The Astrophysical Journal, 954:28 (13pp), 2023 September 1 https://doi.org/10.3847/1538-4357/acdfc7

Domain Adaptation for Simulation-based Dark Matter Searches with Strong Gravitational Lensing

Stephon Alexander[1], Sergei Gleyzer[2], Hanna Parul[2], Pranath Reddy[3], Marcos Tidball[4], and Michael W. Toomey[1]
[1] Brown Theoretical Physics Center and Department of Physics, Brown University, Providence, RI 02912, USA; michael_toomey@brown.edu
[2] Department of Physics & Astronomy, University of Alabama, Tuscaloosa, AL 35401, USA
[3] University of Florida, Gainesville, FL 32611, USA
[4] Departamento de Física, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
Received 2021 December 23; revised 2023 June 13; accepted 2023 June 18; published 2023 August 21

Abstract

The identity of dark matter has remained surprisingly elusive. While terrestrial experiments may be able to nail down a model, an alternative method is to identify dark matter based on astrophysical or cosmological signatures. A particularly sensitive approach is based on the unique signature of dark matter substructure in galaxy–galaxy strong lensing images. Machine-learning applications have been explored for extracting this signal. Because of the limited availability of high-quality strong lensing images, these approaches have exclusively relied on simulations. Due to the differences with the real instrumental data, machine-learning models trained on simulations are expected to lose accuracy when applied to real data. Here domain adaptation can serve as a crucial bridge between simulations and real data applications. In this work, we demonstrate the power of domain adaptation techniques applied to strong gravitational lensing data with dark matter substructure. We show with simulated data sets representative of Euclid and Hubble Space Telescope observations that domain adaptation can significantly mitigate the losses in the model performance when applied to new domains. Lastly, we find similar results utilizing domain adaptation for the problem of lens finding by adapting models trained on a simulated data set to one composed of real lensed and unlensed galaxies from the Hyper Suprime-Cam. This technique can help domain experts build and apply better machine-learning models for extracting useful information from the strong gravitational lensing data expected from the upcoming surveys.

Unified Astronomy Thesaurus concepts: Dark matter (353); Strong gravitational lensing (1643)

1. Introduction

One of the great achievements of astrophysics in the last century was the realization by Zwicky, Rubin, and others that the observed baryonic mass (stars, galaxies, etc.) was not consistent with the dynamics of galaxies and clusters. A natural solution to this problem was to consider unseen dark matter compensating for this discrepancy. Presently, all efforts aimed at extracting a nongravitational signature of dark matter have come up empty (Goodman & Witten 1985; Drukier et al. 1986; Fermi LAT Collaboration 2015; Geringer-Sameth et al. 2015; Graham et al. 2015; The Super-Kamiokande Collaboration 2015; MAGIC Collaboration 2016; Akerib et al. 2017; Cui et al. 2017; IceCube Collaboration et al. 2017; Sirunyan et al. 2017; Archambault et al. 2017; Albert et al. 2018; Aprile et al. 2018; Du et al. 2018; Aaboud et al. 2019; Buch et al. 2020; Froborg & Duffy 2020; Kannike et al. 2020; Rico 2020). While this does not mean that dark matter cannot communicate with standard model (SM) particles, as its SM couplings may be strongly suppressed, there is also the possibility that such interactions do not exist.

Since its discovery, subsequent evidence for particle dark matter from its coupling to gravity is almost irrefutable (Heymans et al. 2012; Anderson et al. 2014; Planck Collaboration et al. 2016). At the same time, the list of possible models that fit current constraints is still quite broad. A particularly well-suited signature that can be used to distinguish among dark matter models is the morphology and distribution of its substructure within dark matter halos. Some promising directions for inferring the properties of substructure include tidal streams (Ngan & Carlberg 2014; Bovy 2016; Carlberg 2016; Erkal et al. 2016; Benito et al. 2020; Shih et al. 2021) and astrometric observations (Feldmann & Spolyar 2015; Sanderson et al. 2016; Van Tilburg et al. 2018; Mishra-Sharma et al. 2020; Vattis et al. 2020; Mishra-Sharma 2022; Pardo & Doré 2021). A particularly sensitive probe is strong gravitational lensing (Buckley & Peter 2018; Drlica-Wagner et al. 2019; Simon et al. 2019), to which we restrict ourselves in this paper.

Strong gravitational lensing has already seen some promising success in extracting information about dark matter substructure from lensed quasars (Mao & Schneider 1998; Dalal & Kochanek 2002; Hsueh et al. 2017), observations with ALMA (Hezaveh et al. 2016), and extended lensing images (Koopmans 2005; Vegetti & Koopmans 2009a, 2009b). Various works have considered the expected signatures and methods to extract information about the underlying distribution of dark matter (Vegetti et al. 2010; Daylan et al. 2018).

More recently, there has been a plethora of applications of machine learning to this challenge, including classification (Alexander et al. 2020b; Diaz Rivero & Dvorkin 2020; Varma et al. 2020), regression (Brehmer et al. 2019), segmentation analysis (Ostdiek et al. 2020a, 2020b), and anomaly detection (Alexander et al. 2020c). To date, all works have exclusively focused on the application of these techniques to simulations, in large part due to the limited availability of strong lensing data, something that is anticipated to change in the near future with the commissioning of the Vera C. Rubin Observatory and

the launch of Euclid (Oguri & Marshall 2010; Verma et al. 2019). However, naively applying models trained on simulations to real data will not likely to be successful, as the data idiosyncrasies will significantly diminish the accuracy of the model. A promising method to bridge the gap between a model trained on simulations and real data is based on the technique of domain adaptation (DA; Ben-David et al. 2010).

A subset of transfer learning, DA is focused on the generalization of the model across different domains or data sets drawn from different underlying distributions. The goal of DA is to adapt a model trained on one data set (source) by generalizing it to another domain (target), where the objective of the model is unchanged. In practice, DA can be realized in several ways, including supervised, semisupervised, and unsupervised approaches (Donahue et al. 2013; Motiian et al. 2017; Farahani et al. 2020).

The DA has been used in a wide variety of applications in computer vision. Examples include adapting a model trained on synthetic images to real images (Peng et al. 2017), simple to complex images (Tzeng et al. 2017), and virtual worlds with controlled data to the real world (Schmidt et al. 2021).

Recently, Ciprijanovic et al. (2021) used unsupervised DA (UDA) to classify merging galaxies. Doing so achieved promising results, with an increase of up to 19% in accuracy compared to a model trained only on simulations. This work showed that models trained without DA achieved poor accuracy on real data. Another work (Ostdiek et al. 2020b) used DA to generalize an image segmentation algorithm to different gravitational lensing systems for subhalo detection.

In this work, we consider DA for dark matter searches in strong gravitational lensing. With the present lack of sufficient real data, we use two data sets to realize mock observations with different surveys, Hubble Space Telescope (HST) and Euclid, of galaxy–galaxy strong gravitational lensing to carefully test the performance of DA prior to its applications to real data. We evaluate the models trained on the source data set to identify various types of dark matter substructure on the target data set. We compare the performance of two DA algorithms based on convolutional neural networks (CNNs) and equivariant neural networks (ENNs) that incorporate a known group symmetry to enhance performance. To demonstrate the usefulness of DA for the transfer of knowledge to real data sets, we apply this technique to the problem of lens finding by performing UDA between a simulated data set and real lensing images from the Hyper Suprime-Cam (HSC).

We begin with a brief review of dark matter substructure and strong lensing signatures in Section 2. We then focus on the details of strong lensing simulations in Section 3, followed by a summary of DA algorithms in Section 4. We present our main results in Section 5, followed by the discussion and outlook in Section 6.

2. Dark Matter Detection and Strong Gravitational Lensing

2.1. Dark Matter Substructure

The Λ cold dark matter (ΛCDM) model envisions nearly scale invariant density fluctuations present in the early Universe serving as seeds of large-scale structure via hierarchical structure formation. Structures such as dark matter halos are formed from the coalescence of smaller halos (Kauffmann et al. 1993). Evidence for such mergers has been observed in our Galaxy (Necib et al. 2020a, 2020b; Chiti et al. 2021) and is a general prediction of N-body simulations, where evidence of mergers should remain largely intact. The distribution of subhalo masses is expected to follow a power-law distribution,

$$\frac{dN}{dm} \propto m^{\beta}, \tag{1}$$

where $\beta \sim -1.9$ has been found from simulations (Madau et al. 2008; Springel et al. 2008).

Comparison between simulations and observations indicates good agreement with ΛCDM on large scales (Heymans et al. 2012; Anderson et al. 2014; Planck Collaboration et al. 2016). However, discrepancies begin to arise on smaller, subgalactic scales. These include the core-versus-cusp (Burkert 1995; Oh et al. 2015), too-big-to-fail (Boylan-Kolchin et al. 2011), missing-satellite (Klypin et al. 1999; Moore et al. 1999; Bullock & Boylan-Kolchin 2017),[5] and diversity problems (Oh et al. 2015). While it may be the case that these problems can be addressed with a better understanding of the astrophysics, e.g., baryonic feedback (Benítez-Llambay et al. 2019), it is imperative that we consider the manifestations of other theories beyond ΛCDM.

Two natural directions to consider are a modification to the general theory of relativity, such as BF coupled (Alexander & Carballo-Rubio 2020a; Alexander et al. 2020d, 2022b) or Chern–Simons gravity (Jackiw & Pi 2003; Alexander & Yunes 2009), or alternatives to cold noninteractive dark matter. In the context of this paper, we will focus on the latter. An example of a dark matter model that addresses several of the tensions mentioned above is condensate dark matter, which can be realized in the context of Bose–Einstein (Sin 1994; Hu et al. 2000; Silverman & Mallett 2002; Sikivie & Yang 2009; Berezhiani & Khoury 2015; Hui et al. 2017; Ferreira et al. 2019) or Bardeen–Cooper–Schreifer (Alexander & Cormack 2017; Alexander et al. 2018, 2021) condensates. A concrete and well-motivated model is the axion. As the Goldstone boson of a broken U(1) symmetry, axions were originally introduced as a solution to the strong CP problem (Peccei & Quinn 1977; Weinberg 1978; Wilczek 1978). Shortly thereafter, it was recognized that they were a promising dark matter candidate (Abbott & Sikivie 1983; Dine & Fischler 1983; Preskill et al. 1983). Very light axions are particularly well suited to address some of the issues with structure on subgalactic scales. Axions with masses of $\sim 1 \times 10^{-23}$ eV have a de Broglie wavelength on kiloparsec scales, which realizes a natural solution to the core-versus-cusp problem. Additionally, light axions can form subhalos but also substructures that are quite different from standard CDM. These include vortices, disks, and interference patterns (Rindler-Daller & Shapiro 2012; Alexander et al. 2019, 2022a; Hui 2021; Hui et al. 2021).

2.2. Strong Lensing Theory

A powerful probe of dark matter substructure is strong gravitational lensing, an effect that is most pronounced near extended lensing arcs. Any over- or underdensities along the line of sight for an observer yields the lens equation, realized as an integral over an induced gravitational potential (Narayan & Bartelmann 1996),

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \frac{2}{c^2}\frac{D_{LS}}{D_S D_L}\nabla \int dz\, \Psi(\boldsymbol{r}), \tag{2}$$

[5] See Kim et al. (2018) for a differing perspective.

Table 1
Parameters with Distributions and Priors Used in the Simulation of Strong Lensing Images

DM Halo			
Param.	Dist.	Priors	Details
θ_x	Fixed	0	x position
θ_y	Fixed	0	y position
z	Fixed	0.5	Redshift
$M_{\rm tot}$	Fixed	1e12	Total halo mass in $M_\odot$
Ext. Shear			
Param.	Dist.	Priors	Details
$\gamma_{\rm ext}$	Uniform	[0.0, 0.3]	Magnitude
$\phi_{\rm ext}$	Uniform	[0, 2π]	Angle
Vortex			
Param.	Dist.	Priors	Details
θ_x	Normal	[0.0, 0.5]	x position
θ_y	Normal	[0.0, 0.5]	y position
l	Uniform	[0.5, 2.0]	Length of vortex
$\phi_{\rm vort}$	Uniform	[0, 2π]	Orientation from y-axis
$m_{\rm vort}$	Uniform	[3.0, 5.5]	% of mass in vortex
NFW Subhalo			
Param.	Dist.	Priors	Details
r	Uniform	[0, 2.0]	Radial distance from center
N	Poisson	$\mu = 25$	Number of subhalos
$\phi_{\rm sh}$	Uniform	[0, 2π]	Orientation from y-axis
$m_{\rm sh}$	Power law	[1e6, 1e10]	Subhalo mass in $M_\odot$
$\beta_{\rm sh}$	Fixed	−1.9	Power-law index

Note. Note that only a single type of substructure was used per image.

where D_{LS}, D_L, and D_S are the angular diameter distances from the lens to the source, the lens to the observer, and the source to the observer. The last term on the right-hand side of Equation (2) is known as the deflection angle, $\boldsymbol{\alpha}$, and is related via a perpendicular derivative to the lensing potential, $\Psi_{;\perp} = \boldsymbol{\alpha}$. The lensing potential can be shown to be related to matter density via a Poisson equation for gravitational lensing,

$$\nabla^2 \Psi = 2\kappa, \tag{3}$$

implying that lensing from separate sources, a dark matter halo and its subhalos, is a sum of deflection angles,

$$\boldsymbol{\alpha} = \sum_{\text{all matter}} \boldsymbol{\alpha}_i, \tag{4}$$

where $\boldsymbol{\alpha}_i$ could represent a dark matter halo, subhalos, vortices, external shear, interlopers, etc.

3. Strong Lensing Simulations

Similar to Alexander et al. (2020b, 2020c), we consider data sets of three substructure classes: no substructure, Navarro–Frenk–White (NFW) subhalos of CDM, and vortex substructure of superfluid-type (axion) dark matter. The parameters for the simulations are shown in Table 1. We construct the simulations with `lenstronomy` (Birrer & Amara 2018) to mimic the characteristics of HST and Euclid using the default instrument and observational settings. For background sources, we use images of galaxies from the Galaxy10 DECals data set (Dey et al. 2019) processed with a Gaussian mask and convolved with a Gaussian of 2 pixels in size. This prevents the lensing of unwanted foreground sources and noise in the image. We choose the apparent magnitude of the background galaxy such that the signal-to-noise ratio (S/N) of the lensing arcs is consistent with real lensing data: S/N $\sim$ 20 (Bolton et al. 2008).

We ensure that the total fraction of mass in the substructure, $f_{\rm sub}$, is of $\mathcal{O}(1\%)$. We constrain the simulations such that the total mass of the halo, including substructure, is always equal to $1 \times 10^{12}\, M_\odot$. This ensures that the classification algorithms do not simply recognize simulations without substructure as less massive on average. When simulating substructure for CDM, we draw subhalo masses from Equation (1) for a total number of sources taken from a Poisson draw with $\mu = 25$, consistent with the expected number of subhalos for our field of view and redshift range (Díaz Rivero et al. 2018). We model vortices of superfluid dark matter as uniform density strings of mass of varying length and orientation, a valid approximation at cosmological distances. Beyond the effects of substructure, we also include the impact of external shear due to large-scale structure.

The inclusion of the effects of substructure in our simulations can be understood from the linearity of the Poisson equation, Equation (3), which implies that the total lensing is just the sum of the individual contributions,

$$\boldsymbol{\alpha} = \boldsymbol{\alpha}_{\rm LSS} + \boldsymbol{\alpha}_{\rm halo} + \boldsymbol{\alpha}_{\text{halo-sub}}, \tag{5}$$

where $\boldsymbol{\alpha}_{\rm LSS}$ is the external shear from large-scale structure, $\boldsymbol{\alpha}_{\rm halo}$ is the lensing from the halo, and $\boldsymbol{\alpha}_{\text{halo-sub}}$ is for the halo substructure. Therefore, the location of an image can be found from a modified form of the lens equation, Equation (2),

$$\beta^i = \theta^i - \boldsymbol{\alpha}^i_{\rm LSS} - \boldsymbol{\alpha}^i_{\rm halo} - \boldsymbol{\alpha}^i_{\text{halo-sub}}. \tag{6}$$

The DA requires at least two data sets: the source and the target. In this work, we will demonstrate DA between models trained on mock HST observations and Euclid (and vice versa). Thus, it will be our goal to successfully adapt and evaluate the algorithms trained on a given source data set to the target data set.

4. Domain Adaptation

The goal of our work is to train a supervised model on a source data set and adapt it to a target data set. For this task, we use a CNN, specifically *EfficientNet* (Tan & Le 2019), as our base architecture. This is the same type of architecture that has achieved top performance in previous applications to lensing data sets (Alexander et al. 2020b, 2020c; Vattis et al. 2020). More generally, CNNs are known to outperform other methods of classification for strong gravitational lenses (Metcalf et al. 2019). Nonetheless, as noted by Ciprijanovic et al. (2021), a model trained on simulations can perform poorly on real data.

To improve the performance of models trained on simulated data, we use UDA, which attempts to mitigate the effects of the domain shift between the source and the target domains. It enables a transfer of knowledge gained from a labeled source data set to a distinct unlabeled target data set within the constraint that the objective remains the same (French et al. 2018). Examples from each data set are shown in Figure 1.

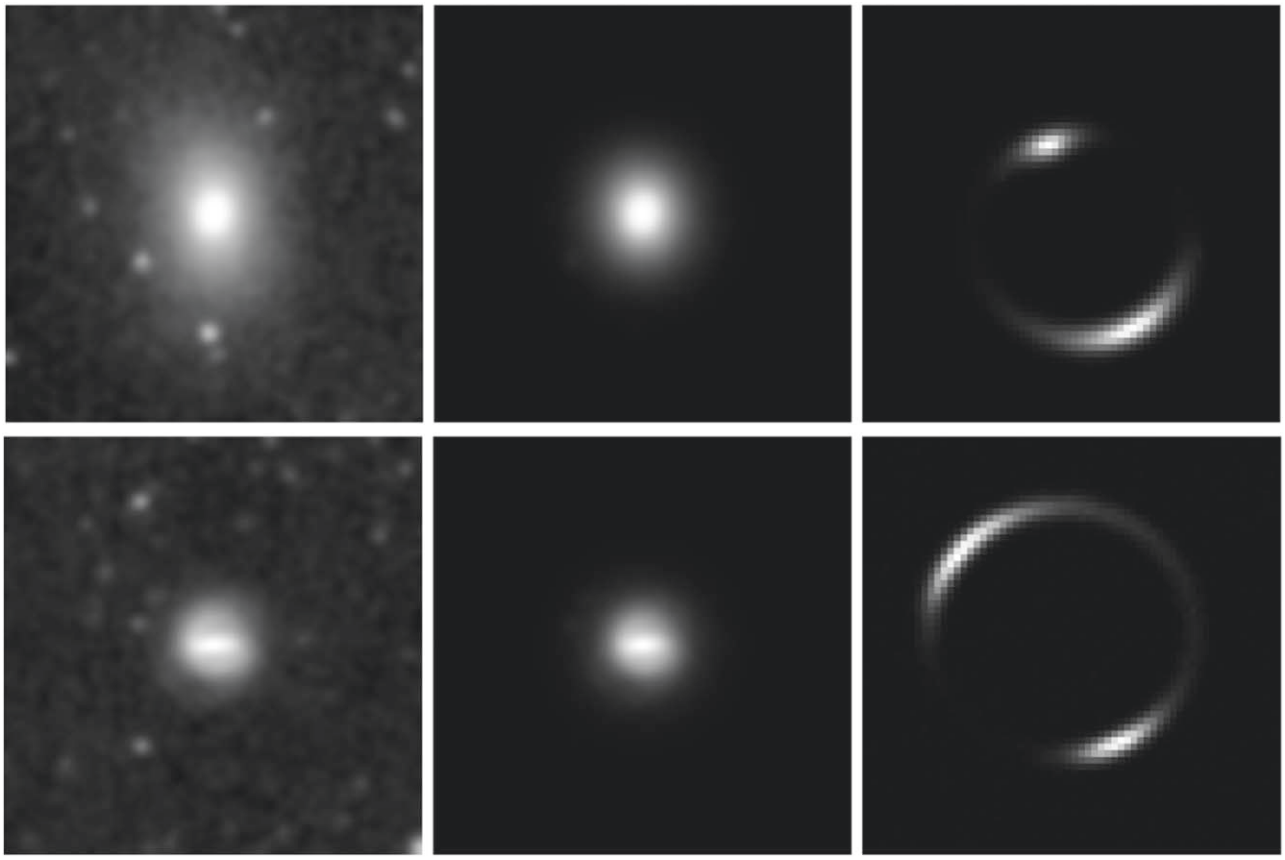

Figure 1. Example of lensing simulation of mock HST (top) and Euclid (bottom) data sets. Shown is an image of a real galaxy from the Galaxy10 DECals data set (left), the image processed with a Gaussian filter to remove unwanted background (middle), and the final lensed image (right).

4.1. Unsupervised Domain Adaptation

We utilize adversarial discriminative DA (ADDA; Tzeng et al. 2017), an adversarial adaptation method, with the goal of minimizing the domain discrepancy distance through an adversarial objective with respect to a discriminator. Ideally, the discriminator will be unable to distinguish between the source and the target distributions. We consider that we have access to source images $\boldsymbol{X}_s$ and labels $\boldsymbol{Y}_s$ that come from a source distribution $p_s(x, y)$, as well as target images $\boldsymbol{X}_t$ from a target distribution $p_t(x, y)$. Our objective is to learn a target encoder M_t and classifier C_t that classifies $\boldsymbol{X}_t$ into K classes.

Due to the fact that it is not possible to perform supervised learning on the target distribution, we learn a source encoder M_s and a source classifier C_s. The encoder learns to map the input samples to a latent vector whose dimensionality is lower than the dimensionality of the input samples. With these networks trained, the distance between $M_s(\boldsymbol{X}_s)$ and $M_t(\boldsymbol{X}_t)$ is minimized, as illustrated in Figure 2. Since we are only minimizing the encoders, we can assume that $C_s = C_t = C$.

We train M_s and C using a standard supervised loss. Then, we train a discriminator D that classifies whether the encoded vector represents an image from the source or target domain using a standard supervised loss, where the labels indicate the origin domain. Finally, we train M_t using D. To evaluate a target image $\boldsymbol{X}_t$, we perform $C(M_t(\boldsymbol{X}_t))$.

In addition to the baseline CNN models, we also consider an ENN (Weiler & Cesa 2021) for substructure classification. An ENN can be thought of as a generalization of a CNN that encodes the representation of a useful symmetry, both global or local, such that its group convolutions are invariant symmetries present in the data. This is useful if there is a known symmetry. As we expect lensing images to have symmetries beyond simple translation, for example, rotations, the flexibility of choosing different group representations is expected to improve the performance.

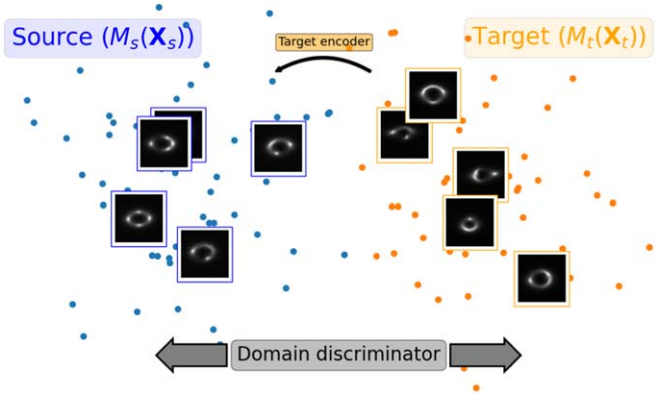

Figure 2. Depiction of target and source representations in the latent space of the encoder in ADDA. Images in blue are from the source domain, and images in orange are from the target domain.

The ENN we use consists of a group equivariant CNN (Cohen & Welling 2016) with six equivariant convolution blocks. We utilize the dihedral group D_2, whose symmetry mappings include the identity, rotations by $\pm\pi$, and horizontal/vertical reflections. The D_2 group structure is visualized in Figure 3. Each block is composed of a convolutional layer, a

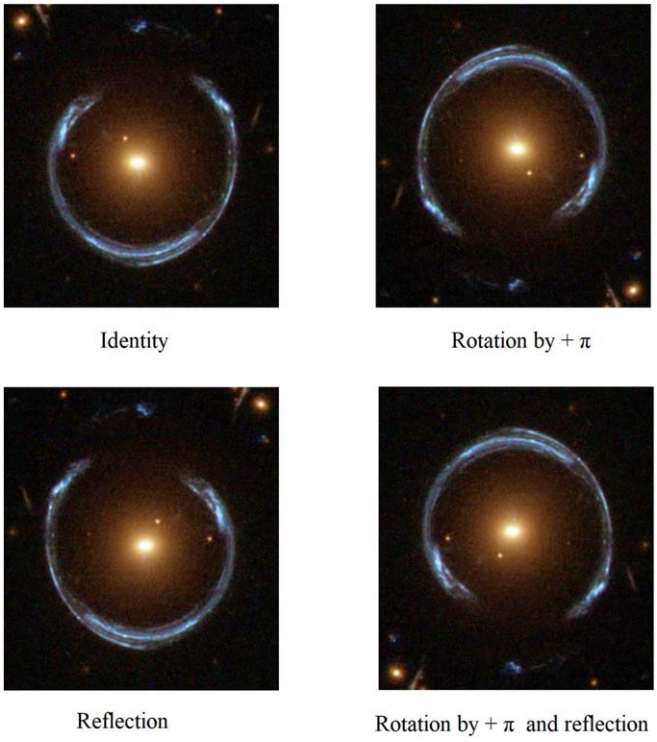

Figure 3. Visualization of the D_2 group structure for a real gravitational lens. Lensing image credit: ESA/Hubble & NASA.

batch normalization layer, and an ReLU activation function. After each pair of layers, we perform channel-wise average pooling, and in the end, we use a fully connected layer for multiclass classification. A schematic of the ENN architecture is presented in Figure 4.

4.2. Network Training

For training, we use 30,000 images for the source domain and 30,000 images for the target domain; in both cases, there are 10,000 images per class. For validation, we use 7500 images for the source domain and 7500 images for the target; in both cases, there are 2500 images per class. We use the *Adam* optimizer (Kingma & Ba 2014) to minimize the loss. We trained both *EfficientNet* and the ENN for 200 epochs, training with a patience of 15 epochs, such that if the accuracy of the model does not improve in 15 epochs, the training is stopped. For the final results, we considered the epoch that achieved the largest accuracy. Learning rate, weight decay, and other hyperparameters were optimized through a hyperparameter search and are available in Table 2.

We used random flip augmentations for both the source and target data set. We also found that the best results were obtained after random zooming (in a range of [0.8, 1.2]) and rotations in a range of [0°, 90°] on the source data set. We utilize the area under the receiver operating characteristic (ROC) curve on the target validation set as the metric for classifier performance for all of the models. All quoted area-under-the-ROC-curve (AUC) values are macro-averaged. All machine-learning models are implemented using `PyTorch` (Paszke et al. 2019b) and run on a single NVIDIA Tesla P100 GPU.

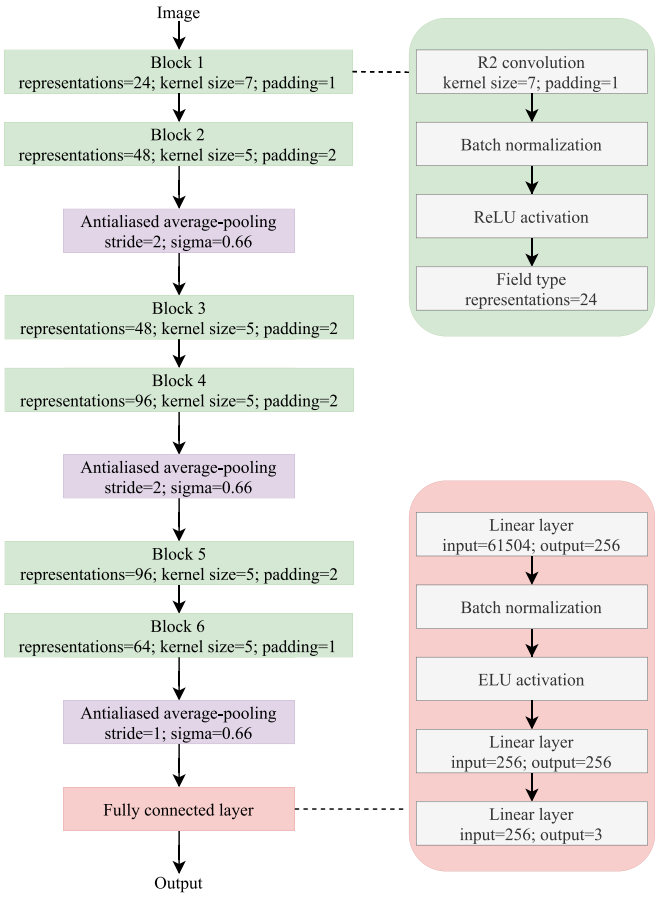

Figure 4. Schematic of the ENN architecture.

Table 2
Hyperparameters and Augmentations Used to Train the UDA Algorithm

Method	Hyperparameters	Hyperparameter Values	Augmentations
Supervised	Learning rate	2×10^{-3}	Horizontal flips
	Weight decay	1×10^{-5}	Vertical flips
	Cyclic scheduler	True	
ADDA	Learning rate target encoder	1×10^{-6}	Horizontal flips
	Learning rate critic	1×10^{-5}	Vertical flips
	Weight decay	2.5×10^{-5}	
	Cyclic scheduler	False	

5. Results

We test the applicability of UDA using ADDA between mock HST and Euclid observations in the context of multiclass classification of three types of substructure: no substructure, NFW subhalos of CDM, and superfluid DM vortices. We employ two different base classifiers: *EfficientNet* and an ENN.[6] The parameters for our data sets are compiled in Table 1. The results from our analysis are shown in Tables 3 and 4 for *EfficientNet* and the ENN, respectively. The ROC curves for both combinations of source/target in addition to the results for different architectures are presented in Figure 5. We

[6] The code used in our analysis can be found at https://github.com/ML4SCI/DeepLense/tree/main/Domain_Adaptation_for_DeepLense_Marcos_Tidball.

Table 3
Macro-averaged Area under the ROC Curve (AUC) for the *EfficientNet* Classifier

	Method	AUC
Euclid to HST		
	ADDA	0.924
	Supervised (source)	0.999
	Supervised (target)	0.685
HST to Euclid		
	ADDA	0.991
	Supervised (source)	0.999
	Supervised (target)	0.789

Table 4
Macro-averaged Area under the ROC Curve (AUC) for the ENN Classifier

	Method	AUC
Euclid to HST		
	ADDA	0.999
	Supervised (source)	0.999
	Supervised (target)	0.915
HST to Euclid		
	ADDA	0.999
	Supervised (source)	0.999
	Supervised (target)	0.973

also include the loss curves (Figure 7) and confusion matrices (Figures 8 and 9) for these models in the Appendix.

5.1. Domain Adaptation

We first train *EfficientNet* on the source data sets where it achieved a macro-averaged AUC ≈ 0.999 for both data sets and an accuracy of $\approx$99.6% for Euclid and $\approx$99.5% for HST. Applying these models to the corresponding target data sets naively, i.e., without DA, results in an AUC of $\approx$0.685 (0.789) when the Euclid (HST) model is applied to the HST (Euclid) data set and an accuracy of $\approx$50% (51%), a significantly degraded performance. This degradation is anticipated, even though the underlying physics is identical, due to different instrument systematics.

Following the application of UDA with ADDA, we observe a significant improvement in the application to the target data set. With the *EfficientNet*-based algorithm, we achieve an $\approx$35% improvement in AUC, at $\approx$0.924, and an accuracy of $\approx$87% adapting from the Euclid to HST data sets. Going the other direction, the AUC is improved to 0.991 and reaches an accuracy of 0.88%. This is a remarkable improvement in performance that can be further appreciated by comparing the ROC curves in Figure 5. The red dashed curves correspond to the naive application of *EfficientNet*, and the solid curves correspond to the performance with ADDA. While there is still some room for improvement in DA from Euclid to HST (left panel), DA from HST to Euclid results in near-perfect classification.

5.2. Equivariant DA

While *EfficientNet* saw a great improvement with ADDA, it is clear there is still a lot on the table, as standard binary classification has an AUC near unity. This informs us that one should be able to further improve on the UDA performance, particularly in the case of going from Euclid to HST. A natural improvement to consider is changing the baseline classifier to an ENN. When we leverage extra known symmetries of lensing, we should see an appreciable bump in performance, as our model does not need to learn this unimportant symmetry. In principle, one could also consider the inclusion of other UDA, but we will not consider that in this work.

Training and testing on source data sets, the ENN achieves an AUC of $\approx$0.999 on both data sets and an accuracy of 99.4% and 99.7% for Euclid and HST, respectively. The naive application to the test data set (i.e., no DA) again results in degraded performance, realized with an AUC for the Euclid- (HST-) trained model applied to HST (Euclid) data of $\approx$0.915 (0.973). This realizes a remarkable performance simply in the naive application of the ENN to the target data set. After training with ADDA, we find that our models are then able to achieve effectively perfect classification with AUCs of $\approx$0.999 for both combinations of source/target and an accuracy of 99.1% (97.5%) from Euclid (HST) to HST (Euclid). We see that the performance with our ENN is near optimal, achieving a significant performance bump over the CNN. This is truly impressive, as our DA algorithm is unsupervised—it never saw the labels from the target data set yet was nearly perfect at adapting to the new domain. This is, of course, exactly the kind of transfer of knowledge one would hope to be able to do between simulations and real data sets.

5.3. Application for Lens Finding

To further demonstrate the prospects of DA, we applied it to the task of lens finding. Identifying new lensed systems in current and upcoming wide-area surveys is an important problem that cannot be tackled efficiently by visual inspection due to the rarity of strong lenses. Previous work (Petrillo et al. 2017; Metcalf et al. 2019) has demonstrated the capability of CNNs in identifying lenses. Due to the relatively small number of confirmed lenses and lens candidates, most studies have used simulated lenses in their training data. However, this has inevitably been shown to result in a degradation in the performance of models when evaluated on real data. Given the increase in performance that we have seen in the last section, its seems likely that the difficulties induced by the differences between simulated and observational data sets could be mitigated by DA.

With this problem in mind, we take our target data set to consist of 10,000 nonlensed galaxies obtained in the HSC Subaru Strategic Program and 2500 real lens candidates covered by HSC observations, where 1000 of the HSC lenses were used in the test data set. The list of lens candidates was compiled based on previous lens searches (Moustakas 2012; Diehl et al. 2017; Pourrahmani et al. 2018; Jacobs et al. 2019; Cañameras et al. 2020, 2021; Huang et al. 2020, 2021; Sonnenfeld et al. 2020; Jaelani et al. 2021; Li et al. 2021; Garvin et al. 2022; Rojas et al. 2022; Shu et al. 2022; Stein et al. 2022; Storfer et al. 2022; Wong et al. 2022). We constructed our source data set to have the same distribution and size as the target data set and included 10,000 nonlensed galaxies and 2500 lenses, all simulated with `lenstronomy`. To obtain cutouts of the nonlenses and galaxies that were used as deflectors in simulations, we queried HSC Wide Public Data Release 2 with the following constraints: magnitude in the *i* band, `i_mag` < 24; spectroscopic redshift (from SDSS), $0.3 < z < 0.7$; and velocity

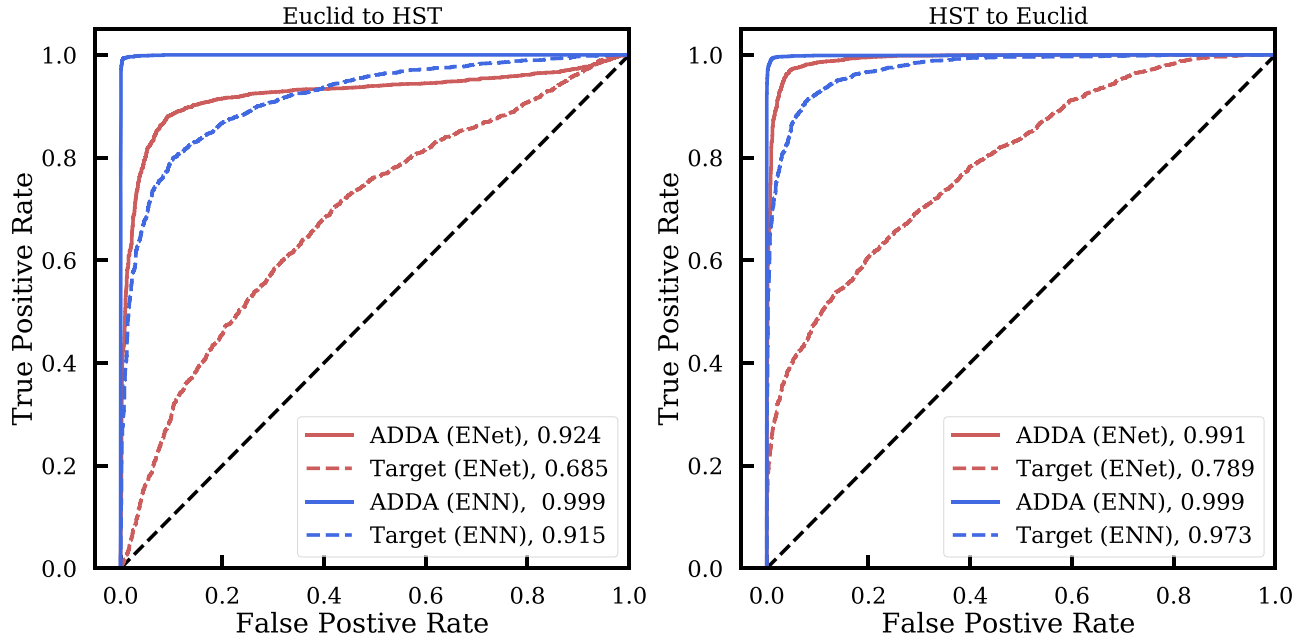

Figure 5. Application of ADDA for DA between two different mock simulations of strong lensing surveys: from Euclid to HST (left) and HST to Euclid (right). Dashed lines represent the naive application of a classifier trained on the source data set to the target, and solid lines represent the result after training with ADDA. Red curves represent models based on *EfficientNet*, and blue curves represent models based on an ENN.

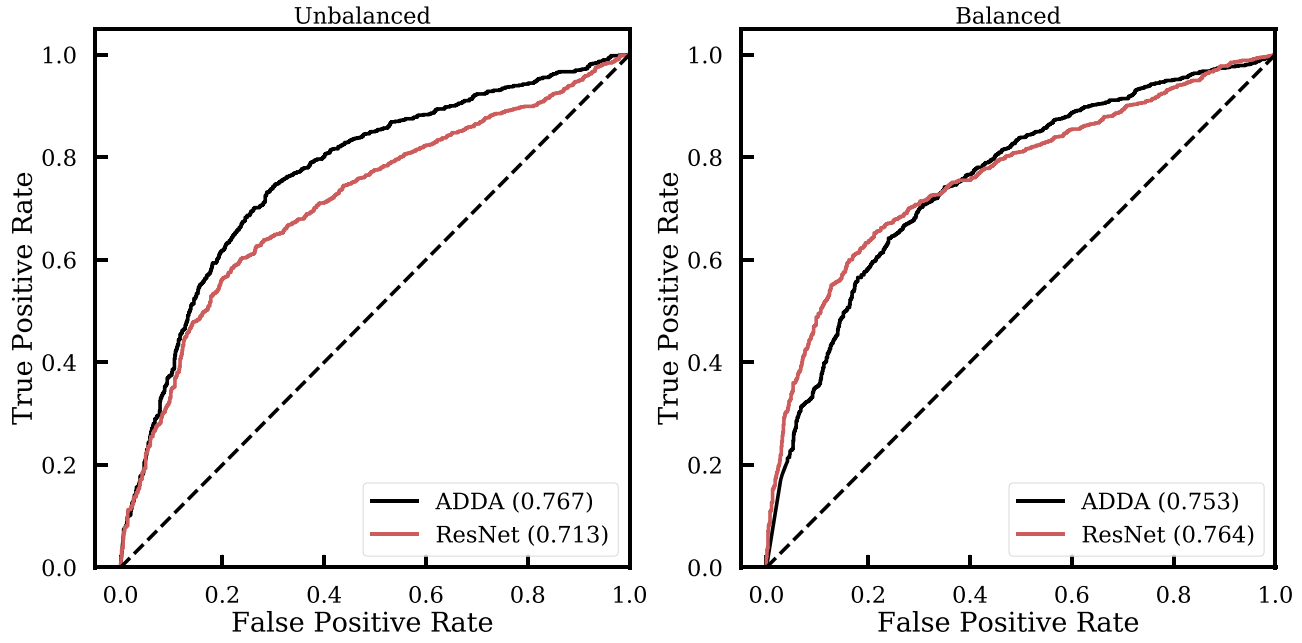

Figure 6. Comparison of ADDA and ResNet-18 performance for classifying objects into lensed and nonlensed galaxies for unbalanced (left) and balanced (right) data sets.

dispersion (from SDSS), $\sigma_V < 400$ km s^{-1}. To construct the simulated systems, we used the following steps.

1. Randomly choose a deflector.
2. Draw a value of Einstein radius from a uniform distribution between $0\farcs75$ and $2\farcs5$.
3. For a given velocity dispersion and redshift of the deflector and Einstein radius of the system, find a redshift of the background source assuming the singular isothermal sphere profile. For that purpose, we used the `lenstronomy` function `sis_sigma_v2theta_E`.
4. Randomly choose a galaxy from the Hubble Zoo Gems catalog (Keel et al. 2022) as a background source and ray trace it with `lenstronomy`, scaling it to the redshift computed in a previous step and assigning the magnitude drawn from a distribution similar to the source magnitude distribution in Collett (2015).

In addition, we built another data set (which we refer to as the balanced data set, as opposed to the unbalanced data set described previously) with a different distribution of lensed and nonlensed systems. In the source data set, we used 10,000

nonlensed galaxies from HSC and 10,000 simulated lenses, and for the target data set, we combined 2500 real lens candidates and 2500 nonlensed objects from HSC. In the balanced data set, there is an equal fraction of lensed and nonlensed systems; however, due to the limited number of observational data, this results in a different size for the source and target data sets, which could have a negative impact on the performance of our ADDA model.

We began with training ResNet-18 on the unbalanced (balanced) source data set (we include the loss curves in Figure 10 in the Appendix). While its accuracy on the simulated data reached 94% (96%) and an AUC of 0.98 (0.99), on the data set with real lenses, the accuracy and AUC dropped to 54% (64%) and 0.71 (0.76), respectively. Moreover, the recall score was only 0.09 (0.28), meaning that we were able to recover only a small fraction of lenses. The DA with ADDA resulted in increasing the accuracy to 68% (69%). We saw an appreciable increase in AUC for the unbalanced data set to 0.76 but a very slight decrease for the balanced data set to 0.75 (Figure 6). The recall score was improved to 0.50 and 0.55 for the unbalanced and balanced data set, respectively. While there is a slight degradation in precision, for both data sets, ADDA shows a large improvement in the number of detected lenses. The confusion matrices in Figure 11 in the Appendix also help us to understand the benefits of ADDA for this problem. We see that for the naive application to the target data sets, the model effectively always guesses a single class, in this case favoring the galaxy label for lenses. We see that the implementation of ADDA, while not perfect, is to alleviate this "confusion" between the class labels.

6. Discussion and Conclusion

With the upcoming arrival of strong gravitational lensing data from Euclid and the Vera C. Rubin Observatory, it is imperative to assess how algorithms trained on simulations can be applied to real-world data. In this work, we studied how UDA algorithms can be used to adapt a model trained on one set of data (the source) to another set of gravitational lensing data (the target). To make a quantitative comparison, we based our work on two sets of realistic lensing simulations, a mock data set from Euclid and another for HST.

We have demonstrated that the naive application of substructure classification models has diminished performance when applied to a different (target) data set. We then tested the implementation of ADDA, an unsupervised DA method, for a popular convolutional (*EfficientNet*) classifier and a symmetry ENN classifier. We found that ADDA consistently increases the performance of all models tested on the target data set. The ENN-based ADDA algorithm achieved top performance relative to the *EfficientNet*-based algorithm, achieving results competitive with the original source trained and evaluated algorithm. Lastly, we implemented ADDA for the problem of lens finding. In this case, we explicitly show that DA can help to alleviate performance losses when adapting a model trained in the controlled simulated domain to real data.

With the upcoming arrival of high-quality strong lensing data, DA techniques will be critical for real-world applications of machine learning–based dark matter analyses. Possible performance degradation for a simulation-trained model naively applied to real data sets can be more significant than what was realized here, particularly for the problem of substructure detection, making the need for further development and application of UDA methods even more critical. While we have restricted ourselves to classification in this work, DA techniques can be additionally useful in the broader context of studying dark matter from regression to image segmentation. In this work, we have not considered the impact of line-of-sight halos, i.e., interlopers (McCully et al. 2017; Despali et al. 2018; Gilman et al. 2019; Çağan Sengül et al. 2020). In some cases, their influence may dominate the signal of substructure, so it is important to take care that they are not incorrectly associated with the dark matter halo when working with real data sets (see, for example, Çağan Şengül et al. 2022). We leave the study of this effect for future work.

Acknowledgments

The authors thank Yashar Hezaveh and Laurence Perreault-Levasseur for helpful discussions. M.T. was a participant in the Google Summer of Code (GSoC) 2021 program. P.R. was a participant in the Google Summer of Code 2020 program. S.G. was supported in part by National Science Foundation award No. 2108645. S.A. and M.W.T. were supported in part by US National Science Foundation award No. 2108866. This work made use of these additional software packages: *Matplotlib* (Hunter 2007), *NumPy* (Harris et al. 2020), *PyTorch* (Paszke et al. 2019a), and *SciPy* (Virtanen et al. 2020).

Appendix
Extra Figures

In the Appendix we include extra figures with details of our models' performance after training. For our substructure models we show the loss curves with ADDA in Figure 7 and the confusion matrices for *EfficientNet* and the ENN in Figures 8 and 9, respectively. Similarly, for our lens identification model we display the loss curve in Figure 10 and confusion matrices in Figure 11.

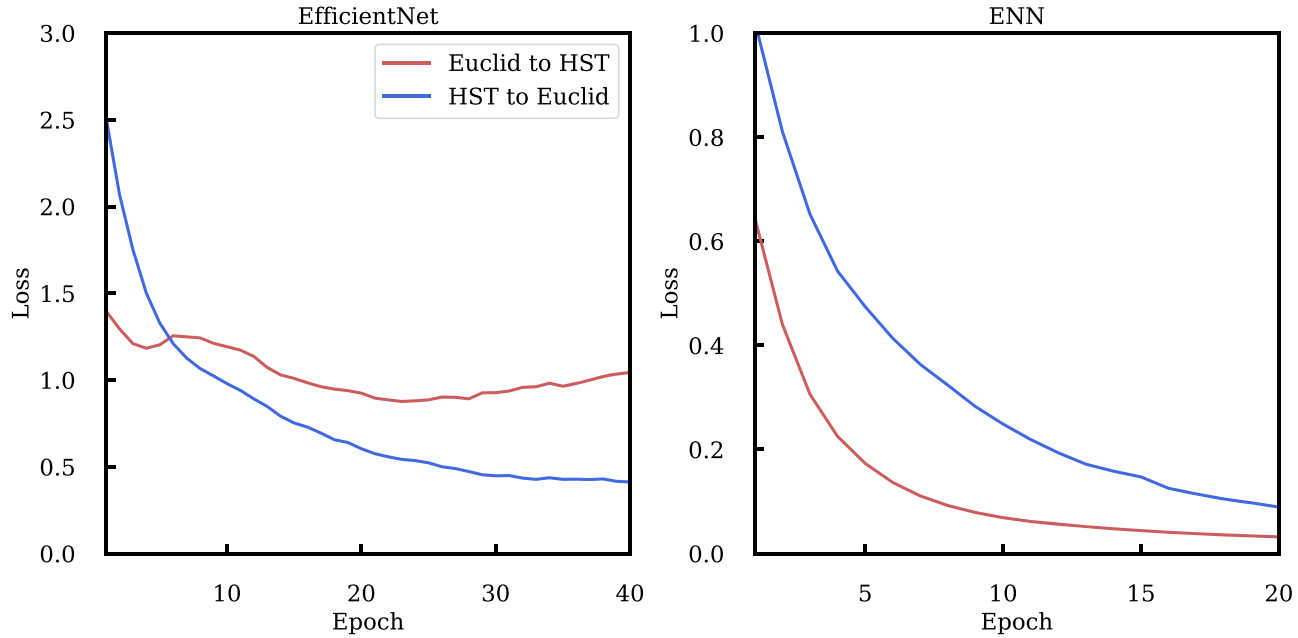

Figure 7. Evolution of loss during training for models with ADDA aimed at substructure identification.

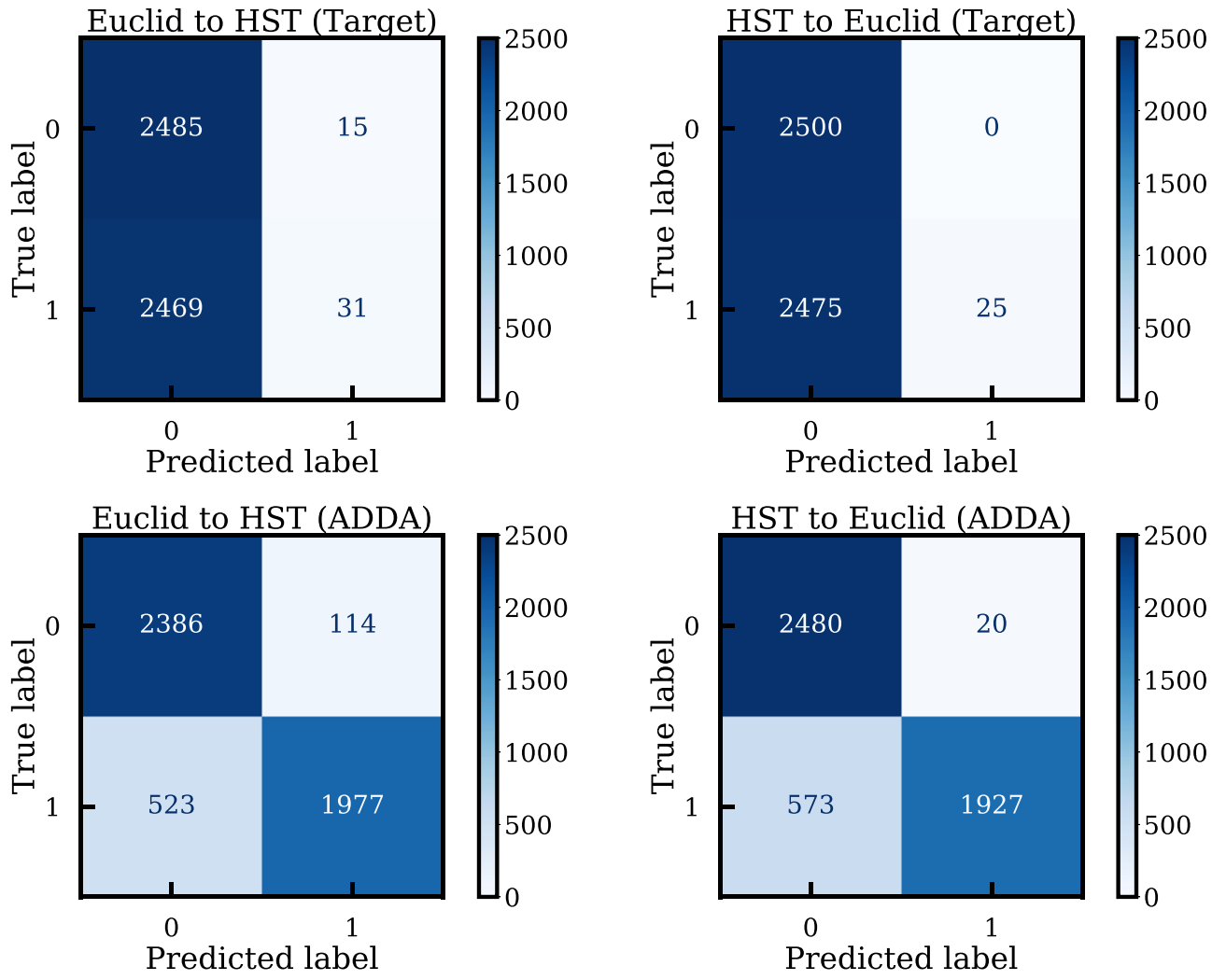

Figure 8. Confusion matrices for *EfficientNet* architecture for substructure classification with and without ADDA. The top row represents the naive application on the target data set, and the bottom row shows after training with ADDA. The left (right) column represents evaluation of a Euclid- (HST-) trained model to HST (Euclid) simulations.

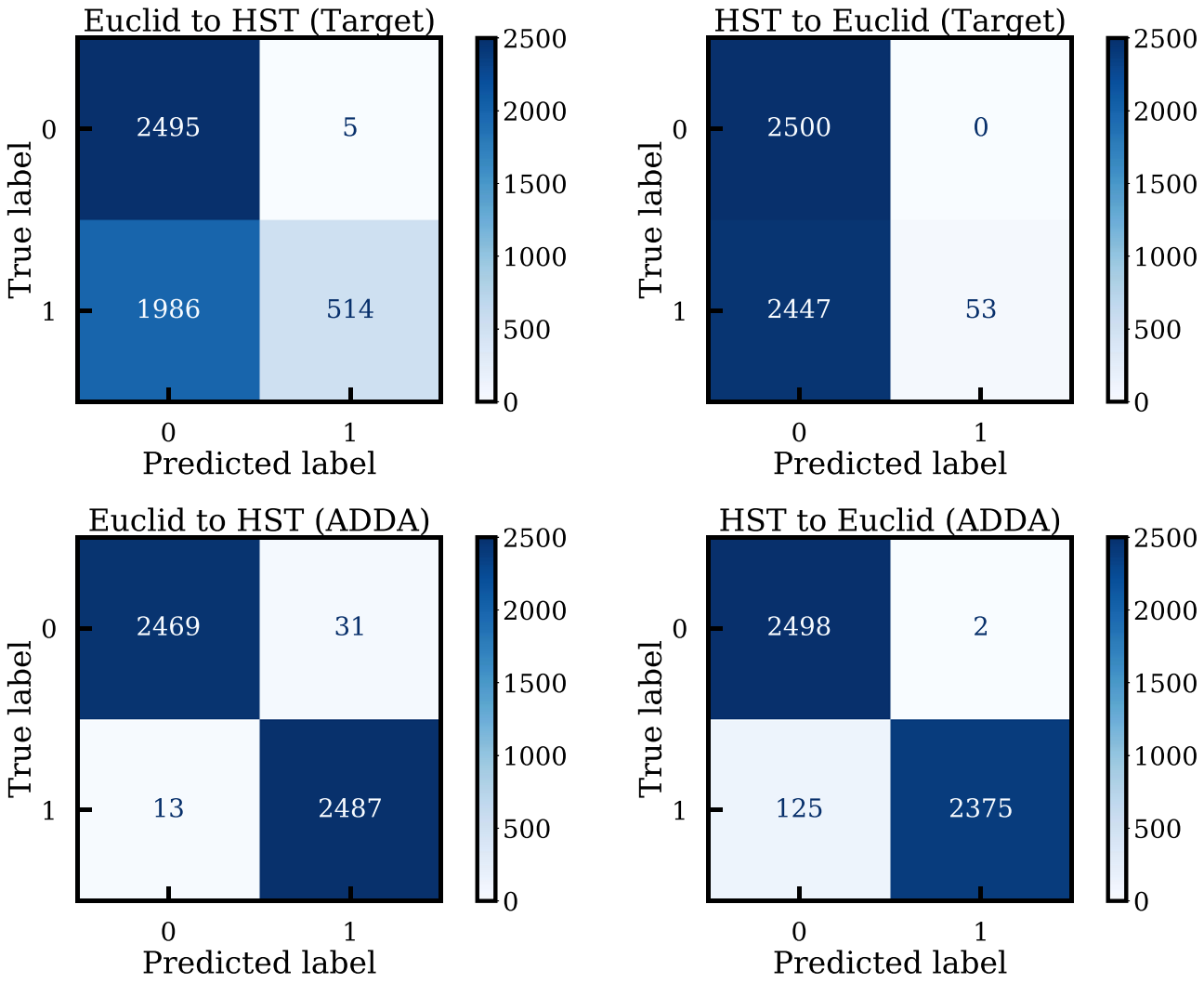

Figure 9. Confusion matrices for the ENN architecture for substructure classification with and without ADDA. The top row represents the naive application on the target data set, and the bottom row shows after training with ADDA. The left (right) column represents evaluation of a Euclid- (HST-) trained model to HST (Euclid) simulations.

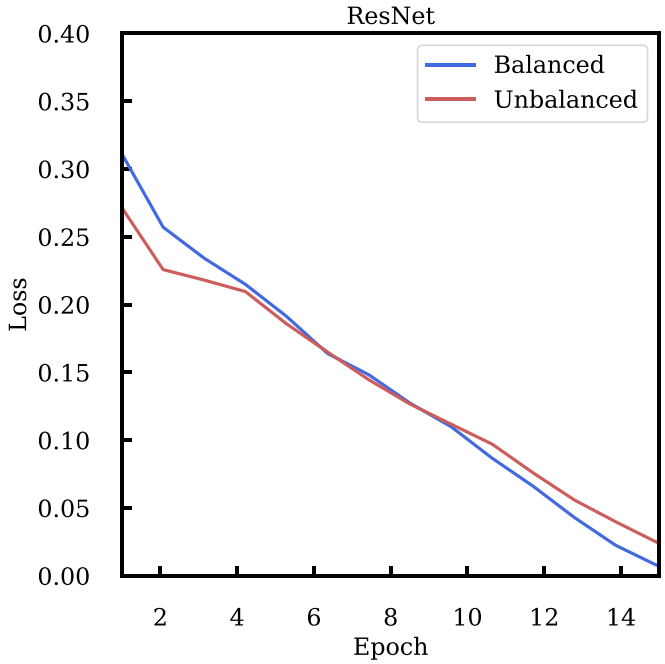

Figure 10. Evolution of loss during training with ADDA for models aimed at lens identification.

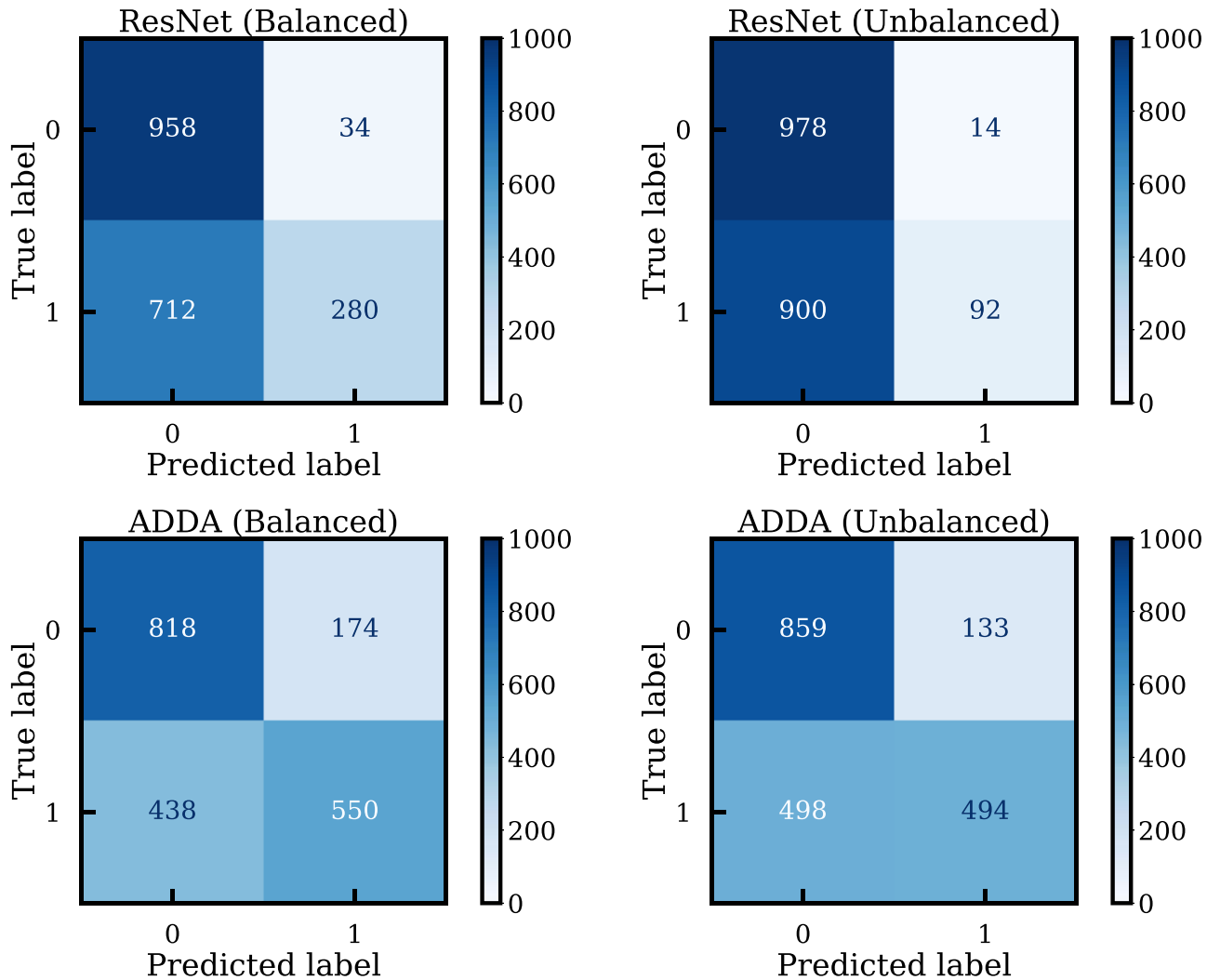

Figure 11. Confusion matrices for the detection of lensing images on real data. The top row represents the naive application on the target data set, and the bottom row shows after training with ADDA. The left and right columns represent training with balanced and unbalanced data sets, respectively.

ORCID iDs

Michael W. Toomey https://orcid.org/0000-0003-1205-4033

References

Aaboud, M., Aad, G., Abbott, B., et al. 2019, JHEP, 05, 142
Abbott, L. F., & Sikivie, P. 1983, PhLB, 120, 133
Akerib, D. S., Alsum, S., Araújo, H. M., et al. 2017, PhRvL, 118, 021303, 021303
Albert, A., Alfaro, R., Alvarez, C., et al. 2018, ApJ, 853, 154
Alexander, S., Bramburger, J. J., & McDonough, E. 2019, PhLB, 797, 134871
Alexander, S., Capanelli, C., Ferreira, E. G. M., & McDonough, E. 2022a, PhLB, 833, 137298
Alexander, S., & Carballo-Rubio, R. 2020a, PhRvD, 101, 024058
Alexander, S., Clark, S. J., Herczeg, G., & Toomey, M. W. 2022b, PhRvD, 105, 044001
Alexander, S., & Cormack, S. 2017, JCAP, 1704, 005
Alexander, S., Gleyzer, S., McDonough, E., Toomey, M. W., & Usai, E. 2020b, ApJ, 893, 15
Alexander, S., Gleyzer, S., Parul, H., et al. 2020c, arXiv:2008.12731
Alexander, S., Herczeg, G., Liu, J., & McDonough, E. 2020d, PhRvD, 102, 083526
Alexander, S., McDonough, E., & Spergel, D. N. 2018, JCAP, 1805, 003
Alexander, S., McDonough, E., & Spergel, D. N. 2021, PhLB, 822, 136653
Alexander, S., & Yunes, N. 2009, PhR, 480, 1
Anderson, L., Aubourg, E., Bailey, S., et al. 2014, MNRAS, 441, 24
Aprile, E., Aalbers, J., Agostini, F., et al. 2018, PhRvL, 121, 111302
Archambault, S., Archer, A., & Benbow, W. 2017, PhRvD, 95, 082001
Ben-David, S., Blitzer, J., Crammer, K., et al. 2010, Mach. Learn., 79, 151
Benítez-Llambay, A., Frenk, C. S., Ludlow, A. D., & Navarro, J. F. 2019, MNRAS, 488, 2387
Benito, M., Criado, J. C., Hütsi, G., Raidal, M., & Veermäe, H. 2020, PhRvD, 101, 103023
Berezhiani, L., & Khoury, J. 2015, PhRvD, 92, 103510
Birrer, S., & Amara, A. 2018, PDU, 22, 189
Bolton, A. S., Burles, S., Koopmans, L. V. E., et al. 2008, ApJ, 682, 964
Bovy, J. 2016, PhRvL, 116, 121301
Boylan-Kolchin, M., Bullock, J. S., & Kaplinghat, M. 2011, MNRAS, 415, L40
Brehmer, J., Mishra-Sharma, S., Hermans, J., Louppe, G., & Cranmer, K. 2019, ApJ, 886, 49
Buch, J., Buen-Abad, M. A., Fan, J., & Chau Leung, J. S. 2020, JCAP, 2020, 051
Buckley, M. R., & Peter, A. H. G. 2018, PhR, 761, 1
Bullock, J. S., & Boylan-Kolchin, M. 2017, ARA&A, 55, 343
Burkert, A. 1995, ApJL, 447, L25
Çağan Şengül, A., Dvorkin, C., Ostdiek, B., & Tsang, A. 2022, MNRAS, 515, 4391
Çağan Şengül, A., Tsang, A., Diaz Rivero, A., et al. 2020, PhRvD, 102, 063502
Cañameras, R., Schuldt, S., Shu, Y., et al. 2021, yCat, J/A+A/653/L6
Cañameras, R., Schuldt, S., Suyu, S. H., et al. 2020, A&A, 644, A163
Carlberg, R. G. 2016, ApJ, 820, 45
Chiti, A., Frebel, A., Simon, J. D., et al. 2021, NatAs, 5, 392
Ciprijanovic, A., Kafkes, D., Downey, K., et al. 2021, MNRAS, 506, 677
Cohen, T., & Welling, M. 2016, in Proc. of the 33rd Int. Conf. on Machine Learning, 48, ed. M. F. Balcan & K. Q. Weinberger (PMLR), 2990
Collett, T. E. 2015, ApJ, 811, 20
Cui, X., Abdukerim, A., Chen, W., et al. 2017, PhRvL, 119, 181302
Dalal, N., & Kochanek, C. S. 2002, ApJ, 572, 25
Daylan, T., Cyr-Racine, F.-Y., Diaz Rivero, A., Dvorkin, C., & Finkbeiner, D. P. 2018, ApJ, 854, 141
Despali, G., Vegetti, S., White, S. D. M., Giocoli, C., & van den Bosch, F. C. 2018, MNRAS, 475, 5424
Dey, A., Schlegel, D. J., Lang, D., et al. 2019, AJ, 157, 168
Diaz Rivero, A., & Dvorkin, C. 2020, PhRvD, 101, 023515
Díaz Rivero, A., Dvorkin, C., Cyr-Racine, F.-Y., Zavala, J., & Vogelsberger, M. 2018, PhRvD, 98, 103517
Diehl, H. T., Buckley-Geer, E. J., Lindgren, K. A., et al. 2017, ApJS, 232, 15
Dine, M., & Fischler, W. 1983, PhLB, 120, 137
Donahue, J., Hoffman, J., Rodner, E., Saenko, K., & Darrell, T. 2013, in IEEE Conf. on Computer Vision and Pattern Recognition, 668
Drlica-Wagner, A., Mao, Y.-Y., Adhikari, S., et al. 2019, arXiv:1902.01055
Drukier, A. K., Freese, K., & Spergel, D. N. 1986, PhRvD, 33, 3495
Du, N., Force, N., Khatiwada, R., et al. 2018, PhRvL, 120, 151301
Erkal, D., Belokurov, V., Bovy, J., & Sanders, J. L. 2016, MNRAS, 463, 102
Farahani, A., Voghoei, S., Rasheed, K., & Arabnia, H. R. 2020, arXiv:2010.03978
Feldmann, R., & Spolyar, D. 2015, MNRAS, 446, 1000
Fermi LAT Collaboration 2015, JCAP, 2015, 008
Ferreira, E. G., Franzmann, G., Khoury, J., & Brandenberger, R. 2019, JCAP, 2019, 027
French, G., Mackiewicz, M., & Fisher, M. 2018, arXiv:1706.05208
Froborg, F., & Duffy, A. R. 2020, JPhG, 47, 094002
Garvin, E. O., Kruk, S., Cornen, C., et al. 2022, A&A, 667, A141
Geringer-Sameth, A., Koushiappas, S. M., & Walker, M. G. 2015, PhRvD, 91, 083535
Gilman, D., Birrer, S., Treu, T., Nierenberg, A., & Benson, A. 2019, MNRAS, 487, 5721
Goodman, M. W., & Witten, E. 1985, PhRvD, 31, 3059
Graham, P. W., Irastorza, I. G., Lamoreaux, S. K., Lindner, A., & van Bibber, K. A. 2015, ARNPS, 65, 485
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357
Heymans, C., Van Waerbeke, L., Miller, L., et al. 2012, MNRAS, 427, 146
Hezaveh, Y. D., Dalal, N., Marrone, D. P., et al. 2016, ApJ, 823, 37
Hsueh, J.-W., Oldham, L., Spingola, C., et al. 2017, MNRAS, 469, 3713
Hu, W., Barkana, R., & Gruzinov, A. 2000, PhRvL, 85, 1158
Huang, X., Storfer, C., Gu, A., et al. 2021, ApJ, 909, 27
Huang, X., Storfer, C., Ravi, V., et al. 2020, ApJ, 894, 78
Hui, L. 2021, ARA&A, 59, 247
Hui, L., Joyce, A., Landry, M. J., & Li, X. 2021, JCAP, 2021, 011
Hui, L., Ostriker, J. P., Tremaine, S., & Witten, E. 2017, PhRvD, 95, 043541
Hunter, J. D. 2007, CSE, 9, 90
IceCube Collaboration, Aartsen, M. G., & Ackermann, M. 2017, arXiv:1705.08103
Jackiw, R., & Pi, S. Y. 2003, PhRvD, 68, 104012
Jacobs, C., Collett, T., Glazebrook, K., et al. 2019, ApJS, 243, 17
Jaelani, A. T., Rusu, C. E., Kayo, I., et al. 2021, MNRAS, 502, 1487
Kannike, K., Raidal, M., Veermäe, H., Strumia, A., & Teresi, D. 2020, PhRvD, 102, 095002
Kauffmann, G., White, S. D. M., & Guiderdoni, B. 1993, MNRAS, 264, 201
Keel, W. C., Tate, J., Wong, O. I., et al. 2022, AJ, 163, 150
Kim, S. Y., Peter, A. H. G., & Hargis, J. R. 2018, PhRvL, 121, 211302
Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
Klypin, A., Kravtsov, A. V., Valenzuela, O., & Prada, F. 1999, ApJ, 522, 82
Koopmans, L. V. E. 2005, MNRAS, 363, 1136
Li, R., Napolitano, N. R., Spiniello, C., et al. 2021, ApJ, 923, 16
Madau, P., Diemand, J., & Kuhlen, M. 2008, ApJ, 679, 1260
MAGIC Collaboration 2016, JCAP, 2016, 039
Mao, S., & Schneider, P. 1998, MNRAS, 295, 587
McCully, C., Keeton, C. R., Wong, K. C., & Zabludoff, A. I. 2017, ApJ, 836, 141
Metcalf, R. B., Meneghetti, M., Avestruz, C., et al. 2019, A&A, 625, A119
Mishra-Sharma, S. 2022, MLS&T, 3, 01LT03
Mishra-Sharma, S., Van Tilburg, K., & Weiner, N. 2020, PhRvD, 102, 023026
Moore, B., Ghigna, S., Governato, F., et al. 1999, ApJL, 524, L19
Motiian, S., Piccirilli, M., Adjeroh, D. A., & Doretto, G. 2017, arXiv:1709.10190
Moustakas, L. 2012, HST Proposal, 12833
Narayan, R., & Bartelmann, M. 1996, arXiv:9606001
Necib, L., Ostdiek, B., Lisanti, M., et al. 2020a, ApJ, 903, 25
Necib, L., Ostdiek, B., Lisanti, M., et al. 2020b, NatAs, 4, 1078
Ngan, W.-H. W., & Carlberg, R. G. 2014, ApJ, 788, 181
Oguri, M., & Marshall, P. J. 2010, MNRAS, 405, 2579
Oh, S.-H., Hunter, D. A., Brinks, E., et al. 2015, AJ, 149, 180
Ostdiek, B., Diaz Rivero, A., & Dvorkin, C. 2020a, arXiv:2009.06639
Ostdiek, B., Diaz Rivero, A., & Dvorkin, C. 2020b, A&A, 657, L14
Pardo, K., & Doré, O. 2021, PhRvD, 104, 103531
Paszke, A., Gross, S., Massa, F., et al. 2019a, in Advances in Neural Information Processing 32, ed. H. Wallach et al. (Red Hook, NY: Curran Associates, Inc.), 8024
Paszke, A., Gross, S., Massa, F., et al. 2019b, arXiv:1912.01703
Peccei, R. D., & Quinn, H. R. 1977, PhRvL, 38, 1440
Peng, X., Usman, B., Kaushik, N., et al. 2017, arXiv:1710.06924
Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2017, MNRAS, 472, 1129
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13
Pourrahmani, M., Nayyeri, H., & Cooray, A. 2018, ApJ, 856, 68
Preskill, J., Wise, M. B., & Wilczek, F. 1983, PhLB, 120, 127
Rico, J. 2020, Galax, 8, 25
Rindler-Daller, T., & Shapiro, P. R. 2012, MNRAS, 422, 135
Rojas, K., Savary, E., Clément, B., et al. 2022, A&A, 668, A73
Sanderson, R. E., Vera-Ciro, C., Helmi, A., & Heit, J. 2016, arXiv:1608.05624

Schmidt, V., Luccioni, A. S., Teng, M., et al. 2021, arXiv:2110.02871
Shih, D., Buckley, M. R., Necib, L., & Tamanas, J. 2021, MNRAS, 509, 5992
Shu, Y., Cañameras, R., Schuldt, S., et al. 2022, A&A, 662, A4
Sikivie, P., & Yang, Q. 2009, PhRvL, 103, 111301
Silverman, M. P., & Mallett, R. L. 2002, GReGr, 34, 633
Simon, J., Birrer, S., Bechtol, K., et al. 2019, BAAS, 51, 153
Sin, S.-J. 1994, PhRvD, 50, 3650
Sirunyan, A. M., Tumasyan, A., Adam, W., et al. 2017, JHEP, 2017, 73
Sonnenfeld, A., Verma, A., More, A., et al. 2020, yCat, J/A+A/642/A148
Springel, V., Wang, J., Vogelsberger, M., et al. 2008, MNRAS, 391, 1685
Stein, G., Blaum, J., Harrington, P., Medan, T., & Lukić, Z. 2022, ApJ, 932, 107
Storfer, C., Huang, X., Gu, A., et al. 2022, arXiv:2206.02764
Tan, M., & Le, Q. V. 2019, arXiv:1905.11946
The Super-Kamiokande Collaboration, Choi, K., Abe, K., et al. 2015, arXiv:1503.04858
Tzeng, E., Hoffman, J., Saenko, K., & Darrell, T. 2017, arXiv:1702.05464
Van Tilburg, K., Taki, A.-M., & Weiner, N. 2018, JCAP, 2018, 041
Varma, S., Fairbairn, M., & Figueroa, J. 2020, arXiv:2005.05353
Vattis, K., Toomey, M. W., & Koushiappas, S. M. 2020, PhRvD, 104, 123541
Vegetti, S., & Koopmans, L. V. E. 2009a, MNRAS, 392, 945
Vegetti, S., & Koopmans, L. V. E. 2009b, MNRAS, 400, 1583
Vegetti, S., Koopmans, L. V. E., Bolton, A., Treu, T., & Gavazzi, R. 2010, MNRAS, 408, 1969
Verma, A., Collett, T., Smith, G. P. & Strong Lensing Science Collaboration, and the DESC Strong Lensing Science Working Group 2019, arXiv:1902.05141
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Weiler, M., & Cesa, G. 2021, arXiv:1911.08251
Weinberg, S. 1978, PhRvL, 40, 223
Wilczek, F. 1978, PhRvL, 40, 279
Wong, K. C., Chan, J. H. H., Chao, D. C. Y., et al. 2022, PASJ, 74, 1209