

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Leveraging large scale data sets: a transfer learning approach for 7T super resolution

Permalink

<https://escholarship.org/uc/item/1jc4w2x1>

Author

Ellis, Ryan

Publication Date

2021

Peer reviewed|Thesis/dissertation

Leveraging large scale data sets: a transfer learning approach for 7T super resolution

by
Ryan Ellis

THESIS
Submitted in partial satisfaction of the requirements for degree of
MASTER OF SCIENCE

in
Biomedical Imaging

in the
GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

Duygu Tosun

Duygu Tosun

A2C3822F2671486...

Chair

DocuSigned by:

Srikantan Nagarajan

Srikantan Nagarajan

DocuSigned by:

Peder Larson

Peder Larson

DocuSigned by:

Ashish Raj

Ashish Raj

13BDA73A066E4E0...

Committee Members

Acknowledgements

I would like to thank Dr. Duygu Tosun-Turgut for her guidance and support, especially for this work, but also for the time I have spent working in her lab. She has been a role model and inspiration, and I am incredibly grateful for her mentorship. I would also like to thank my committee members Dr. Srikantan Nagarajan, Dr. Peder Larson, and Dr. Ashish Raj for their helpful feedback that guided this work.

Abstract

Brain morphometry on data from multi-scanner and multi-site studies can suffer from non-biological variance due to scanner and acquisition differences. Harmonization methods, such as ComBat, have been introduced to remove unwanted variance in structural neuroimaging data. Statistical methods for harmonizing structural data however operate on derived morphological measurements to remove site related effects rather than operating at the voxel level to remove scanner related effects. This study works towards a deep learning-based image harmonization method by training and evaluating a generative adversarial model for transforming 3T images to a standard 7T-like image quality. 7T MRI can achieve better tissue contrast and tissue segmentation results but lacks the widespread availability of 3T MRI, resulting in limited dataset sizes for deep learning. Transfer learning from a 3T synthesis task to a 7T synthesis task was hypothesized to improve synthesis results by greatly increasing dataset size and diversity with multi-site longitudinal data. The 7T synthesis dataset was comprised of 9 subjects each with a 3T MPRAGE and 7T MP2RAGE T1-weighted scan. Leave one out cross validation was used and performance evaluation metrics were reported as the mean across all validation cross folds. The transfer learning dataset consisted of 419 total subjects and 1124 T1 weighted images with a wide variety of sites, scanners, and acquisition sequences. An independent testing set of 17 subjects with paired 1.5T and 3T scans from the transfer learning dataset were used for evaluating the 3T synthesis task. Image similarity metrics such as Structural Similarity Index Measure (SSIM) and Peak Signal to Noise Ratio (PSNR) were used to evaluate synthesis performance. Dice Similarity Coefficient (DSC) and Jaccard Similarity Coefficient (JSC) were used to evaluate the synthesized and 3T segmentations results using 7T segmentation as ground truth. The 7T synthesis network with transfer learning weights achieved an SSIM of 0.950 ± 0.02 and PSNR of 25.44 ± 0.61 , improved over the 3T image which had SSIM of 0.909 ± 0.01 and PSNR of 21.83 ± 0.92 . The DSC for grey matter regions of interest was 0.810 ± 0.02 and $0.916 \pm$

0.004 for white matter regions of interest for the synthesized validation images, an improvement of 0.053 DSC ($p = 0.011$) and 0.017 DSC ($p = 0.0039$) over the 3T results respectively. The JSC for grey matter regions of interest was 0.693 ± 0.03 and 0.842 ± 0.01 for white matter regions of interest, an improvement of 0.066 DSC ($p = 0.011$) and 0.026 DSC ($p = 0.0039$) respectively. Future work will evaluate the ability of the 7T synthesis models at removing non-biological variance, particularly in longitudinal studies where imaging protocol or scanners were updated.

Table of Contents

Chapter 1: Introduction.....	1
Chapter 2: Methods.....	3
Chapter 3: Results.....	10
Chapter 4: Discussion.....	15
References.....	18
Appendix I.....	22
Appendix II.....	23
Appendix III.....	24
Appendix IV.....	25

List of Figures

Figure 2.1.....	8
Figure 3.1.....	10
Figure 3.2.....	10
Figure 3.3.....	10
Figure 3.4.....	11
Figure 3.5.....	14
Figure 3.6.....	14
Figure 3.7.....	15
Figure 3.8.....	15

List of Tables

Table 2.1.....6
Table 2.26
Table 3.1.....12
Table 3.2.....12
Table 3.3.....13

Chapter 1: Introduction

Quantifying measures of morphometric changes in the brain from structural neuroimaging data is of particular interest for studying neurodegenerative diseases. Our understanding of the natural history of normal and pathological morphological brain changes can be improved by leveraging multiple datasets from multi-site longitudinal imaging initiatives. However, non-biological variation may be introduced due to scanner and imaging protocol variability across different sites [1]. Longitudinal studies are particularly affected as imaging protocol, scanner hardware, and scanner software can be updated overtime which further introduce variability in morphometry results. Statistical methods, such as ComBat, have been shown to harmonize multi-site structural neuroimaging data by operating on morphological measurements (i.e. volume, cortical thickness, surface area) derived from automated algorithms to remove non-biological variance due to site differences [2-5]. Site based harmonization techniques correct for variation due to dataset source rather directly addressing the cause of measurement differences due to ‘scanner effects’, such as acquisition sequence, scanner model, and vendor differences. Wrobel et al. proposed an intensity normalization method ‘mica’ to reduce scanner effects directly at the image level by warping image contrast instead of harmonizing data on a covariate such as site. Mica however needs a baseline measurement of variability using a subset of subjects scanned at all sites which may not be possible for all studies [6]. Image level harmonization methods that generalize to unseen data using deep learning have also been proposed, removing the need for traveling subjects. [7-10]. Dewey et al. uses a U-Net based architecture to standardize structural neuroimages to a reference protocol to improve the consistency of volume quantification between protocols [7]. Zuo et al. performs MRI harmonization using unsupervised domain adaption, which adopts a model to unseen data during testing, by using encoder and decoder networks to transfer contrast differences between protocols while retaining anatomical structure [8]. These approaches demonstrate promising deep learning harmonization results, but they require multiple contrasts per subject and the selection of a reference site or protocol, which adopts contrast of a specific site rather than removing variation. Instead of adapting MRI data to a specific site, harmonization may

instead be accomplished through by transforming image data to a tissue contrast more representative of the ground truth anatomy. Ultra-high field 7T MRI can achieve higher signal to noise ratio and better contrast, which has been shown to improve brain parcellation results especially in superior grey matter and subcortical grey matter regions [11, 12]. Synthesizing 7T MRI structural neuroimages from lower field strengths has been the goal of many works using deep learning methods [13-16]. Qu et al. performed 7T super resolution by using an encoder-decoder CNN that injected features from the wavelet domain into the encoder network and achieved Structural Similarity Index (SSIM) of 0.87 and Peak Signal to Noise Ratio (PSNR) of 28.27 [13]. Zhang et al. used dual convolutional neural networks operating on spatial and frequency information to synthesize 7T contrast from 3T images [14]. These works use non-adversarial convolutional neural networks that can train only on paired 3T and 7T images and are limited to image-based loss functions such as mean squared error or mean absolute error. The model architecture used in our approach was based on Cycle-Consistent Adversarial Networks (CycleGAN) which allows for unpaired image to image translation and uses adversarial loss to train competing networks to generate synthesized examples that are indistinguishable from the real inputs [17]. CycleGAN has been widely used on medical image synthesis tasks because the input and target domain training images do not have to be paired or aligned and can come from different subjects [18]. Ali et al. used a CycleGAN to harmonize clinical multi-site MRI data which improved low grade glioma subtype prediction [19]. Do et al. use CycleGAN to perform 3T and 7T super resolution, however their task is to recover the original resolution from downsized 3T and 7T images separately rather than transforming 3T images to 7T-like contrast [20]. The goal of this research is to evaluate the ability of generative adversarial deep learning models to transform 3T T1 weighted neuroimages to a standardized 7T-like image quality as a step towards deep image harmonization. Deep learning methods typically require many examples to learn generalizable features, but studies for 3T to 7T synthesis are hampered by a lack of clinical biomedical imaging data availability from 7T MR systems [21, 22]. Transfer learning is hypothesized to improve the shortcomings of limited dataset size and diversity of the 7T dataset by reusing the knowledge learned from a 3T synthesis task. A review by Valverde et al. found a widespread application of transfer learning to benefit

deep learning tasks in brain MRI, such as brain tumor segmentation and disease classification [23]. First, CycleGAN was trained to synthesize 3T contrast from 1.5T images using data from a large multi-site study to develop a robust feature extractor. A 7T synthesis model was then trained reusing the feature extractor of the 3T synthesis model. We assessed the performance of 7T synthesis models using image similarity metrics such as Structural Similarity Index (SSIM) and Peak Signal to Noise Ratio (PSNR) to compare our results with related works. Dice similarity coefficient and Jaccard similarity coefficient of ROI segmentations were used to validate the ability of the 7T synthesis models to reproduce segmentation results from the ground truth 7T images.

Chapter 2: Methods

2.1 Datasets

Two separate datasets were used, one for the 7T synthesis task and the other for transfer learning. The dataset for the 7T synthesis task consisted of 9 subjects with a 3T scan and a 7T scan. The 3T scans were taken with a Siemens Magnetom Skyra 3T scanner using an T1 weighted MPRAGE sequence with 0.8 mm³ isotropic resolution (TR=2400 ms, TE=2.24 ms, TI=1060 ms, Flip angle=8 deg, GRAPPA acceleration factor=2). The 7T scans were taken with a Siemens Magnetom 7T scanner using an T1 weighted MP2RAGE sequence with 0.7 mm³ isotropic resolution (TR=4500 ms, TE=3.37 ms, TI1 = 900 ms, TI2 = 2750 ms, Flip angle 1=5 deg, Flip angle 2=3 deg, GRAPPA acceleration factor=3). The 7T synthesis dataset consisted of 7 males and 2 females, had an average age of 41 +/- 11 years, and all subjects had a diagnosis of mild traumatic brain injury (TBI).

The dataset for the transfer learning task was sampled from the Alzheimer's Disease Neuroimaging Initiative (ADNI), a longitudinal multi-site study where imaging protocol was updated over time [24, 25]. The images were T1-weighted with field strengths of 1.5T or 3T and were taken from 60 sites with 24 unique scanner models taken over the span of 14 years. The acquisition sequences consisted of MPRAGE, IR-SPGR, and accelerated MP-RAGE (GRAPPA and SENSE). Participants in the ADNI study were screened for significant neurological disease and disorders other than Alzheimer's Disease (AD) [24]. Images were sampled from healthy control subjects by screening for subjects with a

cognitively normal diagnosis to avoid introducing an AD-specific task bias into the transfer learning dataset [9, 26]. A paired dataset of 35 control subjects with scans at both 1.5T and 3T was used to validate and test the model trained for the transfer learning task. The remaining dataset of unpaired 1.5T and 3T scans was used for training and consisted of with 783 3T images (335 unique individuals) and 271 1.5T images (55 unique individuals). The unpaired training dataset contained longitudinal scans and had no overlap between the 1.5T and 3T subjects. The average age of the ADNI dataset was 74 +/- 6.8 years and was 47% female.

2.2 Image processing

The paired datasets for both the 7T synthesis and transfer learning datasets were first aligned to a common space through affine registration. 3T and 7T images of the 7T synthesis dataset were registered to a 7T reference image with 0.7 mm³ isotropic resolution. 1.5T and 3T images in the transfer learning dataset were registered to a reference 3T image with 1 mm³ isotropic resolution. Registration was performed with FSL FLIRT using affine registration (9 degrees of freedom) and a mutual information cost function [27]. The registered brain images from the 7T synthesis dataset were skull stripped using FSL BET, an automated tool for brain extraction [28]. The resulting brain masks were manually edited on 7T images to provide accurate results. The images from the ADNI dataset were skull stripped using Multi-atlas Skull Stripping (MASS), a robust automated skull stripping tool. MASS handles variations in contrast by performing non-linear registrations to several atlases and combining the results with a spatial voting algorithm [29]. The resulting automated masks for the transfer learning dataset were visually inspected but not manually edited. The skull stripped 3T images of the 7T synthesis dataset were non-linearly registered to the 7T images using Advanced Normalization Tools (ANTs) Symmetric diffeomorphic algorithm to correct for geometric distortions due to susceptibility differences. MUSE, a multi-atlas anatomical parcellation algorithm, was used to delineate anatomical regions of interest (ROIs) for the original 7T, geometry distortion adjusted 3T, and synthesized 7T images. MUSE generates robust ROI segmentation results by nonlinearly registering 11 selected atlases to input and fusing the results with a spatially adaptive weighting [30].

2.3 Model Architecture

The model framework was based on Cycle-Consistent Adversarial Network (CycleGAN) which consists of two generative adversarial networks (GANs). CycleGAN performs image transformation by simultaneously training both a forward mapping GAN and backward mapping GAN [17]. In this work, the forward mapping GAN transforms 3T images to 7T-like images while the backward mapping GAN transforms 7T images to 3T-like images. The GANs were made up of two types of networks: a generator model for producing synthetic images and a discriminator model for classifying the generated images as real or synthetic.

The generator model is a 2D convolution network that has 3 input convolutions, 9 residual blocks, and 3 output convolutions. Details on the convolution filter parameters are provided in Table 2.1. The convolutions in the encoder part of the generator have a stride of 2, meaning that the convolutional filters skip 2 pixels during the forward pass which decreases the spatial feature map size by $\frac{1}{4}$. The depth of the convolutional filters doubles each time the feature map size is reduced. The residual blocks are made up of two convolutional filters which are each followed by instance normalization and RELU activation. The residual blocks contain a skip connection, where the input feature maps to the residual block are pointwise added to the output feature maps of the residual block. The decoder part of the generator consists of three transposed convolutions with a stride of 2, which up sample the feature maps to the original input size while decreasing the depth of the convolutional filters. The output of the generator is a single channel image with the same height and width as the input image. The generator model has a total of 11.4 million trainable parameters.

Table 2.1: Generator Network Architecture details.

Layer	Output Size	Output Depth	Filter Size	Stride
Input Conv	128 x 128	64	7 x 7	1 x 1
Stride Conv 0	64 x 64	128	3 x 3	2 x 2
Stride Conv 1	32 x 32	256	3 x 3	2 x 2
Residual Block 1-9	32 x 32	256	3 x 3	1 x 1
Transpose Conv 0	64 x 64	128	3 x 3	2 x 2
Transpose Conv 1	128 x 128	64	3 x 3	2 x 2
Output Conv	128 x 128	1	7 x 7	2 x 2

The discriminator model is a simple fully convolutional network with strided convolutions that down sample the feature maps throughout the forward pass. All convolutions of the discriminator network were followed by instance normalization and RELU activation. Table 2.2 contains implementation details of the discriminator network. The output prediction of the discriminator network was obtained by global average pooling of the final convolutional features, which produces a single scalar value. The discriminator model had 2.8 million trainable parameters.

Table 2.2: Discriminator Network Architecture details.

Layer	Output Size	Output Depth	Filter Size	Stride
Conv 0	128 x 128	64	4 x 4	2 x 2
Conv 1	64 x 64	128	4 x 4	2 x 2
Conv 2	32 x 32	256	4 x 4	2 x 2
Conv 3	16 x 16	512	4 x 4	2 x 2
Conv 4	8 x 8	512	4 x 4	2 x 2
Average Pooling	1	1	4 x 4	1 x 1

2.4 Loss Function

$$Loss = L_{Generator} + L_{Discriminator} + \alpha L_{cycle} + \beta L_{identity}$$

The loss function consisted of four terms: generator loss, discriminator loss, cycle consistency loss, and identity loss. Alpha and beta are hyperparameters for weighting cycle loss and identity loss and were set to 10 and 5 respectively. Generator loss is given by $L_{Generator} = (D(G(X)) - 1)^2$, where the objective is to minimize the least squares loss of the discriminator (D) prediction on the synthesized output of the generator (G(X)) where X is the input image. This objective trains the generators to produce synthetic images that the discriminator network classifies as real images. The generator loss is the same for both the

forward and backward generators. The objective of the discriminators is to minimize the least squares loss between their output predictions and the labels of the input image, where real images are labeled as 1 and synthetic images are labeled as 0. The discriminator networks are trained on a queue of the last 50 generated images rather than the generated images of the immediate batch to promote training stability.

$$L_{Discriminator} = (D(X) - 1)^2 + (D(G(X)))^2$$

Cycle consistency is enforced by minimizing the mean absolute error of the reconstructed input images after they are fed through their respective forward and backward generators, where $L_{cycle} =$

$|G_A(G_B(A), A) + |G_B(G_A(B)), B|$. The cycle consistency step is as follows: a 3T input image is fed through the 7T generator which produces a synthetic 7T image. The synthetic 7T image is then fed through the 3T generator to reconstruct the original input contrast. The cycle consistency loss is also computed for the 3T input image.

Identity loss penalizes the generator networks for making unnecessary changes, given by $L_{identity} = |G_A(A), A| + |G_B(B), B|$ where the mean absolute error is taken between the input images and their reconstructed results.

2.5 Transfer Learning

The weights from encoder of the 1.5T generator (backward mapping for 3T synthesis task) were used to initialize the 7T generator (forward mapping for 7T synthesis task). The transferred weights in the 7T generator were frozen so they were not updated while training the on the 7T dataset. The 1.5T and 7T generators both take 3T images as input but the 1.5T generator was trained on 3T images from a large multi-site dataset. The transfer learning model was trained on the large ADNI dataset for the task of 1.5T to 3T synthesis. 18 subjects of the paired dataset (where subjects had a 1.5T and 3T scan) was used for model validation while the other 17 were used for final testing. The epoch with the lowest GAN loss ($L_{Generator} + L_{Cycle} + L_{identity}$) on the validation set was chosen as the model for final testing and weight transfer.

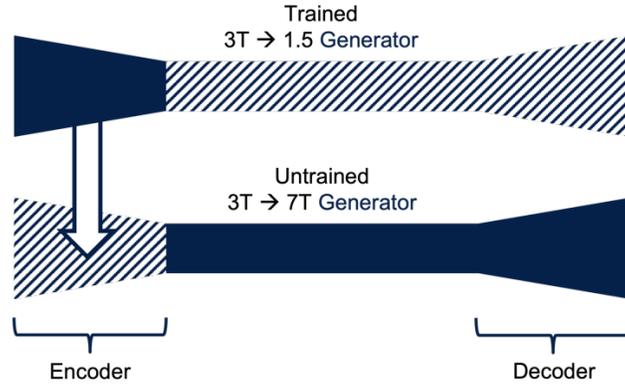


Figure 2.1: Transfer learning. The weights from the encoder 1.5T generator trained on the transfer learning dataset are used to initialize the weights of the untrained 7T generator.

2.6 Model Training

The models were implemented in pytorch version 1.8 and trained using the Adam optimizer with a learning rate of 0.0002. A batch size of 16 was used to train each model. The 3T synthesis model on the transfer learning dataset was trained for 10 epochs, where learning rate decayed after 5 epochs. The transfer learning model was trained on fewer epochs but had a similar number of training steps due to the larger dataset size. The base 7T synthesis model and 7T synthesis model initialized with transfer learning were trained for 100 epochs, where learning rate decayed linearly after 50 epochs.

The input images to the models were 128 x 128 2D patches of axial slices. The volumes are first normalized to have an intensity range between 0 and 1 to ensure numerical stability of the backpropagation algorithm, as large intensities values would lead to an explosion of gradients at the input layer. Data augmentation was performed during training through random rotation (+/- 30 degrees), scaling (80-120%), cropping (128 x 128), and horizontal flipping. Data augmentation effectively increases the number of training samples that the model is exposed to which helps reduce overfitting. Uncropped axial slices were used during validation and testing of the trained models.

2.7 Slice inhomogeneities

Intensity compensation adopted from Malandain et al. was used to correct for slice inhomogeneities in the synthesized images [31]. This strategy was proposed to correct for intensity differences in 3D reconstructed volumes of 2D histological MR data. The synthesized images in our

approach were produced in a slice wise manner which introduced intensity variations in the axial direction. The intensity compensation method works by matching the intensity probability density function of each axial slice to a reference slice, which was chosen to be the middle axial slice. Intensity compensation was performed on the synthesized images for both the base model and model with transfer learning weights prior to MUSE ROI segmentation.

2.8 Performance Evaluation

Leave one out cross validation was used when training on the 7T synthesis task for both the base model and the model initialized with transfer learning weights. Eight of the paired images were used for training while one was left out for validation. Results are reported as the average metrics across the validation images. Structural Similarity Index Measure (SSIM) as a perceptual image quality metric to assess the degradation of image quality was estimated [28]. Peak signal to noise ratio was calculated by $\log_{10}\left(\frac{Max\ Intensity^2}{MSE(True\ Image, Test\ Image)}\right)$ where MSE is mean squared error and max intensity is the maximum possible intensity of the image. PSNR and SSIM were computed using implementations from scikit-image version 0.18.2. Dice similarity coefficient (DSC) and Jaccard similarity coefficient (JSC) were used to measure the consistency of the tissue segmentation results produced from the original 3T and 7T synthesized images relative to the ground truth 7T segmentation. The DSC and JSC from the segmentation results were statistically compared using the Wilcoxon signed rank test.

Chapter 3: Results

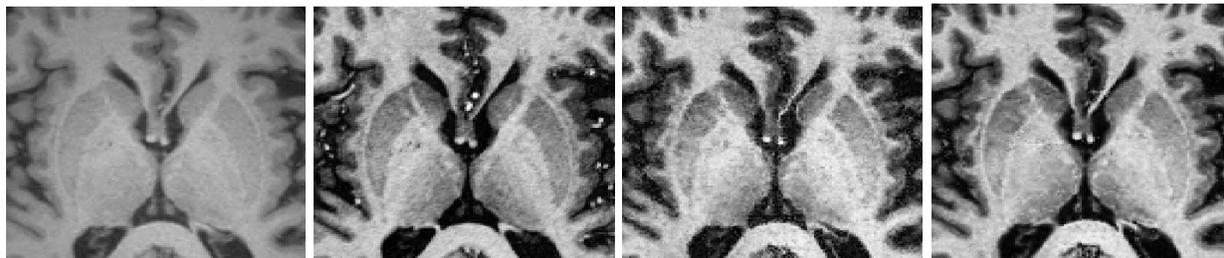


Figure 3.1: Axial patches from a random validation case. From left: input 3T, ground truth 7T, synthesized 7T, synthesized 7T (transfer learning model)



Figure 3.2: Sagittal patches from a random validation case. From left: input 3T, ground truth 7T, synthesized 7T, synthesized 7T (transfer learning model)

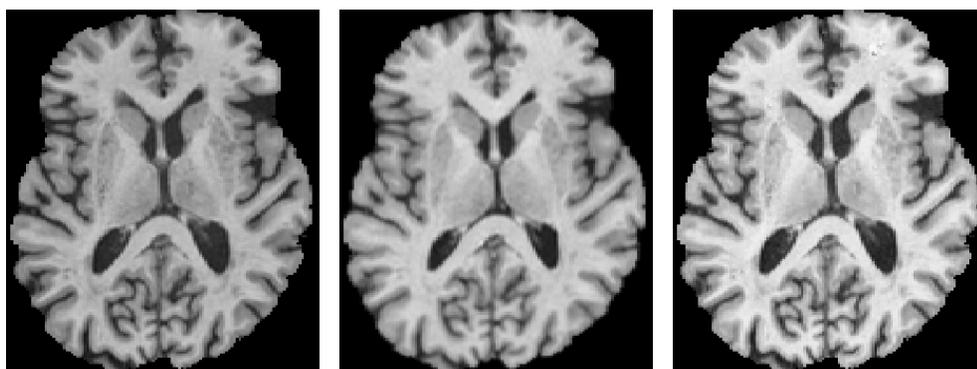


Figure 3.3: Axial slices of 1.5T synthesis results. Left: input 3T image. Middle: 1.5T ground truth. Right: 1.5T synthesized image

Qualitative results for the 7T synthesis models are presented in Figure 3.1 and 3.2. Qualitative results for the 3T synthesis task on the transfer learning dataset are shown in Figure 3.3. Additional qualitative results for 7T synthesis and tissue segmentations results can be found in Appendix I-III. Figure 3.8 shows results of the application of the slice intensity correction. Transfer learning helped the generator networks

converge faster and to a lower training loss, shown by the plots in Figure 3.1. The mean base model generator loss in the final 500 training steps was 1.28 ± 0.1 while the mean transfer learning model generator loss was 1.06 ± 0.06 . The discriminator networks with transfer learning converge faster to a mean loss of 0.48 ± 0.016 compared to the base model mean loss of 0.45 ± 0.02 .

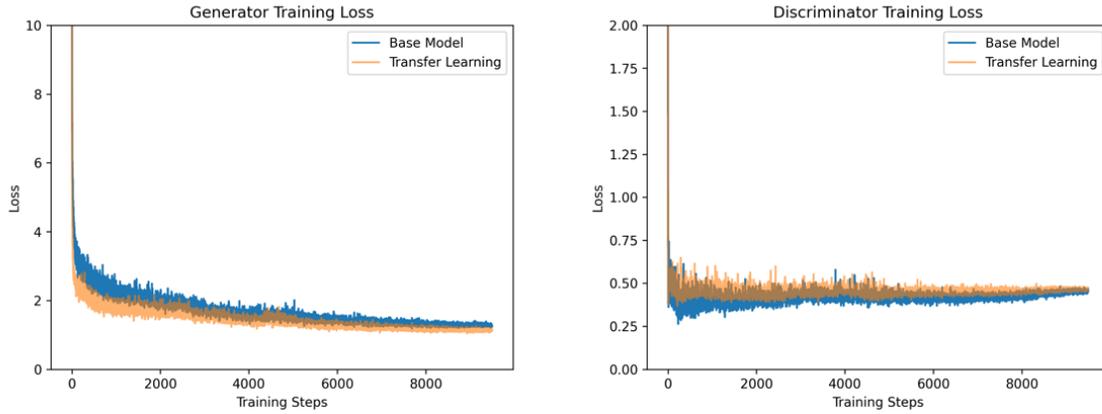


Figure 3.4: Average Generator and Discriminator loss across all validation cross folds. The generator and discriminator losses are the sum of loss for both the forward and backward GANs.

3.1 Image Similarity Metrics

The 3T synthesis task trained on the ADNI transfer learning dataset achieved a SSIM of 0.952 ± 0.05 , PSNR of 23.5 ± 4.5 on the independent testing dataset, a 0.01 point improvement over the 1.5T image (Table 3.1). The transfer learning model improved SSIM over the based model ($p < 0.005$) while PSNR was not significantly improved ($p = 0.054$). The synthetic 7T images improved in both SSIM (+0.05 units) and PSNR (+3.61 dB) compared to the 3T image ($p < 0.005$) (Table 3.2).

Table 3.1: Image similarity metrics for the 3T synthesis task on the transfer learning dataset. Metrics were computed on an independent testing set of 17 subjects with paired 1.5T and 3T scans.

Metric	1.5T input	3T Synth
Structural Similarity	0.940 (0.07)	0.952 (0.05)
PSNR	26.9 (10.1)	23.5 (4.5)

Table 3.2: Image similarity metrics for the 7T synthesis task. All metrics were computed using the 7T image as ground truth. The mean and standard deviation of the metrics across all the leave-one-out validation cross folds are reported.

Metric	3T Input	7T synth	7T synth (transfer)	Qu et al. [13]	Zhang et al. [14]
Structural Similarity	0.909 (0.011)	0.943 (0.006)	0.950 (0.006)	0.8782	0.8438
Peak SNR	21.83 (0.92)	25.18 (0.55)	25.44 (0.61)	28.27	29

3.2 Tissue Segmentation

The Dice and Jaccard similarity coefficient results are reported in Table 3.3. Dice similarity coefficient was improved in both grey matter ROIs (0.053 units; $p = 0.011$) and white matter ROIs (0.017 units; $p = 0.0039$) in the synthesized images using transfer learning weights compared to the 3T input. Jaccard Similarity Coefficient was improved for grey matter ROIs (0.066 units; $p = 0.011$) and white matter ROIs (0.026 units; $p = 0.0039$). The 7T synthesis results with transfer learning were not significantly different from the 7T synthesis base model in GM DSC ($p = 0.43$), GM JSC ($p = 0.43$), WM DSC ($p = 0.055$), or WM JSC ($p = 0.055$) compared to 7T synthesis base model. The DSC across all ROIs was 0.76 ± 0.023 for the 3T image, 0.813 ± 0.02 for the 7T synthesized image, and 0.823 ± 0.021 for the 7T synthesized image using transfer learning weights. Total grey matter and white matter volumes from the segmentation results are compared in Figure 3.5. Differences in DSC with respect to different

lobes of the brain are given in Figure 3.6 and differences in JSC with respected to lobes are given in Figure 3.7.

Table 3.3: Dice similarity coefficients for segmentation results. GM: Grey Matter, WM: White Matter, DSC: Dice Similarity Coefficient, JSC: Jaccard Similarity Coefficient. Synthesized results were compared statistically to the 3T results using the Wilcoxon signed rank test.

	3T	7T synth	p- value	7T synth (transfer)	p- value	Qu et al. [13]	Zhang et al. [14]
Total	0.760	0.813	0.020	0.823 (0.021)	0.0039	N/A	N/A
DSC	(0.023)	(0.020)					
Total	0.641	0.701	0.020	0.714 (0.029)	0.0078	N/A	N/A
JSC	(0.029)	(0.026)					
GM	0.757	0.800	0.027	0.810 (0.02)	0.011	0.85	0.78
DSC	(0.03)	(0.02)					
GM	0.627	0.679	0.027	0.693 (0.03)	0.011	N/A	N/A
JSC	(0.03)	(0.03)					
WM	0.899	0.912	0.0078	0.916 (0.004)	0.0039	0.90	0.89
DSC	(0.011)	(0.006)					
WM	0.820	0.842	0.0078	0.846 (0.01)	0.0039	N/A	N/A
JSC	(0.02)	(0.01)					

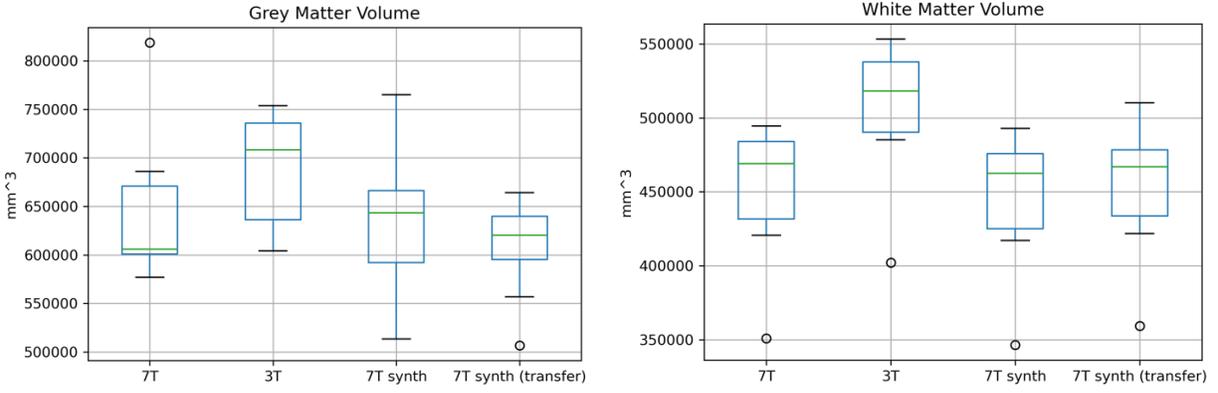


Figure 3.5: Box and whisker plots of the total grey and white matter volume for the tissue segmentation results. Synthesized images were intensity compensated to correct for slice inhomogeneities in the axial direction prior to segmentation. WM volume for the 3T results was significantly different from the 7T ground truth ($p < 0.005$) while it was not significantly different for the synthesized 7T WM volume ($p = 0.359$). GM volume was slightly different in the 3T image compared to 7T ground truth ($p = 0.098$). GM volume for the 7T synthesized results were not significantly different from the ground truth 7T GM volume ($p = 0.57$ base model, $p = 0.16$ with transfer learning).

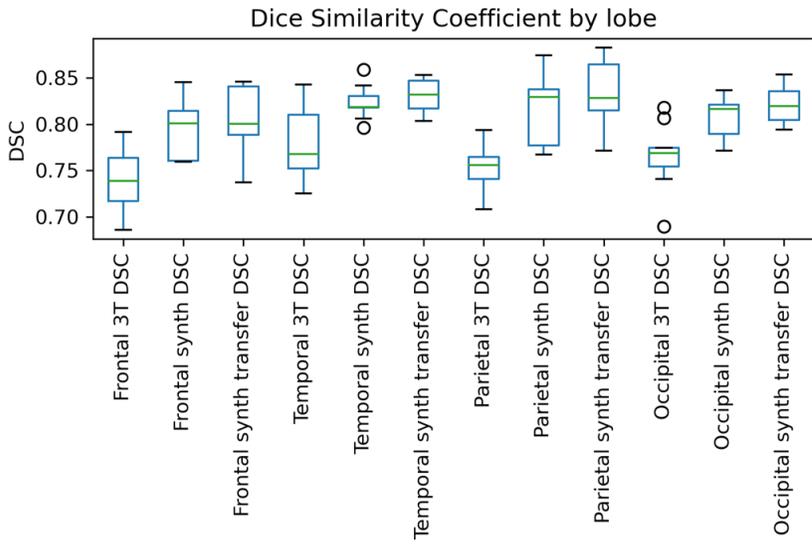


Figure 3.6: Box and whisker plot of DSC for regions of interest separated by lobe. Synthesized 7T DSC increased by 0.054 ($p = 0.020$) in the frontal lobe, 0.045 ($p = 0.020$) in the temporal lobe, 0.065 ($p = 0.054$) in the parietal lobe, and 0.039 ($p = 0.008$) in the occipital lobe over 3T DSC. Synthesized 7T with transfer learning DSC increased by 0.062 ($p = 0.020$) in the frontal lobe, 0.052 ($p = 0.008$) in the temporal lobe, 0.080 ($p = 0.011$) in the parietal lobe, and 0.055 ($p = 0.008$) in the occipital lobe over 3T DSC.

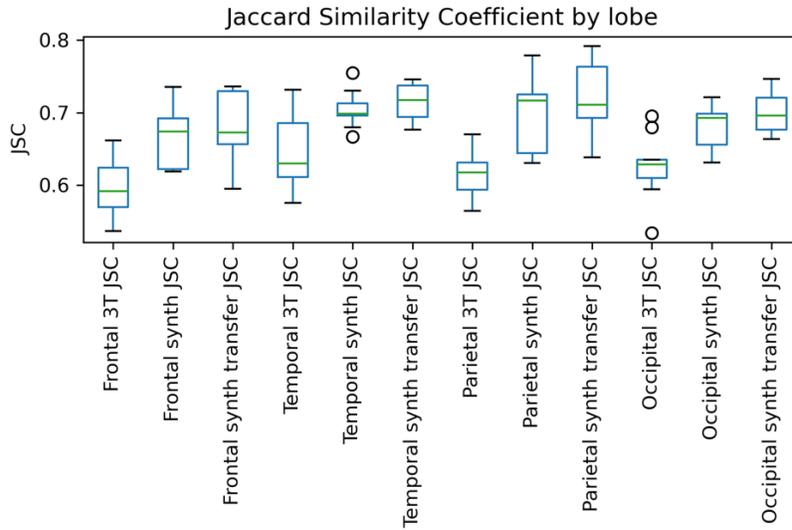


Figure 3.7: Box and whisker plot of JSC for regions of interest separated by lobe. Synthesized 7T JSC increased by 0.067 ($p = 0.027$) in the frontal lobe, 0.060 ($p = 0.020$) in the temporal lobe, 0.083 ($p = 0.027$) in the parietal lobe, and 0.052 ($p = 0.054$) in the occipital lobe over 3T DSC. Synthesized 7T with transfer learning JSC increased by 0.079 ($p = 0.020$) in the frontal lobe, 0.069 ($p = 0.008$) in the temporal lobe, 0.10 ($p = 0.020$) in the parietal lobe, and 0.073 ($p = 0.008$) in the occipital lobe over 3T JSC.

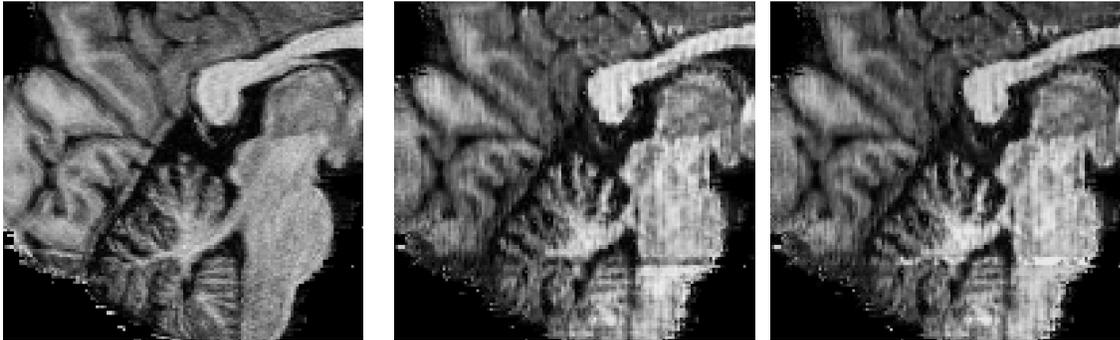


Figure 3.8: Sagittal and coronal views showing slice inhomogeneities. Left image is the 7T ground truth, middle image is the uncorrected synthesized 7T, and right image is the intensity corrected synthesized 7T.

Chapter 4: Discussion

Our work has shown that deep generative models such as CycleGAN are capable of transforming 3T images to 7T-like contrast with high structural similarity. The synthesized 7T images benefit from improved Dice and Jaccard similarity coefficients over 3T images in both grey and white matter ROIs, using the 7T segmentation results as ground truth. The 7T synthesis model with transfer learning had faster training converge and improved SSIM over the base model however this did not lead to

improvements in DSC or JSC for the ROI segmentation results compared to the base model. CycleGAN outperforms the non-adversarial CNN approaches of Qu et al. and Zhang et al. in structural similarity index measure but underperforms in peak signal to noise ratio (see Table 3.1). SSIM was developed to provide a quantitative measure of perceived image quality while PSNR is a function of the bit depth and mean squared error [29]. CycleGAN may have resulted in lower PSNR than other CNN approaches that directly optimize mean squared error such as Zhang et al. [14].

4.1 Transfer Learning Task

The transfer learning weights helped to improve SSIM in the 7T synthesis tasks but ultimately did not lead to better segmentation results. The objective of the generator for the 1.5T generator in the 3T synthesis task on the ADNI dataset was to reduce the quality of the input 3T image and produce 1.5T-like contrast. However, features learned by the encoder portion of the 1.5T synthesis task were assumed to be transfer knowledge to the 7T synthesis task, where the objective was to improve image quality. Image synthesis may not have been the best task to train the generator for transfer learning. Tissue segmentation, classification, and unsupervised image reconstruction tasks have been used for MRI transfer learning and may lead to more generalizable features than the image synthesis task used in this work [23]. The difference in diversity and size of datasets was the focus of transfer learning for this task, as the encoder portion of the generator network of the 3T synthesis model was exposed to multi-site data, while the 7T synthesis dataset only had data from a single site.

4.2 Limitations

One limitation in our study is the 2D patch wise training approach because it limits the context of the images while training. Patch wise training on 2D axial slices limits the network to learn local features within the axial plane. The 2D slice approach also introduced intensity variations between slices because the network lacked 3D context of neighboring slices. 3D generative approaches have been shown to reduce slice continuities but required larger models and longer training time, and lost sharpness compared to the 2D model [33]. 2.5D approaches may address the limited context of 2D models by including sagittal, axial, and coronal planes during training while avoiding the increased computational cost of 3D

models [34, 35]. Another limitation of this study is the use of leave one out validation cross validation for evaluating the 7T synthesis models. Leave one out validation was used to maximize available training data but an independent testing set would be more suited for testing the generalizability of the models. Kwon et al. showed that GANs trained on disease free images can fail to reconstruct images with significant disease [26]. The 7T synthesis may fail to generalize to images with significant disease such as brain tumors or lesions since examples were not provided for training in either dataset. While diversity of data was introduced with the transfer learning dataset, diversity with respect to demographic information such as sex, age, and race was not considered in this study.

4.3 Future Work

This study worked towards a deep learning harmonization tool by validating the results of automated brain morphometry algorithms on images with synthesized 7T contrast. Additional tasks for transfer learning could be evaluated in future work, where the 3T generator network is trained to perform tissue segmentation rather than image synthesis. The 7T synthesis networks could potentially be used as a contrast normalization method for the purpose of image harmonization. The goal of future work is to determine if the 7T synthesized contrast can remove non-biological variance in MR data in longitudinal multi-site study.

References:

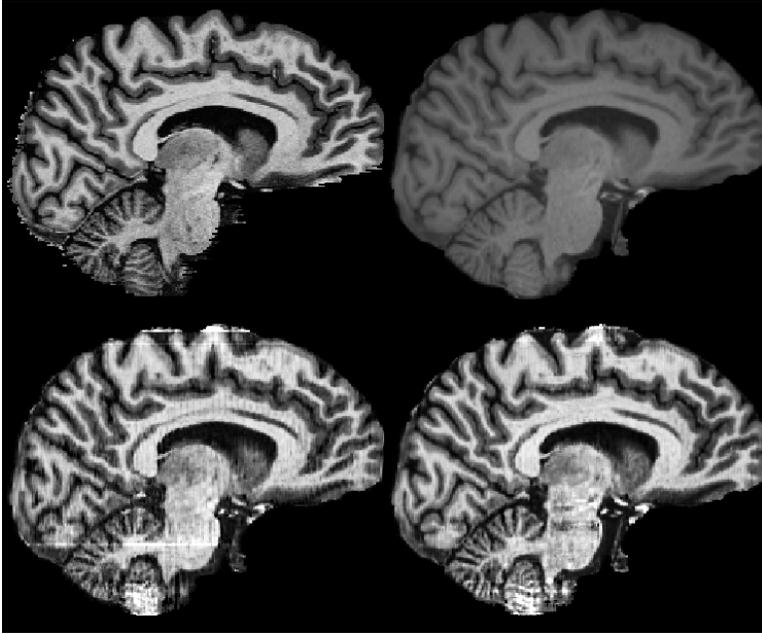
1. Takao, Hidemasa et al. "Effect of scanner in longitudinal studies of brain volume changes." *Journal of magnetic resonance imaging : JMRI* vol. 34,2 (2011): 438-44. doi:10.1002/jmri.22636
2. Pomponio, Raymond et al. "Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan." *NeuroImage* vol. 208 (2020): 116450. doi:10.1016/j.neuroimage.2019.116450
3. Fortin, Jean-Philippe et al. "Harmonization of cortical thickness measurements across scanners and sites." *NeuroImage* vol. 167 (2018): 104-120. doi:10.1016/j.neuroimage.2017.11.024
4. Fortin, Jean-Philippe et al. "Removing inter-subject technical variability in magnetic resonance imaging studies." *NeuroImage* vol. 132 (2016): 198-212. doi:10.1016/j.neuroimage.2016.02.036
5. Radua, Joaquim et al. "Increased power by harmonizing structural MRI site differences with the ComBat batch adjustment method in ENIGMA." *NeuroImage* vol. 218 (2020): 116956. doi:10.1016/j.neuroimage.2020.116956
6. Wrobel, J et al. "Intensity warping for multisite MRI harmonization." *NeuroImage* vol. 223 (2020): 117242. doi:10.1016/j.neuroimage.2020.117242
7. Dewey, Blake E et al. "DeepHarmony: A deep learning approach to contrast harmonization across scanner changes." *Magnetic resonance imaging* vol. 64 (2019): 160-170. doi:10.1016/j.mri.2019.05.041
8. Zuo, Lianrui, et al. "Information-Based Disentangled Representation Learning for Unsupervised MR Harmonization." *International Conference on Information Processing in Medical Imaging*. Springer, Cham, 2021.
9. Liu, Mengting, et al. "Style Transfer Using Generative Adversarial Networks for Multi-Site MRI Harmonization." *bioRxiv* (2021).
10. Dinsdale, Nicola K et al. "Deep learning-based unlearning of dataset bias for MRI harmonisation and confound removal." *NeuroImage* vol. 228 (2021): 117689. doi:10.1016/j.neuroimage.2020.117689

11. Seiger, Rene et al. "Voxel-based morphometry at ultra-high fields. a comparison of 7T and 3T MRI data." *NeuroImage* vol. 113 (2015): 207-16. doi:10.1016/j.neuroimage.2015.03.019
12. Isaacs, Bethany R et al. "3 versus 7 Tesla magnetic resonance imaging for parcellations of subcortical brain structures in clinical settings." *PloS one* vol. 15,11 e0236208. 24 Nov. 2020, doi:10.1371/journal.pone.0236208
13. Qu, Liangqiong et al. "Synthesized 7T MRI from 3T MRI via deep learning in spatial and wavelet domains." *Medical image analysis* vol. 62 (2020): 101663. doi:10.1016/j.media.2020.101663
14. Zhang, Yongqin et al. "Dual-domain convolutional neural networks for improving structural information in 3 T MRI." *Magnetic resonance imaging* vol. 64 (2019): 90-100. doi:10.1016/j.mri.2019.05.023
15. Nie, Dong et al. "Medical Image Synthesis with Deep Convolutional Adversarial Networks." *IEEE transactions on bio-medical engineering* vol. 65,12 (2018): 2720-2730. doi:10.1109/TBME.2018.2814538
16. Bahrami, Khosro et al. "7T-guided super-resolution of 3T MRI." *Medical physics* vol. 44,5 (2017): 1661-1677. doi:10.1002/mp.12132
17. Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." *Proceedings of the IEEE international conference on computer vision*. 2017.
18. Wang, Tonghe et al. "A review on medical imaging synthesis using deep learning and its clinical applications." *Journal of applied clinical medical physics* vol. 22,1 (2021): 11-36. doi:10.1002/acm2.13121
19. Ali, Muhaddisa Barat et al. "Domain Mapping and Deep Learning from Multiple MRI Clinical Datasets for Prediction of Molecular Subtypes in Low Grade Gliomas." *Brain sciences* vol. 10,7 463. 18 Jul. 2020, doi:10.3390/brainsci10070463
20. Do, Huy, et al. "7T MRI super-resolution with Generative Adversarial Network." *IS&T Electronic Imaging 2021 Symposium*. 2021.

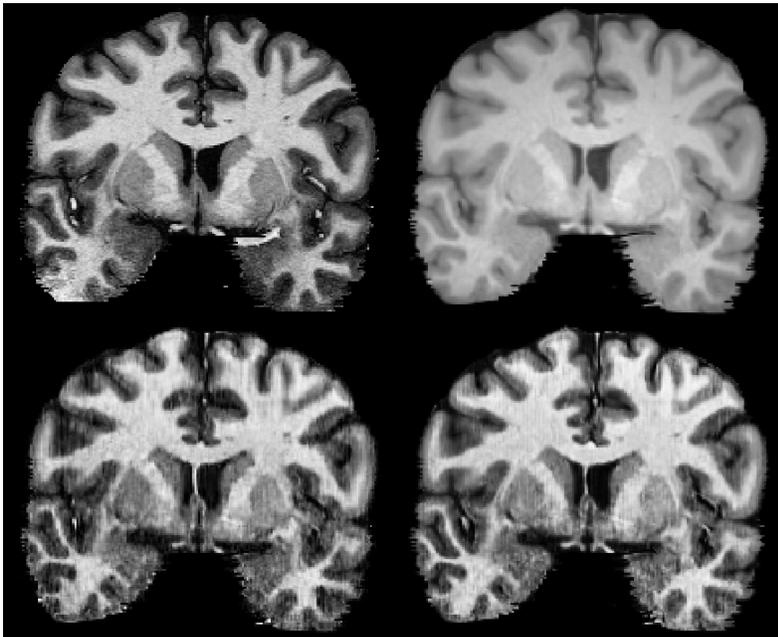
21. "FDA Clears First 7T Magnetic Resonance Imaging Device." U.S. Food and Drug Administration, FDA, 12 Oct. 2017, www.fda.gov/news-events/press-announcements/fda-clears-first-7t-magnetic-resonance-imaging-device.
22. Karamat, Muhammad Irfan et al. "Opportunities and Challenges of 7 Tesla Magnetic Resonance Imaging: A Review." *Critical reviews in biomedical engineering* vol. 44,1-2 (2016): 73-89. doi:10.1615/CritRevBiomedEng.2016016365
23. Valverde, Juan Miguel et al. "Transfer Learning in Magnetic Resonance Brain Imaging: A Systematic Review." *Journal of Imaging* vol. 7,4 66. 1 Apr. 2021, doi:10.3390/jimaging7040066
24. Jack, Clifford R Jr et al. "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods." *Journal of magnetic resonance imaging : JMRI* vol. 27,4 (2008): 685-91. doi:10.1002/jmri.21049
25. Jack, Clifford R Jr et al. "Update on the magnetic resonance imaging core of the Alzheimer's disease neuroimaging initiative." *Alzheimer's & dementia : the journal of the Alzheimer's Association* vol. 6,3 (2010): 212-20. doi:10.1016/j.jalz.2010.03.004
26. Kwon, Gihyun, Chihye Han, and Dae-shik Kim. "Generation of 3D brain MRI using auto-encoding generative adversarial networks." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2019.
27. Smith, Stephen M. "Fast robust automated brain extraction." *Human brain mapping* vol. 17,3 (2002): 143-55. doi:10.1002/hbm.10062
28. Jenkinson, M, and S Smith. "A global optimisation method for robust affine registration of brain images." *Medical image analysis* vol. 5,2 (2001): 143-56. doi:10.1016/s1361-8415(01)00036-6
29. Doshi, Jimit et al. "Multi-atlas skull-stripping." *Academic radiology* vol. 20,12 (2013): 1566-76. doi:10.1016/j.acra.2013.09.010
30. Doshi, Jimit et al. "MUSE: MUlti-atlas region Segmentation utilizing Ensembles of registration algorithms and parameters, and locally optimal atlas selection." *NeuroImage* vol. 127 (2016): 186-195. doi:10.1016/j.neuroimage.2015.11.073

31. Malandain G, Bardinet E. Intensity compensation within series of images. In International Conference on Medical Image Computing and Computer-Assisted Intervention 2003 Nov 15 (pp. 41-49). Springer, Berlin, Heidelberg.
32. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process.* 2004 Apr;13(4):600-12. doi: 10.1109/tip.2003.819861. PMID: 15376593.
33. Nepl, Sebastian et al. "Evaluation of proton and photon dose distributions recalculated on 2D and 3D Unet-generated pseudoCTs from T1-weighted MR head scans." *Acta oncologica (Stockholm, Sweden)* vol. 58,10 (2019): 1429-1434. doi:10.1080/0284186X.2019.1630754
34. Milletari, Fausto, et al. "Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound." *Computer Vision and Image Understanding* 164 (2017): 92-102.
35. Hering, Alessa et al. "Memory-efficient 2.5D convolutional transformer networks for multi-modal deformable registration with weak label supervision applied to whole-heart CT and MRI scans." *International journal of computer assisted radiology and surgery* vol. 14,11 (2019): 1901-1912. doi:10.1007/s11548-019-02068-z

Appendix I:

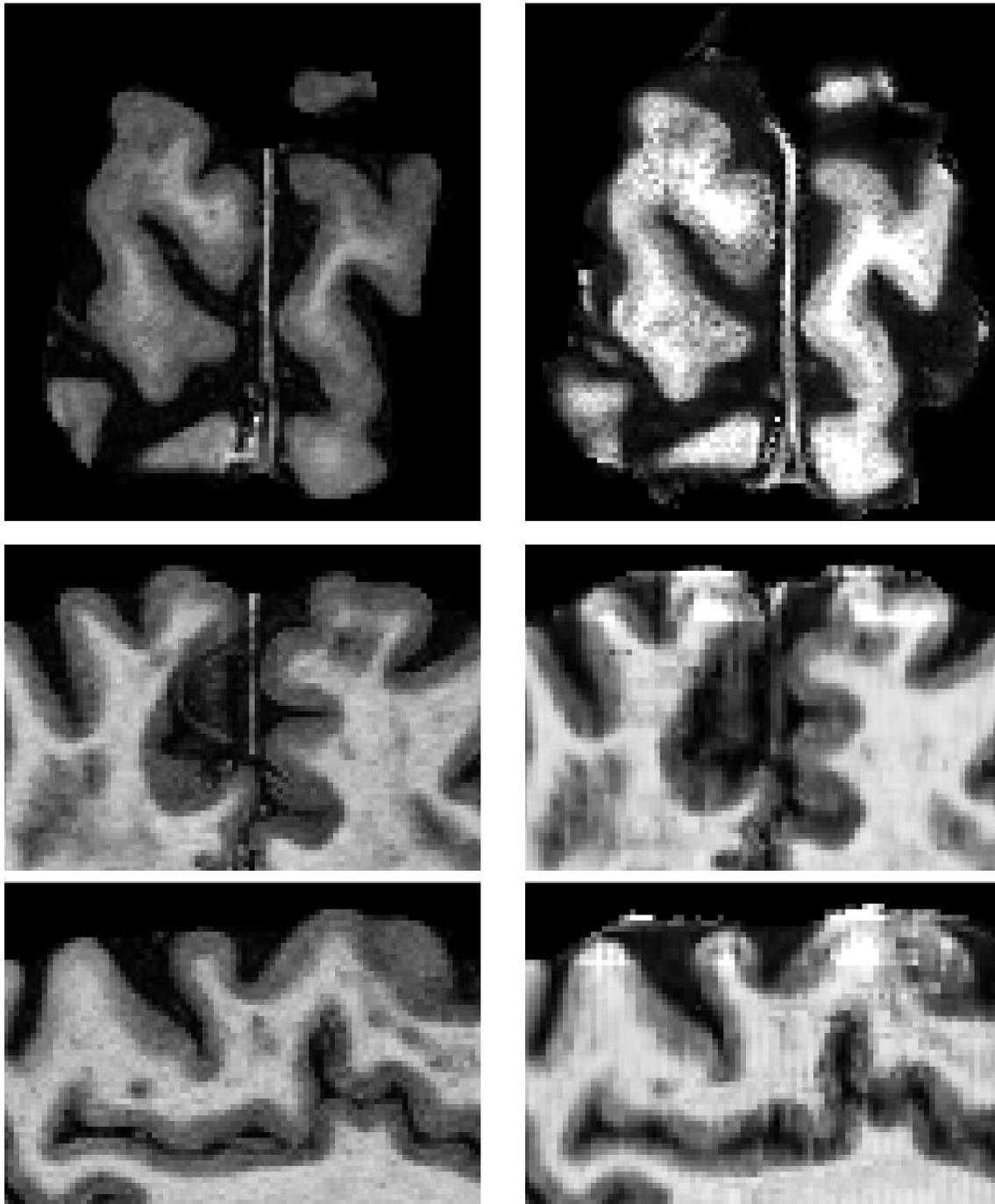


Sagittal views of a validation image. Clockwise from top left: 7T ground truth, 3T input image, synthesized 7T (transfer learning), synthesized 7T



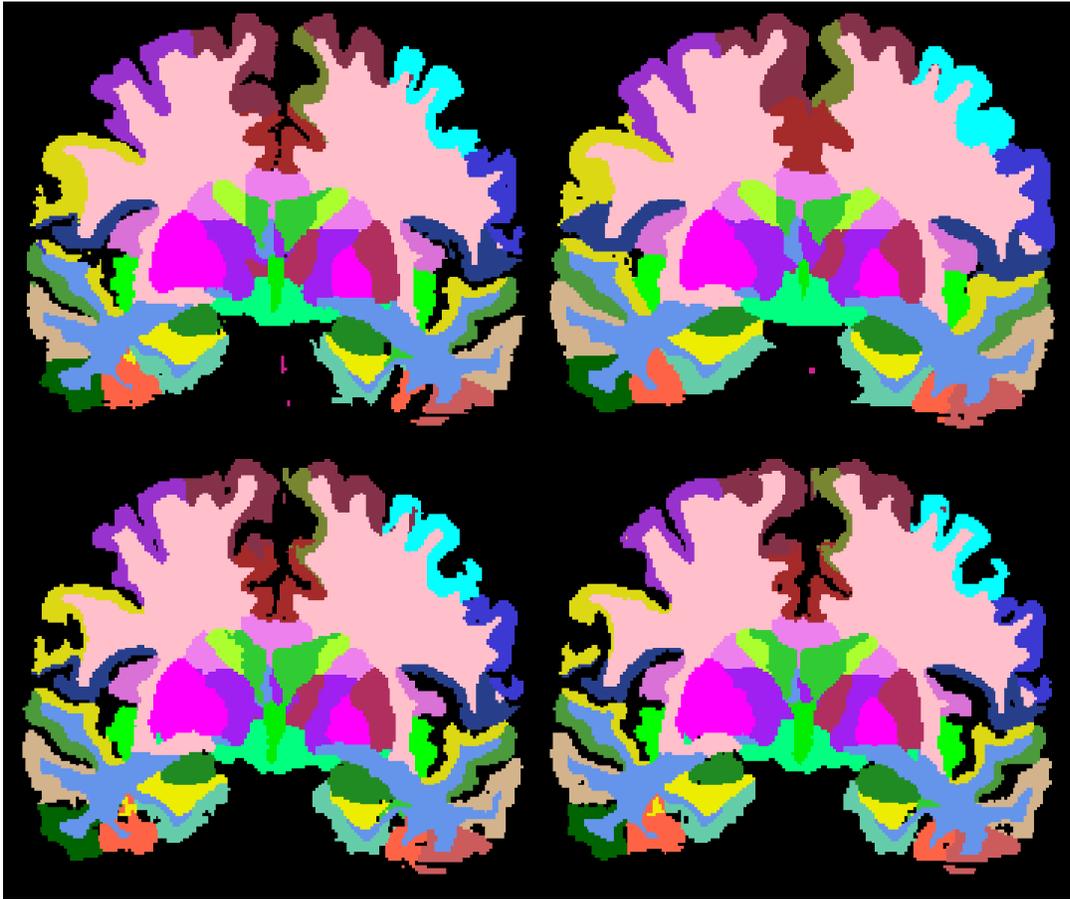
Coronal views of a validation image. Clockwise from top left: 7T ground truth, 3T input image, synthesized 7T (transfer learning), synthesized 7T

Appendix II



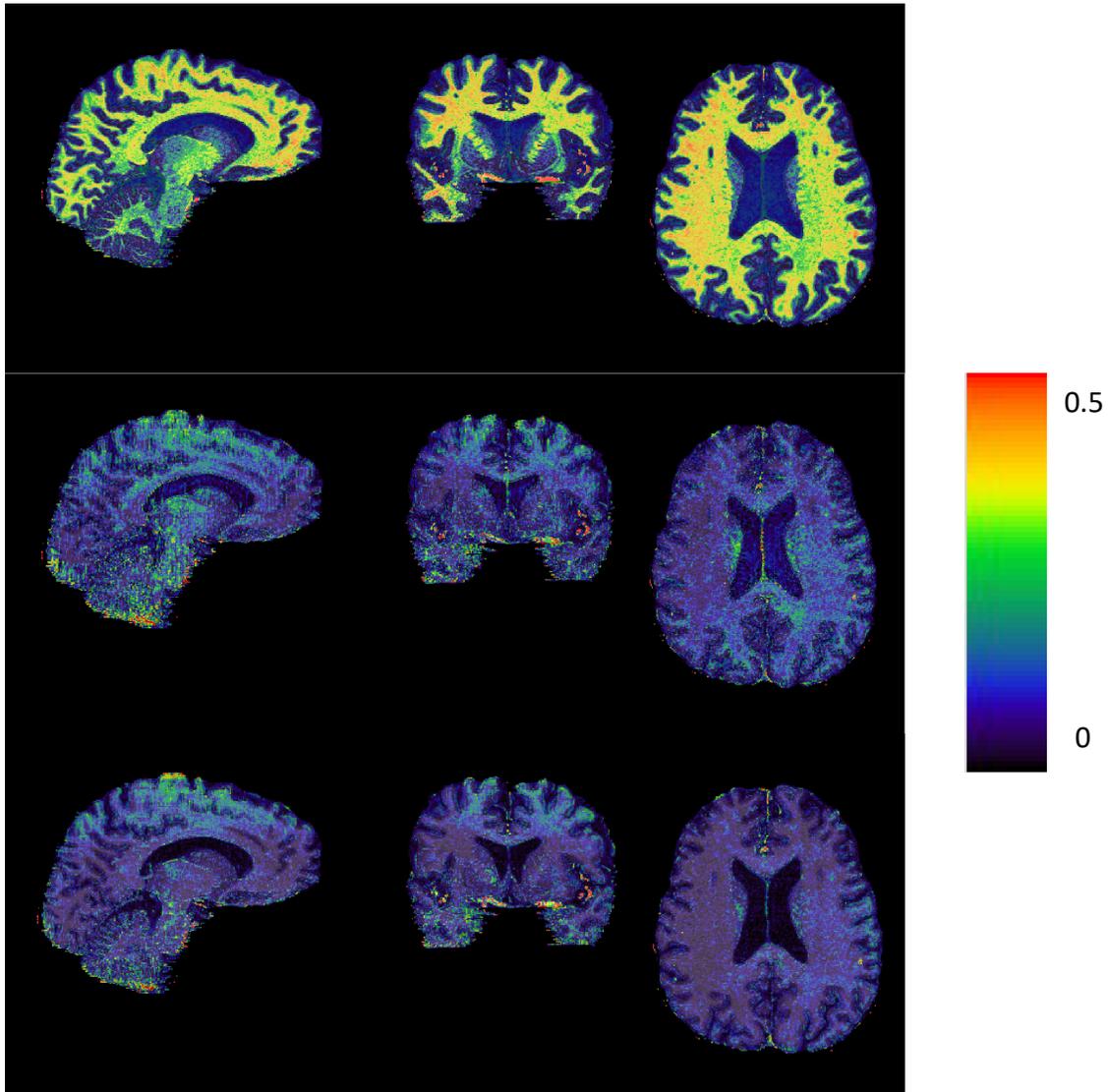
Examples of GAN produced artifacts on a validation image. Left column of images show an axial, coronal, and sagittal patch of superior brain tissue in the ground truth 7T image. Right columns shows the synthesized 7T images

Appendix III



Coronal views of a MUSE ROI segmentation on a validation image. Clockwise from top left: 7T ground truth, 3T input image, synthesized 7T (transfer learning), synthesized 7T

Appendix IV



Mean absolute error maps of the 3T image (top row), 7T synth (middle row), and 7T synth with transfer learning (bottom row) overlaid on the ground truth 7T image.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

Ryan Ellis

F9BD8B6ABED94DB...

Author Signature

8/31/2021

Date