Perceptually Improved T1-T2 MRI Translations Using Conditional Generative Adversarial Networks

Anurag Vaidya[a], Joshua V. Stough[a], and Aalpen A. Patel[b]

[a]Computer Science, Bucknell University, Lewisburg, PA;
[b]Imaging Science and Innovation, Geisinger, Danville, PA

ABSTRACT

Magnetic Resonance Imaging (MRI) encompasses a set of powerful imaging techniques for understanding brain structure and diagnosing pathology. Various MRI sequences including T1- and T2-weighted provide rich complementary information. However, significant equipment costs and acquisition times have inhibited uptake of this critical technology, adversely impacting health equity globally. To ameliorate these costs associated with brain MRIs, we present *pTransGAN*, a generative adversarial network (GAN) capable of translating both healthy and unhealthy T1 scans into T2 scans, thereby obviating T2 acquisition. Extending prior GAN-based image translation, we show that the addition of non-adversarial perceptual losses, like style and content loss, improves the translations provided, especially making the generated images sharper, and making the model more robust. Additionally in previous studies, separate models have been created for healthy and unhealthy brain MRI. Thus here, we also present a novel simultaneous training protocol that allows *pTransGAN* to concurrently train on healthy and unhealthy data sampled from two open brain MRI datasets. As measured by novel metrics that closely match perceptual similarity of human observers, our simultaneously trained *pTransGAN* model outperforms the models individually trained on just healthy or unhealthy data. These encouraging results should be further validated with independent paired and unpaired clinical datasets.

Keywords: Generative/adversarial learning, Image synthesis, Deep learning

1. INTRODUCTION

Recent techniques allow synthesis of multiple sequences of images from one k-Space sampling (the 2D or 3D Fourier transform of the MR image being measured). For example, a typical MRI sequence consists of T1-weighted sequence (favorable for observing large brain structures), T2-weighted sequence (useful for pathology), and T2-FLAIR scan (useful for pathology with suppression of signal from water) (Figure 1A). While these scans provide complementary and critical information (for example a tumor may not be seen in T1 but could show up in T2 scans), they also make the MRI a time consuming (exam times lasting between 45 minutes and 1 hour) and expensive imaging modality (average MRI cost is $2,600 in the US). With more than 40 million MRI scans done each year in the US alone, is there a way to take a T1 scan and use machine learning algorithms to predict the associated T2 scan, thus reducing acquisition times and improving MRI scanner throughput?[1]

Within modality translation, various approaches have been proposed. Rousseau et al.[2] proposed to pick nearest neighbors with similar properties from low-resolution images brain MRI and produce high-resolution MRI using Markov random fields. Regression forest[3] and convolutional network[4] based approaches have also been proposed. Recently, generative adversarial networks (GAN)[5] have gained prominence in the field of image translation. While the original GAN maps random noise to a target domain, a conditional GAN (cGAN) allows images from an input domain to be mapped to a target domain. Armanious et al.[6] and Dar et al.[7] have used cGAN for supervised image translation (images in the source and target domain come from the same subject). cGAN have shown promise for unsupervised translation (images in the source and target domains are unpaired).[8]

Current literature models performing MRI translation utilize loss functions which assume that pixels are independent,[7] thus not allowing higher level features, like texture, to be accurately translated. In this study,

Corresponding author: `anurag.vaidya@bucknell.edu, joshua.stough@bucknell.edu`
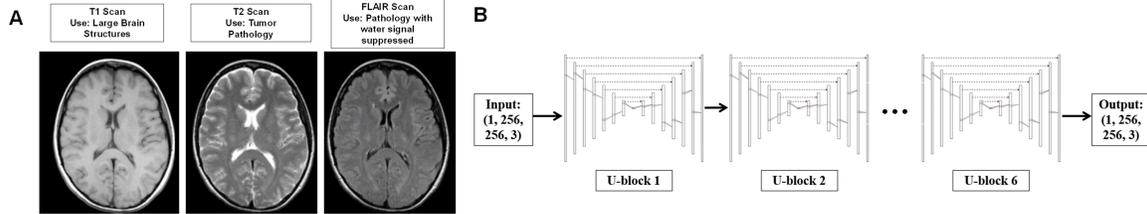
Figure 1. **A** A typical MRI sequence consists of T1-weighted, T2-weighted, and T2-FLAIR sequences. **B** 6 U-blocks are sequentially connected to create the generator of *pTransGAN* that progressively refines the generated images. The dotted lines are skip connections.

we present *pTransGAN*, a cGAN that minimizes individual pixel losses and non-adversarial losses to accurately transfer texture between domains. In the medical domain, translating both healthy and unhealthy images is critical. While individual algorithms can translate healthy and unhealthy algorithms separately,[7] no single algorithm can perform equally well translation of both healthy and unhealthy T1 scans, which limits the clinical use of such machine learning algorithms. Thus, we also present a novel simultaneous training protocol allowing *pTransGAN* to perform well on both healthy MRI images from the IXI dataset[9] and unhealthy MRI images from the Brain Tumor Segmentation Challenge Dataset[10] concurrently.

2. METHODS

Architectures: The *pTransGAN* model consists of a discriminator and a generator. The discriminator is a version of the PatchGAN[11] with a 70x70 receptive field that discriminates real pairs (real T1 and real T2) and fake pairs (real T1 and generated T2). Input images are concatenated channel-wise and passed through 6 convolutions with 64, 128, 256, 512, 512, and 1 spatial filters with kernel size of 4x4. The stride for the first four convolutions is 2, and stride for the rest is 1. The convolution layers are followed by batch normalization and Leaky-ReLU ($\alpha = 0.2$). Sigmoid activation is applied to the output to get a probability map. The generator of *pTransGAN* consists of 6 sequentially connected U-blocks. The U-block's down-sampling path consists of seven convolutions with filters 64, 128, 256, 512, 512, 512, 512. The bottleneck consists of 512 filters and a ReLU activation. The decoding path has seven convolutions with filters 512, 1024, 1024, 1024, 512, 128, and 64. All convolutions have stride of 2, kernel of 4x4, and are followed by batch normalization and LeakyReLU ($\alpha = 0.2$). In order to reduce overfitting, the first three layers of the decoding path have a dropout layer associated with them (rate=0.5). All of the layers were initialized with a Glorot initializer.[12] The U-block also includes skip connections that connects the corresponding encoding and decoding channels. In this study, all models were developed using TensorFlow.

The discriminator trains on an adversarial and a pixel reconstruction loss, whereas the generator trains on non-adversarial losses, namely the style and content loss: $\mathcal{L}_{pTransGAN} = \lambda_{cGAN}\mathcal{L}_{cGAN} + \lambda_{L1}\mathcal{L}_{L1} + \lambda_{style}\mathcal{L}_{style} + \lambda_{content}\mathcal{L}_{content}$. The λ weigh the different loss components and were set to $\lambda_{cGAN} = 1$,[11] $\lambda_{L1} = 100$,[11] $\lambda_{style} = 1e-3$, $\lambda_{content} = 1e-5$. The adversarial loss allows *pTransGAN* to create plausible translations of T1 images in the T2 domain whereas the pixel reconstruction loss (an L1 loss in this study), reduces the pixel to pixel error. The implementation of adversarial and L1 losses were the same as Armanious et al.[6] and Isola et al.[11] The style and content loss use internal feature map activations of a pretrained VGG-19.[13] To calculate the stylistic features of an image, an image is passed to VGG-19 and then Gram matrix of the feature maps of the first layers of blocks 2, 3, and 4 is calculated and is normalized by the dimensions of the feature map. The style loss is the square of the Frobenius norm of the difference between the feature correlations of the generated image, $\hat{y}$, and ground truth target image, y, over all the selected convolutional blocks:[6] $\mathcal{L}_{style} = \sum_{i=1}^{Total\ Blocks} \lambda_{style,\ i} * \frac{1}{4d_i^2} * ||Gram_i(y) - Gram_i(\hat{y})||_F^2$. Here, $\lambda_{style,i} > 0$, weighs contribution of the ith convolutional block to the overall stylistic loss, were set so that the 2nd, 3rd, and 4th convolution blocks had the greatest contributions (in that order). The content loss is defined as the Frobenius norm of the differences in the feature maps of the first layers of all convolutional blocks:

$\mathcal{L}_{content} = \sum_{i=1}^{Total\ Blocks} \lambda_{content,\ i} * || F_i(y) - F_i(\hat{y})||_F^2$. $\lambda_{content,i} > 0$ were set so that all but the last convolution block had equal contributions and the last block had 1/10 the contribution as the others. Addition of non-adversarial losses destabilizes the training of the discriminator, thus spectral normalization was applied to the weights of the discriminator.[14] Through grid search The learning rate of the generator was found to be 0.0002 whereas that of the discriminator was set to 0.0008.[15]

Data: Healthy and unhealthy paired datasets (T1 and T2 MRI belong to the same patient) were used. Healthy scans were acquired from the IXI dataset, which contained non-skull stripped axial T1 and T2 MRI for 577 patients. 461 scans were randomly chosen for training, 58 for hyperparameter optimization, and 58 for testing. Each scan had 180 slices in the axial direction but only the middle 11 were chosen. The images were re-sampled to (1,1,1) spacing and were reordered to be closest to canonical (RAS+) orientation. The MNI mask[16] was then applied to the images so that all images have the same size. For unhealthy MRI, the BRaTS2020 dataset was used. The dataset had 494 patients with brain tumors. The T1 and T2 images were skull stripped and registered. Data from 369 patients were reserved for training purposes and 62 patients were kept for testing. Each patient has 150 scans in the axial direction; 14 scans from the middle of the brain were used for each patient. No data from the unhealthy dataset was used for hyperparameter optimization. All images were sized to 256x256x3 pixels.

Experiments: In this study, three experiments were carried out. Firstly, to see the effect of non-adversarial losses, three variations $pTransGAN$ model were trained and tested on healthy data with different loss configurations: (A) adversarial and L1 loss, (B) adversarial, L1, and style loss, (C) adversarial, L1, style, and content loss. Models (A) and (C) were also tested on unhealthy data. In a second experiment, $pTransGAN$ with adversarial and non-adversarial losses was trained and tested on unhealthy data. This model was also tested on the healthy data to see if one model could be used for both domains. In a final experiment, the authors aimed to test a simultaneous training protocol that could train a single model for both the healthy and unhealthy domains. For each iteration of simultaneous training, a $pTransGAN$ with adversarial, L1, style, and content loss was trained three times on a healthy image and then three times on an unhealthy image, which is then repeated for all training images. In order to compare the generated original images, six different metric were used. These comprised of three traditional metrics that have been used in numerous studies (Peak Signal to Noise Ratio or PSNR,[6] Structural Similarity Index or SSIM,[17] and Mean Squared Error or MSE [6]). These metrics have previously failed to match human perceptual levels,[18] ,[?] so three novel metrics were also used (Learned Perceptual Image Patch Similarity or LPIPS ,[18] Universal Quality Index or UQI ,[17] and Visual Information Fidelity or VIF [19]). Comparison of metrics was performed via single tailed paired Wilcoxon signed rank test.[20] All models were trained for 100 epochs.

3. EXPERIMENTAL RESULTS

The first experiment trained and tested $pTransGAN$ on healthy data with three loss configurations to determine how the addition of non-adversarial losses would affect the translations. The addition of style and content loss showed small improvements in the image comparison metrics (Table 1). The models with perceptual losses outperformed the baseline models in the SSIM and LPIPS metrics (p-value ¡ 0.001). Since the LPIPS loss is significantly lower for models with perceptual losses, this indicates that their image translations are likely to be more perceptually similar to ground truth images for a human observer since LPIPS most closely matches human visual similarity (Figure 2). With the addition of perceptual losses, we noticed that the translated features were sharper and more distinctive in the generated images (Figure 2). Moreover, when the models with and without perceptual losses trained on healthy data were tested on the unseen unhealthy data, it was found that the model with perceptual losses outperformed the model without perceptual losses (Table 1). For example, the SSIM metric increased by 32.7% and the LPIPS metric decreased by 36% with the addition of non-adversarial tests.

However, the translations of unhealthy images by $pTransGAN$ trained on healthy images are not satisfactory. When another $pTransGAN$ model (with adversarial and non-adversarial losses) was trained and tested on unhealthy data, it was found that the translation quality significantly improves (Table 1, Supplemental Figure1). $pTransGAN$ is capable of accurately producing the tumor shape and boundary as well as the healthy portions of the MRI. Nevertheless, when this model was tested on healthy data, the performance of the model worsened significantly. The $pTransGAN$ model trained concurrently on healthy and unhealthy datasets through the novel

Table 1. Metrics for *pTransGAN* trained and tested on healthy and unhealthy data (s: style loss, c: content loss, sim: Simultaneous training). Adversarial (adv) and L1 losses are used in all experiments.

| Train set	Healthy	Healthy	Healthy	Healthy	Unhealthy	Sim.	Sim.
Test set	Healthy	Healthy	Unhealthy	Unhealthy	Unhealthy	Healthy	Unhealthy
Loss	adv, L1	+ s, c	adv, L1	+ s, c	+ s, c	+ s, c	+ s, c
PSNR (dB)	24.7	24.9	17.9	18.2	19.5	24.1	19.9
SSIM	0.875	0.883	0.572	0.685	0.831	0.851	0.825
MSE	0.003	0.004	0.020	0.017	0.012	0.004	0.011
LPIPS	0.061	0.060	0.256	0.197	0.113	0.065	0.132
UQI	0.941	0.952	0.501	0.551	0.835	0.885	0.862
VIF	0.882	0.875	0.778	0.601	0.752	0.821	0.762
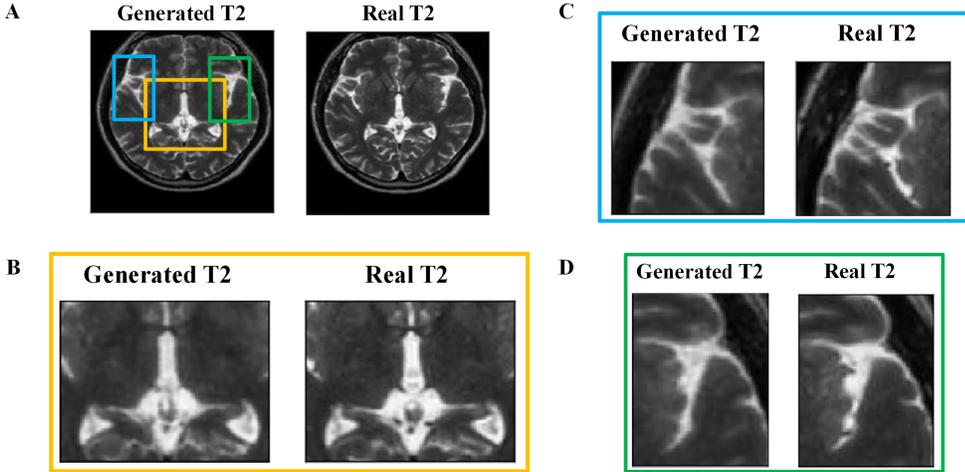
Figure 2. **A** An example original and generated healthy T2 scan produced by *pTransGAN* trained on healthy data. **B, C, D** Three zoomed in versions of the generated and original T2 scans. The *pTransGAN* is able to translate global image properties as well as minute anatomical features.

simultaneous training protocol, however, performed equally well on the healthy and unhealthy datasets (Table 1). In some metrics, for example PSNR and UQI, the simultaneous model was able to outperform *pTransGAN* models trained individually on healthy and unhealthy data.

4. DISCUSSION AND CONCLUSION

In this study, we present an end-to-end conditional generative adversarial framework, *pTransGAN*, that is capable of translating both healthy and unhealthy T1 brain MRI into T2 MRI with high fidelity. In addition to using adversarial losses for training, our framework leverages non-adversarial perceptual losses, specifically style and content losses, to create sharper translated images. The non-adversarial losses also improve the robustness and generalizability of *pTransGAN* in other domains. Overall, *pTransGAN* was able to translate global structures and accurate minute anatomical features between the healthy and unhealthy T1 and T2 domains. The image comparison metrics (Table 1) are comparable to those found by Dar et al.[6] A simultaneous training protocol was designed so that a single *pTransGAN* model could perform well on both healthy and unhealthy MR scans. *pTransGAN* requires further improvements before it can be used in a clinical setting. Firstly we aim to adapt *pTransGAN* to 3D data since a lot of brain imaging data is 3D and volumetric. Secondly, the unhealthy dataset used in this study is limited to brain tumors, so to increase *pTransGAN*'s robustness, we aim to train it on a more diverse dataset. Finally, the image registration has shown to be critical in image translation.[7] In future studies, both healthy and unhealthy datasets would be registered against a common brain MRI mask.

REFERENCES

[1] Stewart, C., "Mri units density by country 2019," (Jul 2021). `https://www.statista.com/statistics/282401/density-of-magnetic-resonance-imaging-units-by-country`.

[2] Rousseau, F., "Brain hallucination," in [*European Conference on Computer Vision*], 497–508, Springer (2008).

[3] Jog, A., Roy, S., Carass, A., and Prince, J. L., "Magnetic resonance image synthesis through patch regression," in [*2013 IEEE 10th International Symposium on Biomedical Imaging*], 350–353, IEEE (2013).

[4] Van Nguyen, H., Zhou, K., and Vemulapalli, R., "Cross-domain synthesis of medical images using efficient location-sensitive deep network," in [*International Conference on Medical Image Computing and Computer-Assisted Intervention*], 677–684, Springer (2015).

[5] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., "Generative adversarial nets," *Advances in neural information processing systems* **27** (2014).

[6] Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., Gatidis, S., and Yang, B., "Medgan: Medical image translation using gans," *Computerized Medical Imaging and Graphics* **79**, 101684 (2020).

[7] Dar, S. U., Yurt, M., Karacan, L., Erdem, A., Erdem, E., and Çukur, T., "Image synthesis in multi-contrast mri with conditional generative adversarial networks," *IEEE transactions on medical imaging* **38**(10), 2375–2388 (2019).

[8] Welander, P., Karlsson, S., and Eklund, A., "Generative adversarial networks for image-to-image translation on multi-contrast mr images-a comparison of cyclegan and unit," *arXiv preprint arXiv:1806.07777* (2018).

[9] "Ixi dataset."

[10] "Brain tumor segmentation (brats) challenge 2020: Scope."

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

[12] Glorot, X. and Bengio, Y., "Understanding the difficulty of training deep feedforward neural networks," in [*Proceedings of the thirteenth international conference on artificial intelligence and statistics*], 249–256, JMLR Workshop and Conference Proceedings (2010).

[13] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556* (2014).

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

[15] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S., "Gans trained by a two time-scale update rule converge to a local nash equilibrium," *Advances in neural information processing systems* **30** (2017).

[16] Evans, A. C., "An mri-based stereotactic atlas from 250 young normal subjects," *Soc. neurosci. abstr, 1992* (1992).

[17] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing* **13**(4), 600–612 (2004).

[18] Zhang, R., Isola, P., Efros, A. A., Shechtman, E., and Wang, O., "The unreasonable effectiveness of deep features as a perceptual metric," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 586–595 (2018).

[19] Sheikh, H. R. and Bovik, A. C., "Image information and visual quality," *IEEE Transactions on image processing* **15**(2), 430–444 (2006).

[20] Pratt, J., "Remarks on zeros and ties in the wilcoxon signed rank procedures," *J. American Statistical Association* **54**, 655–667 (1959).
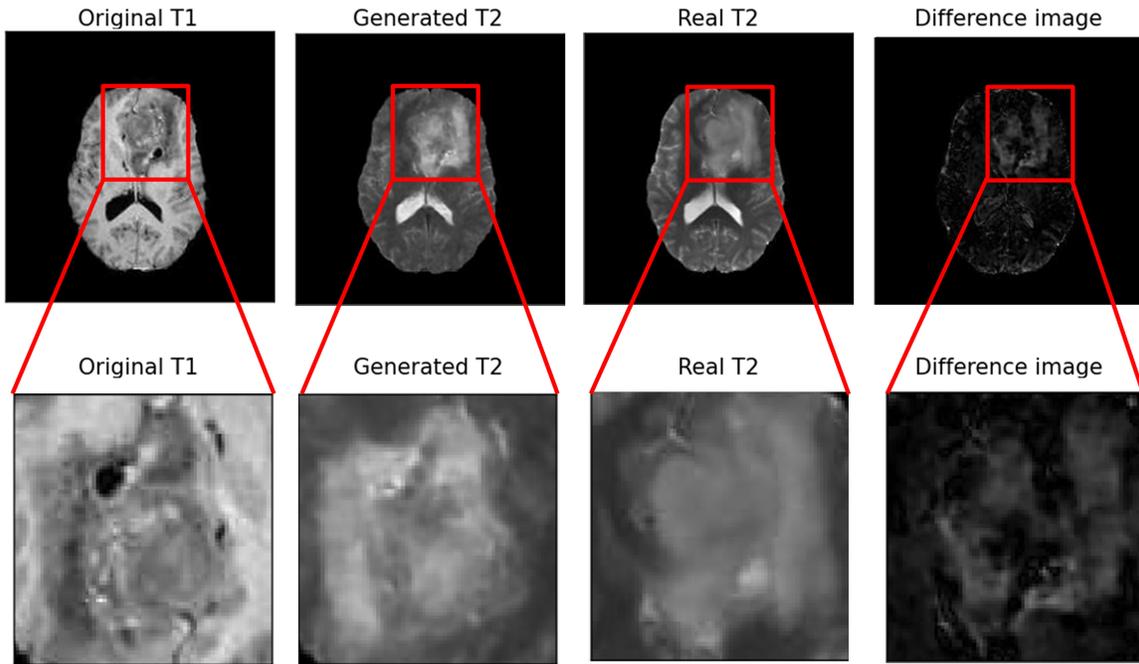
5. SUPPLEMENTAL FIGURES

Figure 1.   An example original and generated unhealthy T2 scan produced by *pTransGAN* trained on unhealthy data. The difference image displays the mean absolute error between the original and generated T2 images. The zoomed in images also show how the tumor boundary, shape, and texture are accurately recreated,