Domain Adaptation for Measurements of Strong Gravitational Lenses

Paxson Swierc[1]
Department of Astronomy and Astrophysics
University of Chicago
pswierc@uchicago.edu

Megan Zhao[1]
Department of Astronomy and Astrophysics
University of Chicago
yifanzyf@uchicago.edu

Aleksandra Ćiprijanović
Computational Science and AI Directorate
Fermi National Accelerator Laboratory;
Department of Astronomy and Astrophysics
University of Chicago
aleksand@fnal.gov

Brian Nord
Computational Science and AI Directorate
Fermi National Accelerator Laboratory;
Department of Astronomy and Astrophysics
University of Chicago;
Kavli Institute for Cosmological Physics
University of Chicago
nord@fnal.gov

Abstract

Upcoming surveys are predicted to discover galaxy-scale strong lenses on the order of 10^5, making deep learning methods necessary in lensing data analysis. Currently, there is insufficient real lensing data to train deep learning algorithms, but the alternative of training only on simulated data results in poor performance on real data. Domain Adaptation may be able to bridge the gap between simulated and real datasets. We utilize domain adaptation for the estimation of Einstein radius (Θ_E) in simulated galaxy-scale gravitational lensing images with different levels of observational realism. We evaluate two domain adaptation techniques - Domain Adversarial Neural Networks (DANN) and Maximum Mean Discrepancy (MMD). We train on a source domain of simulated lenses and apply it to a target domain of lenses simulated to emulate noise conditions in the Dark Energy Survey (DES). We show that both domain adaptation techniques can significantly improve the model performance on the more complex target domain dataset. This work is the first application of domain adaptation for a regression task in strong lensing imaging analysis. Our results show the potential of using domain adaptation to perform analysis of future survey data with a deep neural network trained on simulated data.

1 Introduction

Strong gravitational lensing is a powerful probe of astrophysics, as well as cosmology. Galaxy-scale strong lensing systems in the Sloan Lens ACS (SLACS) [5] survey gave insights into the dark and baryonic matter profiles in massive elliptical galaxies [6, 8]. In cosmology, strong lensing systems have been used as an independent probe of the Hubble constant through time delays in multiple images in the system [10, 33] and for constraining the dark energy equation of state [15, 32].

Upcoming large-scale astronomical surveys, including the Rubin Observatory's Legacy Survey of Space and Time (LSST)[23], the Hyper Suprime-Cam Subaru Strategic Program (HSC)[3], Euclid[27], and the Nancy Grace Roman Space Telescope [2] , will observe an unprecedented number

[1]Equal contribution.

[2]https://www.jpl.nasa.gov/missions/the-nancy-grace-roman-space-telescope

Machine Learning and the Physical Sciences Workshop, NeurIPS 2023.

of lensed systems. Collett [14] predicted that $> 10^5$ galaxy scale strong lensing systems will be observed by LSST and Euclid. The large amount of data from these surveys will require much more efficient analysis techniques, such as those utilizing machine learning (ML) techniques. Currently discovered lensing systems are on the order of a thousand, which is inadequate to train many ML algorithms, forcing researchers to use simulated datasets of various forms, such as in [38, 29, 36]. For classification tasks, it has been shown that the inevitable gap between training and testing data causes complex, large-parameter ML models trained on labeled simulation data to perform poorly (i.e., systematic biases) on new, unlabeled observational data [e.g., 7]. The same issues with classification are present in regression tasks, requiring an advancement to mitigate biases in deep learning analyses.

Domain adaptation (DA) is a class of techniques used when the distribution of training data differs from that of the testing data [16]. In DA, the training distribution of data is the source data, while the data it will be tested on is the target data. We are specifically evaluating unsupervised domain adaptation [31], which is the problem scenario where one has access to labeled source data, but target data is unlabeled. Unsupervised DA techniques adjust training to include unlabeled target data, detect the difference between the source and the target domain distributions, and aim to minimize it. In astronomy, DA has been applied to multiple scenarios of classification [e.g., 11, 13, 12, 4, 34]), generative models [e.g., 24, 26], and regression in one scenario [21].

We apply two DA techniques to strong gravitational lensing analysis in this work: adversarial training using Domain Adversarial Neural Networks (DANNs) [20] and a distance-based method called Maximum Mean Discrepancy (MMD) [22]. We apply both techniques to the task of estimating the Einstein radius (Θ_E) of galaxy-scale gravitational lensing systems. Both source and target domain data are simulated using `deeplenstronomy`[1] [25]. The source data is simulated strong lenses, with no noise. For target data, the same strong lenses are simulated, but with emulated Dark Energy Survey (DES)[1] conditions. We show that both techniques improve the model performance on the target domain, demonstrating the capacity of using DA to bridge the gap between simulated and real data.

Our manuscript is organized as follows. In Section 2 we describe the two DA methods utilized in this work; in Section 3 we describe the data; in Section 4 we describe the neural network model; in Section 5 we present our results; and in Section 6 we discuss our findings.

2 Methods

Both DA approaches used are unsupervised techniques: while they require labeled data from the source domain, they leverage only unlabeled data from the target domain. This mirrors the problem setting of training on labeled simulated data, while only having access to unlabeled real data. We assume that the two domains have the same conditional probability distributions of a label y, given some data x, $p_s(y|x) = p_t(y|x)$, but that the marginal probability distribution of the data x within the domains themselves is not equal: $p_s(x) \neq p_t(x)$. This is known as a covariate shift between domains [19]. Under this assumption, feature-based DA techniques such as DANN and MMD aim to close the gap by learning a transformation that maps source data into target data [19]. DANN and MMD find this transformation by utilizing domain invariant features. In practice, a Convolutional Neural Network (CNN) feature extractor learns to produce invariant features that are identical between the two domains.

We use a total loss: $\mathcal{L}_{Total} = \mathcal{L}_{MSE} + \lambda_{DA}\mathcal{L}_{DA}$, where $\mathcal{L}_{MSE}$ is the Mean Squared Error (MSE) loss used to minimize the error on the task the model is trained to perform (estimate the Einstein radius of a gravitational lens). $\mathcal{L}_{DA}$ is either the MMD loss or the DANN loss, and it minimizes the variance of extracted features between domains. This loss is weighted by constant λ_{DA}.

DANN [20] uses a two-headed network, with an adversarial approach to find invariant features. All data is fed into a feature extractor made of convolutional layers, which is then forwarded to the two heads, both made of linearly connected layers. The first head is the label predictor, which takes features and makes predictions of the labels associated with them. This head is only trained with features extracted from labeled source data and minimizes $\mathcal{L}_{MSE}$. The second head is the domain classifier, which takes features from both domains and makes predictions as to which domain those features came from. We use a Negative Log-Likelihood Loss function (NLL) loss as $\mathcal{L}_{DA}$. With DANNs, the goal is to learn domain-invariant features, which will maximize $\mathcal{L}_{DA}$ loss and confuse

[1] https://github.com/deepskies/deeplenstronomy

the domain classifier. To simultaneously optimize the the feature extractor to find features that will enable classification of labeled data in the label predictor head, and confuse the domain classifier head (rendering it unable to differentiate between the domains), a gradient reversal layer is used in between the feature extractor and the domain classifier. This layer does not affect the data in forward propagation, but multiplies the gradient by -1 in back propagation, flipping the gradient.

MMD [22] is a distance-based domain adaptation method that identifies invariant features by minimizing the distance between the source and target domain distributions. We employ kernel methods to calculate MMD distance, which we use as our DA loss:

$$\mathcal{L}_{DA} = \frac{1}{N(N-1)} \sum_{i!=j} k\left(x_{\mathrm{s}}(i), x_{\mathrm{s}}(j)\right) - k\left(x_{\mathrm{s}}(i), x_{\mathrm{t}}(j)\right) - k\left(x_{\mathrm{t}}(i), x_{\mathrm{s}}(j)\right) + k\left(x_{\mathrm{t}}(i), x_{\mathrm{t}}(j)\right),$$

where k is a kernel between samples x_{s} from the source domain and samples x_{t} from the target domain, and N is the total number of samples across both domains. We follow [11, 39] and use a combination of multiple Gaussian Radial Basis Functions (RBF) kernels.

3 Data

We use the `deeplenstronomy` [25] package, which is built on `lenstronomy` [9], to generate galaxy-scale strong lens datasets with 40×40-pixel images in a source domain and in a target domain. For the source domain, we generate images with no pixel noise. For the target domain, we add noise based on DES[1] observing conditions, which include sky brightness and seeing, drawn from empirical distributions of these values [2]. Both domains contain $50,000$ lenses with Einstein radius varying from 0.5 arc seconds to 3.0 arc seconds drawn from a uniform distribution. We divide the dataset into training, validation, and testing with a ratio of 70%:10%:20%. Figure 1 provides example lensing systems from the source and target domain datasets.

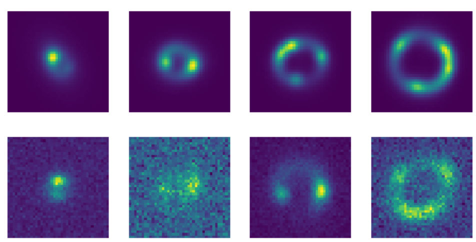

Figure 1: Examples of our source (top) and target (bottom) domain images. The Einstein radius increases from left to right.

4 Neural Network Architecture

We employ a baseline CNN with three convolution layers (each layer followed by ReLu activation, batch normalization, and max pooling) and two linear layers. Details of the CNN architecture are in Table 2 in the appendix. MSE loss is used for the main regression task, and DA is performed using a DANN or MMD loss. For both the DANN and MMD variations, latent space features are gathered after the final max pooling layer. For the DANN, we use a gradient reversal layer followed by the same two-linear layer architecture repeated for the domain classifier. A softmax activation is applied to the domain classifier's output and NLL is used as the domain classifier loss. We use the Adam optimizer for both the DANN and the MMD. Optimal hyperparameters were found through cross-validation. For all models, the batch size is fixed at 32. For training without DA we used a learning rate of 10^{-5}, while both DANN and MMD used 6×10^{-5}. For MMD, we used DA loss weight $\lambda_{DA} = 1.4$. The DA loss weight for DANN was set adaptively as proposed in [20]. It is initiated at 0 and gradually reaches 1 over the course of training, using the weighting function $\lambda_{DA} = \frac{2}{1+\exp(-10\cdot p)} - 1$, where p is fraction of training process completed, from 0 to 1. For all models, convergence was reached after 30 epochs.

	Source Domain	**Target Domain**
No DA	0.98 ± 0.0011	0.50 ± 0.23
DANN	0.97 ± 0.0052	$\mathbf{0.94 \pm 0.011}$
MMD	0.98 ± 0.013	$\mathbf{0.95 \pm 0.018}$

Table 1: R^2 score for training without DA in the top row, with DANN in the middle row, and with MMD in the bottom row. Each score is the mean of ten identical networks, initiated with different random seeds and trained under the same procedure. Standard deviation is also given across these models.

To evaluate the performance of the CNN, we use the R-squared (R^2) score [37]. It is defined as $R^2 = 1 - \frac{SS_{Regression}}{SS_{Total}}$, where $SS_{Regression} = \sum(\hat{y_i} - y_i)^2$ and $SS_{Total} = \sum(\bar{y} - y_i)^2$. Here, $\hat{y_i}$ is the model's prediction, y_i is the corresponding true value, and $\bar{y}$ is the mean of all true values. This metric shows how well the model can predict a dependent variable (Einstein radius). A larger R^2 indicates a better fit of the data, with a perfect fit of $R^2 = 1.0$.

5 Results

We report the R^2 scores in the source and target test sets in Table 1. Furthermore, in Figure 2 we visualise predictions against actual values, as well as fractional residuals, which are calculated by $Residual_{fract} = (\Theta_{E,Predicted} - \Theta_{E,True})/\Theta_{E,Predicted}$. The fractional residuals help visualize how each prediction is potentially biased.

Figure 2 and the R^2 scores in Table 1 show that both the DANN and MMD are able to significantly improve model performance on a more complex target domain when trained with a simpler source domain (up to 0.45 improvement in target R^2 scores, making both source and target scores close to 1) . The fractional residuals in Figure 2 support these results, with both DANN and MMD attaining target domain fractional residuals substantially smaller in magnitude than the model without DA. It should also be noted from the residuals that both DANN and MMD struggle the most with lower values of Einstein radius in the target domain. The model without DA and MMD similarly struggle on the low end of Einstein radius in the source domain, while DANN alone performs similarly across the whole source domain. With the inclusion of DA, DANN suffers a small trade-off in accuracy on the source domain. MMD may not lose accuracy on the source data, compared to regular training without DA, but has a higher degree of uncertainty due to its increased standard deviation.

6 Discussion

These results show that both of the tested DA-based methods have the potential to facilitate the analysis of real survey data in the future – i.e., reducing bias when using labeled simulated data for training and new unlabeled real data as the test set.

Currently, the existing number of real lens images are likely not enough to use as a target domain for DANN or MMD. However, as future large-scale surveys are released, many more real lens images will become available, with the help of automated identification using deep learning. Studies such as [17], [28], and [30] have demonstrated the viability of these methods. At that point, DANN and MMD may be used for further analysis.

It should be noted that if noise were to be the only difference between simulated and real data, it may be viable to just simulate noise in training data. However, other domain shifts are likely to exist between real and simulated data, which DANN and MMD would be employed to mitigate. To this end, a further investigation into domain shifts beyond DES-survey noise emulation is needed to ensure the gap from simulated to real data can be closed. Both DANN and MMD would likely struggle to overcome the domain shift caused by differences in data distributions of the parameters the network is trying to infer. For instance, if the source and target domain lenses are drawn from different distribution of Einstein radii, the network may have difficulty estimating this parameter correctly. This calls for investigation into other domain adaptation techniques, such as instance-based techniques [18, 19], which will be a topic of our future investigations.

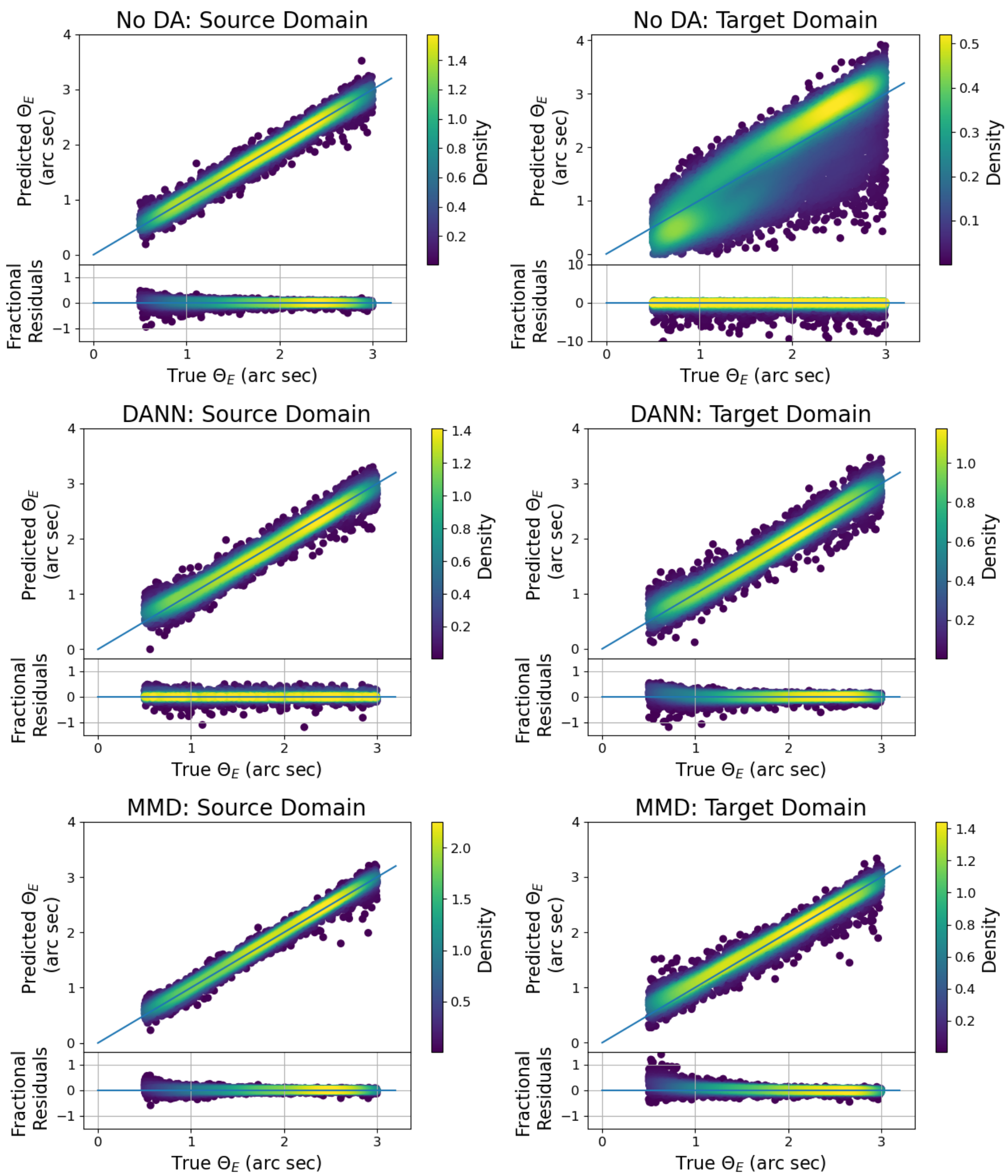

Figure 2: Predictions against actual values of test data sampled from the source domain (left panels) and target domain (right panels). Below each plot are the fractional residuals for visualization. Brighter color represents a higher density of samples, found using a kernel-density estimate with Gaussian kernels [35]. Results are given without DA (top row), with DANN (middle row), and with MMD (bottom row).

Acknowledgments and Disclosure of Funding

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics. The authors of this paper have committed themselves to performing this work in an equitable, inclusive, and just environment, and we hold ourselves accountable, believing that the best science is contingent on a good research environment.

We acknowledge the Deep Skies Lab as a community of multi-domain experts and collaborators who have facilitated an environment of open discussion, idea-generation, and collaboration. This community was important for the development of this project.

Furthermore, we also thank the anonymous referees whose comments helped improve this manuscript.

References

[1] T. M. C. Abbott, F. B. Abdalla, A. Alarcon, J. Aleksić, S. Allam, S. Allen, A. Amara, J. Annis, J. Asorey, S. Avila, and et. al. Dark energy survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing. *Phys. Rev. D*, 98:043526, Aug 2018.

[2] T. M. C. Abbott, F. B. Abdalla, S. Allam, A. Amara, J. Annis, J. Asorey, S. Avila, O. Ballester, M. Banerji, W. Barkhouse, and et al. The dark energy survey: Data release 1. *The Astrophysical Journal Supplement Series*, 239(2):18, nov 2018.

[3] Hiroaki Aihara, Nobuo Arimoto, Robert Armstrong, Sté phane Arnouts, Neta A Bahcall, Steven Bickerton, James Bosch, Kevin Bundy, Peter L Capak, James H H Chan, and et. al. The hyper suprime-cam SSP survey: Overview and survey design. *Publications of the Astronomical Society of Japan*, 70(SP1), sep 2017.

[4] Stephon Alexander, Sergei Gleyzer, Hanna Parul, Pranath Reddy, Marcos Tidball, and Michael W. Toomey. Domain Adaptation for Simulation-based Dark Matter Searches with Strong Gravitational Lensing. , 954(1):28, September 2023.

[5] M. W. Auger, T. Treu, A. S. Bolton, R. Gavazzi, L. V. E. Koopmans, P. J. Marshall, L. A. Moustakas, and S. Burles. The Sloan Lens ACS Survey. X. Stellar, Dynamical, and Total Mass Correlations of Massive Early-type Galaxies. , 724(1):511–525, November 2010.

[6] MW Auger, T Treu, AS Bolton, R Gavazzi, LVE Koopmans, PJ Marshall, LA Moustakas, and S Burles. The sloan lens acs survey. x. stellar, dynamical, and total mass correlations of massive early-type galaxies. *The Astrophysical Journal*, 724(1):511, 2010.

[7] Camille Avestruz, Nan Li, Hanjue Zhu, Matthew Lightman, Thomas E Collett, and Wentao Luo. Automated lensing learner: automated strong lensing identification with a computer vision technique. *The Astrophysical Journal*, 877(1):58, 2019.

[8] Matteo Barnabe, Oliver Czoske, Léon VE Koopmans, Tommaso Treu, and Adam S Bolton. Two-dimensional kinematics of slacs lenses–iii. mass structure and dynamics of early-type lens galaxies beyond z 0.1. *Monthly Notices of the Royal Astronomical Society*, 415(3):2215–2232, 2011.

[9] Simon Birrer and Adam Amara. lenstronomy: Multi-purpose gravitational lens modelling software package. *Physics of the Dark Universe*, 22:189–201, 2018.

[10] Simon Birrer, Suhail Dhawan, and Anowar J. Shajib. The hubble constant from strongly lensed supernovae with standardizable magnifications. *The Astrophysical Journal*, 924(1):2, jan 2022.

[11] A. Ćiprijanović, D. Kafkes, K. Downey, S. Jenkins, G. N. Perdue, S. Madireddy, T. Johnston, G. F. Snyder, and B. Nord. DeepMerge - II. Building robust deep learning algorithms for merging galaxy identification across domains. , 506(1):677–691, September 2021.

[12] A. Ćiprijanović, A. Lewis, K. Pedro, S. Madireddy, B. Nord, G. N. Perdue, and S. M. Wild. DeepAstroUDA: semi-supervised universal domain adaptation for cross-survey galaxy morphology classification and anomaly detection. *Machine Learning: Science and Technology*, 4(2):025013, June 2023.

[13] Aleksandra Ćiprijanović, Diana Kafkes, Gregory Snyder, F. Javier Sánchez, Gabriel Nathan Perdue, Kevin Pedro, Brian Nord, Sandeep Madireddy, and Stefan M. Wild. DeepAdversaries: examining the robustness of deep learning models for galaxy morphology classification. *Machine Learning: Science and Technology*, 3(3):035007, September 2022.

[14] Thomas E. Collett. THE POPULATION OF GALAXY–GALAXY STRONG LENSES IN FORTHCOMING OPTICAL IMAGING SURVEYS. *The Astrophysical Journal*, 811(1):20, sep 2015.

[15] Thomas E Collett and Matthew W Auger. Cosmological constraints from the double source plane lens sdssj0946+ 1006. *Monthly Notices of the Royal Astronomical Society*, 443(2):969–976, 2014.

[16] Gabriela Csurka. Domain Adaptation for Visual Applications: A Comprehensive Survey. *arXiv e-prints*, page arXiv:1702.05374, February 2017.

[17] Andrew Davies, Stephen Serjeant, and Jane M Bromley. Using convolutional neural networks to identify gravitational lenses in astronomical images. *Monthly Notices of the Royal Astronomical Society*, 487(4):5263–5271, May 2019.

[18] Antoine de Mathelin, Guillaume Richard, François Deheeger, Mathilde Mougeot, and Nicolas Vayatis. Adversarial weighting for domain adaptation in regression. In *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 49–56. IEEE, 2021.

[19] Abolfazl Farahani, Sahar Voghoei, Khaled Rasheed, and Hamid R Arabnia. A brief review of domain adaptation. *Advances in data science and information engineering: proceedings from ICDATA 2020 and IKE 2020*, pages 877–894, 2021.

[20] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1):2096–2030, 2016.

[21] Sankalp Gilda, Antoine de Mathelin, Sabine Bellstedt, and Guillaume Richard. Unsupervised Domain Adaptation for Constraining Star Formation Histories. *arXiv e-prints*, page arXiv:2112.14072, December 2021.

[22] Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012.

[23] Ž. Ivezić, S. M. Kahn, J. A. Tyson, B. Abel, E. Acosta, R. Allsman, D. Alonso, Y. AlSayyad, S. F. Anderson, J. Andrew, and et al. LSST: From Science Drivers to Reference Design and Anticipated Data Products. , 873:111, March 2019.

[24] Qiufan Lin, Dominique Fouchez, and Jerome Pasquet. Galaxy image translation with semi-supervised noise-reconstructed generative adversarial networks. In *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, January 2021.

[25] Robert Morgan, Brian Nord, Simon Birrer, Joshua Yao-Yu Lin, and Jason Poh. deeplenstronomy: A dataset simulation package for strong gravitational lensing. *arXiv preprint arXiv:2102.02830*, 2021.

[26] Teaghan O'Briain, Yuan-Sen Ting, Sébastien Fabbro, Kwang M. Yi, Kim Venn, and Spencer Bialek. Cycle-starnet: Bridging the gap between theory and data by leveraging large data sets. *The Astrophysical Journal*, 906(2):130, January 2021.

[27] Will Percival, Michael Balogh, Dick Bond, Jo Bovy, Raymond Carlberg, Scott Chapman, Patrick Cote, Nicolas Cowan, Sebastien Fabbro, Laura Ferrarese, and et al. The Euclid Mission. In *Canadian Long Range Plan for Astronomy and Astrophysics White Papers*, volume 2020, page 20, October 2019.

[28] C E Petrillo, C Tortora, S Chatterjee, G Vernardos, L V E Koopmans, G Verdoes Kleijn, N R Napolitano, G Covone, L S Kelvin, and A M Hopkins. Testing convolutional neural networks for finding strong gravitational lenses in kids. *Monthly Notices of the Royal Astronomical Society*, October 2018.

[29] S Rezaei, J P McKean, M Biehl, W de Roo, and A Lafontaine. A machine learning based approach to gravitational lens identification with the international LOFAR telescope. *Monthly Notices of the Royal Astronomical Society*, 517(1):1156–1170, jul 2022.

[30] C. Schaefer, M. Geiger, T. Kuntzer, and J.-P. Kneib. Deep convolutional neural networks as strong gravitational lens detectors. *Astronomy amp; Astrophysics*, 611:A2, March 2018.

[31] Manuel Schwonberg, Joshua Niemeijer, Jan-Aike Termöhlen, Jörg P. Schäfer, Nico M. Schmidt, Hanno Gottschalk, and Tim Fingscheidt. Survey on Unsupervised Domain Adaptation for Semantic Segmentation for Visual Perception in Automated Driving. *arXiv e-prints*, page arXiv:2304.11928, April 2023.

[32] M. Sereno. Probing the dark energy with strong lensing by clusters of galaxies. *Astronomy &amp Astrophysics*, 393(3):757–764, oct 2002.

[33] SH Suyu, PJ Marshall, MW Auger, S Hilbert, RD Blandford, LVE Koopmans, CD Fassnacht, and T Treu. Dissecting the gravitational lens b1608+ 656. ii. precision measurements of the hubble constant, spatial curvature, and the dark energy equation of state. *The Astrophysical Journal*, 711(1):201, 2010.

[34] Ricardo Vilalta, Kinjal Dhar Gupta, Dainis Boumber, and Mikhail M. Meskhi. A general approach to domain adaptation with applications in astronomy. *Publications of the Astronomical Society of the Pacific*, 131(1004):108008, sep 2019.

[35] Stanisław Węglarczyk. Kernel density estimation and its application. In *ITM web of conferences*, volume 23, page 00037. EDP Sciences, 2018.

[36] Joshua Wilde, Stephen Serjeant, Jane M Bromley, Hugh Dickinson, Lé on V E Koopmans, and R Benton Metcalf. Detecting gravitational lenses using machine learning: exploring interpretability and sensitivity to rare lensing configurations. *Monthly Notices of the Royal Astronomical Society*, 512(3):3464–3479, feb 2022.

[37] Sewall Wright. Correlation and causation. 1921.

[38] E Zaborowski, A Drlica-Wagner, F Ashmead, JF Wu, R Morgan, CR Bom, AJ Shajib, S Birrer, W Cerny, L Buckley-Geer, et al. Identification of galaxy-galaxy strong lens candidates in the decam local volume exploration survey using machine learning. *arXiv preprint arXiv:2210.10802*, 2022.

[39] Yinghua Zhang, Yu Zhang, Ying Wei, Kun Bai, Yangqiu Song, and Qiang Yang. Fisher deep domain adaptation, 2020.

Appendix

Layers	Properties	Output Shape	Parameters
input	1*40*40	(1, 40, 40)	0
Convolution (2D)	Filters: 8 Kernel: 3 3 Activation: ReLu	(8, 42, 42)	80
Batch Normalization	-	(8, 42, 42)	16
MaxPooling	Kernel: 2 2 Stride: 2	(8, 21, 21)	0
Convolution (2D)	Filters: 16 Kernel: 3 3 Activation: ReLu	(16, 21, 21)	1,168
Batch Normalization	-	(16, 21, 21)	32
MaxPooling	Kernel: 2 2 Stride: 2	(16, 10, 10)	0
Convolution (2D)	Filters: 32 Kernel: 3 3 Activation: ReLu	(32, 10, 10)	4,640
Batch Normalization	-	(32, 10, 10)	64
MaxPooling	Kernel: 2 2 Stride: 2	(32, 5, 5)	0
Flatten	-	(800)	-
Fully Connected	Activation: ReLu	(128)	102,528
Fully Connected	Activation: ReLu	(1)	129

Table 2: CNN architecture