Exploring domain adaptation with generative adversarial networks for galaxy images

Dezso Ribli

September 15, 2019

Kavli Summer Program in Astrophysics
Supervisors: Helena Domínguez Sáanchez
Marc -Huertas- Company
Yuan-Sen Ting

1 Introduction

The popularity of machine learning applications is rising in astrophysics and astronomy as well as generally in sciences. The recent interest is partially due to the ever increasing size of measurement data as well and recent advances in algorithms and parallel computing resources such as graphics processing units. The primary goal of supervised machine learning is to make predictions of previously unseen data points. In an idealized scenario, the training dataset is indistinguishable from the unseen dataset, in other words both datasets come from the same underlying data distribution. However, in many applications such an ideal set-up can not be achieved, and the two datasets are generated in a somewhat different fashion. For example the training dataset may come from one region of the sky with a limited set of observing conditions which may not perfectly match the observing conditions during predictions. In an other very common machine learning scenario one trains on simulated data and applies the trained machine learning model on observational data. We refer to datasets generated by different processes as *domains*, and when the training data and the data used during inference come from different domains we talk about domain shift. Domain shift is generally undesired as it can result in reduced prediction accuracy and biases predictions (Domínguez Sánchez et al., 2018).

Numerous approaches try to tackle the problem of different training and predictions data domains, these efforts are usually referred to as domain adaptation (Ganin et al., 2016). In the general problem we have a labeled dataset in domain A, and we want to predict the labels of unseen data points in domain B.

One successful approach to domain adaptation relies on transferring examples from domain A to domain B, while maintaining the key characteristics of the transferred examples, especially those characteristics which correspond to the labels. Such transferred examples then can be used to construct a labeled training dataset in domain B, using the original labels of the transferred examples. Apart from domain adaptation for supervised classification or regression tasks, one can also use these transferred examples for content creation, such as transforming a doodle into a photo realistic image.

Very successful solutions exists to domain transfer, when paired realizations of the same example are available in both domains (Isola et al., 2017). When large number of such examples are available, domain adaptation is somewhat trivial, as one can just use the images of these paired examples in domain B as a labeled training and testing dataset. However, paired examples may be scarce, hard to generate or even non-existent for numerous scenarios, and a more challenging question arises when one only has access to unpaired sets of examples in both domains. A prime example for unpaired datasets in astrophysics is the case when domain A consists of simulated data, and domain B consists of observations. The simulated data points have many known properties which we may want to infer from observational data, such as galaxy magnitudes (Boucaud et al., 2019), therefore transferring between the domains is desirable. It is also possible that one wants to use machine learning to assess the quality of simulations with a model trained on observations (Huertas-Company et al., 2019). Note that solving domain transfer via neural networks approach is complementary to physics based realism, and could be most valuable when underlying connection between the domains is very complex or not completely known.

A challenging domain adaptation question for machine learning is astrophysics is the morphological classification of galaxies in numerous surveys with ever improving depth and resolution. The laborious visual inspection and classification of large numbers of galaxies for every single survey to generate a large training datasets is not a very compelling solution. A notable dataset is the GalaxyZoo (Lintott et al., 2008; Willett et al., 2013), where citizen scientist projects classified more than 280000 galaxies using Sloan Digital Sky survey images. A few other classified datasets exist for different surveys and classification schemes, but the list is far from comprehensive. Domínguez Sánchez et al. (2018) showed that a deep learning morphology classifier trained on the SDSS images performs significantly worse on Dark Energy Survey (DES) galaxy images due to domain shift, even though the latter is a significantly deeper survey with superior seeing. That study was able to improve the performance of the classifier by further training it on a small set of visually classified sample of DES images. A universal method which works across any surveys remains an elusive goal for automated galaxy morphology classification.

In this present work we focus on domain transfer via generative adversarial networks (GANs) (Goodfellow et al., 2014; Zhu et al., 2017). We consider domain adaptation for galaxy images of Illustris simulations to observations with the SDSS and the DES survey. We also evaluate transfer between different surveys with little overlap, SDSS and DES. The outline of the report is as follows. In §2 we give a brief overview of domain transfer via generative adversarial networks, and we present our methodology for the evaluation of the quality of the generated examples. In §3 we present the datasets used for the study. In §4 we explore the use of GANs to generate examples in each domain to draw a baseline of GAN capabilities for galaxy image generation. In §4.1 we explore the use of cycle consisted GANs for domain trainser, and we report the results of our domain transfer experiments. In §5 we draw the conclusions of our results, and lay out future direction in S 6.

2 Domain transfer via cycle consistent generative adversarial networks

A successful approach to domain transfer relies on generative models, specifically generative adversarial networks. GANs are able to generate images in a single domain, but the framework is very flexible and extending it with a sort of cycle consistency one can use it for transferring examples between domain. Here, we first review the basics on GANs and then the modifications needed for domain transfer applications.

Deep convolutional generative adversarial networks revolutionized the field of image generation in the last 5 years. The GAN approach relies on a large sample of images and two competing neural networks, one attempts to generate realistic images from pure noise vectors as inputs, while the other neural network is presented with a mix of real and generated images and it attempts to classify the origin of these samples. The first neural network is called the generator, the second one is called the discriminator. During training the discriminator is trained to recognize the features which discriminate real data from the generated examples. One can calculate the gradient of the discriminators judgment with respect to the pixel values of generated images, to calculate the changes needed to make the generated examples look slightly more realistic as judged by the discriminator. These gradients can be used as the supervised signal to train the generator to make more realistic images. Therefore the neural networks are not exactly adversarial, as the discriminator provides instructions to teach the generator. The GAN framework turned out to be very fruitful, and its evolution led to generators which are sometimes capable of producing very compelling high resolution natural images (Karras et al., 2017; Brock et al., 2018). A few studies explored generative adversarial network is astronomy, for generating galaxy images (Ravanbakhsh et al., 2017; Fussell and Moews, 2019) or weak lensing maps (Rodríguez et al., 2018; Mustafa et al., 2019).

However, GANs are notoriously hard to train, they suffer from multiple pathologies which can disrupt training and they require a fine balance between generator and the discriminator. A too sophisticated discriminator often produces useless gradients for a bad generator, as small changes on the input image would not make the image any more realistic. In the more common failure case, called mode collapse, the generator ends up producing a single example. Note that even when successfully trained, the generators of GANs produce images with relatively low diversity compared to natural images (Razavi et al., 2019). Apparently it is hard to enforce large diversity of the generator in the GAN framework.

A large variety of generative adversarial networks emerged in the last years, which attempted to solve the shortcomings of the initial GAN framework (Radford et al., 2015; Salimans et al., 2016; Arjovsky et al., 2017). A major improvement on the stability of GAN training was achieved with replacing the original binary cross-entropy objective of the discriminator with the Wasserstein distance based objective (Arjovsky et al., 2017). GANs are probabilistic generative models which implicitly define a distribution via mapping a low dimensional diagonal Gaussian distribution to the pixel space with a parametric function, the neural network. During training, examples of this distribution are evaluated and the parameters are optimized to fool the discriminator network, which is equivalent to minimizing the Kullback-Leibler divergence (Goodfellow et al., 2014) between the true and the generated data distributions. However, when two distributions are sufficiently different, minimizing the KL divergence may not provide reasonable gradients to train the generator (Arjovsky et al., 2017). Following this insight one can attempt to minimize the Wasserstein distance often called the earth-mover distance between two distributions, which is the cost of the optimal transport plan between two distributions. The Wasserstein distance between two distributions is always defined and could provide meaningful gradients to train the generator even when the generated distribution is quite different than the true data distribution. Naturally, the Wasserstein distance is intractable but one can provide a reasonable approximation using the Kantorovich-Rubinstein duality (Arjovsky et al., 2017).

$$W(P_d, P_g) = \sup_f E_{x\sim P_d}[f(x)] - E_{x\sim P_g}[f(x)], \tag{1}$$

where f are all the possible 1-Lipschitz functions, E denotes the expected value, and $P_{g,d}$ denote the true and the generated data distributions respectively. The idea behind the Wasserstein GAN is to approximate the above mentioned f function with the discriminator neural network. Restricting our discriminator to be a 1-Lipschitz function we can estimate the Wasserstein distance between the generated and the true data distribution by training the discriminator to maximize the mean difference between its output on the real and the generated examples. We attempt to approximate the supremum in the definition of the Wasserstein distance by training the discriminator to near optimality by training it for multiple iterations between each training iteration of the generator. A straightforward way to enforce a Lipschitz constraint on the discriminator is to clip its weights to a certain interval, The weight clipping approach is simple and effective for small neural networks, however it can seriously hamper the training of more complex neural networks, and therefore limit the complexity of the usable architectures (Arjovsky et al., 2017). A 1-Lipschitz function also has bounded gradients, and an improved training procedure of WGANs tries to enforce the 1-Lipschitz constraint via penalizing the gradients of the discriminator with respect to the image (Gulrajani et al., 2017). Wasserstein GANs were shown to be much more stable, and easy to train than the usual GAN formulation. WGAN with gradient penalty allows the training of complex models such as residual networks with more than a 100 layers (Gulrajani et al., 2017). The current state of the art GANs (Zhang et al., 2018; Brock et al., 2018) strongly build on the idea of WGANs, and these solutions also use 1-Lipschitz constrained neural networks achieved via spectral normalization (Miyato et al., 2018), however the Wasserstein objective is often replaced with the original GAN objective when performance is preferred over stability.

Apart from single domain image generation, GANs are successfully used for paired image to image translation problems (Isola et al., 2017), due to their ability to generate realistic samples with multi-modal distributions instead of generating the mean of the distribution, such as generating a high resolution flow of water from a low resolution input image in the context of super resolution (Ledig et al., 2017). Domain transfer with unpaired sets of images can also be attempted using a generalization of the GAN framework (Zhu et al., 2017). Given two sets of images, one can use two neural networks to translate between the domains, these network are the analogues of generators. In order to achieve realistic examples in the target domains, one needs to add discriminator networks to both domains, these networks judge the quality of the samples generated just as in a standard single image GAN. However, an example of domain A could in principle be transformed by the neural network to a very different example in domain B, potentially changing the content we wanted to preserve during the transfer. In order to enforce the transfer of the domain independent content of the image,we transform back the generated examples into their original domain using the already mentioned generators and we penalize the difference between the original images and their cycled versions. The loss function which penalizes the deviations from the original image is called the cycle consistency loss, it is generally an L1 loss. There is a large variety of frameworks built on the idea of cycle consistency, the one described above is called CycleGAN (Zhu et al., 2017), and it produces the most visually compelling transferred images (Lee et al., 2018). Training the complete CycleGAN means optimizing the 6 different loss terms: the parameters of the two discriminators are optimized to differentiate between real and fake examples, the parameters of the generators are optimized to increase the difference of the discriminators output on fake examples and the parameters of the generators are also optimized to reduce the difference between the original images and the cycled samples. Note that the original formulation of the cycleGAN uses the GAN objective, which we replaced with the Wasserstein distance based objective to potentially improve stability. For a WGAN based CycleGAN the minimized loss functions can be written as follows.

$$L_{disc} = E[D_1(G_1(x_2)] - E[D_1(x_1)] + E[D_2(G_2(x_1)] - E[D_2(x_2)] \quad (2)$$

$$L_{gen} = -L_{disc} + |G_2(G_1(x_2) - x_2| + |(G_1(G_2(x_1)) - x_1|, \quad (3)$$

where, G, D are generator and the discriminator functions and x_1, x_2 are examples from the two domains. We modify the original WGAN implementation [1] for our experiments and we follow the details of the implementation of the cycleGAN from the original authors [2].

3 Data

We consider three domains in our experiments, Illustris, SDSS and DES galaxy images. In each domain, the images were transformed with an asinh function to reduce the dynamic range of the pixel values. We found it convenient to help the visual inspection of the generated samples.

- In the first domain we use images of Illustris and IllustrisTNG galaxies spatially down sampled by a factor of 2 to 64×64 pixel size.
- In the second domain we consider 64×64 pixel postage stamp cutouts of SDSS DR7 main sample galaxies in the r band.
- In the third domain we consider 96×96 pixel postage stamp cutouts of DES DR1 galaxies resampled to 64×64 pixels. The DES galaxies are selected to be brighter than 18 magnitudes, with a half light radius larger than 10 pixels but smaller than 48 pixels.

4 Generating galaxy images using GANs

To establish a baseline for our domain transfer experiments, we first attempt to generate samples in each domain using single domain GANs. Interestingly, we find that one of the most advanced and stable GAN variant the WGAN with gradient penalty is not able to generate realistic images, due to instability during training. We find that the generators loss starts to wildly oscillate after a few thousand or a few tens of thousands of iterations, and the visual quality of the samples degrade after that. We attempted to mitigate the behavior by enlarging the coefficient of the gradient penalty with no success. Note that the behavior is not due to the high noise in our observational data, as we also experience the same behavior with noiseless simulated Illustris galaxies. Also, note that using the same procedure we are able to conduct stable training on natural images, such as MNIST digits re-scaled to 64 pixels or small imagenet

[1] https://github.com/igul222/improved_wgan_training
[2] https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix

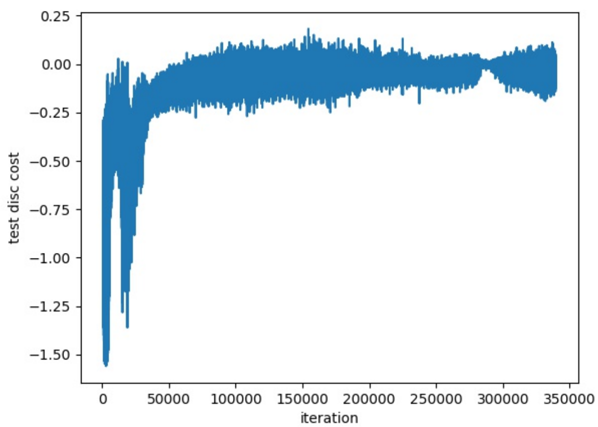

Figure 1: The discriminator loss converges to 0 with a held out test set, demonstrating that the generator is able to produce SDSS galaxies which are statistically indistinguishable from real ones. Similar loss curves are obtained for both the Illustris and the DES domains.

images. A detailed investigation of the issue could yield interesting insights into the stability of WGANs with gradient penalty, however, we leave it for future work. Interestingly, we find that replacing the gradient penalty with weight clipping leads to stable training and good quality samples, although this method is regarded as inferior to gradient penalty. However, other methodologies such as spectral normalization could potentially be better solutions to the problem.

We use a 128 dimensional latent space vector as input to the generators. The spatial extent of the first layer is 4x4 units which grows two-fold via every transposed convolution and finally reaches the 64x64 pixel size of the images. The first layer has 512 filters which is reduced two-fold via every transposed convolution and finally reaches the 64 filters before the final layer which outputs the generated galaxy image. The generator network consists of 4 blocks of transposed convolutions followed by a batch normalization and a ReLU. The discriminator CNN is a mirrored version of the generator. We use a batch size of 64 for each experiment.

After approximately 2×10^5 iterations the discriminators loss function converges to 0 on held out test sets for each of the 3 datasets, indicating that the discriminator is unable to tell apart generated and true examples. Note that the discriminators are relatively simple neural networks, and their weights are clipped, therefore their discrimination ability is limited. It could be useful to train another more powerful complex neural network, without weight clipping

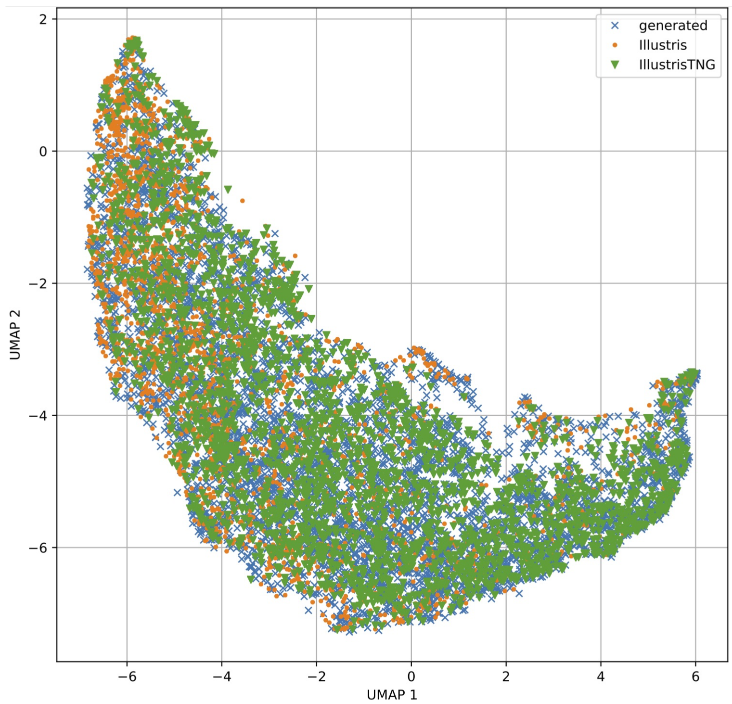

Figure 2: The distribution of true and WGAN generated Illustris galaxies cover the same space when their representation in the last hidden layer of the discriminator is embedded in two dimensions via UMAP.

to evaluate more thoroughly whether the generated examples are truly indistinguishable from the real ones.

As another test we also embedded the true and the generated galaxy samples with UMAP into two dimensions using both the pixel space and the representations of the last hidden layer of the discriminator. The generated and the true samples have roughly the same distribution in the two dimensional embedded space. Note that this is hardly surprising, because if they would have different distributions in the two dimensional embedding, the discriminator could probably tell them apart from each other. Nevertheless these tests further strengthen the point that it is indeed hard to discriminate between the true and the generated samples. We show the representation of the last hidden layer embedded with UMAP for the Illustris galaxies on Fig 2.

We show a truly random selection of real and generated galaxies for each dataset on Fig 3, Fig 4 and 5. The examples generally look visually satisfactory.

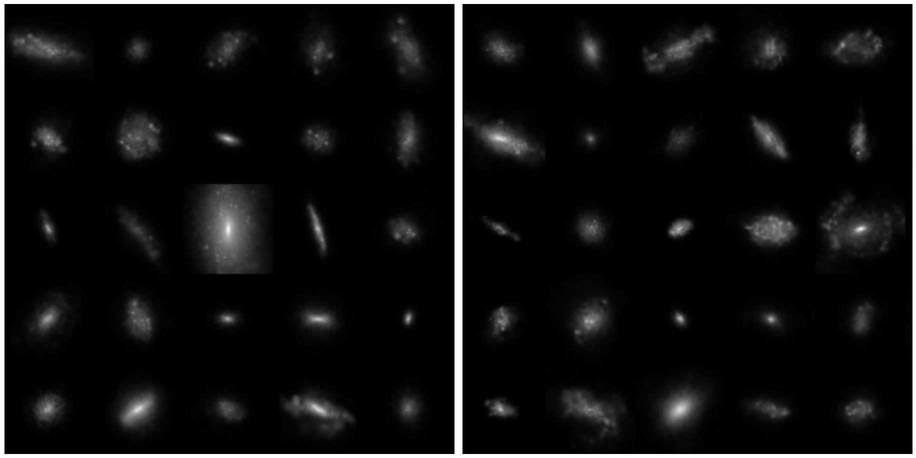

Figure 3: True (left) and WGAN generated (right) galaxy images in the Illustris simulation domain.

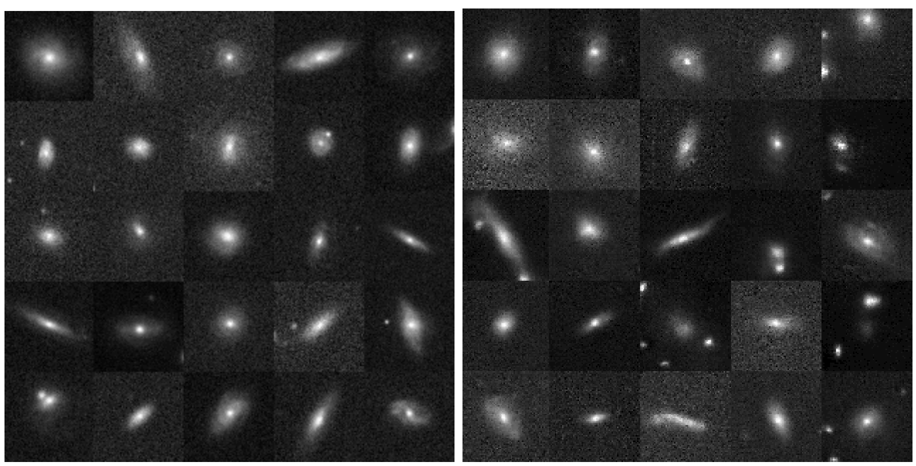

Figure 4: True (left) and WGAN generated (right) galaxy images in the SDSS observation domain.

In summary, we established that we we are able to generate reasonably satisfactory galaxy images using GANs in any of the three domains used.

4.1 Transferring galaxy images between domains

We train cycleGANs to transfer between the following pairs of domains:

- simulated noiseless Illustris galaxies and SDSS observations

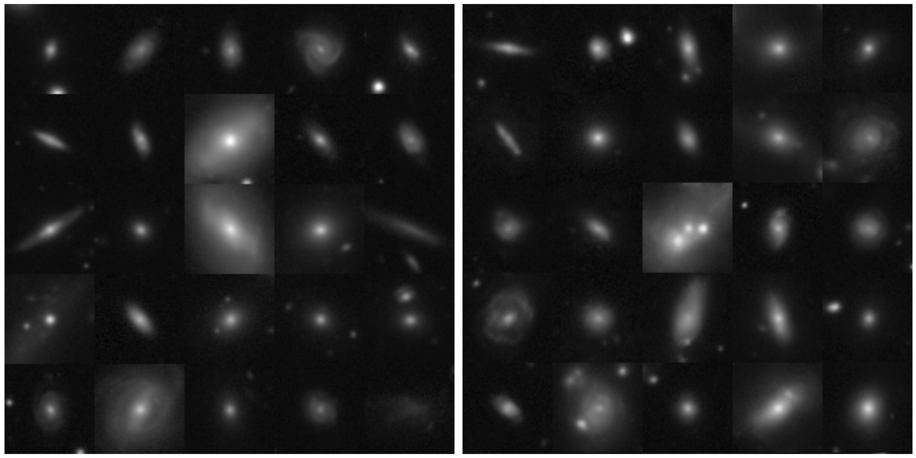

Figure 5: True (left) and WGAN generated (right) galaxy images in the DES observation domain.

- simulated noiseless Illustris galaxies and DES observations
- SDSS and DES observations

Interestingly, even after long training with a million iterations, we find that the output of the discriminator is significantly different for the true samples and the generated, transferred examples 6. In other words, our transferred images do not fool the discriminator, unlike in the single image GAN scenario. We note that this is not special to the domains between the Illustris and SDSS, but we obtain similar results for each of our domain transfer experiments. In the light of the results of our single GAN experiments the failure to fool the discriminators here is somewhat surprising. The single GAN results show that a convolutional neural network is able to reproduce each domain sufficiently using noise vectors as inputs, and apparently the restrictions in the cycleGAN framework pose too strong restrictions on the generator.

We also evaluated the embedded representations of the generated and the true samples. We find that our SDSS dataset has a large number of pipeline errors, resulting in off-center or non-existing objects on the images, which are missing from the other two datasets. The cycleGAN apparently was not able to handle these objects, which may partially explain why the the discriminators are not completely fooled. The experiments with SDSS data need to be reevaluated with an improved dataset which is free of the above mentioned erroneous objects. Interestingly while the discriminators of the Illustris-DES transfer are not fooled by the generated examples, the embedded representations of the last hidden layer of the respective discriminators show no difference between the true and the generated samples 7. These results are promising, and we have to further investigate what attributes allow the distinguish between the true and

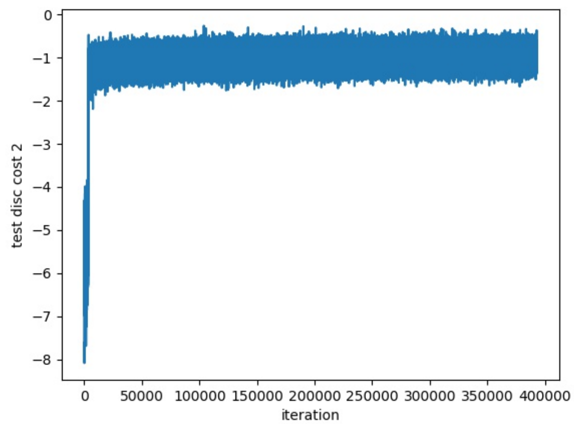

Figure 6: The discriminators are not fooled during transfer. The discriminator loss curve plotted for the SDSS domain during Illustris-SDSS transfer. The loss functions do not reach 0 during training, which means that the discriminator networks are capable of telling apart true and generated samples. Similar loss curves are obtained for every single domain or experiment during transfer.

the generated samples.

In what follows, we discuss the factors which possibly limit the ability of our generator to produce samples which are indistinguishable from the true examples. First, we use a relatively simple neural network for transfer with 2 convolutional layers with spatial down-sampling by a factor of two and 2 transposed convolutions for up-sampling. However the neural networks used in the single GAN scenario are not much more complex, as they consist of 4 transposed convolutional layers for the generator. The discriminators in the two setups are identical. A large difference between the single GAN and the cycleGAN generators is the number of spatial hierarchies used. In the early experiments, we attempted to train a cycleGAN with more down-sampling and up-sampling transformations, but we found found that the produced transferred images bear little resemblance to the inputs. Note that given large enough, freedom the cycleGAN may be able to decode our original images into the new domain without keeping the structure we wish to transfer, and the original cycleGAN implementations also do not use more than 2 down-sampled spatial resolution feature maps. However, due to our limited time and computing resources we have not conducted an exhaustive neural network architecture search, and we can not rule our that a more complex neural network could be capable of producing realistic transferred examples.

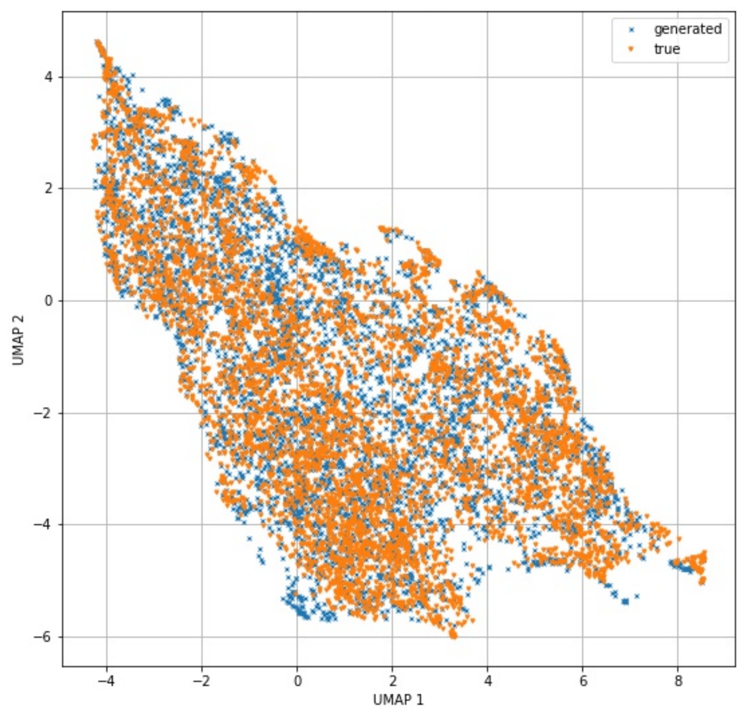

Figure 7: The distribution of true and generated DES galaxies cover the same space when their representation in the last hidden layer of the discriminator is embedded in two dimensions via UMAP. The DES galaxies are generated via a cycleGAN from noiseless Illustris galaxies.

Another potential problem may arise when there are large differences between the datasets we want to transfer between. Zhu et al. (2017) also note that only very close domains distances can be bridged with a cycleGAN, transformations which only need to make small changes on the image. For example they are unable to transfer between images of cat and dogs, as the necessary deformations on the images are too large. The datasets used here have notable differences: the Illustris galaxies are noiseless, while observations have pixel noise. Moreover each dataset has varying degrees of visible background objects and companions.

We show transferred examples produced at the end of the training schedule together with their original counterparts on Fig 8, Fig 9, Fig 10. Visual inspection of the samples show that the training was relatively successful and the transferred examples look similar to the target domains while maintaining some of the morphological properties of the source galaxies. However, closer inspection reveals some noise artifacts, which may be eradicated with different

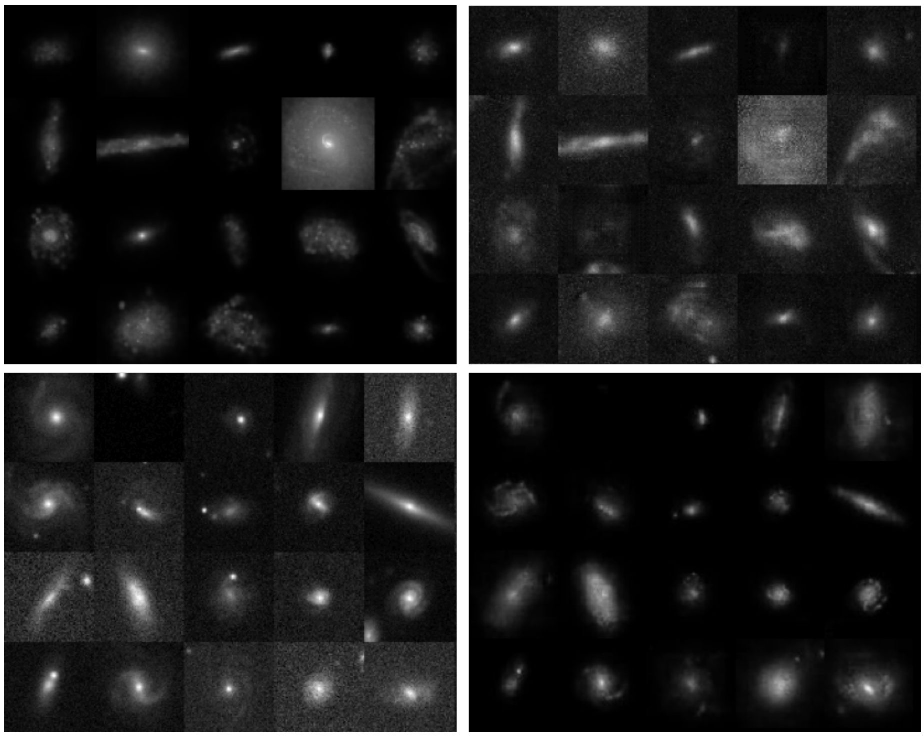

Figure 8: True (left) and generated (right) examples in the context of transfer between Illustris (top) and SDSS (bottom) galaxies.

neural network architectures or different GAN frameworks such as spectral normalization. A peculiar source of bias is also apparent, namely bright galaxies are often transformed into noisy galaxies, potentially due to the fact there is no noise injected into the generator and transforming signal into noise is a potentially easy solution to generating noisy samples. A potential remedy to this problem could be the injection of noise into the generators.

The DES and the SDSS surveys have a small overlapping region around the celestial equator, where we can find paired examples. We evaluated the trained cycleGAN on these examples, and we show one interesting galaxy where spiral arms are also resolved in SDSS on Fig 11. Naturally, the counterfeit example in the DES domain which is the SDSS image transferred to the DES domain does not reproduce the the detailed structure of the galaxy, which is only resolved in DES. Note that the goal of the transfer is not the accurate reproduction of single galaxies, as that is practically impossible when starting with an unresolved and noisy image. However, the generated DES examples not only does not reproduce the exact realization of the galaxy in DES, but lacks any detailed structure.

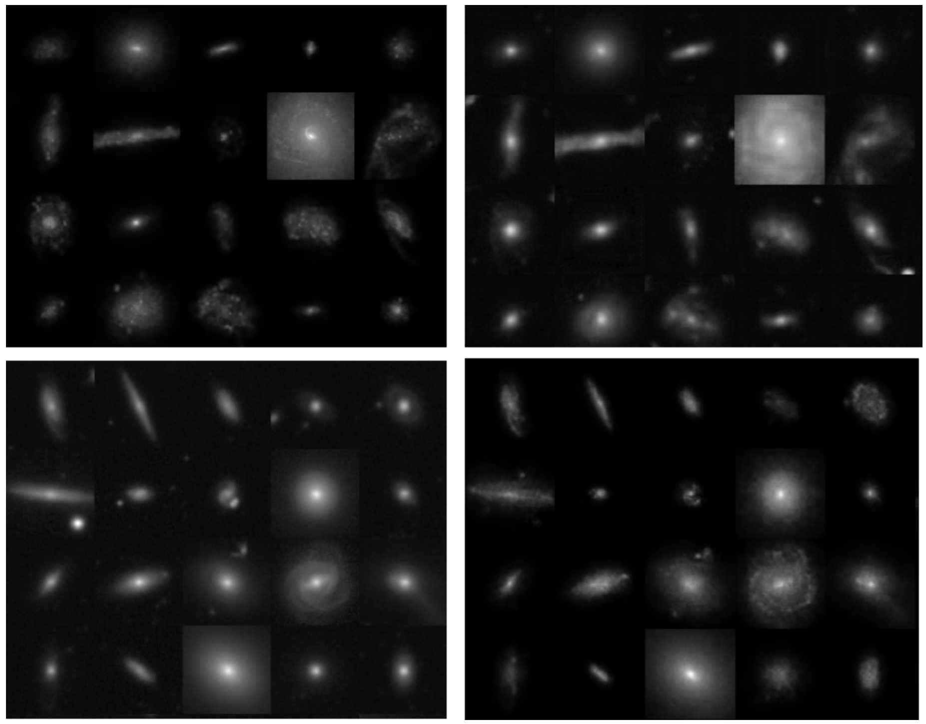

Figure 9: True (left) and generated (right) examples in the context of transfer between Illustris (top) and DES (bottom) galaxies.

We find that the lack of detailed spiral structure is a general problem of our GANs both in single domains and transfers. More complex approaches which enlarge the field of view of neuron such as self-attention may be able to mitigate the problem of lacking structure. We also find that block like noise artifacts appear on the generated example, which may be due to the relatively simple architecture, with up-sampling staged very close to the final image. Changes in the networks architecture may be able to mitigate this problem. The generated SDSS image of the galaxy looks like a fairly realistic realization, of the galaxy as seen in SDSS. Reassuringly the structure of the galaxy remains similar, and the class of the galaxy is not changed during transfer.

5 Discussion

We explored cycle consistent generative adversarial networks to transfer astronomical objects between different domains, namely the Illustris simulations, the Sloan Digital Sky survey and the Dark Energy Survey. Acquiring labeled

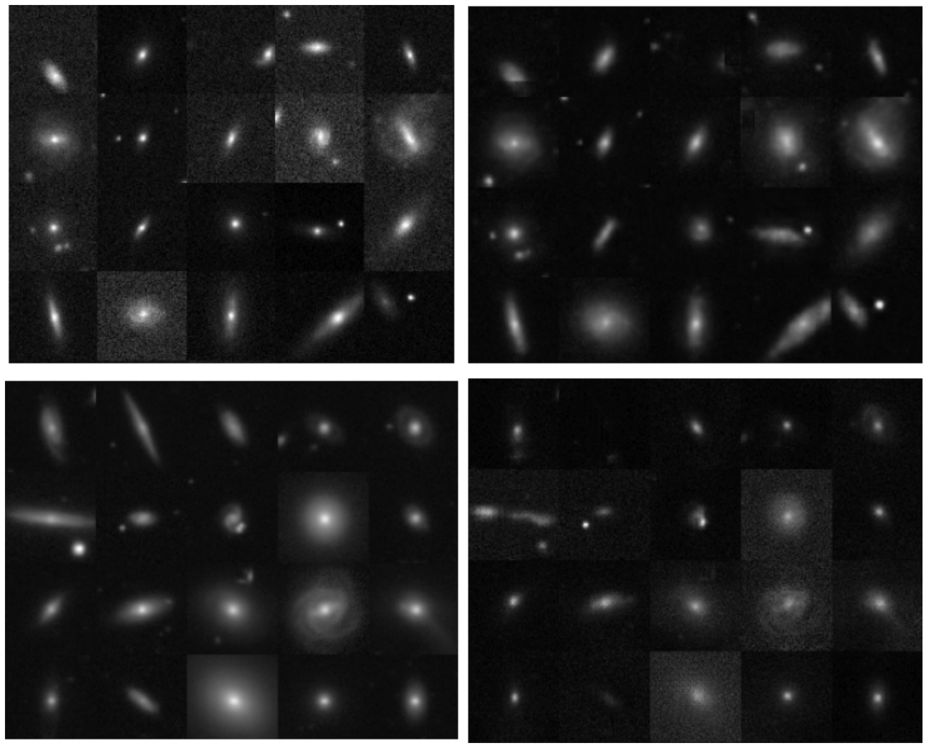

Figure 10: True (left) and generated (right) examples in the context of transfer between SDSS (top) and DES (bottom) galaxies. The galaxies on right are transferred version of the galaxies on the left.

training data for machine learning tools is either laborious or sometimes simply not possible in observations. Transferring labeled data between different domains enables the construction of training datasets in domains where no such labeled training data is available. When transferring between simulations and observations cycleGANs could be used to paint realism on simulated data as an alternative to hand crafted realism processes. Transferring between observations could adjust the instrumental and atmospheric effects between surveys. Note that the use cases presented here could in principle be also realized with meticulous reconstruction of the underlying physical processes, however a simple shortcut solution with machine learning has obvious benefits. When physical transfer codes are available, comparing the two solutions could lead to interesting insights about the transfer process.

We demonstrated that generative adversarial networks are capable of generating realistic samples of galaxy images in both simulation and observation domains. The quality of the samples is high enough to fool a simple discrim-

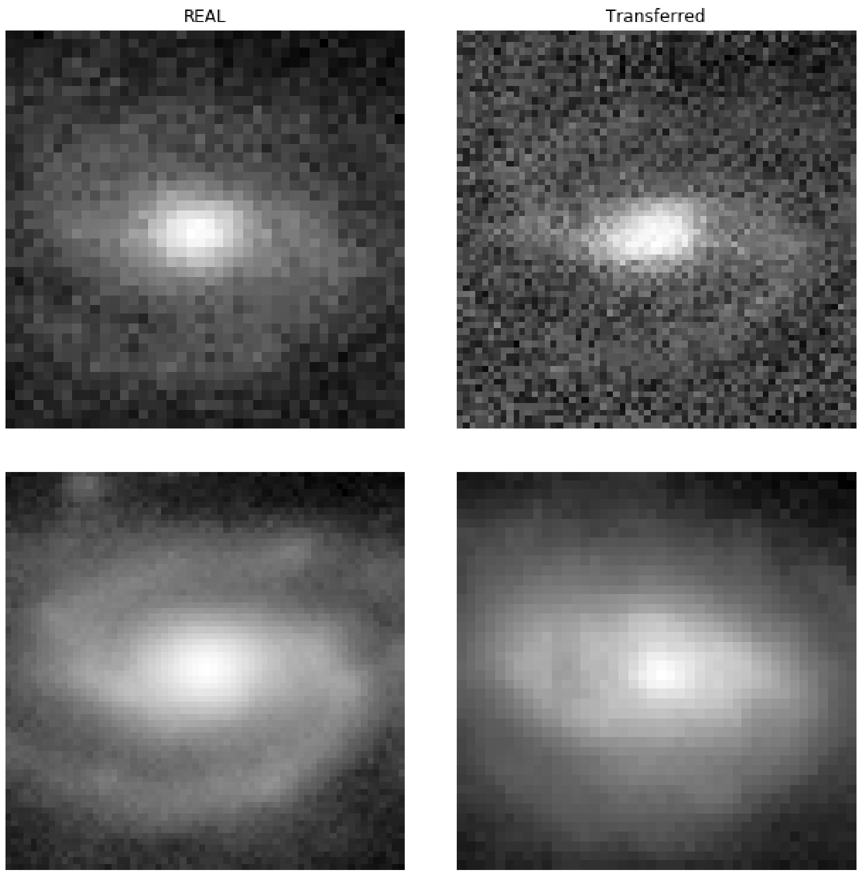

Figure 11: True (left) and generated (right) example in the context of transfer between SDSS (top) and DES (bottom) galaxies for a galaxy which is both observed in SDSS and DES.

inator network, however complex CNNs could be trained to further validate whether the generated and true samples are truly indistinguishable. After qualitative visual inspection of the samples we also find it hard or impossible to tell apart generated and true galaxies. One notable exception is that large spiral galaxies with well resolved spiral arms are missing from our generated samples, possibly showing the general phenomenon of mode dropping often encountered when generating data with GANs. Further improvements of GANs which are able to produce such examples may be desirable for certain purposes.

We demonstrate that it is indeed possible to transfer galaxy images with cycleGANs between different domains such as simulations or different observations. Qualitative evaluation of the quality of transferred examples using the overlapping footprint of the SDSS and the DES surveys reveals promising results. Interestingly the discriminator networks are not completely fooled by our generated samples, but this may be simply due to the intrinsic differences between the datasets.

6 Future work

The project is work in progress, and we attempt to summarize the potential future directions here. The most promising and straightforward question we identified is the transfer from SDSS galaxy images to DES galaxy images, with the goal of building a morphological classifier of DES galaxies, and we will further explore this particular question. Note that the DES survey can be replaced with any single survey in the proposed setup. The SDSS dataset we used is based on the DR7 data release, and we are going to move on to use both imaging and catalog data from DR12 to improve both the catalog and the imaging data. The differences between the datasets may limit the efficiency of the transfer, and we will attempt to make the task easier by trying to match the properties of the datasets e.g.: transforming the intensities or pixel scale of the datasets. Currently we were not able to completely fool the discriminators, and in order to understand the flaws of our generated examples, we will attempt to use attribution methods such as layer wise relevance propagation (LRP) of integrated gradients (**??**) to highlight regions, and features of the images which allow the discriminators to tell apart true and generated samples. If we can better understand the problems of the generated samples, we may be able to refine the framework to produce higher quality counterfeit samples. There are newer and more powerful GAN frameworks which could potentially improve the quality of the produced examples, namely spectral normalization and self attention (Miyato et al., 2018; Zhang et al., 2018). Finally, a training dataset from labeled galaxies generated by the transfer needs to be created to ultimately evaluate the merit of the domain transfer approach.

7 Acknowledgments

We acknowledge funding from the sponsors of the Kavli summer program in astrophysics, The Kavli Foundation, The National Science Foundation and UC Santa Cruz. RD acknowledges funding by the National Research, Development and Innovation Office of Hungary via grant OTKA NN 129148 and computational resources from the National Quantum Technologies Program.

References

Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein gan. *arXiv preprint arXiv:1701.07875*.

Boucaud, A., Huertas-Company, M., Heneka, C., Ishida, E. E., Sedaghat, N., de Souza, R. S., Moews, B., Dole, H., Castellano, M., Merlin, E., et al. (2019). Photometry of high-redshift blended galaxies using deep learning. *arXiv preprint arXiv:1905.01324*.

Brock, A., Donahue, J., and Simonyan, K. (2018). Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*.

Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., Kaviraj, S., Fischer, J., Abbott, T., Abdalla, F., Annis, J., Avila, S., Brooks, D., et al. (2018). Transfer learning for galaxy morphology from one survey to another. *Monthly Notices of the Royal Astronomical Society*, 484(1):93–100.

Fussell, L. and Moews, B. (2019). Forging new worlds: high-resolution synthetic galaxies with chained generative adversarial networks. *Monthly Notices of the Royal Astronomical Society*, 485(3):3203–3214.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., and Lempitsky, V. (2016). Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1):2096–2030.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. (2017). Improved training of wasserstein gans. In *Advances in neural information processing systems*, pages 5767–5777.

Huertas-Company, M., Rodriguez-Gomez, V., Nelson, D., Pillepich, A., Bottrell, C., Bernardi, M., Domínguez-Sánchez, H., Genel, S., Pakmor, R., Snyder, G. F., et al. (2019). The hubble sequence at z 0 in the illustristng simulation with deep learning. *Monthly Notices of the Royal Astronomical Society*, 489(2):1859–1879.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134.

Karras, T., Aila, T., Laine, S., and Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690.

Lee, H.-Y., Tseng, H.-Y., Huang, J.-B., Singh, M., and Yang, M.-H. (2018). Diverse image-to-image translation via disentangled representations. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 35–51.

Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M. J., Nichol, R. C., Szalay, A., Andreescu, D., et al. (2008). Galaxy zoo: morphologies derived from visual inspection of galaxies from the sloan digital sky survey. *Monthly Notices of the Royal Astronomical Society*, 389(3):1179–1189.

Miyato, T., Kataoka, T., Koyama, M., and Yoshida, Y. (2018). Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*.

Mustafa, M., Bard, D., Bhimji, W., Lukić, Z., Al-Rfou, R., and Kratochvil, J. M. (2019). Cosmogan: creating high-fidelity weak lensing convergence maps using generative adversarial networks. *Computational Astrophysics and Cosmology*, 6(1):1.

Radford, A., Metz, L., and Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

Ravanbakhsh, S., Lanusse, F., Mandelbaum, R., Schneider, J., and Poczos, B. (2017). Enabling dark energy science with deep generative models of galaxy images. In *Thirty-First AAAI Conference on Artificial Intelligence*.

Razavi, A., Oord, A. v. d., and Vinyals, O. (2019). Generating diverse high-fidelity images with vq-vae-2. *arXiv preprint arXiv:1906.00446*.

Rodríguez, A. C., Kacprzak, T., Lucchi, A., Amara, A., Sgier, R., Fluri, J., Hofmann, T., and Réfrégier, A. (2018). Fast cosmic web simulations with generative adversarial networks. *Computational Astrophysics and Cosmology*, 5(1):4.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X. (2016). Improved techniques for training gans. In *Advances in neural information processing systems*, pages 2234–2242.

Willett, K. W., Lintott, C. J., Bamford, S. P., Masters, K. L., Simmons, B. D., Casteels, K. R., Edmondson, E. M., Fortson, L. F., Kaviraj, S., Keel, W. C., et al. (2013). Galaxy zoo 2: detailed morphological classifications for 304 122 galaxies from the sloan digital sky survey. *Monthly Notices of the Royal Astronomical Society*, 435(4):2835–2860.

Zhang, H., Goodfellow, I., Metaxas, D., and Odena, A. (2018). Self-attention generative adversarial networks. *arXiv preprint arXiv:1805.08318*.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232.