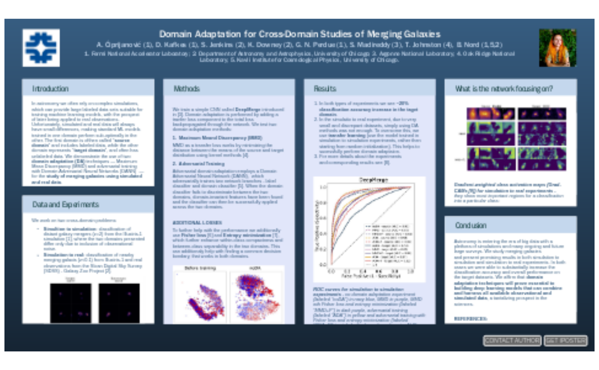
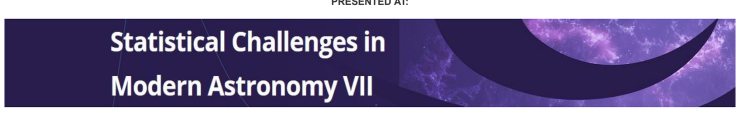
Domain Adaptation for Cross-Domain Studies of Merging Galaxies



A. Ćiprijanović (1), D. Kafkes (1), S. Jenkins (2), K. Downey (2), G. N. Perdue (1), S. Madireddy (3), T. Johnston (4), B. Nord (1,5,2)

1. Fermi National Accelerator Laboratory; 2. Department of Astronomy and Astrophysics, University of Chicago; 3. Argonne National Laboratory; 4. Oak Ridge National Laboratory; 5. Kavli Institute for Cosmological Physics, University of Chicago.

PRESENTED AT:



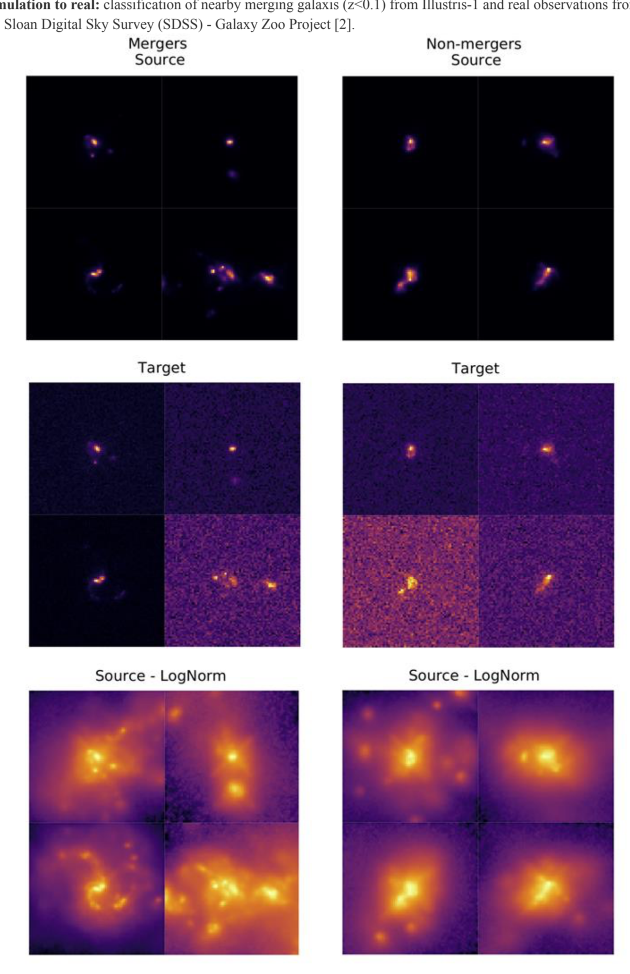
INTRODUCTION

In astronomy we often rely on complex simulations, which can provide large labeled data sets suitable for training machine learning models, with the prospect of later being applied to real observations. Unfortunately, simulated and real data will always have small differences, making standard ML models trained in one domain perform sub-optimally in the other. The first domain is ofthen called "**source domain**" and includes labeled data, while the other domain represents "**target domain**", and often has unlabeled data. We demonstrate the use of two **domain adaptation (DA)** techniques — Maximum Mean Discrepancy (MMD) and adversarial training with Domain Adversarial Neural Networks (DANN) — for the **study of merging galaxies using simulated and real data**.

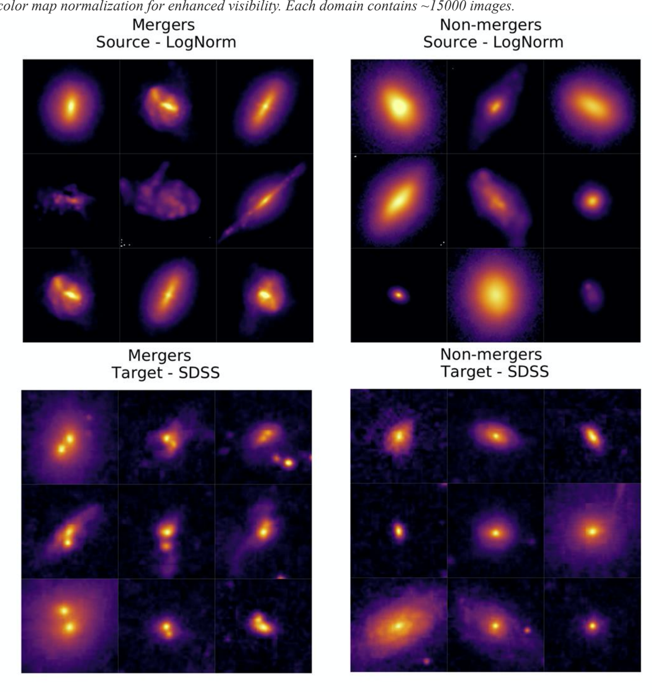
DATA AND EXPERIMENTS

We work on two cross-domain problems:

- **Simultion to simulation:** classification of distant galaxy mergers (z=2) from the Illustris-1 simulation [1], where the two domains presented differ only due to inclusion of observational noise.
- **Simulation to real:** classification of nearby merging galaxis (z<0.1) from Illustris-1 and real observations from the Sloan Digital Sky Survey (SDSS) - Galaxy Zoo Project [2].



Simulation to simulation: *Galaxy images from Illustris-1. Mergers - left; Non-mergers - right. The same objects are repeated across rows, with the top showing the source domain, the middle showing the target domain (added noise mimicking the Hubble Space Telescope), and the bottom displaying the source objects with logarithmic color map normalization for enhanced visibility. Each domain contains ~15000 images.*



Simulatio to real: *Galaxy images from Illustris-1 mimicking SDSS observations (top row) and real SDSS images (bottom row). Mergers - left; non-mergers - right. Source domain images in the top row were plotted with a logarithmic colormap to make features more visible. We can see that the source domain contains more relaxed systems, while the target contains easier examples, with two bright clearly visible cores. This makes the two domains very discrepant. Each domain contains 6000 images.*

METHODS

We train a simple CNN called **DeepMerge** introduced in [3]. Domain adaptation is performed by adding a tranfer loss component to the total loss backpropagated through the network. We test two domain adaptation methods:

1. Maximum Meand Discrepancy (MMD)

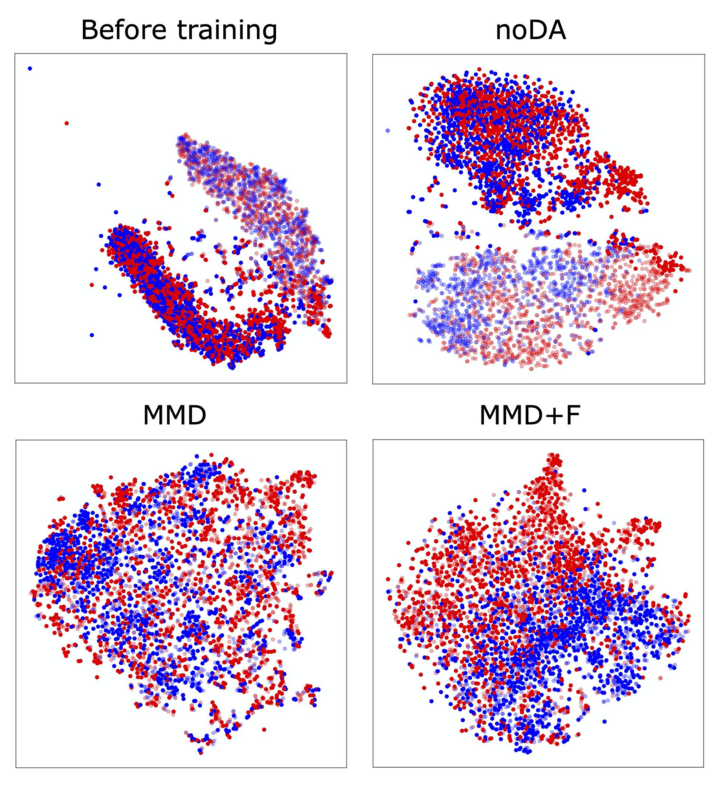
MMD as a transfer loss works by minimizing the distance between the means of the source and target distribution using kernel methods [4].

2. Adversarial Training

Adversarial domain adaptation employs a Domain Adversarial Neural Network (DANN), which adversarially traines two network branches - label classifier and domain classifier [5]. When the domain classifier fails to discriminate between the two domains, domain-invariant features have been found and the classifier can then be successfully applied across the two domains.

ADDITIONAL LOSSES

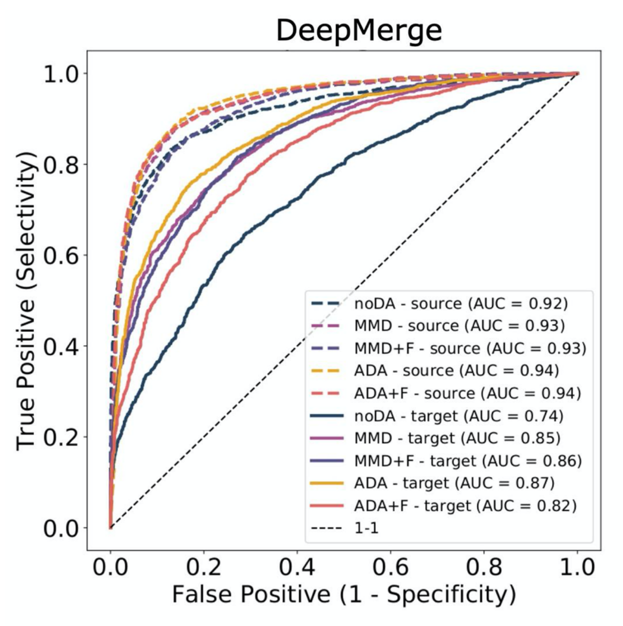
To further help with the performance we additionally use **Fisher loss** [6] and **Entropy minimization** [7], which further enhance within-class compactness and between-class separability in the two domains. This can additionaly help with finding a common decision bondary, that works in both domains.



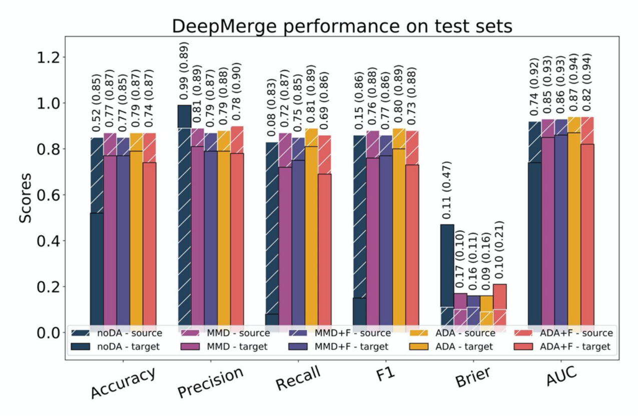
tSNE plots - 2D projections of the latent feature space. *Classes: colors; Domains: opacity. Before training and when training without domain adaptation (first row left and right) two domains remain separate. Training with MMD and MMD with Fisher loss and Entropy minimization (second row left and right) leads to domain overlap, with Fisher loss and Entropy minimization hepling further separte the two classes.*

RESULTS

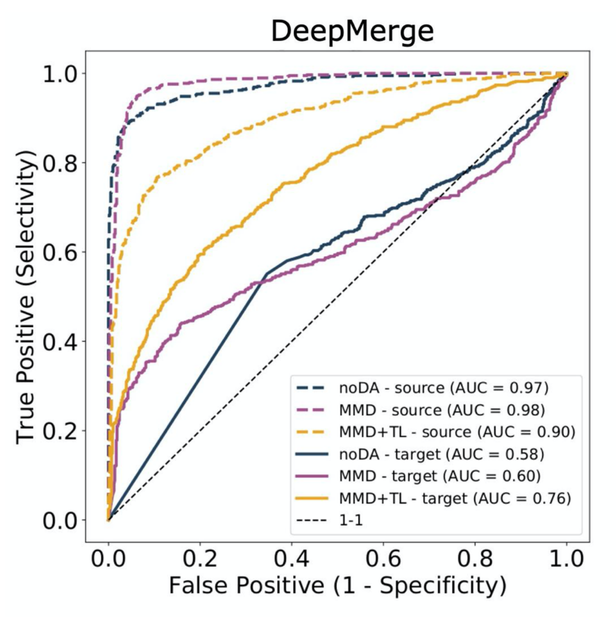
1. In both types of experiments we see **~20% classification accuracy increase in the taget domain**.
2. In the simulatio to real experiment, due to very small and discrepant datasets, simply using DA methods was not enough. To overcome this, we use **transfer learning** (use the model trained in simulation to simulation experiments, rather then starting from random initialization). This helps to sucessfully perform domain adaptaion.
3. For more details about the experiments and corresponding results see [8].



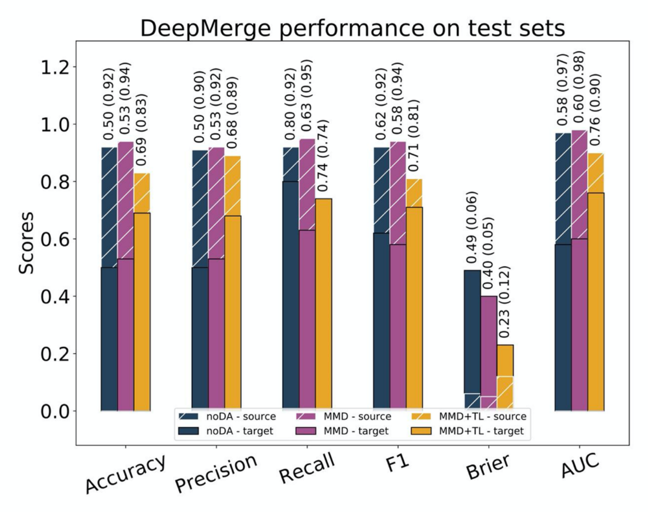
ROC curves for simulation to simulation experiments - *no domain adaptation experiment (labeled "noDA") in navy blue, MMD in purple, MMD wih Fisher loss and entropy minimization (labeled "MMD+F") in dark purple, adversarial training (labeled "ADA") in yellow and adversarial training with Fisher loss and entropy minimization (labeled "ADA+F") in pink. In the legend we also give AUC values for all five experiments.*



Simulation to simulation experiments: *accuracy, precision, recall, F1 score, Brier score, and AUC. Dashed bars - source domain; Solid colored bars - target domain. Color coding is the same as for the corresponding ROC curves.*

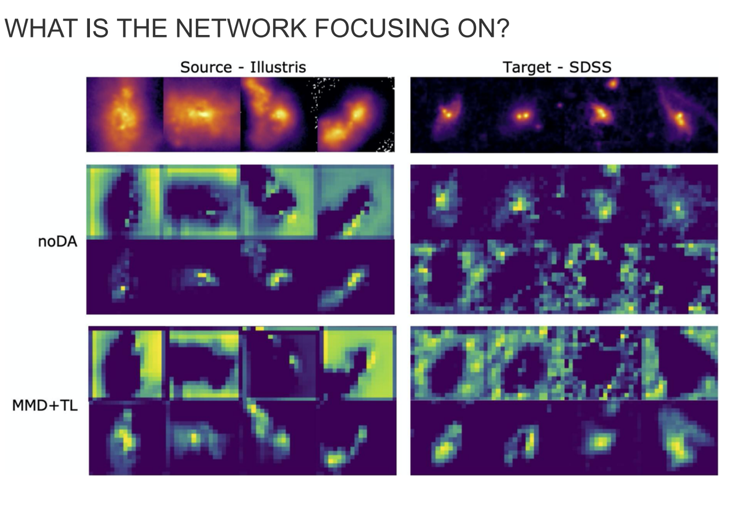


ROC curves for simulation to real experiments *- no domain adaptation experiment in navy blue, MMD in purple, and MMD with transfer learning in yellow.*



Simulation to real experiments: *accuracy, precision, recall, F1 score, Brier score, and AUC.*

WHAT IS THE NETWORK FOCUSING ON?



Gradient-weighted class activation maps (Grad-CAMs [9]) for simulation to real experiments - *they show most important regions for a classification into a particular class:*

- *The top row of images shows examples of true mergers from simulated source dataset (left) and SDSS target dataset (right).*
- *The second row shows Grad-CAMs for the designated example images for classification into the merger class, while row three shows Grad-CAM for the same image for the non-merger class. In case of the source domain, the peripheries are important for positive classification in the merger class, while central regions are important for positive classification as a non-merger. In the case of the target domain, both mergers and non-mergers look very different, so Grad-CAMs become noisy and display inverted behavior compared with the source domain. This leads to the classifier not working in the target domain.*
- *Finally, the third and fourth rows show merger and non-merger Grad-CAMs for the model trained with MMD and transfer learning. The successful domain adaptation is apparent, as the network performs both source and target domain classification in a similar manner as in the case of source domain classification without DA.*

CONCLUSION

Astronomy is entering the era of big data with a plethora of simulations and many ongoing and future large surveys. We study merging galaxies and present promising results in both simulation to simulation and simulation to real experiments. In both cases we were able to substantially increase the classification accuracy and overall performance on the target datasets. We affirm that **domain adaptation techniques will prove essential to building deep learning models that can combine and harness all available observational and simulated data**, a tantalizing prospect in the sciences.

REFERENCES:

[1] Vogelsberger M., et al. (2014), MNRAS, 444, 1518.

[2] Lintott C. J., et al., (2008), MNRAS, 389, 1179.

[3] Ćiprijanović, A., et al. (2020), Astron. Comput., 32, 100390.

[4] Gretton, A., et al. (2008), arXiv:0805:2368.

[5] Ganin, Y., et al. (2016), Journal of Machine Learning Research, 17, 1.

[6] Zhang Y., et al. (2020), arXiv:2003:05636.

[7] Grandvalet Y. and Bengio Y. (2005), Advances in Neural Information Processing Systems17. MIT Press, pp 529–536.

[8] Ćiprijanović, A., et al. (2021), arXiv:2103.01373.

[9] Selvaraju R. R., et al. (2020), Int. J. Comput. Vis., 128. 336.