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

Stephon Alexander and Michael W. Toomey
Department of Physics and Brown Theoretical Physics Center
Brown University
Providence, RI 02912

Sergei Gleyzer
Department of Physics & Astronomy
University of Alabama
Tuscaloosa, AL 35401

Pranath Reddy
University of Florida
Gainesville, FL 32611

Marcos Tidball
Universidade Federal do Rio Grande do Sul
Porto Alegre, Brazil

Abstract

The application of machine learning for quantifying dark matter substructure is growing in popularity. However, 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 (HST) observations that domain adaptation can significantly mitigate the losses in the model performance when applied to new domains.

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 non-gravitational signature of dark matter have come up empty. While this does not mean that dark matter can not 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 [1, 2, 3]. A particularly sensitive probe is strong gravitational lensing [4, 5, 6] for which we restrict ourselves in this article.

Strong gravitational lensing has already seen some promising success in extracting information about dark matter substructure. More recently, there has been a plethora of applications of machine

Machine Learning and the Physical Sciences workshop, NeurIPS 2022.

learning to this challenge, ranging from classification [7, 8, 9], regression [10], segmentation analysis [11, 12], and anomaly detection [13]. 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 [14, 15]. 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) [16]. A subset of transfer learning, domain adaptation is focused on the generalization of the model across different domains or data sets drawn from different underlying distributions. The goal of domain adaptation 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, domain adaptation can be realized in several ways, including supervised, semi-supervised, and unsupervised approaches [17, 18, 19].

In this work, we consider domain adaptation 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, HST and Euclid, of galaxy-galaxy strong gravitational lensing to carefully test the performance of domain adaptation 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 domain adaptation algorithms based on convolutional neural networks and equivariant neural networks that incorporate a known group symmetry to enhance performance.

2 Dark Matter Detection and Strong Gravitational Lensing

The Λ Cold Dark Matter (ΛCDM) model envisions near-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 [20]. Evidence for such merges has been observed in our Galaxy [21, 22, 23] and is a general prediction of N-body simulations where evidence of mergers should remain largely intact. Comparison between simulation and observation indicates good agreement with ΛCDM on large-scales [1, 2, 3]. However, discrepancies begin to arise on smaller, sub-galactic scales. These include the core-vs-cusp, too big to fail, missing satellite, and diversity problems.

Similar to [7, 13] we consider data sets of three substructure classes; no substructure, NFW subhalos of cold dark matter, and vortex substructure of superfluid type (axion) dark matter. We construct the simulations with `lenstronomy` [24] 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 [25] processed with a Gaussian mask and convolved with a Gaussian of size 2 pixels. 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 the signal-to-noise ratio (SNR) of the lensing arcs are consistent with real lensing data – $SNR \sim 20$ [26].

Domain adaptation requires at least two data sets, the source and the target. In this work we will demonstrate domain adaptation 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.

3 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 Convolutional Neural Network (CNN), specifically *EfficientNet* [27], as our base architecture. This is the same type of architecture that has achieved top performance in previous applications to lensing data sets [7, 13, 28]. More generally, CNNs are known to outperform other methods of classification for strong gravitational lenses [29], nonetheless, as noted by [30], a model trained on simulations can perform poorly on real data.

To improve the performance of models trained on simulated data, we use unsupervised domain adaptation, 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

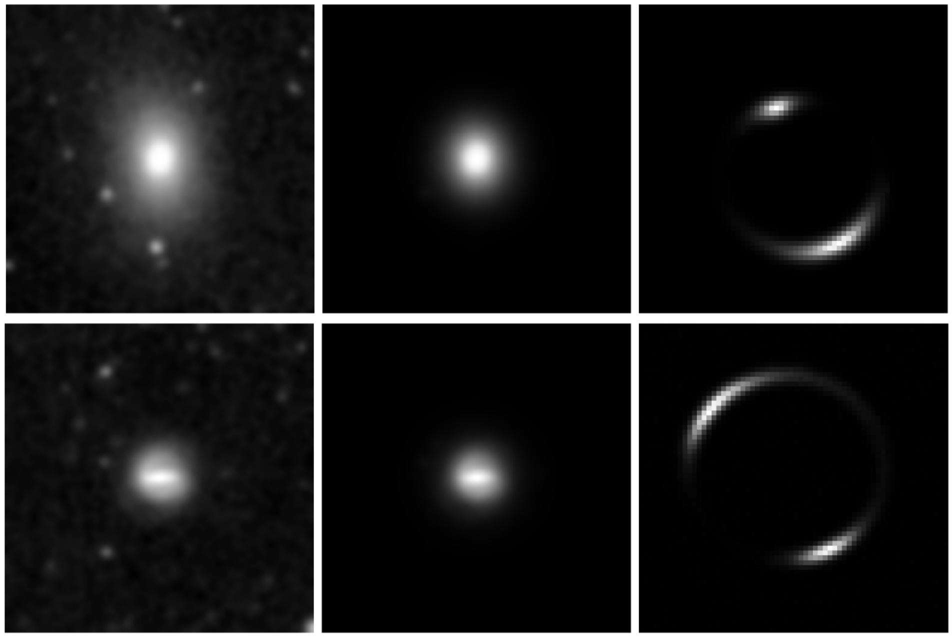

Figure 1: Example of lensing simulation of mock HST (top) and Euclid (bottom) data sets. Image of real galaxy from the Galaxy10 DECals data set (left), image processed with a Gaussian filter to remove unwanted background (center), and final lensed image (right).

unlabeled target data set, within the constraint that the objective remains the same [31]. Examples from each data set are shown in Figure. 1.

We utilize Adversarial Discriminative Domain Adaptation (ADDA) [32], 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 $\mathbf{X}_s$ and labels $\mathbf{Y}_s$ that come from a source distribution $p_s(x, y)$ and also target images $\mathbf{X}_t$ from a target distribution $p_t(x, y)$. Our objective is to learn a target encoder M_t and classifier C_t that classifies $\mathbf{X}_t$ into K classes.

In addition to the baseline CNN models, we also consider an Equivariant Neural Network (ENN) [33] for substructure classification. ENNs can be thought of a generalization of a CNN that encode 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.

The ENN we use consists of a group equivariant convolutional neural network [34] 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. Each block is composed of a convolutional layer, a batch normalization layer and a 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.

4 Results

We test the applicability of unsupervised domain adaptation using ADDA between mock HST and Euclid observations in the context of multi-class 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. 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

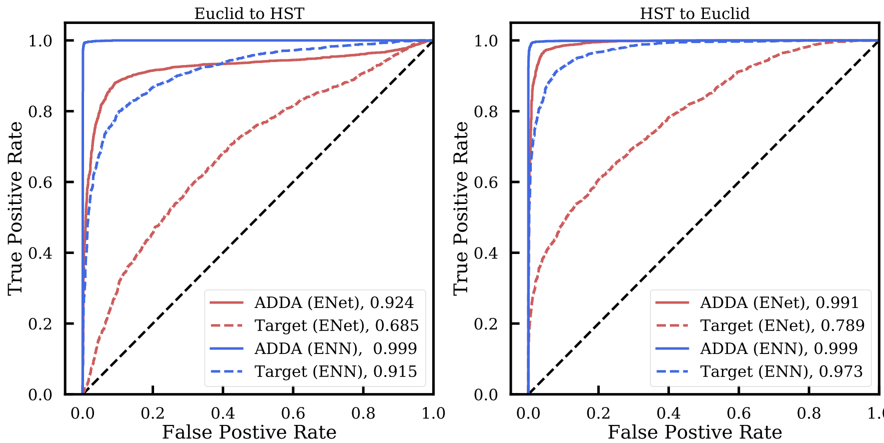

Figure 2: Application of ADDA for domain adaptation between two different mock simulations of strong lensing surveys: from Euclid to HST (left) and HST to Euclid (right). Dashed lines represent the naive application of classifier trained on source data set to target and solid represents result after training with ADDA. Red curves represent models based on *EfficientNet* and blue based on an ENN.

validation we use $7,500$ images for the source domain and $7,500$ images for the target; in both cases there are $2,500$ images per class. We use the *Adam* optimizer [35] 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. Learning rate, weight decay and other hyperparameters were optimized through a hyperparameter search. We have set the learning rate to 2×10^{-3} and weight decay to 1×10^{-5}. We utilize the area under the ROC curve (AUC) on the target validation set as the metric for classifier performance for all the models. All quoted AUC values are macro-averaged. All machine learning models were implemented using `PyTorch` [36] and are run on a single NVIDIA Tesla P100 GPU. The ROC curves for both combinations of source/target in addition to results for different architectures are presented in Figure 2.

We first train *EfficientNet* on the source data sets where it achieved a macro-averaged AUC $\approx$ 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* domain adaptation, results in an AUC of ≈ 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 unsupervised domain adaptation with ADDA, we observe a significant improvement in the application to the target data set. With the *EfficientNet* based algorithm we achieve a $\approx$ 35% improvement in AUC, at ≈ 0.924, and accuracy of $\approx$ 87% adapting from 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 which can be further appreciated by comparing the ROC curves in Figure. 2. The red dashed curves correspond to the naive application of the *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 (the left figure), DA from HST to Euclid results in near-perfect classification.

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 equivariant neural network. 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 principal 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 domain adaption) again results in degraded performance, realized with an AUC for the Euclid (HST) trained model applied to HST (Euclid) data of ≈ 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 ≈ 0.999 for both combinations of source/target and accuracy an 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 Discussion & Conclusion

With the upcoming arrival of strong gravitational lensing data from Euclid and the Vera 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 unsupervised domain adaptation 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. While we have restricted ourselves to substructure classification in this work, domain adaptation techniques can be additionally useful in the broader context of studying dark matter, from regression to image segmentation, in applications to real world strong gravitational lensing data sets.

Broader Impact

This work serves to augment the understanding and application of machine learning in cosmology - which is still very much in its initial stages. This work serves to increase the accessibility to those interested in applications of machine learning for strong lensing applications around the globe as our simulation data set and analysis pipeline is open sourced. Given the computational requirements of our implementation, those who have limited access to computing power may be at a disadvantage.

References

[1] Planck Collaboration. Planck 2015 results. XIII. Cosmological parameters. *A&A*, 594(A13):63, 2016. arXiv.

[2] L. Anderson, E. Aubourg et al. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples . *MNRAS*, 441(1):24–62, 2014. MNRAS.

[3] C. Heymans, L. van Waerbeke et al. CFHTLenS: the Canada–France–Hawaii Telescope Lensing Survey . *MNRAS*, 427(1):146–166, 2012. MNRAS.

[4] Matthew R. Buckley and Annika H. G. Peter. Gravitational probes of dark matter physics. *Phys. Rept.*, 761:1–60, 2018.

[5] Alex Drlica-Wagner et al. Probing the Fundamental Nature of Dark Matter with the Large Synoptic Survey Telescope. *arXiv e-prints*, page arXiv:1902.01055, February 2019.

[6] Joshua Simon et al. Testing the Nature of Dark Matter with Extremely Large Telescopes. *Bull. Am. Astron. Soc*, 51(3):153, May 2019.

[7] Stephon Alexander, Sergei Gleyzer, Evan McDonough, Michael W. Toomey, and Emanuele Usai. Deep Learning the Morphology of Dark Matter Substructure. *Astrophys. J.*, 893:15, 2020.

[8] Ana Diaz Rivero and Cora Dvorkin. Direct Detection of Dark Matter Substructure in Strong Lens Images with Convolutional Neural Networks. *Phys. Rev. D*, 101(2):023515, 2020.

[9] Sreedevi Varma, Malcolm Fairbairn, and Julio Figueroa. Dark Matter Subhalos, Strong Lensing and Machine Learning. *arXiv e-prints*, page arXiv:2005.05353, May 2020.

[10] Johann Brehmer, Siddharth Mishra-Sharma, Joeri Hermans, Gilles Louppe, and Kyle Cranmer. Mining for Dark Matter Substructure: Inferring subhalo population properties from strong lenses with machine learning. *Astrophys. J.*, 886(1):49, Nov 2019.

[11] Bryan Ostdiek, Ana Diaz Rivero, and Cora Dvorkin. Extracting the Subhalo Mass Function from Strong Lens Images with Image Segmentation. 9 2020.

[12] Bryan Ostdiek, Ana Diaz Rivero, and Cora Dvorkin. Detecting Subhalos in Strong Gravitational Lens Images with Image Segmentation. 9 2020.

[13] Stephon Alexander, Sergei Gleyzer, Hanna Parul, Pranath Reddy, Michael W. Toomey, Emanuele Usai, and Ryker Von Klar. Decoding Dark Matter Substructure without Supervision. *arXiv:2008.12731 [astro-ph, physics:hep-ph]*, August 2020. arXiv: 2008.12731.

[14] Aprajita Verma, Thomas Collett, Graham P. Smith, Strong Lensing Science Collaboration, and the DESC Strong Lensing Science Working Group. Strong Lensing considerations for the LSST observing strategy. *arXiv e-prints*, page arXiv:1902.05141, February 2019.

[15] Masamune Oguri and Philip J. Marshall. Gravitationally lensed quasars and supernovae in future wide-field optical imaging surveys. *MNRAS*, 405(4):2579–2593, July 2010.

[16] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Vaughan. A theory of learning from different domains. *Machine Learning*, 79:151–175, 2010.

[17] Saeid Motiian, Marco Piccirilli, Donald A. Adjeroh, and Gianfranco Doretto. Unified Deep Supervised Domain Adaptation and Generalization. *arXiv e-prints*, page arXiv:1709.10190, September 2017.

[18] Jeff Donahue, Judy Hoffman, Erik Rodner, Kate Saenko, and Trevor Darrell. Semi-supervised domain adaptation with instance constraints. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 668–675, 2013.

[19] Abolfazl Farahani, Sahar Voghoei, Khaled Rasheed, and Hamid R. Arabnia. A Brief Review of Domain Adaptation. *arXiv e-prints*, page arXiv:2010.03978, October 2020.

[20] G Kauffmann, Simon D. M. White, and B. Guiderdoni. The Formation and Evolution of Galaxies Within Merging Dark Matter Haloes. *MNRAS*, 264:201, 1993.

[21] Anirudh Chiti, Anna Frebel, Joshua D. Simon, Denis Erkal, Laura J. Chang, Lina Necib, Alexander P. Ji, Helmut Jerjen, Dongwon Kim, and John E. Norris. An extended halo around an ancient dwarf galaxy. *Nature Astronomy*, 5:392–400, January 2021.

[22] Lina Necib, Bryan Ostdiek, Mariangela Lisanti, Timothy Cohen, Marat Freytsis, and Shea Garrison-Kimmel. Chasing Accreted Structures within Gaia DR2 using Deep Learning. *Astrophys. J.*, 903(1):25, 2020.

[23] Lina Necib, Bryan Ostdiek, Mariangela Lisanti, Timothy Cohen, Marat Freytsis, Shea Garrison-Kimmel, Philip F. Hopkins, Andrew Wetzel, and Robyn Sanderson. Evidence for a Vast Prograde Stellar Stream in the Solar Vicinity. *Nature Astron.*, 4(11):1078–1083, 2020.

[24] Simon Birrer and Adam Amara. Lenstronomy: multi-purpose gravitational lens modelling software package. 3 2018.

[25] Arjun Dey, David J. Schlegel, Dustin Lang, Robert Blum, Kaylan Burleigh, Xiaohui Fan, Joseph R. Findlay, Doug Finkbeiner, David Herrera, Stéphanie Juneau, Martin Landriau, Michael Levi, Ian McGreer, Aaron Meisner, Adam D. Myers, John Moustakas, Peter Nugent, Anna Patej, Edward F. Schlafly, Alistair R. Walker, Francisco Valdes, Benjamin A. Weaver, Christophe Yèche, Hu Zou, Xu Zhou, Behzad Abareshi, T. M. C. Abbott, Bela Abolfathi, C. Aguilera, Shadab Alam, Lori Allen, A. Alvarez, James Annis, Behzad Ansarinejad, Marie Aubert, Jacqueline Beechert, Eric F. Bell, Segev Y. BenZvi, Florian Beutler, Richard M. Bielby, Adam S. Bolton, César Briceño, Elizabeth J. Buckley-Geer, Karen Butler, Annalisa Calamida, Raymond G. Carlberg, Paul Carter, Ricard Casas, Francisco J. Castander, Yumi Choi, Johan Comparat, Elena Cukanovaite, Timothée Delubac, Kaitlin DeVries, Sharmila Dey, Govinda Dhungana, Mark Dickinson, Zhejie Ding, John B. Donaldson, Yutong Duan, Christopher J. Duckworth, Sarah Eftekharzadeh, Daniel J. Eisenstein, Thomas Etourneau, Parker A. Fagrelius, Jay Farihi, Mike Fitzpatrick, Andreu Font-Ribera, Leah Fulmer, Boris T. Gänsicke, Enrique Gaztanaga, Koshy George, David W. Gerdes, Satya Gontcho A. Gontcho, Claudio Gorgoni, Gregory Green, Julien Guy, Diane Harmer, M. Hernandez, Klaus Honscheid, Lijuan Wendy Huang, David J. James, Buell T. Jannuzi, Linhua Jiang, Richard Joyce, Armin Karcher, Sonia Karkar, Robert Kehoe, Jean-Paul Kneib, Andrea Kueter-Young, Ting-Wen Lan, Tod R. Lauer, Laurent Le Guillou, Auguste Le Van Suu, Jae Hyeon Lee, Michael Lesser, Laurence

Perreault Levasseur, Ting S. Li, Justin L. Mann, Robert Marshall, C. E. Martínez-Vázquez, Paul Martini, Hélion du Mas des Bourboux, Sean McManus, Tobias Gabriel Meier, Brice Ménard, Nigel Metcalfe, Andrea Muñoz-Gutiérrez, Joan Najita, Kevin Napier, Gautham Narayan, Jeffrey A. Newman, Jundan Nie, Brian Nord, Dara J. Norman, Knut A. G. Olsen, Anthony Paat, Nathalie Palanque-Delabrouille, Xiyan Peng, Claire L. Poppett, Megan R. Poremba, Abhishek Prakash, David Rabinowitz, Anand Raichoor, Mehdi Rezaie, A. N. Robertson, Natalie A. Roe, Ashley J. Ross, Nicholas P. Ross, Gregory Rudnick, Sasha Safonova, Abhijit Saha, F. Javier Sánchez, Elodie Savary, Heidi Schweiker, Adam Scott, Hee-Jong Seo, Huanyuan Shan, David R. Silva, Zachary Slepian, Christian Soto, David Sprayberry, Ryan Staten, Coley M. Stillman, Robert J. Stupak, David L. Summers, Suk Sien Tie, H. Tirado, Mariana Vargas-Magaña, A. Katherina Vivas, Risa H. Wechsler, Doug Williams, Jinyi Yang, Qian Yang, Tolga Yapici, Dennis Zaritsky, A. Zenteno, Kai Zhang, Tianmeng Zhang, Rongpu Zhou, and Zhimin Zhou. Overview of the DESI Legacy Imaging Surveys. *The Astronomical Journal*, 157(5):168, May 2019.

[26] Adam S. Bolton, Scott Burles, Léon V. E. Koopmans, Tommaso Treu, Raphaël Gavazzi, Leonidas A. Moustakas, Randall Wayth, and David J. Schlegel. The Sloan Lens ACS Survey. V. The Full ACS Strong-Lens Sample. *ApJ*, 682(2):964–984, August 2008.

[27] Mingxing Tan and Quoc V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv e-prints*, page arXiv:1905.11946, May 2019.

[28] Kyriakos Vattis, Michael W. Toomey, and Savvas M. Koushiappas. Deep learning the astrometric signature of dark matter substructure. *arXiv e-prints*, page arXiv:2008.11577, August 2020.

[29] R. Benton Metcalf, M. Meneghetti, Camille Avestruz, Fabio Bellagamba, Clécio R. Bom, Emmanuel Bertin, Rémi Cabanac, F. Courbin, Andrew Davies, Etienne Decencière, Rémi Flamary, Raphael Gavazzi, Mario Geiger, Philippa Hartley, Marc Huertas-Company, Neal Jackson, Eric Jullo, Jean-Paul Kneib, Léon V. E. Koopmans, François Lanusse, Chun-Liang Li, Quanbin Ma, Martin Makler, Nan Li, Matthew Lightman, Carlo Enrico Petrillo, Stephen Serjeant, Christoph Schäfer, Alessandro Sonnenfeld, Amit Tagore, Crescenzo Tortora, Diego Tuccillo, Manuel B. Valentín, Santiago Velasco-Forero, Gijs A. Verdoes Kleijn, and Georgios Vernardos. The Strong Gravitational Lens Finding Challenge. *A&A*, 625:A119, May 2019. arXiv: 1802.03609.

[30] 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. *Monthly Notices of the Royal Astronomical Society*, 506(1):677–691, July 2021. arXiv: 2103.01373.

[31] Geoffrey French, Michal Mackiewicz, and Mark Fisher. Self-ensembling for visual domain adaptation. *arXiv:1706.05208 [cs]*, September 2018. arXiv: 1706.05208.

[32] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial Discriminative Domain Adaptation. *arXiv:1702.05464 [cs]*, February 2017. arXiv: 1702.05464.

[33] Maurice Weiler and Gabriele Cesa. General $E(2)$-Equivariant Steerable CNNs. *arXiv:1911.08251 [cs, eess]*, April 2021. arXiv: 1911.08251.

[34] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999. PMLR, 2016.

[35] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*, January 2017. arXiv: 1412.6980.

[36] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv:1912.01703 [cs, stat]*, December 2019. arXiv: 1912.01703.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [Yes]

(d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]

2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Section 4.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See Section 4.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]