

Strong Lens Discoveries in DESI Legacy Imaging Surveys DR10 with Two Deep Learning Architectures

JOSE CARLOS INCHAUSTI,¹ CHRISTOPHER J. STORFER,^{2,3} XIAOSHENG HUANG,^{4,3} YUAN-MING HSU,⁵ BRANDT KAUFMANN,⁶
 CHAITANYA PASUPALA,⁷ S. BANKA,⁸ A. DEY,⁹ D. LANG,¹⁰ A. MEISNER,⁹ J. MOUSTAKAS,¹¹ A.D. MYERS,¹²
 E. F. SCHLAFLY,¹³ AND D.J. SCHLEGEL³

¹Department of Physics & Astronomy, University of San Francisco, San Francisco, CA 94117, USA

²Institute for Astronomy, University of Hawai‘i, 2680 Woodlawn Drive, Honolulu, HI 96822, USA

³Physics Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA, 94720

⁴Department of Physics & Astronomy, University of San Francisco, San Francisco, CA 94117-1080

⁵Department of Physics, National Taiwan University, No. 1, Section 4, Roosevelt Road, Taipei 106319, Taiwan

⁶Department of Mathematics & Statistics, University of San Francisco, San Francisco, CA 94117, USA

⁷Department of Economics, University of San Francisco, San Francisco, CA 94117, USA

⁸Department of Electrical Engineering & Computer Sciences, University of California, Berkeley, Berkeley, CA 94720

⁹NSF’s National Optical-Infrared Astronomy Research Laboratory, 950 N. Cherry Ave., Tucson, AZ 85719

¹⁰Perimeter Institute for Theoretical Physics, Waterloo, ON N2L 2Y5, Canada

¹¹Department of Physics and Astronomy, Siena College, 515 Loudon Rd., Loudonville, NY 12211

¹²Department of Physics & Astronomy, University of Wyoming, 1000 E. University, Dept 3905, Laramie, WY 82071

¹³Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

ABSTRACT

We have conducted a search for strong gravitational lensing systems in the Dark Energy Spectroscopic Instrument (DESI) Legacy Imaging Surveys Data Release 10 (DR10). This paper is the fourth in a series of searches (following Huang et al. 2020; Huang et al. 2021; Storfer et al. 2024, Paper I, II, & III respectively). This is the first catalog of lens candidates covering nearly the entirety of the extragalactic sky south of declination $\delta \approx +32^\circ$, all of it observed by the DECam, covering $\sim 14,000$ deg². We impose a z -band magnitude cut of < 20 in AB magnitude. We deploy a Residual Neural Network and EfficientNet as an ensemble trained on a compilation of known lensing systems and high-grade candidates as well as nonlenses in the same footprint. The predictions from these two base models are aggregated using a meta-learner. After applying our ensemble to the survey data, we exclude known candidates and systems, and use our own visual inspection portal to rank images in the top 0.01 percentile of all neural network recommendations. We have found 811 new lens candidates. These include 484 new candidates in the Legacy Surveys DR9 footprint, all parts of which have been searched for strong lenses at least once before, either by our group or others. Combining the discoveries from this work with those from Paper I (335), II (1210), and III (1512), we have discovered a total of 3868 new candidates in the DESI Legacy Surveys.

Keywords: galaxies: high-redshift – gravitational lensing: strong

1. INTRODUCTION

Strong gravitational lensing systems are a powerful tool for cosmology. They have been used to study dark matter in galaxies and clusters (e.g., Kochanek 1991; Broadhurst et al. 2000; Koopmans & Treu 2002; Bolton et al. 2006; Koopmans et al. 2006; Massey et al. 2010; 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 (CDM) model beyond the local universe (e.g., Vegetti & Koopmans 2009; Huang et al. 2009; Vegetti

et al. 2010; Meneghetti et al. 2020, 2023; Çağan Şengül et al. 2021; Wagner-Carena et al. 2022; Fagin et al. 2024). Searches and spectroscopic follow-up of lensed quasars (e.g., Dawes et al. 2023; Sheu et al. 2024; He et al. 2023) have enabled time-delay cosmography, offering an independent route to measure H_0 . Multiply lensed supernovae (SNe) are also ideal for measuring time delays and H_0 because of their well-characterized light curves, and in the case of Type Ia, with the added benefit of standardizable luminosity (e.g., Refsdal 1964; Treu 2010; Oguri & Marshall 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 supernovae, 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. Sheu et al. (2023) conducted a retrospective search for strongly lensed supernovae 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 supernovae (e.g., Pierel & Rodney 2019; Suyu et al. 2020; Huber et al. 2021; Pascale et al. 2025), combined with measurements from distance ladders (e.g., Freedman et al. 2020; Riess et al. 2021) and lensed quasars (e.g., Wong et al. 2019; Millon et al. 2020; Birrer 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 (CMB; Planck Collaboration et al. 2020). In addition, magnified (but not multiply-lensed) SNe Ia were identified (Patel et al. 2014; Nordin 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., Jullo et al. 2010; Li et al. 2024), especially using systems with multiple source planes (e.g., Collett & Auger 2014; Linder 2016; Sharma & Linder 2022; Caminha et al. 2022).

The introduction of neural networks to identify gravitational lens candidates in imaging surveys has been transformative and continues to show success in identifying lensing candidates (e.g., Jacobs et al. 2017; Metcalf et al. 2018; Jacobs et al. 2019a,b; Cañameras et al. 2020, 2021; Rojas et al. 2022; Shu et al. 2022; Zaborowski et al. 2023; González et al. 2025). In our recent work, we discovered over 3000 new strong lenses (Huang et al. 2020; Huang et al. 2021; Storfer et al. 2024, hereafter referred to as Papers I, II, and III, 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 fourth search for strong lenses in the Legacy Surveys, now using DR10. We provide an overview of the observations in §2. In §3, we describe the construction of the training sample, our neural network ensemble, and our optimized visual inspection method. New lens candidates are presented in §4. We discuss our discoveries in §5 and conclude in §6.

2. DATA

2.1. *The Tractor*

First, we briefly describe *The Tractor* (Lang et al. 2016), 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 25¹. *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 de Vaucouleurs (DEV; Sérsic index $n = 4$) or 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 9 over DEV/EXP.

2.2. *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 $\sim 9000 \text{ deg}^2$ of the sky in the range of $-18^\circ \lesssim \delta \lesssim +32^\circ$. As in DR9, DR10 also contains DECam data reprocessed from the Dark Energy Survey (DES; Dark Energy Survey Collaboration et al. 2016) for $\delta \lesssim -18^\circ$, which provides a footprint of $\sim 5000 \text{ deg}^2$. The newly observed area in DR10 consists of other DECam surveys: the DELVE (Drlica-Wagner et al. 2021) and DeROSITAS (Zenteno et al. 2025) programs. The DECam surveys will hereafter be referred to in their entirety as DECaLS.² BASS/MzLS are observed in the g and r bands by the 90Prime camera (Williams et al. 2004) on the Bok 2.3-m

¹ For more details, see <https://www.legacysurvey.org/dr10/description/>

² Note that the footprint of DECaLS has expanded from DR7 to DR10, as we show later in this work.

telescope and in the z -band by the Mosaic3 camera (Dey et al. 2019) on the 4-meter Mayall telescope. Together, BASS/MzLS cover the same ~ 5000 deg 2 of the northern subregion of the Legacy Surveys. Since DR10 includes new imaging data only from DECam, and not in this northern sub-region, *this work performs a lens search only in DECaLS*.

The Legacy Surveys is imaged with a median 5σ PSF depths of 23.95 (g), 23.54 (r), and 22.50 (z) AB mag. The median FWHMs for the delivered images are: 1.29'' (g), 1.18'' (r), and 1.11'' (z) for DECaLS (Dey et al. 2019).

Figure 1 displays the z -band depth map of objects classified as SER with $z < 20.0$ mag in the Legacy Survey DR10. The figure also outlines the coverage of the DECaLS footprint in DR9, illustrating the overlapping regions as well as the additional areas newly observed in DR10. 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 the DECaLS region.

Table 1. Object Counts in DECaLS DR10 $z < 20$ mag

SER	DEV	REX	EXP	Total by Region
21,482,748	7,774,926	9,514,266	4,774,571	43,546,511

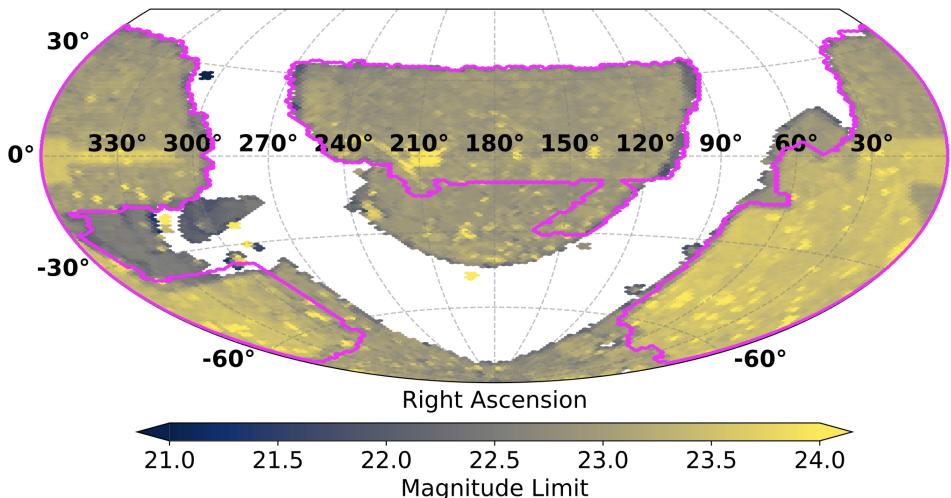


Figure 1. The DECaLS survey footprint in DR10 shown in an equal-area Aitoff projection in equatorial coordinates using a hexagonal binning format. The color gradient of the hexagonal bins represents the average z -band 5σ magnitude limit for galaxies (0.45'' round exponential profile) within a given bin. These averages are computed from randomly sampled points across the DR10 DECaLS footprint. The DR9 DECaLS footprint is outlined in purple for reference.

3. THE TRAINING SAMPLE, NEURAL NETWORK ENSEMBLE, AND DEPLOYMENT

3.1. Training Sample

As in Papers I, II and III, we continue to use observed images of both lenses and nonlenses for training. Our training sample consists of 1372 lenses and high quality candidates, 869 of which come from Paper I, II, and III, with the rest selected from the following searches arranged by publication date: Zitrin et al. (2012); More et al. (2012); Anguita et al. (2012); Brownstein et al. (2012); Gavazzi et al. (2014); More et al. (2016); Carrasco et al. (2017); Diehl et al. (2017); Wong et al. (2018); Johnson et al. (2018); Sonnenfeld et al. (2018, 2019); Jacobs et al. (2019b); Petrillo et al. (2019); Jaelani et al. (2020); Cañameras et al. (2020); Sonnenfeld et al. (2020); Chan et al. (2020); Stein et al. (2022); Talbot et al. (2021); Rojas et al. (2021); Cañameras et al. (2021); Li et al. (2021); Jaelani et al. (2021); Wong et al. (2022); O'Donnell et al. (2022); Shu et al. (2022).

To select the lenses for the training sample, we inspect the $\sim 12,000$ known lenses and candidates reported in literature that overlap with the DECam ($\delta < +32^\circ$). We apply a uniform set of selection criteria (similar to those in Paper I, II, and III) 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.

We apply cuts to the training sample based on z -band magnitude ($z < 20.0$ mag), location ($\delta < +32^\circ$), and depth (at least three exposures³ in the g , r , and z bands). The z -band magnitude cut was initially tested in Paper I, where we discovered that including objects fainter than the 20th magnitude yields diminishing returns for identifying lens candidates.

Paper III excluded EXP objects from the training sample since lens galaxies rarely fall under that type. However, the model in Paper III was able to perform out-of-sample classifications and correctly identified EXP candidates during deployment that meet the selection criteria for the training sample in this work. Hence, we now include all types (SER, DEV, REX, and EXP) in the training sample. Table 2 breaks down the counts per *The Tractor* type present in the training sample.

Table 2. Lenses Counts in the Training Sample

SER	DEV	REX	EXP
960	266	112	34

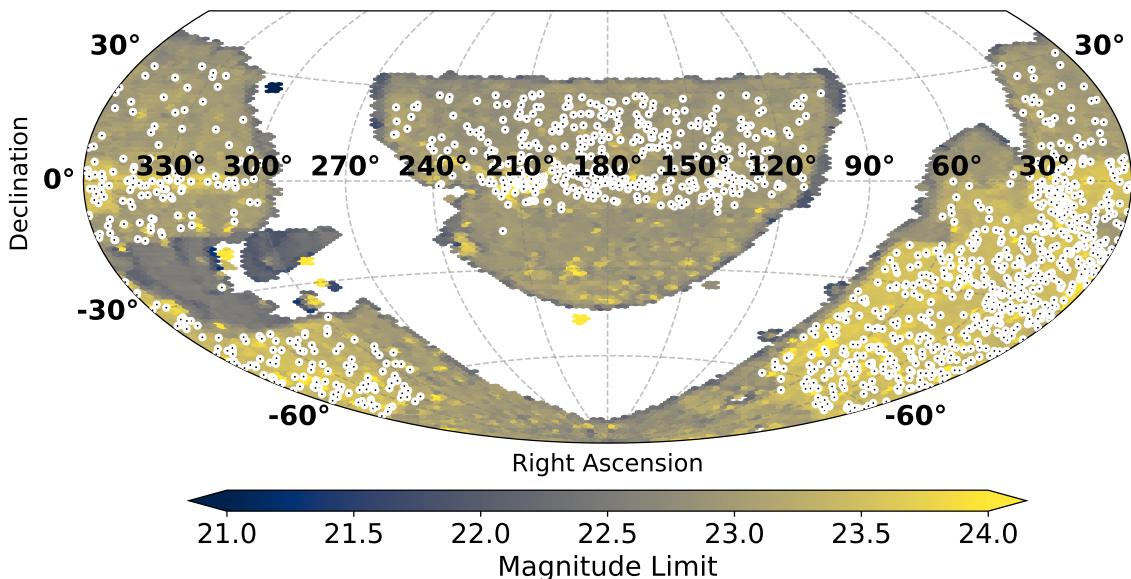


Figure 2. The 1372 lenses included in the training sample over the depth map of the Legacy Surveys DR10 shown in Figure 1.

For the selection of nonlenses, for each type (SER, DEV, REX) in DECaLS, we bin by the number of exposures in z -band. This is to prevent potential bias by the neural net based on the number of exposures (see Paper II). We then select nonlenses randomly, but keep the proportionality to the lenses the same in each bin ($\sim 100:1$). This ratio has been steadily increased from Paper I through Paper III (where this ratio was $\sim 33:1$ in Paper III) to better reflect the expected nonlens-to-lens ratio during deployment, which is estimated to be $\sim 10,000:1$. We cross-reference this nonlens selection with the $\sim 12,000$ known lenses and candidates found in the literature to remove potential contaminants

³ Papers I-III refer to “exposures” as “passes”. In this work, we use “exposures” instead for consistency with Legacy Survey documentation.

(i.e., cutouts that contain a lens or a lens candidate). This is made possible implementing `spherimatch`⁴ (Hsu 2025), a package intended for cross-referencing catalogs in spherical coordinates. The package has a time complexity of $\mathcal{O}(N \log N)$, making it well-suited for handling large catalogs, as required in this work. In an effort to remove possible additional contamination, we deploy the trained ResNet from Paper III to this selection and inspect the images that receive a probability greater than 0.4 (the same threshold used in Paper III). After visually inspecting the recommendations from this model, we remove 9 potential contaminants from this selection. The contaminants we remove are comparable to or better than Grade D in Paper III (also, see § 3.3.1). This results in a total of 134,182 nonlenses in the training sample. This is 10 \times larger than training sample in our first lens search (Paper I).

The final dataset of lenses and nonlenses is randomized and divided into 70% for training and 30% for validation, while maintaining the $\sim 100:1$ nonlens-to-lens ratio in each split. For both the training and validation catalogs, we extract image cutouts of 101×101 pixels, corresponding to an angular size of $\sim 26'' \times 26''$. The exact same training and validation sets are used to train our two models described below in § 3.2.

3.2. The Neural Network Models

In this section, we describe the two neural network models used in our lens search, a residual network (ResNet; § 3.2.1) and a so-called “EfficientNet” (§ 3.2.2).

3.2.1. ResNet

The “shielded” residual network (ResNet) model, as described in Paper II, is an improvement upon the architecture presented in Lanusse et al. (2018, L18). The “shielding” layers (Szegedy et al. 2014) perform 1×1 convolutions that reduce the dimensionality (i.e., the number of layers) of the output from each ResNet block.

We use the cross-entropy loss function,

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

where y_i is the label for the i th 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 batch size, given below). During training, the ResNet attempts to minimize the loss while the Area Under the Curve (AUC) metric is used throughout training to identify the best-performing model. Figure 3 shows the loss minimization process alongside the Receiver Operating Characteristic (ROC) curve for the model that achieved the highest validation AUC during training.

The ResNet model contains a total of 194,433 trainable parameters and was trained with a batch size of 2048 ($\times 4$ more than the EfficientNet, § 3.2.2), an initial learning rate (LR) of 5×10^{-4} , and a learning rate decay schedule that reduced the LR by half at the 80th epoch. The ResNet is trained on 4 NVIDIA A100 GPUs on one GPU node on the *Perlmutter* supercomputer at the National Energy Research Scientific Computing Center (NERSC)⁵. Using TensorFlow’s MirroredStrategy, training on 4 GPUs is straightforward. The multi-GPU implementation results in a speed gain of 3.7 \times (remarkably close to the naive expectation of 4 \times): the training time per epoch is reduced from 78 to 21 seconds.

In addition to the cross-entropy loss, we further assess the performance of our trained ResNet using the ROC, which plots the True Positive Rate (TPR) against the False Positive Rate (FPR), where a “Positive” (P) indicates a lens and a “Negative” (N) a nonlens. The definitions of TPR and FPR are given as:

$$TPR = \frac{TP}{P} = \frac{TP}{TP + FN} \quad (2)$$

and

$$FPR = \frac{FP}{N} = \frac{FP}{FP + TN} \quad (3)$$

where True Positives (TP) are lenses correctly identified, False Positives (FP) are nonlenses incorrectly identified as lenses, True Negatives (TN) are nonlenses correctly rejected, and False Negatives (FN) are lenses incorrectly rejected.

⁴ <https://github.com/technic960183/spherimatch>

⁵ A U.S. Department of Energy Office of Science user facility, <https://www.nersc.gov/>.

Random classifications produce a diagonal line on the ROC curve with an AUC of 0.5, while a perfect classifier achieves an AUC of 1. During training, the validation AUC is calculated at each epoch, and the model with the highest validation AUC is selected for deployment.

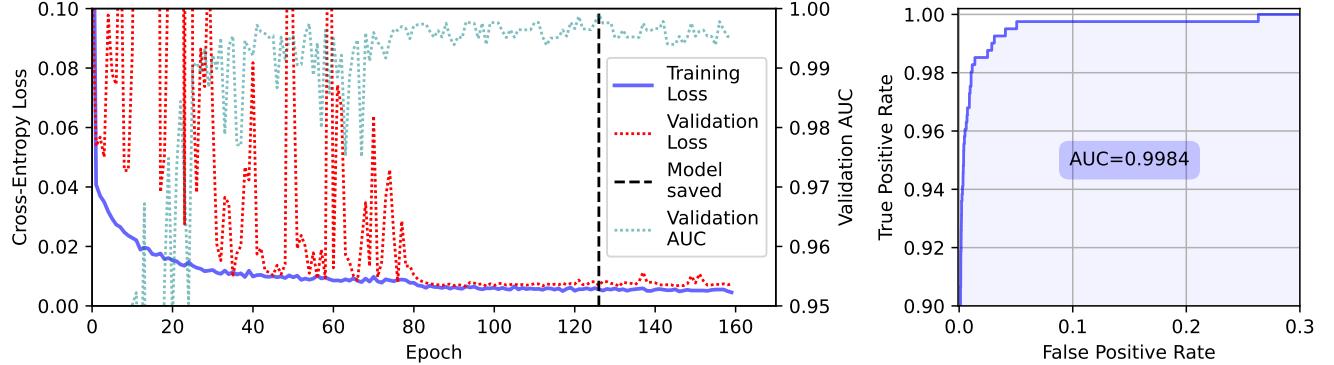


Figure 3. Left: Cross-entropy loss for the training and validation sets (left y-axis) and validation AUC (right y-axis) vs. training epochs. The vertical dashed black line at epoch 126 marks the point where the model achieved its highest AUC at 0.9984. This trained model was selected for deployment. Right: ROC curve for the validation set using the best-performing model from epoch 126.

3.2.2. EfficientNet Model

For the first time with respect to Papers I through III, we integrate the EfficientNetV2, referred to in this work as the “EfficientNet,” which improved the EfficientNetV1 presented in Tan & Le (2019). EfficientNetV1 comes from the search for a principled method for scaling Convolutional Neural Network (CNN) parameters (width of layers, depth of layers, resolution scaling) in order to achieve high accuracy, even with a small model that can be run on mobile devices. The primary building block for EfficientNetV1 is the mobile inverted bottleneck MBConv from Sandler et al. (2018), with a squeeze and excitation layer. The MBConv block’s input and output are thin “bottle-necked” layers as opposed to the standard residual block, which uses expanded representations in the input. In an MBConv block, the input is first expanded with a 1×1 convolution which increases the number of channels. Lightweight depthwise convolution is performed on each channel of the expanded input separately, followed by a pointwise (1×1) convolution, which builds features via linear combinations of channels. The output is sent into a squeeze-and-excitation block, an attention mechanism which gives weights to each feature map. The squeeze and excitation block’s output is put through another pointwise (1×1) convolution, this time to reduce the number of channels back down (This is similar to the “shielding” layers we added to the ResNet model; § 3.2.1). Compared to a standard convolution, this process requires less memory and fewer computations. However, because depthwise convolutions suffer from poor parallelization, they are slower to train using modern GPUs. To improve speed, EfficientNetV2 introduces Fused-MBConv blocks, which replace the expansion and depthwise convolution with a regular 3×3 convolution. While this is more expensive in terms of parameters and computations, because of parallelization, using FusedMBConv in early layers results in a $\sim 40\%$ increase in training speed when using a GPU, and only a $\sim 4\%$ increase in parameters compared to EfficientNetV1.

Unlike our ResNet model, the EfficientNet was pre-trained. In this work, we fine-tune the EfficientNet model using the same cross-entropy loss function as for the ResNet (Equation 1). The training results are shown in Figure 4.

The EfficientNet model contains a total of 20,542,883 trainable parameters and was trained with a batch size of 512, an initial learning rate (LR) of 3.88×10^{-4} , and a learning rate decay schedule that reduced the LR by half at the 130th epoch. Training was also conducted on 4 GPUs at NERSC (the same as in § 3.2.1). With a per GPU batch size of 512, the training time per epoch improves from 72 to 23 seconds, or, an approximately $3 \times$ speed gain, slightly less than the naive expectation of $4 \times$. This results in a 3 hour reduction of training time, from 4.7 hours to train a 160 epoch model down to 1.3 hours.

Figure 4 shows the AUC and loss progression for the training and validation sets, and the ROC curve for our best-performing model, with an AUC of 0.9987, which is comparable to the ResNet’s best AUC.

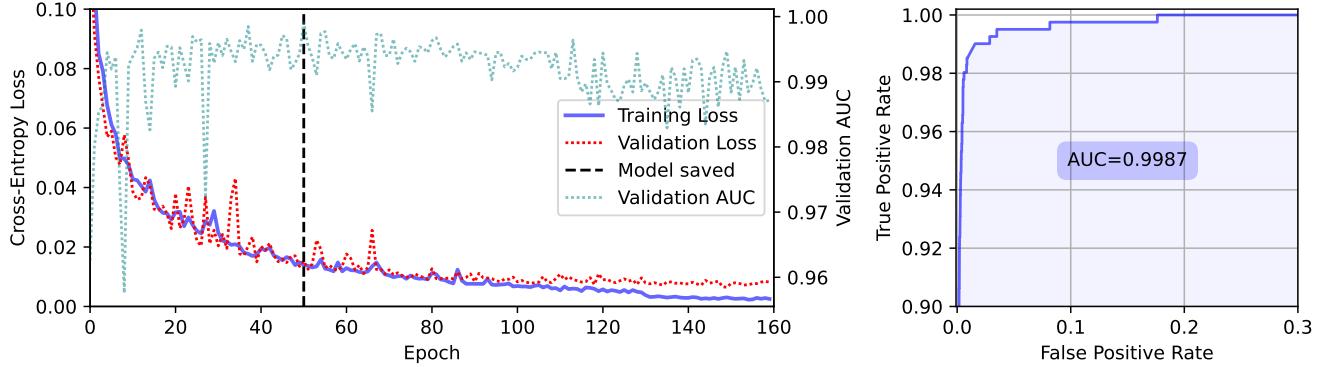


Figure 4. Left: Cross-entropy loss for the training and validation sets (left y-axis) and validation AUC (right y-axis) vs. the training epochs. The vertical dashed black line at epoch 50 marks the point where the model achieved its highest AUC. This trained model was selected for deployment. Right: ROC curve for the validation set using the best-performing model from epoch 50.

The divergent behavior of the loss and AUC curves in Figure 4 is worth noting. After epoch 50, where the model achieves the highest validation AUC, the training and validation loss continue to improve (i.e., decrease), unlike the validation AUC, which deteriorates. The validation AUC fluctuates near 0.995 before starting to decline around epoch 80. Figure 5 addresses this discrepancy by comparing the probability distributions of an EfficientNet at 50 epochs vs. at 160 epochs, showing that the 160-epoch model puts 283 more images in its smallest prediction probability bin (0, 0.04) than the 50-epoch model. Due to the significant class imbalance in the training sample (negative examples outnumber positive examples by 100:1), the model is able to achieve an overall lower loss by pushing these images into the lowest probability bin. As indicated by the lower panel of Figure 6, most of the additional 283 images came from images assigned a probability in the adjacent (low) probability bins. The vast majority of images that were not in the top or bottom bin (predicted probability between 0.04 and 0.96) were already correctly identified as TNs by the 50-epoch model (300 of 404 images were TNs). Thus, pushing them into the lowest bin improves (lowers) the loss but does not improve the AUC. On the other hand, 104 of these images are TPs, so pushing them to the lowest bin *should* negatively affect our performance. The lowest bin of the 50-epoch model is 99.98% TNs, but the same bin of the 160-epoch model is only 99.94% TNs. This reduction in purity will increase the loss. However, given the much larger number of TNs that got pushed into the lowest bin, the overall effect of redrawing the boundary by the 160-epoch model will improve (lower) the loss, but hurt (lower) the AUC. Similarly, some lenses that were assigned to the second highest probability bin by the 50-epoch model are now assigned to the highest probability bin by the 160-epoch model. This will improve (lower) the loss but will not affect the AUC. This logic is why the AUC is the definitive metric to select the best-performing model for deployment.

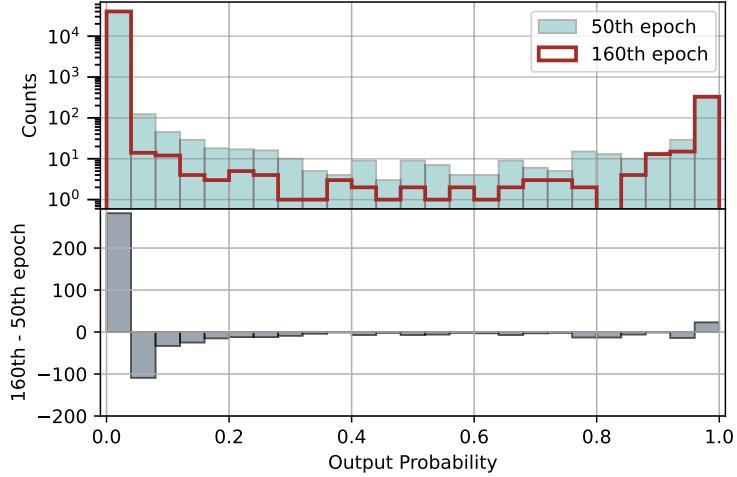


Figure 5. Upper panel: The probability distributions on the validation set from the EfficientNet at the 50th epoch (clear blue) and at the 160th epoch (maroon outline), where the counts are shown in logarithmic scale. Lower panel: differences between counts at the 160th and 50th epoch (160th epoch - 50th epoch counts) in linear scale. The histogram shows a noticeable drop in the bins between the first and last in the 160-epoch model compared to the 50-epoch model (note that these differences are difficult to see in the upper panel, given its logarithmic scale). Many images that received middle-range predictions from the 50-epoch model are pushed into the lowest bin by the 160-epoch model as it tries to minimize the loss. But this actually decreases the AUC and therefore makes the performance worse (see text). To a much lesser extent, a similar trend can be observed for the highest probability bin, which will improve (lower) the loss but not affect the AUC.

3.2.3. The Meta-learner

The ResNet and EfficientNet models, described in § 3.2.1 and § 3.2.2 respectively, generate independent predictions for each image during deployment. To combine these predictions, we employ feature weighted stacking ([Coscrato et al. 2020](#)), which consists of a meta-learner that aggregates the probabilities produced by the two base models. This meta-learner is implemented as a simple one-layer neural network with 300 nodes, designed to take the probabilities of the base models as its input and output a single probability.

The meta-learner is trained using the probability outputs of the base models on the training dataset. During validation, the meta-learner is evaluated by feeding it the predictions of the base models on the validation set. The primary objective of the meta-learner is to find an optimal combination of the probabilities of the two base models.

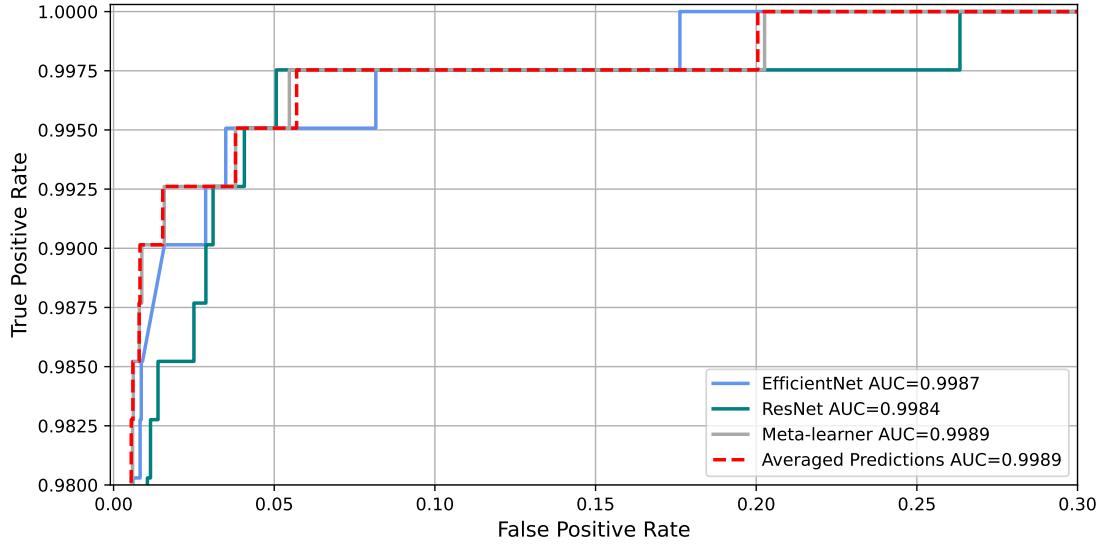


Figure 6. ROCs for the base models individually (ResNet and EfficientNet), the meta-learner, and simple averaging of the base models' predictions. The legend displays the AUC for each of these ROCs.

Given that the two base models are trained on the same dataset, their predictions are inherently correlated. This correlation constrains the meta-learner's ability to substantially outperform simple averaging of the base models' predictions, as illustrated by the AUCs in Figure 6. However, the meta-learner still optimizes the weighting of the base model contributions, which provides a more systematic approach than manual averaging. Figure 6 shows that the meta-learner achieves a higher overall AUC compared to each individual base model.

To further enhance the use of two or more models, strategies such as training the base models on different subsets of the data could be explored, as doing so may produce more diverse predictions by the base models. For example, Shu et al. (2022) trained models using lenses at different redshifts, which resulted in the models finding different candidates.

3.3. Visual Inspection

3.3.1. Visual Inspection Scheme

We choose a probability threshold of ≥ 0.9867 from the meta-learner, resulting in 5680 recommendations from approximately 43 million images (the top 0.01 percentile). Out of 5680 pre-selected images, we remove 1102 recommendations present in the $\sim 12,000$ known candidates and systems from literature. Our chosen probability threshold and elimination of known systems from visual inspection ensures that only the most confident predictions are considered and only *new* discoveries are reported in this paper. This final set of 4578 recommendations is then visually inspected to identify small blue or red galaxy/galaxies next to (typically) a red galaxy/galaxies that exhibit the following traits:

- are typically 1 - 5'' away from the red galaxy/galaxies
- have low surface brightness
- curve toward the red galaxy/galaxies
- have counter/multiple images with similar colors (especially in Einstein-cross like configuration)
- are elongated (including semi- or nearly full rings)

We assign grades based on the number of traits and the level of clarity of each trait exhibited by a recommendation. The criteria used by human graders in this work, built upon the ones used in Papers I through III, are the following:

- Grade A: We have a high level of confidence in 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, also typically blue. However, there are clear cases with red arcs as well.

- 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 counter images).
- Grade C: They generally have features that are even fainter and/or smaller than what is typical for Grade B candidates, but are nevertheless suggestive of lensed arclets. Counter images 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, to attain a higher level of certainty for some of these candidates, higher spatial resolution, deeper data, or spectroscopic observations would be required.

Co-authors JI, BK, CS, and XH performed generous initial inspection of the recommended images using the grading scheme above. CS and XH provided final inspection grades, and we report the average and difference of the results of these two graders.

3.3.2. Visual Inspection Portal

To optimize the grading process, we have developed a visual inspection portal. Given a list of recommendations, the portal displays each co-added image along with its individual *grz* channels with “Previous”/“Next” buttons that enable scrolling through individual recommendations. The cutout image for each recommendation is displayed with text-boxes to enter a grade and an optional comment, which are written into a CSV file. Each image is also linked to the Legacy Surveys’ sky viewer⁶ for zoom and other options, e.g., to provide larger scale context for a possible group/cluster scale lens. Figure 7 shows the user interface of the portal.

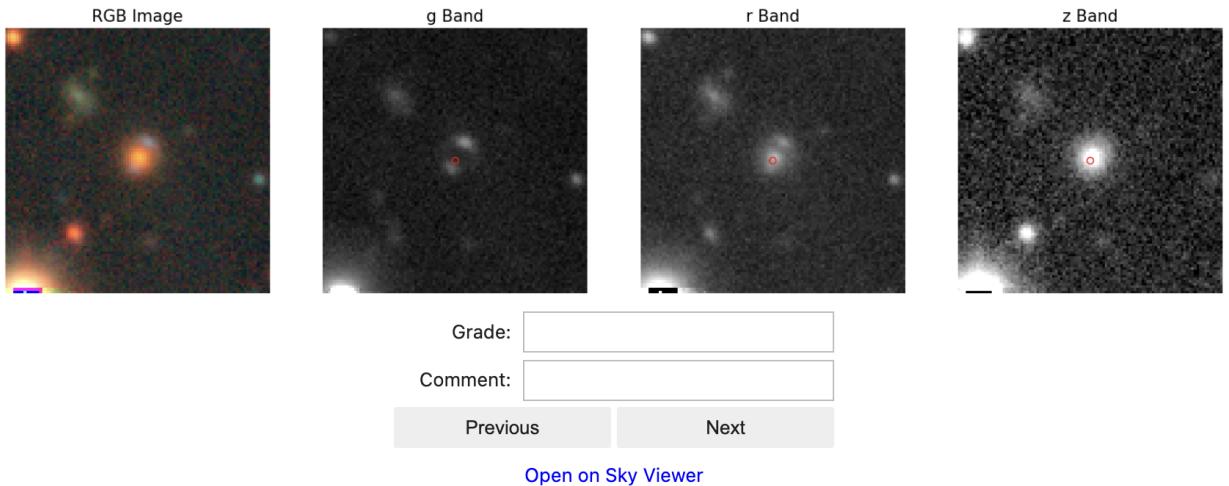


Figure 7. User interface of the visual inspection portal, displaying images from left to right: co-added image, and the *g*, *r*, and *z* bands. The candidate system DESI-316.3086-07.3386, identified for the first time on this paper, is shown as an example. In this case, the arc, and especially the counter-arc are more clearly visible in the *g*-band without the putative lens (the red galaxy at the center of the cutout, highlighted by the red circle).

With respect to Papers I–III, this is the first time that individual *grz* bands are used in addition to the co-added images to visually inspect recommendations. Typical lenses are early-type galaxies with redder spectral energy distributions (SEDs), often peaking at longer wavelengths, which makes them more prominent in the *z* band. Conversely, lensed sources, which are often star-forming galaxies or quasars, tend to have bluer SEDs, making their arcs or multiple images more prominent in the *g* band. The *r* band provides an intermediate perspective, capturing both the redder lenses and bluer sources with moderate contrast. By examining the separate bands, it is possible to detect subtle features that may otherwise be overlooked in a combined RGB image, as it has been noted by other lens searches such as Storfer et al. (2025).

4. RESULTS

In this section, we first present the lens candidates identified from the visual inspection in this work (§ 4.1). We then provide a summary of all lens candidates discovered by our group in the Legacy Surveys DR7, 8, 9, and 10 (§ 4.2).

⁶ <https://www.legacysurvey.org/viewer>

4.1. Lens Candidates in DR10

In this section, we present the strong lens candidates discovered in Legacy Surveys DR10 from this work. We deployed our trained models on ~ 43 million cutouts with ≥ 3 exposures in g, r, z -bands, centered on all galaxy types (non-PSF) with z -band mag < 20.0 mag in the DECaLS footprint.

As described in Section 3.3.1, the top 5680 recommendations with highest probabilities are selected, and then 1102 known candidates and systems are removed, which come from: Stein et al. (2022); O’Donnell et al. (2022); Jaelani et al. (2020); Cañameras et al. (2020); Petrillo et al. (2019); Huang et al. (2020); Huang et al. (2021); Wong et al. (2022); Sonnenfeld et al. (2018); Wong et al. (2018); Li et al. (2021); Sonnenfeld et al. (2019); Rojas et al. (2021); Jacobs et al. (2019a); Talbot et al. (2021); Storfer et al. (2024); More et al. (2016, 2012); Sonnenfeld et al. (2020); Diehl et al. (2017); Gavazzi et al. (2014); Savary et al. (2022); Wong et al. (2022); Cañameras et al. (2021); Carrasco et al. (2017); Chan et al. (2020)

Resulting from visual inspection, we report 811 *new* lens candidates. These candidates contain 90 Grade As, 104 Bs, and 617 Cs. Figure 8 shows three examples for each of the four categories of new lens candidates. The locations in the sky of all new candidates are shown in Figure 10 and Table 3 breaks them down by *The Tractor* type.

Table 3. Strong Lens Candidates by Type

Grade	A	B	C	Total by Type
SER	65	71	357	493
DEV	9	12	100	121
REX	11	18	133	162
EXP	5	3	27	35
Total by Grade	90	104	617	811

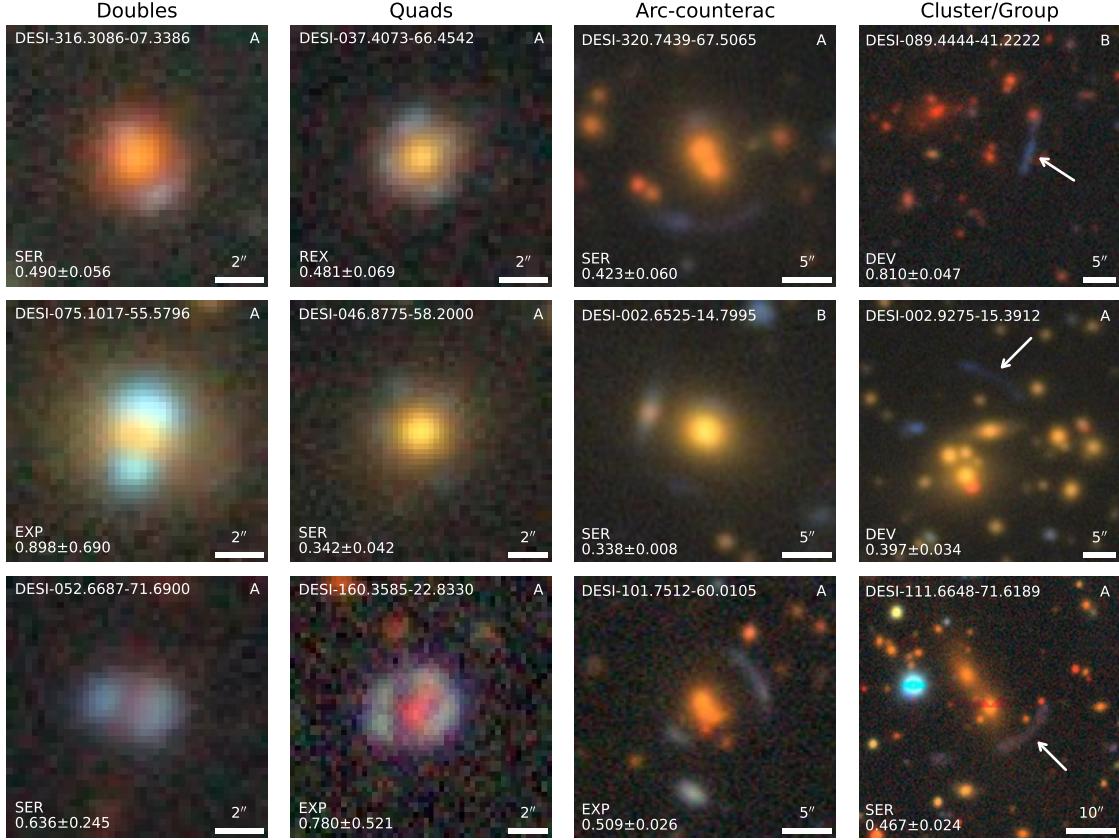


Figure 8. Twelve of the 811 new candidates found in this work, categorized into four groups: doubles, quads, arc-counterarc systems, and cluster/group lenses. The naming convention follows R.A. and decl. in decimal format. All images have north up and east to the left. The grade is shown at the top-right corner of each image, while the bottom-left corner indicates *The Tractor* type and the photometric redshift of the lens (z_d). **First column:** doubly lensed systems. **Second column:** quadruply lensed systems. **Third column:** systems with an arc–counterarc pair. **Fourth column:** groups/cluster lenses, where putative arcs are indicated with white arrows.

Our search demonstrates that we can find a large number of new strong lens candidates in a footprint that has been mined for lenses before in certain sub-regions, repeatedly. Remarkably, $\sim 60\%$ of our 811 new candidates lie within the DR9 footprint, which is an area that has been the focus of multiple previous lens searches. Table 4 breaks down the distribution of these candidates between the previously searched DR9 region and the newly added area in DR10. Figure 9 shows a subset of the highest-quality (A-grade) candidates found within DR9. The abundance of new high-quality candidates from DR9 in this work may be owing to our use of separate bands for visual inspection (see §3.3.2), which previous searches within DR9 lack.

Table 4. Strong Lens Candidates

Grade	A	B	C	Total by Type
Within DR9	21	51	412	484
Outside DR9	69	53	205	327
Total by Grade	90	104	617	811

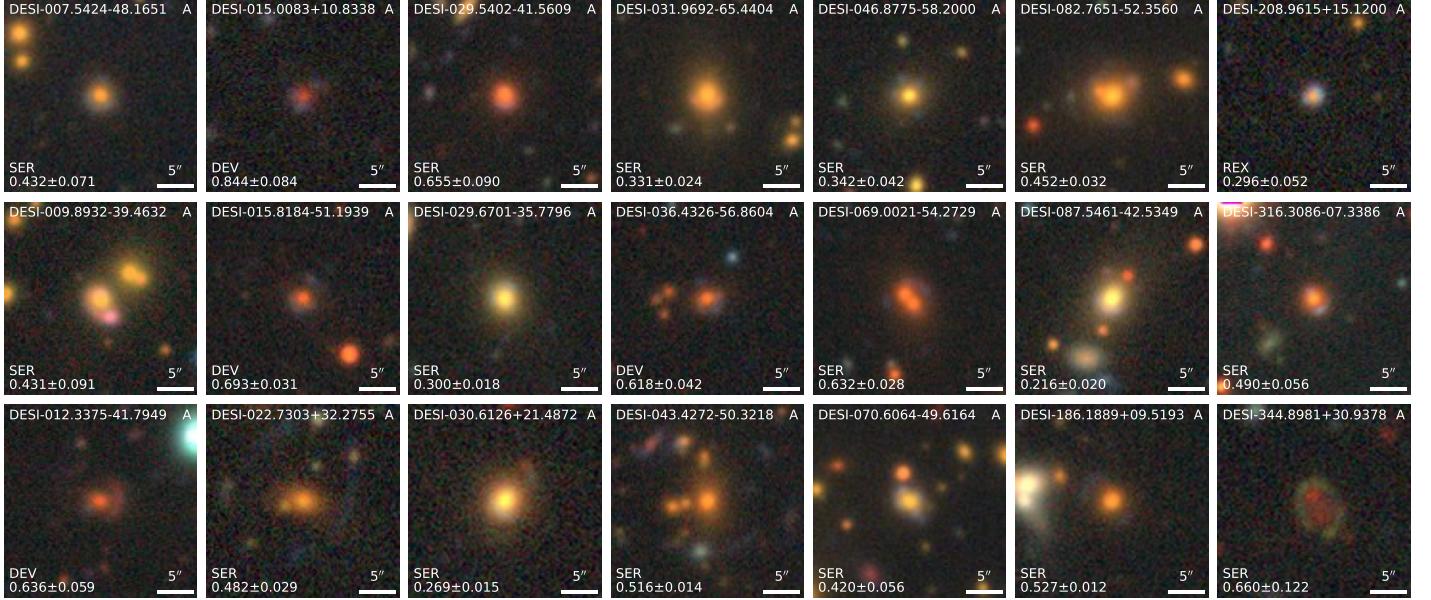


Figure 9. All 21 newly identified A-grade strong lens candidates located within the DR10 data but falling inside the previously mined DR9 footprint.

The purity of systems by type is shown in Table 5. These values are computed by scaling the grade distribution of *new* candidates to known ones, under the assumption that both new and rediscovered candidates follow the same grade distribution. This is due to the fact that only recommendations that did not match with known candidates and systems were inspected. The SER, DEV, and REX types exhibit improved purity compared to Paper III, where, respectively, one in every 23, 22, and 27 recommendations was a candidate. The only exception to this trend is galaxies typed as EXP by *The Tractor*, for which Paper III reported a purity of 39, whereas this work finds a purity of 80.

Table 5. Purity for Our Neural Network Ensemble

<i>The Tractor</i> Type	SER	DEV	REX	EXP
Purity by Type	5	18	18	80

NOTE—Purities are shown as the number of Neural Network Ensemble recommendations inspected in order to find a lens candidate.

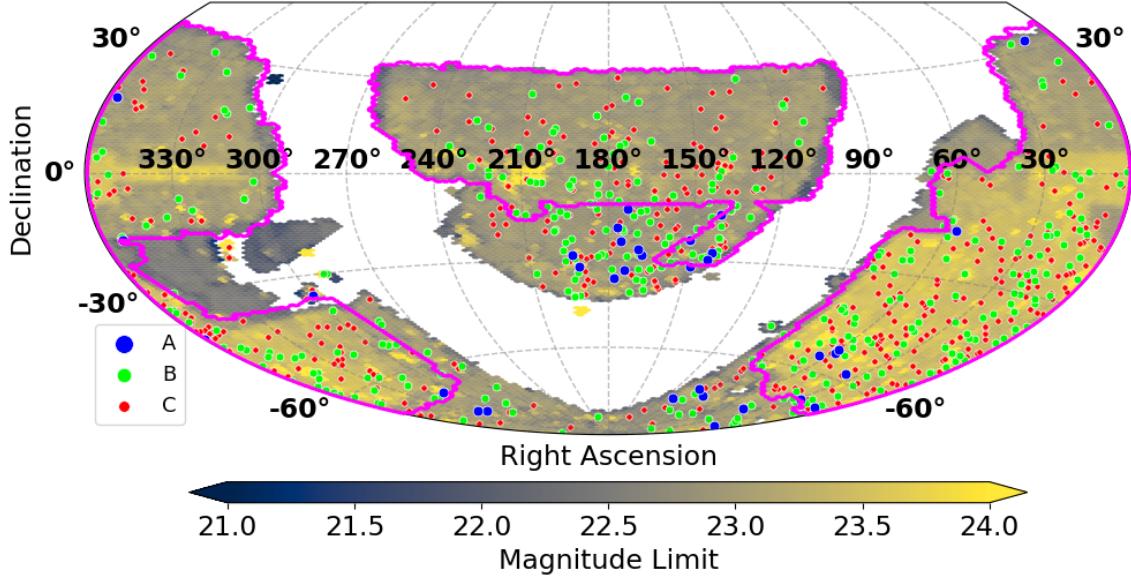


Figure 10. The 811 new candidate lensing systems discovered in this work by grades over the depth map of the Legacy Surveys DR10 shown in Figure 1. The magenta contours enclose the DECaLS region of DR9 for comparison.

4.2. Lens Candidates in DR7, 8, 9, and 10

From our four searches in the Legacy Surveys, we have found a grand total of 4,869 *new* candidates (Paper I in DR7 among DEV and COMP in DECaLS only, Paper II in DR8, Paper III in DR9, and this work in DR10). The entire catalog of these lenses can be found on our project website⁷. The positions of all candidates on the sky are shown in Figure 11.

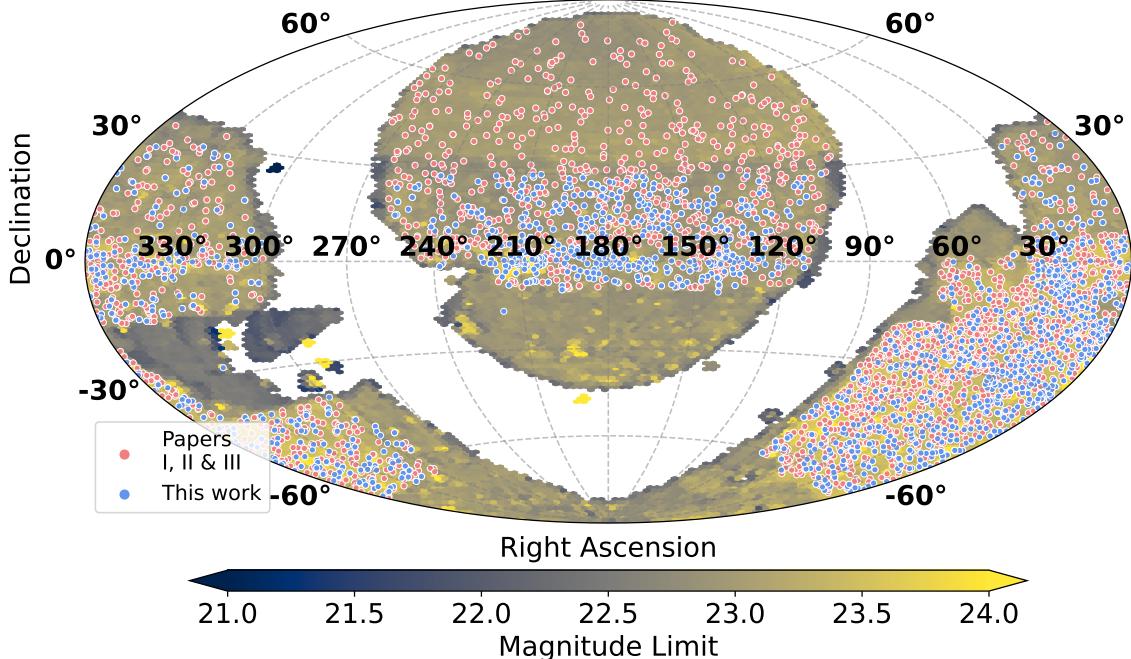


Figure 11. All 4,869 candidate lensing systems discovered in the Legacy Surveys reported in Papers I through III (red), and this work (blue), over the hexagonal bin depth map of the Legacy Surveys DR10, including the footprint of the BASS/MzLS surveys.

⁷ <https://sites.google.com/usfca.edu/neuralens/publications/lens-candidates-inchausti-et-al-2025>

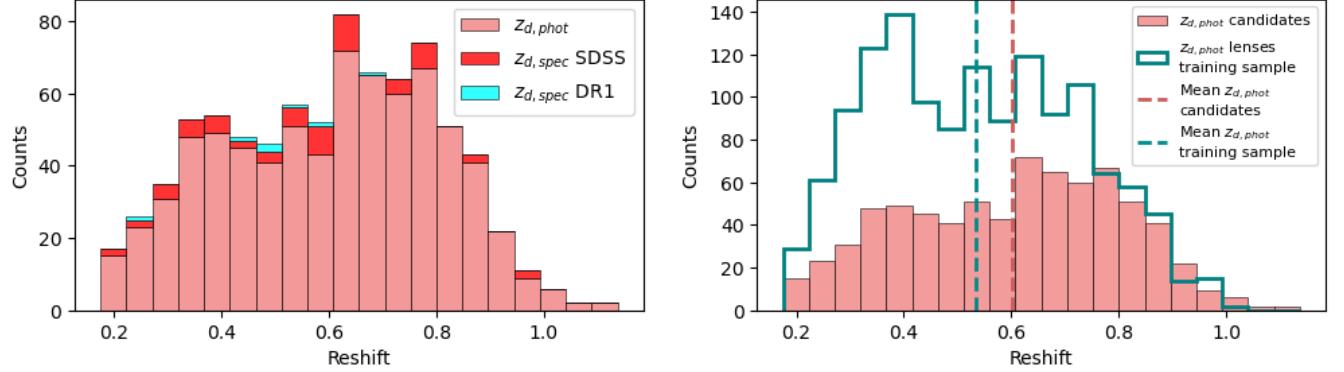


Figure 12. **Left:** lens spectroscopic redshift, $z_{d,\text{spec}}$, stacked distributions of the candidates found in this work (SDSS Data Release 17 in dark red and DESI DR1 [DESI Collaboration et al. \(2025\)](#) in cyan), and photometric redshift, $z_{d,\text{phot}}$, distribution from [Zhou et al. \(2021\)](#) in light red. When DR1 redshift is not available, SDSS redshift is reported. In the absence of DR1 and SDSS redshifts, photometric redshift is provided. A DR1 redshift outlier of $z = 1.669$ corresponding to candidate DESI-006.2804+03.3820 was omitted from this histogram because visual inspection of its spectrum revealed no spectroscopic features to support the *Redrock* (see <https://github.com/desihub/redrock/releases/tag/0.15.4>) redshift. **Right:** lens redshift distributions of the new candidates discovered in this work (light red) and the lenses included in the training sample (teal outline). Dashed vertical lines indicate the average photometric redshift of the training sample and the candidates.

5. DISCUSSION

This search in DR10 employs two neural network architectures: a custom ResNet (§3.2.1) and an EfficientNet (§3.2.2). Table 6 compares the sizes, training times, and AUCs of these two models. Notably, both trained in approximately the same amount of time (~ 1 hour), despite the EfficientNet having ~ 100 times more trainable parameters than ResNet. This stark difference in size, combined with the similar training durations, highlights EfficientNet’s scalability. Conversely, our custom ResNet is impressive in that it achieves virtually identical AUC performance as EfficientNet despite having far fewer trainable parameters.

Table 6. Model size, training time, and AUC comparison.

	EfficientNet	ResNet
Trainable Parameters	20,542,883	194,433
Training Time [mins]	61	56
Highest AUC	0.9987	0.9984

Figure 13 compares the output probabilities assigned by the ResNet and the EfficientNet for new candidates graded A, B, and C. Although both models achieve similar AUCs, they exhibit slight differences in their probability distributions. The EfficientNet assigns a broader range of probabilities, extending down to ~ 0.86 , whereas the ResNet probabilities only drop to ~ 0.92 . This suggests that the ResNet is more confident in its predictions, while the EfficientNet is more conservative. Despite these differences, the marginal histograms show that both models concentrate candidates in the upper probability range, with all grades peaking at the highest probability bin.

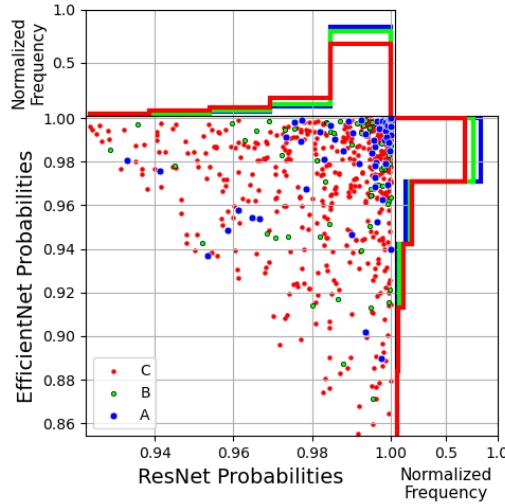


Figure 13. Comparison of the ResNet and EfficientNet output probabilities for candidates by grade. Each point represents a candidate, with its position indicating the probability assigned by ResNet (x-axis) and EfficientNet (y-axis). The marginal histograms show the probability distributions for each grade, normalized such that the total frequency for each grade sums to 1.

While other lens searches, such as Shu et al. (2022), have implemented two different classifiers simultaneously, we advance this technique by using a meta-learner to aggregate the probabilities of our base models (see §3.2.3). This method of combining probabilities results in an overall AUC higher than that of the individual models, as shown in Figure 6. While in this case, the performance of the meta-learner is the same as the simple average of the predictions

from the two models, this nevertheless provides a systematic approach to optimally combine the results from multiple models in future lens searches.

The use of individual *grz* bands for visual inspection (see §3.3.2) has proven to be highly effective. As demonstrated in Figure 7, different bands can help distinguish the lens from the source. This capability is particularly valuable in cases where the Einstein radius is comparable to the seeing.

6. CONCLUSIONS

We have carried out a search for strong gravitational lensing systems in the DESI Legacy Surveys Data Release 10 (DR10). This is the fourth paper in a series of lens searches in the DESI Legacy Surveys, following Paper I (Huang et al. 2020, DR7 DECaLS, DEV and COMP only), Paper II (Huang et al. 2021, DR8, DEV, COMP, and REX), and Paper III (Storfer et al. 2024, DR9, SER, DEV, REX, and EXP). We use a neural network ensemble consisting of a custom ResNet (Lanusse et al. 2018; Paper II) and an EfficientNet (Tan & Le 2019), trained on observed lenses and nonlenses. We apply our trained models to ~ 43 million non-PSF (i.e., SER, DEV, REX, and EXP, extracted and typed by *The Tractor*) cutout images with at least three exposures in each of the *grz* bands and a *z*-band magnitude cut of < 20.0 mag for the galaxy at the center of each image.

We hold a high standard in grading these candidate systems. We report 811 *new* lens candidates with a grade breakdown of 90 As, 104 Bs, and 617 Cs. Combining all four searches, we have found a grand total of new 3868 strong lens candidates. Grade D systems are not included in the counts, but they can be found on our project website (URL provided in §4.2).

In this search we have made improvements in four major areas: machine learning methods, training sample, neural network training hardware, and visual inspection:

- We have implemented a neural network ensemble composed of models with two distinct architectures whose predictions are optimally combined using a meta-learner, as opposed to a single architecture (Papers I–III).
- We continue to use observed images for training as in Papers I–III, but we have increased the size of our training sample by increasing the nonlens-to-lens ratio from 33:1 to 100:1, respectively, from Paper III to this work. The total training sample size thus has increased from 66,545 to 135,554 between Paper III and this work.
- We have trained both models on four NVIDIA A100 GPUs on a single GPU node, whereas for Papers I–III, the training was done on one GPU. Using 4 GPUs results in approximately 3.7 and 3 \times speed gain for the ResNet and EfficientNet, respectively. Our training and hyperparameter optimization for two models with a training sample that is twice as large as for Paper III takes advantage of the speed afforded by the parallelized training.
- We use both the combined RGB and individual *grz* band images during visual inspection. This makes it easier to identify systems that have small Einstein radius and/or faint lensed arc(s), but sufficient color difference between the lens and lensed arcs. This work also uses a specialized portal that streamlines the process by enabling the inspection and annotation of candidates efficiently.

With the improvements and innovations outlined above, we find an increase in purity in this work where 1 in every 5 SER, 18 DEV, 18 REX, and 80 EXP neural network recommendations is a lens candidate.

To our knowledge, the additional footprint covered by DR10 compared to DR9 (see Figure 1) has not been searched for lenses. In this sub-region, we have found 327 lens candidates. However, in all other parts of the DECaLS footprint, multiple lens searches have been performed, by our group or others. And yet, we have found 484 *new* candidates in these parts. Therefore, systematic comparisons of lens searches to investigate the demographics (the “selection function”) of lens candidates (especially those with A and B grades) found and missed by various searches covering the same footprint would be beneficial to astrophysics and machine learning. Work in this direction has begun (e.g., Knabel et al. 2020; More et al. 2024) and the search results from this work make it clear that more such studies are needed.

This work, together with other searches, has clearly demonstrated that machine learning approaches are highly effective in discovering large numbers of high-quality strong lensing candidates. This will certainly continue to be the case for future surveys such as the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST), Euclid, and the Nancy Grace Roman Space Telescope. The discovery of strong lenses will therefore continue to accelerate and the potential for using strong lensing to tackle major astrophysical and cosmological questions has never been greater.

7. ACKNOWLEDGEMENT

This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231 and the Computational HEP program in The Department of Energy's Science Office of High Energy Physics provided resources through the "Cosmology Data Repository" project (Grant #KA2401022) X.H. acknowledges the University of San Francisco Faculty Development Fund. A.D.'s research is supported by National Science Foundation's National Optical-Infrared Astronomy Research Laboratory, which is operated by the Association of Universities for Research in Astronomy (AURA) under cooperative agreement with the National Science Foundation.

This paper is based on observations at Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory (NOAO Prop. ID: 2014B-0404; co-PIs: D. J. Schlegel and A. Dey), which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

This project used data obtained with the Dark Energy Camera (DECam), which was constructed by the Dark Energy Survey (DES) collaboration. Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at The Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft, and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Enérgéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the OzDES Membership Consortium the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University.

REFERENCES

- Anguita, T., Barrientos, L. F., Gladders, M. D., et al. 2012, ApJ, 748, 129, doi: [10.1088/0004-637X/748/2/129](https://doi.org/10.1088/0004-637X/748/2/129)
- Birrer, S., Shajib, A. J., Galan, A., et al. 2020, arXiv e-prints, arXiv:2007.02941. <https://arxiv.org/abs/2007.02941>
- Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., & Moustakas, L. A. 2006, ApJ, 638, 703, doi: [10.1086/498884](https://doi.org/10.1086/498884)
- Broadhurst, T., Huang, X., Frye, B., & Ellis, R. 2000, ApJL, 534, L15, doi: [10.1086/312651](https://doi.org/10.1086/312651)
- Brownstein, J. R., Bolton, A. S., Schlegel, D. J., et al. 2012, ApJ, 744, 41, doi: [10.1088/0004-637X/744/1/41](https://doi.org/10.1088/0004-637X/744/1/41)
- Cañameras, R., Schuldt, S., Suyu, S. H., et al. 2020, arXiv e-prints, arXiv:2004.13048. <https://arxiv.org/abs/2004.13048>
- Cañameras, R., Schuldt, S., Shu, Y., et al. 2021, A&A, 653, L6, doi: [10.1051/0004-6361/202141758](https://doi.org/10.1051/0004-6361/202141758)
- Caminha, G. B., Suyu, S. H., Grillo, C., & Rosati, P. 2022, A&A, 657, A83, doi: [10.1051/0004-6361/202141994](https://doi.org/10.1051/0004-6361/202141994)
- Carrasco, M., Barrientos, L. F., Anguita, T., et al. 2017, ApJ, 834, 210, doi: [10.3847/1538-4357/834/2/210](https://doi.org/10.3847/1538-4357/834/2/210)
- Çağan Şengül, A., Dvorkin, C., Ostdiek, B., & Tsang, A. 2021, arXiv e-prints, arXiv:2112.00749. <https://arxiv.org/abs/2112.00749>
- Chan, J. H. H., Suyu, S. H., Sonnenfeld, A., et al. 2020, A&A, 636, A87, doi: [10.1051/0004-6361/201937030](https://doi.org/10.1051/0004-6361/201937030)
- Chen, W., Kelly, P. L., Oguri, M., et al. 2022, Nature, 611, 256, doi: [10.1038/s41586-022-05252-5](https://doi.org/10.1038/s41586-022-05252-5)
- Collett, T. E., & Auger, M. W. 2014, MNRAS, 443, 969, doi: [10.1093/mnras/stu1190](https://doi.org/10.1093/mnras/stu1190)
- Coscrato, V., de Almeida Inácio, M. H., & Izbicki, R. 2020, Neurocomputing, 399, 141, doi: <https://doi.org/10.1016/j.neucom.2020.02.073>

- Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270, doi: [10.1093/mnras/stw641](https://doi.org/10.1093/mnras/stw641)
- Dawes, C., Storfer, C., Huang, X., et al. 2023, ApJS, 269, 61, doi: [10.3847/1538-4365/ad015a](https://doi.org/10.3847/1538-4365/ad015a)
- DESI Collaboration, Abdul-Karim, M., Adame, A. G., et al. 2025, arXiv e-prints, arXiv:2503.14745, doi: [10.48550/arXiv.2503.14745](https://doi.org/10.48550/arXiv.2503.14745)
- Dey, A., Schlegel, D. J., Lang, D., et al. 2019, The Astronomical Journal, 157, 168, doi: [10.3847/1538-3881/ab089d](https://doi.org/10.3847/1538-3881/ab089d)
- Diehl, H. T., Buckley-Geer, E. J., Lindgren, K. A., et al. 2017, ApJS, 232, 15, doi: [10.3847/1538-4365/aa8667](https://doi.org/10.3847/1538-4365/aa8667)
- Ding, X., Liao, K., Birrer, S., et al. 2021, MNRAS, 504, 5621, doi: [10.1093/mnras/stab1240](https://doi.org/10.1093/mnras/stab1240)
- Drlica-Wagner, A., Carlin, J. L., Nidever, D. L., et al. 2021, ApJS, 256, 2, doi: [10.3847/1538-4365/ac079d](https://doi.org/10.3847/1538-4365/ac079d)
- Fagin, J., Vernardos, G., Tsagkatakis, G., et al. 2024, arXiv e-prints, arXiv:2403.13881, doi: [10.48550/arXiv.2403.13881](https://doi.org/10.48550/arXiv.2403.13881)
- Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, AJ, 150, 150, doi: [10.1088/0004-6256/150/5/150](https://doi.org/10.1088/0004-6256/150/5/150)
- Freedman, W. L., Madore, B. F., Hoyt, T., et al. 2020, ApJ, 891, 57, doi: [10.3847/1538-4357/ab7339](https://doi.org/10.3847/1538-4357/ab7339)
- Gavazzi, R., Marshall, P. J., Treu, T., & Sonnenfeld, A. 2014, ApJ, 785, 144, doi: [10.1088/0004-637X/785/2/144](https://doi.org/10.1088/0004-637X/785/2/144)
- González, J., Holloway, P., Collett, T., et al. 2025, arXiv e-prints, arXiv:2501.15679, doi: [10.48550/arXiv.2501.15679](https://doi.org/10.48550/arXiv.2501.15679)
- Grillo, C., Suyu, S. H., Rosati, P., et al. 2015, ApJ, 800, 38, doi: [10.1088/0004-637X/800/1/38](https://doi.org/10.1088/0004-637X/800/1/38)
- He, Z., Li, N., Cao, X., et al. 2023, A&A, 672, A123, doi: [10.1051/0004-6361/202245484](https://doi.org/10.1051/0004-6361/202245484)
- Hsu, Y.-M. 2025, spherimatch: Cross-matching and self-matching in spherical coordinates, Astrophysics Source Code Library, ascl:2507.022
- Huang, X., Morokuma, T., Fakhouri, H. K., et al. 2009, ApJ, 707, L12, doi: [10.1088/0004-637X/707/1/L12](https://doi.org/10.1088/0004-637X/707/1/L12)
- Huang, X., Storfer, C., Ravi, V., et al. 2020, The Astrophysical Journal, 894, 78, doi: [10.3847/1538-4357/ab7ff](https://doi.org/10.3847/1538-4357/ab7ff)
- Huang, X., Storfer, C., Gu, A., et al. 2021, ApJ, 909, 27, doi: [10.3847/1538-4357/abd62b](https://doi.org/10.3847/1538-4357/abd62b)
- Huber, S., Suyu, S. H., Noebauer, U. M., et al. 2021, A&A, 646, A110, doi: [10.1051/0004-6361/202039218](https://doi.org/10.1051/0004-6361/202039218)
- Jacobs, C., Glazebrook, K., Collett, T., More, A., & McCarthy, C. 2017, MNRAS, 471, 167, doi: [10.1093/mnras/stx1492](https://doi.org/10.1093/mnras/stx1492)
- Jacobs, C., Collett, T., Glazebrook, K., et al. 2019a, MNRAS, 484, 5330, doi: [10.1093/mnras/stz272](https://doi.org/10.1093/mnras/stz272)
- . 2019b, ApJS, 243, 17, doi: [10.3847/1538-4365/ab26b6](https://doi.org/10.3847/1538-4365/ab26b6)
- Jaelani, A. T., More, A., Oguri, M., et al. 2020, MNRAS, 495, 1291, doi: [10.1093/mnras/staa1062](https://doi.org/10.1093/mnras/staa1062)
- Jaelani, A. T., Rusu, C. E., Kayo, I., et al. 2021, Monthly Notices of the Royal Astronomical Society, 502, 1487, doi: [10.1093/mnras/stab145](https://doi.org/10.1093/mnras/stab145)
- Johnson, L. E., Irwin, J. A., White, III, R. E., et al. 2018, ApJ, 856, 131, doi: [10.3847/1538-4357/aab430](https://doi.org/10.3847/1538-4357/aab430)
- Jullo, E., Natarajan, P., Kneib, J. P., et al. 2010, Science, 329, 924, doi: [10.1126/science.1185759](https://doi.org/10.1126/science.1185759)
- Knabel, S., Steele, R. L., Holwerda, B. W., et al. 2020, AJ, 160, 223, doi: [10.3847/1538-3881/abb612](https://doi.org/10.3847/1538-3881/abb612)
- Kochanek, C. S. 1991, ApJ, 373, 354, doi: [10.1086/170057](https://doi.org/10.1086/170057)
- Koopmans, L. V. E., & Treu, T. 2002, ApJL, 568, L5, doi: [10.1086/340143](https://doi.org/10.1086/340143)
- Koopmans, L. V. E., Treu, T., Bolton, A. S., Burles, S., & Moustakas, L. A. 2006, ApJ, 649, 599, doi: [10.1086/505696](https://doi.org/10.1086/505696)
- Lang, D., Hogg, D. W., & Mykytyn, D. 2016, The Tractor: Probabilistic astronomical source detection and measurement, Astrophysics Source Code Library, <http://ascl.net/1604.008>
- Lanusse, F., Ma, Q., Li, N., et al. 2018, MNRAS, 473, 3895, doi: [10.1093/mnras/stx1665](https://doi.org/10.1093/mnras/stx1665)
- Li, R., Napolitano, N. R., Spinelli, C., et al. 2021, ApJ, 923, 16, doi: [10.3847/1538-4357/ac2df0](https://doi.org/10.3847/1538-4357/ac2df0)
- Li, T., Collett, T. E., Krawczyk, C. M., & Enzi, W. 2024, MNRAS, 527, 5311, doi: [10.1093/mnras/stad3514](https://doi.org/10.1093/mnras/stad3514)
- Linder, E. V. 2016, PhRvD, 94, 083510, doi: [10.1103/PhysRevD.94.083510](https://doi.org/10.1103/PhysRevD.94.083510)
- Massey, R., Kitching, T., & Richard, J. 2010, Reports on Progress in Physics, 73, 086901, doi: [10.1088/0034-4885/73/8/086901](https://doi.org/10.1088/0034-4885/73/8/086901)
- Meneghetti, M., Davoli, G., Bergamini, P., et al. 2020, Science, 369, 1347, doi: [10.1126/science.aax5164](https://doi.org/10.1126/science.aax5164)
- Meneghetti, M., Cui, W., Rasia, E., et al. 2023, A&A, 678, L2, doi: [10.1051/0004-6361/202346975](https://doi.org/10.1051/0004-6361/202346975)
- Metcalf, R. B., Meneghetti, M., Avestruz, C., et al. 2018, arXiv e-prints, arXiv:1802.03609, <https://arxiv.org/abs/1802.03609>
- Millon, M., Galan, A., Courbin, F., et al. 2020, A&A, 639, A101, doi: [10.1051/0004-6361/201937351](https://doi.org/10.1051/0004-6361/201937351)
- More, A., Cabanac, R., More, S., et al. 2012, ApJ, 749, 38, doi: [10.1088/0004-637X/749/1/38](https://doi.org/10.1088/0004-637X/749/1/38)
- More, A., Verma, A., Marshall, P. J., et al. 2016, MNRAS, 455, 1191, doi: [10.1093/mnras/stv1965](https://doi.org/10.1093/mnras/stv1965)
- More, A., Cañameras, R., Jaelani, A. T., et al. 2024, MNRAS, 533, 525, doi: [10.1093/mnras/stae1597](https://doi.org/10.1093/mnras/stae1597)
- Nordin, J., Rubin, D., Richard, J., et al. 2014, MNRAS, 440, 2742, doi: [10.1093/mnras/stu376](https://doi.org/10.1093/mnras/stu376)

- O'Donnell, J. H., Wilkinson, R. D., Diehl, H. T., et al. 2022, ApJS, 259, 27, doi: [10.3847/1538-4365/ac470b](https://doi.org/10.3847/1538-4365/ac470b)
- Oguri, M., & Marshall, P. J. 2010, MNRAS, 405, 2579, doi: [10.1111/j.1365-2966.2010.16639.x](https://doi.org/10.1111/j.1365-2966.2010.16639.x)
- Pascale, M., Frye, B. L., Pierel, J. D. R., et al. 2024, arXiv e-prints, arXiv:2403.18902, doi: [10.48550/arXiv.2403.18902](https://doi.org/10.48550/arXiv.2403.18902)
- Pascale, M., Frye, B. L., Pierel, J. D. R., et al. 2025, The Astrophysical Journal, 979, 13, doi: [10.3847/1538-4357/ad9928](https://doi.org/10.3847/1538-4357/ad9928)
- Patel, B., McCully, C., Jha, S. W., et al. 2014, ApJ, 786, 9, doi: [10.1088/0004-637X/786/1/9](https://doi.org/10.1088/0004-637X/786/1/9)
- Petrillo, C. E., Tortora, C., Verner, G., et al. 2019, MNRAS, 484, 3879, doi: [10.1093/mnras/stz189](https://doi.org/10.1093/mnras/stz189)
- Pierel, J. D. R., & Rodney, S. 2019, ApJ, 876, 107, doi: [10.3847/1538-4357/ab164a](https://doi.org/10.3847/1538-4357/ab164a)
- Pierel, J. D. R., Arendse, N., Ertl, S., et al. 2023, ApJ, 948, 115, doi: [10.3847/1538-4357/acc7a6](https://doi.org/10.3847/1538-4357/acc7a6)
- Pierel, J. D. R., Newman, A. B., Dhawan, S., et al. 2024, arXiv e-prints, arXiv:2404.02139, doi: [10.48550/arXiv.2404.02139](https://doi.org/10.48550/arXiv.2404.02139)
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6, doi: [10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910)
- Refsdal, S. 1964, MNRAS, 128, 307, doi: [10.1093/mnras/128.4.307](https://doi.org/10.1093/mnras/128.4.307)
- Riess, A. G., Casertano, S., Yuan, W., et al. 2021, ApJL, 908, L6, doi: [10.3847/2041-8213/abdbaf](https://doi.org/10.3847/2041-8213/abdbaf)
- Rodney, S. A., Strolger, L. G., Kelly, P. L., et al. 2016, ApJ, 820, 50, doi: [10.3847/0004-637X/820/1/50](https://doi.org/10.3847/0004-637X/820/1/50)
- Rojas, K., Savary, E., Clément, B., et al. 2021, Strong lens systems search in the Dark Energy Survey using Convolutional Neural Networks. <https://arxiv.org/abs/2109.00014>
- Rojas, K., Savary, E., Clément, B., et al. 2022, A&A, 668, A73, doi: [10.1051/0004-6361/202142119](https://doi.org/10.1051/0004-6361/202142119)
- Rubin, D., Hayden, B., Huang, X., et al. 2018, ApJ, 866, 65, doi: [10.3847/1538-4357/aad565](https://doi.org/10.3847/1538-4357/aad565)
- Sandler, M., Howard, A. G., Zhu, M., Zhmoginov, A., & Chen, L. 2018, CoRR, abs/1801.04381
- Savary, E., Rojas, K., Maus, M., et al. 2022, A&A, 666, A1, doi: [10.1051/0004-6361/202142505](https://doi.org/10.1051/0004-6361/202142505)
- Sharma, D., & Linder, E. V. 2022, arXiv e-prints, arXiv:2204.03020. <https://arxiv.org/abs/2204.03020>
- Sheu, W., Huang, X., Cikota, A., et al. 2024, ApJ, 973, 24, doi: [10.3847/1538-4357/ad5dad](https://doi.org/10.3847/1538-4357/ad5dad)
- . 2023, arXiv e-prints, arXiv:2301.03578, doi: [10.48550/arXiv.2301.03578](https://doi.org/10.48550/arXiv.2301.03578)
- Shu, Y., Cañameras, R., Schuldt, S., et al. 2022, A&A, 662, A4, doi: [10.1051/0004-6361/202243203](https://doi.org/10.1051/0004-6361/202243203)
- Shu, Y., Brownstein, J. R., Bolton, A. S., et al. 2017, ApJ, 851, 48, doi: [10.3847/1538-4357/aa9794](https://doi.org/10.3847/1538-4357/aa9794)
- Sonnenfeld, A., Wang, W., & Bahcall, N. 2019, A&A, 622, A30, doi: [10.1051/0004-6361/201834260](https://doi.org/10.1051/0004-6361/201834260)
- Sonnenfeld, A., Chan, J. H. H., Shu, Y., et al. 2018, PASJ, 70, S29, doi: [10.1093/pasj/psx062](https://doi.org/10.1093/pasj/psx062)
- Sonnenfeld, A., Verma, A., More, A., et al. 2020, arXiv e-prints, arXiv:2004.00634. <https://arxiv.org/abs/2004.00634>
- Stein, G., Blaum, J., Harrington, P., Medan, T., & Lukic, Z. 2022, ApJ, 932, 107, doi: [10.3847/1538-4357/ac6d63](https://doi.org/10.3847/1538-4357/ac6d63)
- Storfer, C., Huang, X., Gu, A., et al. 2024, ApJS, 274, 16, doi: [10.3847/1538-4365/ad527e](https://doi.org/10.3847/1538-4365/ad527e)
- Storfer, C. J., Magnier, E. A., Huang, X., et al. 2025, arXiv e-prints, arXiv:2505.05032, doi: [10.48550/arXiv.2505.05032](https://doi.org/10.48550/arXiv.2505.05032)
- Suyu, S. H., Huber, S., Cañameras, R., et al. 2020, arXiv e-prints, arXiv:2002.08378. <https://arxiv.org/abs/2002.08378>
- Szegedy, C., Liu, W., Jia, Y., et al. 2014, arXiv e-prints, arXiv:1409.4842. <https://arxiv.org/abs/1409.4842>
- Talbot, M. S., Brownstein, J. R., Dawson, K. S., Kneib, J.-P., & Bautista, J. 2021, MNRAS, 502, 4617, doi: [10.1093/mnras/stab267](https://doi.org/10.1093/mnras/stab267)
- Tan, M., & Le, Q. V. 2019, ArXiv, abs/1905.11946
- Treu, T. 2010, ARA&A, 48, 87, doi: [10.1146/annurev-astro-081309-130924](https://doi.org/10.1146/annurev-astro-081309-130924)
- Vegetti, S., Czoske, O., & Koopmans, L. V. E. 2010, MNRAS, 407, 225, doi: [10.1111/j.1365-2966.2010.16952.x](https://doi.org/10.1111/j.1365-2966.2010.16952.x)
- Vegetti, S., & Koopmans, L. V. E. 2009, MNRAS, 392, 945, doi: [10.1111/j.1365-2966.2008.14005.x](https://doi.org/10.1111/j.1365-2966.2008.14005.x)
- Wagner-Carena, S., Aalbers, J., Birrer, S., et al. 2022, arXiv e-prints, arXiv:2203.00690. <https://arxiv.org/abs/2203.00690>
- Williams, G. G., Olszewski, E., Lesser, M. P., & Burge, J. H. 2004, in Proc. SPIE, Vol. 5492, Ground-based Instrumentation for Astronomy, ed. A. F. M. Moorwood & M. Iye, 787–798
- Wong, K. C., Chan, J. H. H., Chao, D. C. Y., et al. 2022, PASJ, 74, 1209, doi: [10.1093/pasj/psac065](https://doi.org/10.1093/pasj/psac065)
- Wong, K. C., Sonnenfeld, A., Chan, J. H. H., et al. 2018, ApJ, 867, 107, doi: [10.3847/1538-4357/aae381](https://doi.org/10.3847/1538-4357/aae381)
- Wong, K. C., Suyu, S. H., Chen, G. C. F., et al. 2019, arXiv e-prints, arXiv:1907.04869. <https://arxiv.org/abs/1907.04869>
- Yahalomi, D. A., Schechter, P. L., & Wambsganss, J. 2017, arXiv e-prints, arXiv:1711.07919. <https://arxiv.org/abs/1711.07919>

- Zaborowski, E. A., Drlica-Wagner, A., Ashmead, F., et al. 2023, *The Astrophysical Journal*, 954, 68, doi: [10.3847/1538-4357/ace4ba](https://doi.org/10.3847/1538-4357/ace4ba)
- Zenteno, A., Kluge, M., Kharkrang, R., et al. 2025, *A&A*, 698, A171, doi: [10.1051/0004-6361/202452440](https://doi.org/10.1051/0004-6361/202452440)
- Zhou, R., Newman, J. A., Mao, Y.-Y., et al. 2021, *MNRAS*, 501, 3309, doi: [10.1093/mnras/staa3764](https://doi.org/10.1093/mnras/staa3764)
- Zitrin, A., Rephaeli, Y., Sadeh, S., et al. 2012, *MNRAS*, 420, 1621, doi: [10.1111/j.1365-2966.2011.20155.x](https://doi.org/10.1111/j.1365-2966.2011.20155.x)