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

Stephon Alexander, 1 Sergei Gleyzer, 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
2 Department of Physics & Astronomy, University of Alabama, Tuscaloosa, AL 35401, USA
3 Birla Institute of Technology & Science, Pilani - Hyderabad Campus, Telangana, India
4 Departamento de Física, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

(Dated: December 23, 2021)

Clues to the identity of dark matter have remained surprisingly elusive, given the scope of experimental programs aimed at its identification. While terrestrial experiments may be able to nail down a model, an alternative, and equally promising, 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 on galaxy-galaxy strong lensing images. Machine learning applications have been explored in detail for extracting just this signal. With limited availability of high quality strong lensing data, these approaches have exclusively relied on simulations. Naturally, 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. This is where 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 of varying complexity, that domain adaptation can significantly mitigate the losses in the model performance. This technique can help domain experts build and apply better machine learning models for extracting useful information from strong gravitational lensing data expected from the upcoming surveys.

I. 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 some unseen matter that compensated for this discrepancy, or so-called dark matter. Presently, all efforts aimed at extracting a non-gravitational signature of dark matter have come up empty 1 [21]. While this does not mean that dark matter can not communicate with Standard Model (SM) particles, as it may be the case that its SM couplings are 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 21 [23]. However, 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 24 [29] and astrometric observations 30 [34]. A particularly sensitive probe will be strong gravitational lensing 37 [39] for which we restrict ourselves in this article.

Strong lensing has already seen some promising success in extracting information about dark matter substructure, from lensed quasars 40 [42], observations with ALMA 43, and extended lensing images 44 [46]. Various works have considered the expected signatures and methods to extract information about the underlying distribution of dark matter 47 [50].

More recently, there has been a plethora of applications of machine learning to this challenge, ranging from classification 51 [53], regression 54, segmentation analysis 55 [56], and anomaly detection 57. 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 58 [59]. However, not unexpectedly, naively applying a model trained on simulations to real data is not likely to be very successful, as the data idiosyncrasies relative to simulations will significantly diminish the accuracy of the model. A promising method to bridge the gap between a model trained on simulations and real world data is based on the technique of domain adaptation (DA) 60.

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. Concretely, 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 fashions 61 [63].

Domain adaptation techniques have been used in a wide variety of applications related to computer vision, such as adapting a model trained on synthetic images to real images 64, simple to complex images 65, and virtual worlds with controlled data to the real world 66.
Recently, \cite{67} used unsupervised domain adaptation (UDA) to classify merging galaxies. Doing so achieved promising results, with an increase of up to 19\% in accuracy when compared to a model trained only on simulations. This work also showed that models trained without domain adaptation, achieved a poor accuracy on real data. In another work \cite{55} used domain adaptation to generalize an image segmentation algorithm to different gravitational lensing systems for subhalo detection.

In this work, we consider several domain adaptation algorithms to establish a proof-of-concept application for dark matter searches in strong gravitational lensing. Given the present lack of sufficient real data, we use two datasets with differing levels of complexity of strong lensing simulations to carefully test the performance of domain adaptation prior to eventual applications to real data. We evaluate the application of 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 several domain adaptation algorithms, to find the best model. We additionally investigate invariant neural network models that incorporate a known group symmetry, to further enhance the performance of domain adaptation.

We begin with a brief review of dark matter substructure and also discuss several lensing signatures in Section \ref{A}. We then focus on the details of strong lensing simulations in Section \ref{B} followed by a summary of domain adaptation algorithms in Section \ref{C}. We present our main results in Section \ref{D}, followed by the discussion and outlook in Section \ref{E}.

II. DARK MATTER DETECTION AND STRONG GRAVITATIONAL LENSING

A. Dark Matter Substructure

The $\Lambda$ Cold Dark Matter ($\Lambda$CDM) model envisions near scale invariant density fluctuations, present in the early universe, serving as seeds of large-scale structure via hierarchical structure formation. Concretely, structures such as dark matter halos are formed from the coalescence of smaller halos \cite{65}. Evidence for such mergers has been observed in our Galaxy \cite{69,71} and is a general prediction of N-body simulations where evidence of mergers should remain largely in tact. The distribution of subhalos masses is expected to follow a power law distribution,

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

where $\beta \sim -1.9$ has been found from simulations \cite{72,73}.

Comparison between simulation and observation indicates good agreement with $\Lambda$CDM on large-scales \cite{21,22}. However, discrepancies begin to arise on smaller, sub-galactic scales. These include the core-vs-cusp \cite{74}, too big to fail \cite{75}, missing satellite \cite{77,79} and diversity problems \cite{75}. While it may be the case that these problems can be addressed with a better understanding of the astrophysics, e.g. baryonic feedback \cite{81}, it is imperative that we consider the manifestations of other theories beyond $\Lambda$CDM.

Two natural directions to consider are a modification to the general theory of relativity, such as BF coupled \cite{82,84} or Chern-Simons gravity \cite{85,86}, or alternatives to cold non-interacting 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 (BEC) \cite{87,88} or Bardeen-Cooper-Schreiffer (BCS) \cite{89,90} 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 \cite{97,98}. Shortly thereafter, it was recognized that they were a promising dark matter candidate \cite{100,101}. Very light axions are particularly well suited to address some of the issues with structure on sub-galactic scales. Axions with masses $\sim 1 \times 10^{-23}$ eV have a de Broglie wavelength on kpc scales, which realizes a natural solution to the core-vs-cusp problem. Additionally, light axions can form subhalos but can also form substructures quite different from standard CDM. These include vortices, disks, and interference patterns \cite{103,104}.

While we will not consider their impact in this work, it is important to also consider line of sight halos, i.e. interlopers \cite{108,109}. 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 \cite{110}).

B. Strong Lensing Theory

A powerful probe of dark matter substructure is strong gravitational lensing; an effect which is most pronounced near extended lensing arcs. Any over or under densities along the line of sight for an observer yields the lens equation, realized as an integral over an induced gravitational potential \cite{111},

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

where $D_{LS}, D_L, D_S$ are the angular diameter distances from the lens to the source, the lens to the observer, and the source to observer. The last term on the r.h.s. of Equation \ref{eq:lens} is known as the deflection angle, $\bar{\alpha}$, and is related via a perpendicular derivative to the the lensing

\footnote{See \cite{80} for a differing perspective.}

\[\int dz \overline{\Psi(r)}\]
potential, $\Psi_{\perp} = \alpha$. The lensing potential can be shown to be related to matter density via a Poisson equation for gravitational lensing,

$$\nabla^2 \Psi = 2\kappa,$$  

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

$$\vec{\alpha} = \sum_{\text{all matter}} \vec{\alpha}_i,$$  

where $\vec{\alpha}_i$ could represent a DM halo, subhalos, vortices, external shear, interlopers etc.

**III. STRONG LENSING SIMULATIONS**

Similar to [51, 57] we consider data sets of three substructure classes; no substructure, subhalos of CDM, and vortex substructure of superfluid type dark matter. The parameters for the simulations are compiled in Table I. The strong lensing images are simulated with PyAutoLens [114, 115]. The data are sized $150 \times 150$ pixels with a scale $0.5''$/pixel. The light from lensed background galaxies is modeled as a simple Sersic profile and the signal-to-noise ratio (SNR) of the lensing arcs are consistent with real lensing data – $\text{SNR} \sim 20$ [116] – by appropriate inclusion of backgrounds and noise. We further consider the modifications induced by a point-spread function (PSF), modeled as an Airy disk with first zero-crossing $\sigma_{\text{psf}} \lesssim 1''$.

We ensure that the total fraction of mass in substructure, $f_{\text{sub}}$, is of $O(1\%)$. 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 is done to ensure that classification algorithms don’t simply recognize simulations without substructure as less massive on average. When simulating substructure for CDM we draw subhalo masses from 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 [117]. 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 the total lensing is just the sum of the individual contributions,

$$\alpha = \alpha_{\text{LSS}} + \alpha_{\text{halo}} + \alpha_{\text{halo-sub}},$$  

where $\alpha_{\text{LSS}}$ is the external shear from large-scale-structure, $\alpha_{\text{halo}}$ the lensing from the halo and $\alpha_{\text{halo-sub}}$ for halo substructure. It then follows that the location of an image can be found from a modified form of the lens equation, Equation 2,

$$\beta^i = \theta^i - \alpha_{\text{LSS}}^i - \alpha_{\text{halo}}^i - \alpha_{\text{halo-sub}}^i.$$

Domain adaptation requires at least two data sets, the source and the target. In this work we consider two distinct simulations as source and target data sets. The biggest difference between source and target data sets is that source images are held at fixed redshift, while the target data set allows the redshifts to float over a range of values for both the lensed and lensing galaxy. Additionally, the SNR in the target data set varies over a larger range of values, $10 \lesssim \text{SNR} \lesssim 30$, consistent with the challenges associated with variable source distances. Thus, the target data set constitutes a more challenging sample for identifying dark matter substructure - this difference is easy to see by eye in Figure 4 for example lenses from both classes. Thus, it will be our goal to successfully
TABLE I. Parameters with distributions and priors used in the simulation of strong lensing images. Parameters with subscript $s$ & $t$ correspond to parameters for source and target data sets respectively. Note that only a single type of substructure was used per image.

| Substructure | Param. | Dist. | Priors | Details |
|--------------|--------|-------|--------|---------|
| DM Halo | $\theta_x$ | fixed | 0 | x position |
|          | $\theta_y$ | fixed | 0 | y position |
|          | $z_s$ | fixed | 0.5 | redshift |
|          | $z_t$ | uniform | $[0.4, 0.6]$ | redshift |
|          | $M_{\text{TOT}}^s$ | fixed | 1e12 | total halo mass in $M_\odot$ |
| External Shear | $\gamma_{\text{ext}}$ | uniform | $[0, 0.3]$ | magnitude |
|          | $\phi_{\text{ext}}$ | uniform | $[0, 2\pi]$ | angle |
| Lensed Gal. | $\phi_k$ | uniform | $[0, 2\pi]$ | orientation from y axis |
|          | $\phi_s$ | fixed | 1.0 | redshift |
|          | $z_t$ | uniform | $[0.8, 1.2]$ | redshift |
|          | $\phi$ | uniform | $[0, 2\pi]$ | orientation to y axis |
|          | $n$ | fixed | 1.5 | Sersic index |
|          | $R$ | uniform | $[0.25, 1]$ | effective radius |
| Vortex | $\theta_2$ | normal | $[0, 0.5]$ | x position |
|          | $\theta_3$ | normal | $[0, 0.5]$ | y position |
|          | $l$ | uniform | $[0.5, 2.0]$ | length of vortex |
|          | $\phi_{\text{vort}}$ | uniform | $[0, 2\pi]$ | orientation from y axis |
|          | $m_{\text{vort}}$ | uniform | $[0.5, 5.5]$ | % of mass in vortex |
| Subhalo | $r$ | uniform | $[0, 2.0]$ | radial distance from center |
|          | $N$ | Poisson | $\mu=25$ | number of subhalos |
|          | $\phi_{\text{sh}}$ | uniform | $[0, 2\pi]$ | orientation from y axis |
|          | $m_{\text{sh}}$ | power law | $[1e6, 1e10]$ | subhalo mass in $M_\odot$ |
|          | $\beta_{\text{sh}}$ | fixed | -1.9 | power law index |

adapt and evaluate the algorithms trained on the source data set to the target data set.

IV. DOMAIN ADAPTATION

Our goal is to train a supervised model on a source data set and adapt it to a target data set. For this task we use ResNet-18 [113], a Convolutional Neural Network (CNN), as our base architecture. This is the same architecture that has achieved top performance in previous applications to lensing data sets [34, 51, 57]. More generally, CNNs are known to outperform other methods of classification for strong gravitational lenses [119], nonetheless, as noted by [67], 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 unlabeled target data set, within the constraint that the objective remains the same [120]. Ex-Nov.

FIG. 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.

A. Unsupervised Domain Adaptation Techniques

We first consider Adversarial Discriminative Domain Adaptation (ADDA) [65], 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 $X_s$ and labels $Y_s$ that come from a source distribution $p_s(x,y)$ and also target images $X_t$ from a target distribution $p_t(x,y)$. Our objective is to learn a target encoder $M_t$ and classifier $C_t$ that classifies $X_t$ 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 $M_s(X_s)$ and $M_t(X_t)$ is minimized, as illustrated in Fig. 2. Since we’re 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 if the encoded vector represents an image from the source domain or from the target domain using a standard supervised loss, where the labels indicate the origin domain. Finally, we train the $M_t$ using $D$. To evaluate a target image $X_t$ we perform $C(M_t(X_t))$.

We additionally evaluate two DA algorithms derived from semi-supervised learning. The first makes use of self-ensembling (Self-Ensemble) [120] and is based on the
mean teacher semi-supervised model [121]. There are two networks in this method: a student network that is trained using gradient descent and a teacher network whose weights are an exponential moving average of the student’s. During training, the labeled source inputs are passed through the student network and the cross-entropy loss is taken. However, the unlabeled target inputs pass through both the student and teacher networks and the self-ensembling loss is used. It is computed as the mean-squared difference between the predictions created by the student and the teacher networks with different augmentations, dropout and noise parameters, and penalizes the difference in class prediction between the student and the teacher. This method also makes use of confidence thresholding and a class balancing loss term. The second semi-supervised method, AdaMatch [122], uses weak and strong augmentations on both the source and target input images. During training the method also uses random logit interpolation, distribution alignment and relative confidence thresholding to achieve a better performance.

In additional to the baseline CNN models, we also consider an Equivariant Neural Network (ENN) [123] 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 in the problem. 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 [124] 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 can be visualized in Fig. 3. 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. A schematic of the ENN architecture is presented in Fig. 4.

### B. 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 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 used the Adam optimizer [125] to minimize our losses. We trained both ResNet-18 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 we stop training. For final results, we considered the epoch that achieved the largest accuracy. Learning rate, weight decay and other hyperparameters were optimized through a hyperparameters search, and are available in Table II.

We used random horizontal and random flips augmentations.

| Method         | Hyperparameters                  | Hyperparameter values | Augmentations                  |
|----------------|----------------------------------|-----------------------|--------------------------------|
| Supervised     | Learning rate                    | $1 \times 10^{-3}$    | Horizontal flips              |
|                | Weight decay                     | $5 \times 10^{-5}$    | Vertical flips                |
|                | Cyclic scheduler                 | True                  |                                |
| ADDA           | Learning rate target encoder     | $1 \times 10^{-5}$    | Horizontal flips              |
|                | Learning rate target encoder     | $1 \times 10^{-4}$    | Vertical flips                |
|                | Weight decay                     | $5 \times 10^{-5}$    |                                |
|                | Cyclic scheduler                 | False                 |                                |
| AdaMatch       | Learning rate                    | $1 \times 10^{-3}$    | Horizontal flips              |
|                | Tau                              | 0.9                   | Vertical flips                |
|                | Weight decay                     | $5 \times 10^{-5}$    | Random autocontrast           |
|                | Cyclic scheduler                 | False                 | Gaussian blur                 |
|                |                                  |                       | Random invert                 |
|                |                                  |                       | Random adjust sharpness        |
|                |                                  |                       | Random solarize                |
|                |                                  |                       | Random affine transformations   |
| Self-Ensemble  | Learning rate                    | $1 \times 10^{-3}$    | Horizontal flips              |
|                | Weight of unsupervised loss      | 3.0                   | Vertical flips                |
|                | Weight decay                     | $5 \times 10^{-5}$    | Gaussian blur                 |
|                | Cyclic scheduler                 | False                 |                                |
FIG. 3. Visualization of the $D_2$ group structure for a real gravitational lens. Lensing image credit: ESA/Hubble & NASA.

FIG. 4. Schematic of the equivariant (ENN) architecture.

V. RESULTS

We compare three different UDA techniques in the context of multi-class classification of three types of substructure: no substructure, CDM subhalos, and superfluid DM vortices. We employ two different base classifiers for each technique - ResNet-18 and an ENN. The parameters for our source and target data sets are shown in Table I. Better results on all methods were achieved when starting UDA training with models that were pre-trained in a supervised fashion using source data, as discussed in [67]. As such, our models were trained starting from models pre-trained on the source simulation data. Results from our analysis are shown in Tables III and IV for ResNet-18 and the ENN respectively. ROC curves for both architectures are presented in Figure 5.

A. Domain Adaptation

We first train ResNet-18 on the source data set where it achieved a macro-averaged AUC $\approx 0.996$ and an accuracy of $\approx 97\%$. Applying this model to the target data set naively, i.e. without domain adaptation, results in an AUC of $\approx 0.880$ and accuracy of $\approx 59\%$, a significantly degraded performance, as anticipated.

Following the application of UDA techniques, we observe a significant improvement when applied to the target data set. AdaMatch and Self-Ensemble achieve an AUC of $\approx 0.919$ and $0.917$, respectively, which shows good improvement over the results obtained without domain adaptation. The top performing ResNet-18 based UDA model is ADDA. With a $\approx 9\%$ improvement in AUC, at $\approx 0.955$, and accuracy of $\approx 86\%$, ADDA significantly improves the performance of our model on the target data set.

Considering the performance for individual classes, we find that ADDA achieves consistent AUC scores of $\approx 0.95$ for both the source and target dataset. We also found that the best results were obtained after random zooming (in a range of $[0.8, 1.2]$) and random rotations (in a range of $[0, 90]$ degrees) on the source dataset. Two of the UDA algorithms, Self-Ensemble and AdaMatch, are also highly dependent on augmentations, and different augmentations were tested to find the optimal configurations. 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 unless stated otherwise. All machine learning models were implemented using PyTorch [127] and are run on a single NVIDIA Tesla P100 GPU.

2 The code used in our analysis can be found [here](#).
for all three classes. This is in contrast to the performance on naive application, where no substructure and subhalos classification each had an AUC of $\approx 0.90$ and classification of vortices was severely degraded to an AUC of $\approx 0.80$. The baseline model applied to the source data set, on the other hand, had consistent performance on a class by class basis at $\approx 0.99$ AUC.

**FIG. 5.** ROC-AUC curves for ResNet-18 (left) and ENN (right) based algorithms applied to our data set.

B. Equivariant Domain Adaptation

With an AUC of $\approx 0.996$ and accuracy of $\approx 91\%$, the ENN performance is commensurate with ResNet-18 on the source data set. The naive application to the test data set (i.e. no domain adaption) again results in degraded performance, realized with an AUC $\approx 0.856$ and accuracy of $\approx 68\%$.

The Self-Ensemble and AdaMatch equivariant models again show improved performance relative to the naive case and with AUC scores of $\approx 0.939$ and $\approx 0.960$ respectively show a notable performance boost relative to the same UDA algorithms with ResNet-18. This is easily visualized by comparing the increased separation between the target ROC curve (blue) with the three UDA algorithms in Figure 5 between ResNet-18 (left) and the ENN (right). Finally, ADDA again achieves impressive top performance with an AUC of $\approx 0.980$. Thus, the equivariant model with ADDA is the top performing algorithm relative to the baseline source model.

**TABLE IV.** Accuracy and macro-averaged area under the ROC curve (AUC) for the ResNet-18 classifier.

| Method     | Accuracy (%) | AUC     |
|------------|--------------|---------|
| ADDA       | 85.84        | 0.955   |
| AdaMatch   | 75.55        | 0.919   |
| Self-Ensemble | 76.71    | 0.917   |
| Supervised (target) | 59.19   | 0.880   |
| Supervised (source)   | 96.84       | 0.996   |

Let us again investigate the performance of the ADDA augmented equivariant model on a class by class basis. Performance of the original classifier was near perfect, achieving AUC scores of $\approx 0.99$ for all three classes (e.g. subhalos vs. vortices and no substructure, etc.) on the source data set. Applied naively to the target data set it achieved a respectable performance of $\approx 0.92$ for subhalos but only managed AUC scores of $\approx 0.83$ and $0.82$ for no substructure and vortices respectively. Note that this is actually worse than the comparable ResNet-18 application. ADDA dramatically increases the algorithm’s performance, managing consistent classification across all classes with AUC scores of $\approx 0.98$. Considering the naive performance was worse than ResNet-18, it is impressive that our equivariant model ENN was augmented more effectively with ADDA, a feature that is shared with the other two UDA methods.
Now let us consider the performance of the ENN relative to ResNet-18. We can see from Figure 5 that there was a notable increase in the ability of AdaMatch and Self-Ensemble to improve performance for the ENN. This is likely related to the fact that both AdaMatch and Self-Ensemble are highly dependent on augmentations. Thus, the additional symmetries manifest in the ENN, relative to CNNs, can be thought of as negating the redundancies of augmentation. Concretely, data augmentation is less important for the ENN, since augmented data (rotations, translation, reflections, etc.) are invariant under a gauge transformation. On the other hand, CNNs only realize translational invariance, which means they are susceptible to idiosyncrasies induced in training.

VI. DISCUSSION & CONCLUSION

With the upcoming arrival of strong lensing data from Euclid and the Vera Rubin Observatory, it is imperative to assess how algorithms trained on simulations can be adapted to study real world data. To this goal, 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, more complex, data set (the target). To make a precise quantitative evaluation, we based it on two sets of simulations, with the more complex simulation as a proxy for real data.

We have demonstrated that the naive application of substructure classification models have diminished performance when applied to a more complex target data set. We then tested the implementation of several UDA techniques (ADDA, AdaMatch, and Self-Ensemble) for a popular convolutional (ResNet-18) classifier and a symmetry equivariant (ENN) classifier. We found that UDA consistently increases the performance of all models on the target data set. The ENN-based ADDA algorithm achieved top performance, achieving performance competitive with the original source trained and evaluated algorithm.

Investigating performance on a class by class basis, we found that classification performance is consistent between classes. This is despite the fact that the naive application of the source-trained ENN has significantly degraded AUC scores for no substructure and vortices relative to subhalos. This observation, UDA aside, is encouraging as the base architectures are relatively robust to the signature of subhalos lensing data, a common observable among many dark matter models. This result is not surprising as the lensing signature of vortices is inherently more difficult relative to subhalos since they induce no magnification of the background source (see [51] and references therein for details). Nonetheless, it is impressive that UDA can completely compensate for this extra difficulty.

With the upcoming arrival of high quality strong lensing data, domain adaptation techniques will be critical for real world applications of machine learning based dark matter analyses. Possible performance degradations for a simulation trained model naively applied to real data sets can be more significant than what was realized here, making the need for further development and application of UDA methods even more critical. 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.

VII. ACKNOWLEDGEMENTS

M. T. was a participant in the Google Summer of Code (GSoC) 2021 program. S. G. was supported in part by the National Science Foundation Award No. 2108645. S. A. and M. W. T. were supported in part by U.S. National Science Foundation Award No. 2108866. This work made use of these additional software packages: Matplotlib [127], NumPy [128], PyTorch [129], and SciPy [130].

[1] A. K. Drukier, K. Freese, and D. N. Spergel, Phys. Rev. D33, 3495 (1986)
[2] M. W. Goodman and E. Witten, Phys. Rev. D31, 3059 (1985).
[3] D. S. Akerib et al. (LUX), Phys. Rev. Lett. 118, 021303 (2017), arXiv:1608.07648 [astro-ph.CO]
[4] X. Cui et al. (PandaX-II), Phys. Rev. Lett. 119, 181302 (2017), arXiv:1708.06917 [astro-ph.CO]
[5] E. Aprile et al. (XENON), Phys. Rev. Lett. 121, 111302 (2018), arXiv:1805.12562 [astro-ph.CO]
[6] F. Froborg and A. R. Duffy, arXiv e-prints , arXiv:2003.04545 (2020), arXiv:2003.04545 [astro-ph.CO]
[7] Fermi LAT Collaboration, arXiv:1501.05464 [astro-ph.CO]
[8] A. Geringer-Sameth, S. M. Koushiappas, and M. G. Walker, 91, 083535 (2015) arXiv:1410.2242 [astro-ph.CO]
[9] A. Albert, R. Alfaro, C. Alvare, J. D. Álvare, R. Arco, et al., ApJ 853, 154 (2018) arXiv:1706.01277 [astro-ph.HE]
[10] VERITAS Collaboration, Phys. Rev. D 95, 082001 (2017) arXiv:1703.04937 [astro-ph.HE]
[11] J. Rico, Galaxies 8, 25 (2020) arXiv:2003.13482 [astro-ph.HE]
[12] MAGIC Collaboration, 2016, 039 (2016) arXiv:1601.06590 [astro-ph.HE]
[13] IceCube Collaboration, arXiv e-prints , arXiv:1705.08103 (2017), arXiv:1705.08103 [hep-ex]
[14] The Super-Kamiokande Collaboration, arXiv e-prints , arXiv:1503.04858 (2015), arXiv:1503.04858 [hep-ex]
[116] A. S. Bolton, S. Burles, L. V. E. Koopmans, T. Treu, R. Gavazzi, L. A. Moustakas, R. Wayth, and D. J. Schlegel, ApJ 682, 964 (2008), arXiv:0805.1931 [astro-ph]

[117] A. Diaz Rivero, C. Dvorkin, F.-Y. Cyr-Racine, J. Zavala, and M. Vogelsberger, Phys. Rev. D 98, 103517 (2018), arXiv:1809.00004 [astro-ph.CO]

[118] K. He, X. Zhang, S. Ren, and J. Sun, arXiv:1512.03385 [cs] (2015), arXiv: 1512.03385.

[119] R. B. Metcalf, M. Meneghetti, C. Avestruz, F. Bellagamba, C. R. Bom, E. Bertin, R. Cabanac, F. Courbin, A. Davies, E. Decenci`ere, R. Flamary, R. Gavazzi, M. Geiger, P. Hartley, M. Huertas-Company, N. Jackson, E. Jullo, J.-P. Kneib, L. V. E. Koopmans, F. Lanusse, C.-L. Li, Q. Ma, M. Makler, N. Li, M. Lightman, C. E. Petrillo, S. Serjeant, C. Schäfer, A. Sonnenfeld, A. Tortora, D. Tuccillo, M. B. Valentin, S. Velasco-Forero, G. A. V. Kleijn, and G. Vernardos, A&A 625, A119 (2019), arXiv: 1802.03609.

[120] G. French, M. Mackiewicz, and M. Fisher, arXiv:1706.05208 [cs] (2018), arXiv: 1706.05208.

[121] A. Tarvainen and H. Valpola, arXiv e-prints , arXiv:1703.01780 (2017), arXiv:1703.01780 [cs,NE]

[122] D. Berthelot, R. Roelofs, K. Sohn, N. Carlini, and A. Kurakin, arXiv:2106.04732 [cs] (2021) arXiv: 2106.04732 version: 1.

[123] M. Weiler and G. Cesa, arXiv:1911.08251 [cs, eess] (2021) arXiv: 1911.08251.

[124] T. Cohen and M. Welling, in International conference on machine learning (PMLR, 2016) pp. 2990–2999.

[125] D. P. Kingma and J. Ba, arXiv:1412.6980 [cs] (2017), arXiv: 1412.6980.

[126] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, arXiv:1912.01703 [cs, stat] (2019) arXiv: 1912.01703.

[127] J. D. Hunter, Computing in Science & Engineering 9, 90 (2007).

[128] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, Nature 585, 357 (2020).

[129] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, in Advances in Neural Information Processing Systems 32 edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019) pp. 8024–8035.

[130] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, Nature Methods 17, 261 (2020).