019). Recently, Gu et al. (2022) introduced GIGA-Lens, which takes advantage of the high level of parallelization of GPUs and automatic differentiation, speeding up lens modeling by 1 to 2 orders of magnitudes. This makes it possible to model the $\mathcal{O}(10^5)$ strong-lensing systems expected to be discovered in the next decade in a reasonable amount of time. The immediate future of using strong lensing to address significant astrophysical and cosmological questions is indeed very bright.

ORCID iDs

C. Storfer ● https://orcid.org/0000-0002-0385-0014
X. Huang ● https://orcid.org/0000-0001-8156-0330
A. Gu ● https://orcid.org/0000-0003-2748-7333
W. Sheu ● https://orcid.org/0000-0003-1889-0227
A. Dey ● https://orcid.org/0000-0002-4928-4003
J. Inchausti Reyes ● https://orcid.org/0009-0009-8667-763X
K. J. Kwon ● https://orcid.org/0000-0001-9802-362X
D. Lang ● https://orcid.org/0000-0002-1172-0754
A. Meisner ● https://orcid.org/0000-0002-1125-7384
J. Moustakas ● https://orcid.org/0000-0002-2733-4559
E. F. Schlafly ● https://orcid.org/0000-0002-3569-7421
D. J. Schlegel ● https://orcid.org/0000-0002-5042-5088

References

Birrer, S., & Amara, A. 2018, PDU, 22, 189
Birrer, S., Shajib, A. J., Galan, A., et al. 2020, A&A, 643, A165
Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., & Moustakas, L. A. 2006, ApJ, 638, 703
Bolton, A. S., Burles, S., Koopmans, L. V. E., et al. 2008, ApJ, 682, 964
Cañameras, R., Schuldt, S., Shu, Y., et al. 2021, A&A, 653, L6
Cañameras, R., Schuldt, S., Suyu, S. H., et al. 2020, A&A, 644, A163
Carrasco, M., Barrientos, L. F., Anguita, T., et al. 2017, ApJ, 834, 210
Chen, W., Kelly, P. L., Oguri, M., et al. 2022, Natur, 611, 256
Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270
Dey, A., Rabinowitz, D., Karcher, A., et al. 2016, Proc. SPIE, 9908, 99082C
Dey, A., Schlegel, D. J., Lang, D., et al. 2019, AJ, 157, 168
Diehl, H. T., Buckley-Geer, E. J., Lindgren, K. A., et al. 2017, ApJS, 232, 15
Ding, X., Liao, K., Birrer, S., et al. 2021, MNRAS, 504, 5621
Fagin, J., Vernardos, G., Tsagkatakis, G., et al. 2024, MNRAS, 532, 2248
Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, AJ, 150, 150
Freedman, W. L., Madore, B. F., Hoyt, T., et al. 2020, ApJ, 891, 57
Grillo, C., Suyu, S. H., Rosati, P., et al. 2015, ApJ, 800, 38
Gu, A., Huang, X., Sheu, W., et al. 2022, ApJ, 935, 49
Hezaveh, Y. D., Dalal, N., Marrone, D. P., et al. 2016, ApJ, 823, 37
Huang, X., Morokuma, T., Fakhouri, H. K., et al. 2009, ApJL, 707, L12
Huang, X., Storfer, C., Gu, A., et al. 2021, ApJ, 909, 27
Huang, X., Storfer, C., Ravi, V., et al. 2020, ApJ, 894, 78
Huber, S., Suyu, S. H., Noebauer, U. M., et al. 2021, A&A, 646, A110
Jacobs, C., Collett, T., Glazebrook, K., et al. 2019a, MNRAS, 484, 5330
Jacobs, C., Collett, T., Glazebrook, K., et al. 2019b, ApJS, 243, 17
Jacobs, C., Glazebrook, K., Collett, T., More, A., & McCarthy, C. 2017, MNRAS, 471, 167
Jaelani, A. T., More, A., Oguri, M., et al. 2020, MNRAS, 495, 1291
Kochanek, C. S. 1991, ApJ, 373, 354
Koopmans, L. V. E., & Treu, T. 2002, ApJL, 568, L5
Koopmans, L. V. E., Treu, T., Bolton, A. S., Burles, S., & Moustakas, L. A. 2006, ApJ, 649, 599
Lang, D., Hogg, D. W., & Mykytyn, D., 2016 The Tractor: Probabilistic Astronomical Source Detection and Measurement, Astrophysics Source Code Library, ascl:1604.008
Lanusse, F., Ma, Q., Li, N., et al. 2018, MNRAS, 473, 3895
Li, R., Napolitano, N. R., Spiniello, C., et al. 2021, ApJ, 923, 16
Linder, E. V. 2016, PhRvD, 94, 083510
Metcalf, R. B., Meneghetti, M., Avestruz, C., et al. 2019, A&A, 625, A119
Millon, M., Galan, A., Courbin, F., et al. 2020, A&A, 639, A101
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
Nightingale, J. W., Massey, R. J., Harvey, D. R., et al. 2019, MNRAS, 489, 2049
Nordin, J., Rubin, D., Richard, J., et al. 2014, MNRAS, 440, 2742
O'Donnell, J. H., Wilkinson, R. D., Diehl, H. T., et al. 2022, ApJS, 259, 27
Oguri, M., & Marshall, P. J. 2010, MNRAS, 405, 2579
Pascale, M., Frye, B. L., Pierel, J. D. R., et al. 2024, arXiv:2403.18902
Patel, B., McCully, C., Jha, S. W., et al. 2014, ApJ, 786, 9
Petrillo, C. E., Tortora, C., Vernardos, G., et al. 2019, MNRAS, 484, 3879
Pierel, J. D. R., Arendse, N., Ertl, S., et al. 2023, ApJ, 948, 115
Pierel, J. D. R., Newman, A. B., Dhawan, S., et al. 2024, ApJL, 967, L37
Pierel, J. D. R., & Rodney, S. 2019, ApJ, 876, 107
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6
Pourrahmani, M., Nayyeri, H., & Cooray, A. 2018, ApJ, 856, 68
Refsdal, S. 1964, MNRAS, 128, 307
Riess, A. G., Casertano, S., Yuan, W., et al. 2021, ApJL, 908, L6
Rodney, S. A., Strolger, L. G., Kelly, P. L., et al. 2016, ApJ, 820, 50
Rojas, K., Collett, T. E., Ballard, D., et al. 2023, MNRAS, 523, 4413
Rojas, K., Savary, E., Clément, B., et al. 2022, A&A, 668, A73
Rubin, D., Hayden, B., Huang, X., et al. 2018, ApJ, 866, 65
Savary, E., Rojas, K., Maus, M., et al. 2022, A&A, 666, A1
Sharma, D., & Linder, E. V. 2022, JCAP, 2022, 033
Sheu, W., Huang, X., Cikota, A., et al. 2023, ApJ, 952, 10
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., Gavazzi, R., Suyu, S. H., Treu, T., & Marshall, P. J. 2013, ApJ, 777, 97
Sonnenfeld, A., Verma, A., More, A., et al. 2020, A&A, 642, A148
Stein, G., Blaum, J., Harrington, P., Medan, T., & Lukic, Z. 2022, ApJ, 932, 107
Suyu, S. H., Huber, S., Cañameras, R., et al. 2020, A&A, 644, A162
Szegedy, C., Liu, W., Jia, Y., et al. 2014, arXiv:1409.4842
Talbot, M. S., Brownstein, J. R., Dawson, K. S., Kneib, J.-P., & Bautista, J. 2021, MNRAS, 502, 4617
Timmis, I., & Shamir, L. 2017, ApJS, 231, 2
Treu, T. 2010, ARA&A, 48, 87
Vegetti, S., & Koopmans, L. V. E. 2009, MNRAS, 392, 945
Wagner-Carena, S., Aalbers, J., Birrer, S., et al. 2023, ApJ, 942, 75
Williams, G. G., Olszewski, E., Lesser, M. P., & Burge, J. H. 2004, Proc. SPIE, 5492, 787
Wong, K. C., Chan, J. H. H., Chao, D. C. Y., et al. 2022, PASJ, 74, 1209
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
Yahalomi, D. A., Schechter, P. L., & Wambsganss, J. 2017, arXiv:1711.07919
Zhou, R., Newman, J. A., Mao, Y.-Y., et al. 2021, MNRAS, 501, 3309