bioRxiv preprint doi: https://doi.org/10.1101/2022.11.15.516349; this version posted November 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

UNSUPERVISED HARMONIZATION OF BRAIN MRI USING 3D CYCLE GANS AND ITS EFFECT ON BRAIN AGE PREDICTION

Dheeraj Komandur[1], Umang Gupta[1,4], Tamoghna Chattopadhyay[1], Nikhil J. Dhinagar[1], Sophia I. Thomopoulos[1], Jiu-Chiuan Chen[2], Dan Beavers[3], Greg ver Steeg[4], Paul M. Thompson[1] and the Alzheimer's Disease Neuroimaging Initiative (ADNI)

[1]Imaging Genetics Center, Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Marina del Rey, CA, USA

[2]Depts. Of Population and Public Health Science and Neurology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA [3] Dept. of Statistical Sciences, Wake Forest University, NC, USA [4] USC Information Sciences Institute, Marina del Rey, CA, USA

ABSTRACT

Deep learning methods trained on brain MRI data from one scanner or imaging protocol can fail catastrophically when tested on data from other sites or protocols - a problem known as *domain shift*. To address this, here we propose a *domain adaptation* method that trains a 3D CycleGAN (cycle-consistent generative adversarial network) to harmonize brain MRI data from 5 diverse sources (ADNI, WHIMS, OASIS, AIBL, and UK Biobank; total N=4,941 MRIs, age range: 46-96 years). The approach uses 2 generators and 2 discriminators to generate an image harmonized to a specific target dataset given an image from the source domain distribution and *vice versa*. We train the CycleGAN to jointly optimize an adversarial loss and cyclic consistency. We use a patch-based discriminator and impose identity loss to further regularize model training. To test the benefit of the harmonization, we show that brain age estimation - a common benchmarking task - is more accurate in GAN-harmonized versus raw data. *t*-SNE maps show the improved distributional overlap of the harmonized data in the latent space.

Index Terms - Generative Adversarial Networks (GANs), CycleGAN, Harmonization, Brain Age Estimation

1. INTRODUCTION

Deep learning methods are now widely applied to brain MRI data for diagnostic classification, disease staging, and prognosis for a range of neurological and neurodevelopmental diseases. Even so, brain MRI protocols vary widely, and models trained on data from one scanner can fail when tested on data from a new site or protocol. To tackle this 'domain shift' problem, domain adaptation methods have been developed to adjust multisite brain MRI data to match a reference dataset or training data to facilitate machine learning on downstream tasks.

Two broad categories of domain adaptation methods have emerged: (1) adversarial methods that map source data into a site-invariant latent space [1, 2, 7], where features are optimized for the main task (e.g., disease detection) but also adapted to defeat an adversary that tries to predict which site the data came from; and (2) synthetic methods that also synthesize a new image to appear as if it came from another scanner, often using neural style transfer methods [4, 5]. Such reconstruction methods can also be extended to cross-modal data synthesis (e.g., simulating PET or CT scans from MRI) or for image enhancement with super-resolution [3].

Several GAN-based approaches have shown promising results for domain harmonization. For instance, Liu et al. [4] used style transfer methods to match brain MRI scans to a reference dataset. Sinha et al. [5] used attention-guided GANs for harmonization and demonstrated improvements in Alzheimer's disease classification with harmonized data. Dinsdale et al. [2] developed a deep learning-based model to remove dataset bias while improving performance on a downstream task of brain age prediction.

Dewey et al. [6] presented DeepHarmony, a UNET-based architecture that requires a paired dataset for training. Many works limit the harmonization training to 2D slices rather than the whole 3D volume [2, 5, 6]. Zuo et al. [7] developed CALAMITI, which uses information bottleneck theory to learn a disentangled latent space that contains both anatomical and contrast information, enabling controllable harmonization in a parametrized protocol space.

Here we introduce an unsupervised CycleGAN method for domain adaptation which does not need any paired data across domains. We use 3D convolutions in our model's generators and discriminators to harmonize full 3D MRI scans. We evaluate the harmonized scans by training machine learning models on a common benchmark downstream task (brain age estimation). We train downstream models for brain age estimation with harmonized and non-harmonized scans from five datasets. Our results illustrate improved brain age prediction after

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.15.516349; this version posted November 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

harmonization, suggesting that harmonization can improve the predictive performance of deep learning models in multisite predictive modeling.

2. DATASETS

Table 1 summarizes the datasets analyzed in our study. We used 5 datasets: UK Biobank [8], ADNI [9], AIBL [10], OASIS-1 [11], and WHIMS [12]. Apart from WHIMS, we used 1,000 samples from each dataset.

Dataset	N	C / P	Age (mean±SD)	F / M
UK Biobank	1,000	1000 / 0*	64.6 ± 7.6	517 / 483
ADNI	1,000	377 / 623	74.7 ± 7.1	558 / 442
AIBL	1,000	743 / 257	73.5 ± 6.7	518 / 482
WHIMS	941	564 / 377	70.0 ± 3.6	941 / 0*
OASIS-1	1,000	790 / 210	66.9 ± 9.4	559 / 441

Table 1. Statistics of the Training Dataset used for CycleGAN Training. N is the total number of samples. C, P, F, and M indicate the number of samples controls, patients, female, and male subjects, respectively. *WHIMS scans women only.

ADNI, AIBL, OASIS1, and WHIMS datasets include both controls and patient scans. The patients are diagnosed with Alzheimer's disease in ADNI, AIBL, and OASIS1 datasets, and with Parkinson's disease in the WHIMS dataset. UK Biobank has only controls. These datasets were used to train 4 CycleGAN models, mapping each of the four datasets (target datasets) ADNI, AIBL, OASIS, and WHIMS to (source dataset) UK Biobank and vice-versa.

All 3D T1-weighted MRI brain scans were pre-processed using a sequence of steps detailed in [13]. These included nonparametric intensity normalization (N4 bias field correction), skull stripping, 6 degrees-of-freedom registration to a template, and isotropic voxel resampling to 2 mm. Pre-processed images of size 91x109x91 were resized to 64x64x64, and intensities were linearly mapped to [0,1] using min-max normalization.

3. METHODS

Figure 1 summarizes our CycleGAN architecture. Similar to [14], it consists of two generators (G_X: $X \rightarrow Y$ and G_Y: $Y \rightarrow X$) and two discriminators (D_X and D_Y) for source domain X and target domain Y. With G_X, we want to generate an image from the target distribution given an image from the source domain distribution and *vice versa* with G_Y. We train the CycleGAN with an adversarial GAN loss and cyclic consistency loss. We use a patch-based discriminator and impose identity loss to regularize model training.

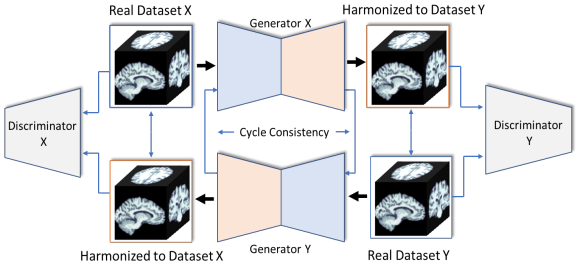


Fig. 1. CycleGAN Architecture. Two generators and discriminators are trained to harmonize scans from dataset X to match the protocol of dataset Y and *vice versa*.

3.1 Model Architecture

The generator first transforms the input using a 3D convolution layer with 32 output channels and then passes it to the downsampling block. The downsampling block consists of two 3D convolution layers with 64 and 128 output channels that project the 64x64x64 input to 16x16x16. This is followed by 5 residual blocks, each having two 3D convolution layers with 128 output channels, and residual connections are made between each block [15]. Next, this output is passed to the upsampling block, which has two 3D transposed convolution layers with 64 and 32-channel outputs. Like UNET, we concatenate the downsampling block's output with the upsampling blocks' input. Finally, we concatenate the output of the upsampling block with the input image and pass it through a convolution layer to compute the output. We use instance normalization and ReLU non-linearity for all layers. For all convolution layers, we use padding and stride of 1, and kernel size 3 except for upsampling and downsampling blocks, for which stride was 2. The network's output has the same dimensions as the input.

The discriminator uses a PatchGAN architecture [16, 17, 18] and has five 3D Convolution layers with 32, 64, 128, 256, and 1 output channel with kernel size 4. All layers except the first are followed by instance normalization and have a stride of 2, and all the layers use ReLU non-linearity except the last, which uses sigmoid activation. The output shape of the discriminator is 6x6x6 when the input dimension of the scan is 64x64x64.

3.2.1 Adversarial Loss

We used the standard adversarial training to train the generator to synthesize real images. Instead of using the negative log-likelihood objective, we used least-squares loss, which provided stable training and better results. We use separate discriminators for each domain to compute the adversarial loss. The job of D_x is to distinguish between samples from the source data distribution, P(X), and the output of G_Y, and similarly, D_Y distinguishes between P(Y) and the output of G_X. The overall adversarial loss is

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.15.516349; this version posted November 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

$$\mathcal{L}_{GAN}(G_X, G_Y, D_X, D_Y) = \mathbb{E}_{y\sim P(Y)}[\ ||D_Y(y) - 1||^2 + ||D_X G_y(Y)||^2\ ] + \mathbb{E}_{x\sim P(X)}[\ ||D_X(x)||^2 + ||D_Y G_X(X)||^2\ ]$$

3.2.2 Cycle-Consistency Loss

To reduce the space of possible mapping functions, we adopt the Cycle Consistency loss [14] which ensures that the two generators are cycle-consistent. That is, the scan translated to the target domain $G_X(x)$ and then translated back to the source domain - i.e., $G_Y(G_X(x))$ - must be mapped to the input from domain X. Effectively, we want $G_Y G_X$ to be identity mapping, i.e., $x \rightarrow G_X(x) \rightarrow G_Y G_X(x) \approx x$ and similarly for $G_x G_Y$. We enforce this with pixel-wise L_1 loss in both forward and backward directions, ensuring cycle consistency.

$$\mathcal{L}_{cyc}(G_X, G_Y) = \mathbb{E}_{x\sim P(X)}[|| G_Y G_X(x)) - x||_1] + \mathbb{E}_{y\sim P(Y)}[|| G_X G_Y(y) - y||_1]$$

3.2.3 Identity Loss

We further regularize the CycleGAN models by enforcing the identity loss similar to Taigman et al. [19], which showed that it helps to restrict the intensity range of the image for the image-to-image translation task. In the case of the grayscale scan, we noticed a slight improvement in the quality of the harmonized scans. The identity loss is computed as the pixel-wise L_1 loss between the output of the generator (e.g., G_X: $X \rightarrow Y$) and the corresponding input ($y\sim P(Y)$):

$$\mathcal{L}_{identity}(G_X, G_Y) = \mathbb{E}_{y\sim P(Y)}\ ||G_X(y) - y||_1 + \mathbb{E}_{x\sim P(X)}\ ||G_Y(x) - x||_1$$

3.2.4 Full Objective

We introduce additional hyperparameters, λ_1 and λ_2 to balance different loss terms. These are set to 10 and 0.1 during training. The overall objective loss for CycleGAN training is:

$$\mathcal{L}(G_X, G_Y, D_X, D_Y) = \mathcal{L}_{GAN}(G_X, G_Y, D_X, D_Y) + \lambda_1 \mathcal{L}_{cycle}(G_X, G_Y) + \lambda_2 \mathcal{L}_{identity}(G_X, G_Y)$$

The patch-based adversarial loss ensures that the generator synthesizes images from distributions similar to the original distribution. The cyclic consistency and identity loss provide regularization on the generators.

3.3 Training

We trained 4 CycleGAN models, with *X* representing the UK Biobank source dataset and *Y* representing each of the ADNI, AIBL, OASIS, and WHIMS target datasets.

Each model was trained on pre-processed scans from the source and target datasets. Both generators and discriminators are trained with the Adam optimizer [20], with a learning rate of 10^{-4} and batch size of 4. The model was trained for 100 epochs with a multi-step learning rate scheduler with a gamma of 0.1 and steps on 35 and 75 epochs. Overall, our model has 16 million (M) parameters. Each generator and discriminator has 5M and 3M parameters, respectively.

4. RESULTS

We evaluated our domain adaptation model (1) via *t*-SNE for visual analysis of the data distributions, and (2) by testing if data harmonization reduced error on the downstream task of brain age estimation.

4.1 *t*-SNE and Clustering

Figure 2 visualizes the results of the harmonization of the source and target domains. Scans from the target domain were harmonized to the source domain. All the scans were resized to 8x8x8 before passing to the *t*-distributed Stochastic Neighbor Embedding algorithm to yield a 2D output for each scan. **Figure 2** shows the *t*-SNE embeddings of the unharmonized source and target domain. Without harmonization, source and target data are readily distinguished: in ADNI, all source domain points are clustered near the center, while target domain points are scattered around the source cluster. WHIMS data initially clustered almost entirely within the UK Biobank reference data. The second row shows *t*-SNE results after harmonizing the target to the source domain; there are now no distinguishable clusters, and source and target distributions overlap. This result is consistent with the goal that the source of the scans is hard to distinguish once they are harmonized.

4.2 Brain Age Prediction

We tested how harmonization affects 'Brain Age' estimation, where a 3D CNN is trained on healthy controls' MRI data to predict a person's age from their scan. We used a model with 5 3D-Convolutional layers (output sizes: 32, 64, 128, 256, and 256), with a max pool in the first 3 layers. These layers were followed by LeakyReLU non-linearity and batch normalization. The last Conv. layer's output was flattened and passed through 2 linear layers.

The CycleGAN can translate source datasets to target and target to source as well. Thus, when pooling the training datasets for the downstream task, we experimented with harmonizing all the target data to the source (UK Biobank in our case) and harmonizing UK Biobank data to the target datasets. We report the MAE score for the transformed dataset (a mixed hold-out test set from the source and the target) after applying harmonization in each direction. **Table 2** compares models trained on the pooled source and target datasets (a) with *no harmonization* (B, for *baseline performance*), or (b) after harmonizing the pooled data to either the source (To Source) or target (To Target).

In **Table 2**, N represents the total number of control subjects, which are split 70:30 ratio for training and evaluation. The datasets used in **Table 2** contain samples that were used to train the unsupervised CycleGAN models. This gives insights into using the same samples for unsupervised harmonization as well as the downstream task. We also report the results of training and evaluation on the samples which were not seen by CycleGAN in **Table 3**

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.15.516349; this version posted November 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

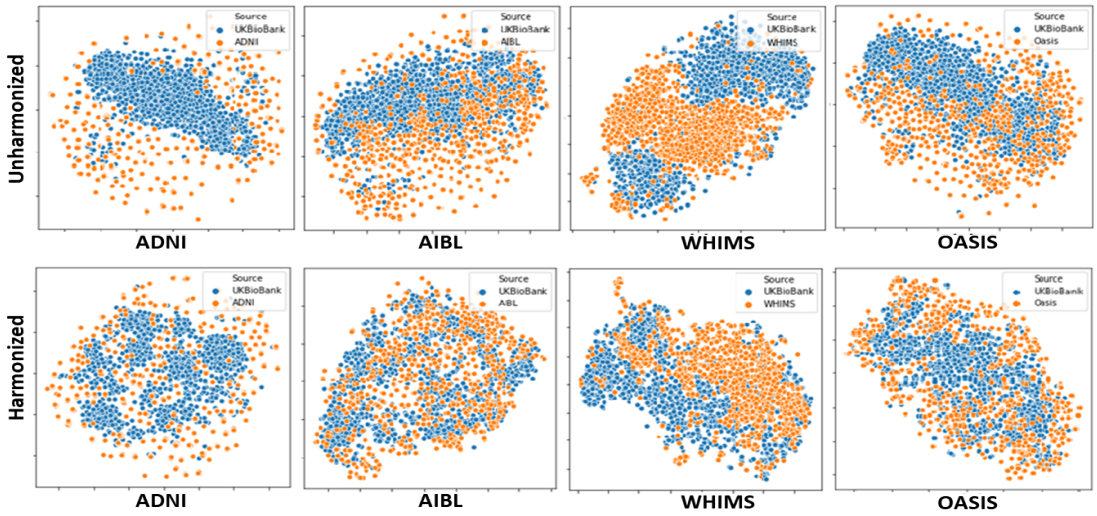


Fig 2. *t*-SNE visualization of harmonized (*bottom row*) and unharmonized (*top row*) scan from source and target domains. After CycleGAN harmonization, source and target scans are harder to distinguish in the *t*-SNE embedding

which gives us insights into the generalization capabilities of our harmonization approach.

	AIBL	ADNI	OASIS	WHIMS
S	72.9 ± 6.5	75.1 ± 6.5	65.37 ± 9.3	69.5 ± 3.6
N	743	377	790	564
B	4.557	3.437	5.109	3.856
To Source	**4.250**	3.772	5.487	3.645
To Target	4.551	**3.149**	**4.770**	**3.552**

Table 2. Brain Age Estimation on healthy controls used for training the unsupervised CycleGAN. S: subject's ages, N: sample size. In the AIBL column, for example, we take 743 AIBL controls and 743 UKB controls, mix them, and split them into 1040 training and 446 test samples. After training, we report the MAE on the test samples with no harmonization (B, *baseline*). Then we harmonize all scans to 'look like' UKB and report the test MAE (to Source) and harmonize all scans to look 'like AIBL' (to Target) and report the test MAE.

Table 2 shows improved brain age MAE on all 4 target domains after harmonization. We see similar trends with mean square error (MSE) but it is omitted due to space. For ADNI, WHIMS, and AIBL, we get better predictions when we translate from source to target when UK Biobank scans are harmonized to the target domain. For OASIS, harmonizing OASIS to UK Biobank was better. Visual insights after training at different numbers of epochs may be found on our GitHub repository (https://github.com/dheeraj-komandur/3D-CycleGAN-based-Harmonization).

Table 3 reports brain age results for ADNI and OASIS datasets on a held-out set not used for training the CycleGAN. Prediction error changes little between **Tables 2 and 3**, suggesting that training on harmonized data can generalize to unseen samples.

	ADNI		OASIS	
S	76.4 ± 6.4		65.4 ± 8.9	
N	1,000		742	
	Before	**After**	**Before**	**After**
MAE	4.387	3.929	4.344	4.339
MSE	30.539	25.428	29.780	31.231

Table 3. Brain age prediction results for ADNI and OASIS on a held-out set that were not used for training the CycleGAN. MAE is lower (better) after harmonization.

5. CONCLUSION

We trained four CycleGAN models for deep learning-based harmonization of multicohort MRI, using the UK Biobank as the source domain and ADNI, AIBL, OASIS, and WHIMS as targets. Each CycleGAN model was trained with a joint adversarial loss, cycle consistency, and identity loss. We visualized the positive effects of harmonization in overlaying distributions by performing *t*-SNE on data before and after harmonization; clusters in the data distributions, found initially, were indistinguishable after harmonization. Further, brain age estimation improved in controls across all target domains after harmonization. In the future, we will assess if GAN-based harmonization improves multisite predictive modeling in AD and MRI-based amyloid level prediction. We also aim to train a model for multisite domain harmonization instead of 1-to-1 source-to-target mapping.

bioRxiv preprint doi: https://doi.org/10.1101/2022.11.15.516349; this version posted November 15, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

6. COMPLIANCE WITH ETHICAL STANDARDS

This study uses previously collected, anonymized, de-identified data. All the anonymized datasets are publicly available, including the UK Biobank (access approval #11559).

7. ACKNOWLEDGEMENTS

This work is supported by NIH (U01AG068057, P01AG055367).

8. REFERENCES

[1] Moyer, D., Ver Steeg, G., Tax, C.M., & Thompson, P.M. (2020). Scanner invariant representations for diffusion MRI harmonization. *Magnetic Resonance in Medicine*, 84(4), 2174-2189.

[2] Dinsdale, N. K., Jenkinson, M., & Namburete, A. I. (2021). Deep learning-based unlearning of dataset bias for MRI harmonisation and confound removal. *NeuroImage*, 228, 117689.

[3] Ali, H., Biswas, R., Ali, F., Shah, U., Alamgir, A., Mousa, O., & Shah, Z. (2022). The role of generative adversarial networks in brain MRI: a scoping review. *Insights into Imaging*, 13(1), 1-15.

[4] Liu, M., Maiti, P., Thomopoulos, S., Zhu, A., Chai, Y., Kim, H., & Jahanshad, N. (2021). Style transfer using generative adversarial networks for multi-site MRI harmonization. *MICCAI 2021* (pp. 313-322).

[5] Sinha, S., Thomopoulos, S. I., Lam, P., Muir, A., & Thompson, P. M. (2021). Alzheimer's disease classification accuracy is Improved by MRI harmonization based on attention-guided generative adversarial networks. In *17th International Symposium on Medical Information Processing and Analysis* (Vol. 12088, pp. 180-189).

[6] Dewey, B. E., Zhao, C., Reinhold, J. C., Carass, A., Fitzgerald, K. C., et al. (2019). DeepHarmony: A deep learning approach to contrast harmonization across scanner changes. *Magnetic Resonance Imaging*, 64, 160-170.

[7] Zuo, L., Dewey, B. E., Carass, A., Liu, Y., He, Y., Calabresi, P. A., & Prince, J. L. (2021). Information-based disentangled representation learning for unsupervised MR harmonization. *In International Conference on Information Processing in Medical Imaging (IPMI)* (pp. 346-359).

[8] Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., et al. (2015). UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*, 12(3), e1001779.

[9] Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., et al. (2017). The Alzheimer's Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. *Alzheimer's & Dementia*, 13(5), 561-571.

[10] Ellis, K. A., Rainey-Smith, S. R., Rembach, A., Macaulay, S. L., Villemagne, V. L., & AIBL Research Group. (2013). Enabling a multidisciplinary approach to the study of ageing and Alzheimer's disease: An update from the Australian Imaging Biomarkers and Lifestyle (AIBL) study. *Int. Review of Psychiatry*, 25(6), 699-710.

[11] Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2007). Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cognitive Neuroscience*, 19(9), 1498-1507.

[12] Resnick, S. M., Espeland, M. A., Jaramillo, S. A., Hirsch, C., Stefanick, M. L., et al. (2009). Postmenopausal hormone therapy and regional brain volumes: the WHIMS-MRI Study. *Neurology*, 72(2), 135-142.

[13] Lam, P., Zhu, A. H., Gari, I. B., Jahanshad, N., & Thompson, P. M. (2020). 3D Grid-Attention Networks for Interpretable Age and Alzheimer's Disease Prediction from Structural MRI. *arXiv preprint* arXiv:2011.09115.

[14] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *In Proc. IEEE International Conference on Computer Vision* (pp. 2223-2232).

[15] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *In Proc. IEEE CVPR* (pp. 770-778).

[16] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *In Proc. IEEE CVPR* (pp. 1125-1134).

[17] Li, C., & Wand, M. (2016). Precomputed real-time texture synthesis with markovian generative adversarial networks. *In European Conference on Computer Vision* (pp. 702-716).

[18] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network. *In Proc. IEEE CVPR* (pp. 4681-4690).

[19] Taigman, Y., Polyak, A., & Wolf, L. (2016). Unsupervised Cross-Domain Image Generation. *In International Conference on Learning Representations.*

[20] Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *In International Conference on Learning Representations.*