

Cross-modality image translation of 3 Tesla Magnetic Resonance Imaging to 7 Tesla using Generative Adversarial Networks

Eduardo Diniz¹ | Tales Santini² | Karim Helmet^{2,3} | Howard J. Aizenstein^{2,3} | Tamer S. Ibrahim²

¹Department of Electrical and Computer Engineering, University of Pittsburgh, Pennsylvania, United States

²Department of Bioengineering, University of Pittsburgh, Pennsylvania, United States

³Department of Psychiatry, University of Pittsburgh, Pennsylvania, United States

Correspondence

Tamer S. Ibrahim, Department of Bioengineering, University of Pittsburgh, Pennsylvania, United States.
Email: tsi2@pitt.edu

Abstract

The rapid advancements in magnetic resonance imaging (MRI) technology have precipitated a new paradigm wherein cross-modality data translation across diverse imaging platforms, field strengths, and different sites is increasingly challenging. This issue is particularly accentuated when transitioning from 3 Tesla (3T) to 7 Tesla (7T) MRI systems. This study proposes a novel solution to these challenges using generative adversarial networks (GANs)—specifically, the CycleGAN architecture—to create synthetic 7T images from 3T data. Employing a dataset of 1112 and 490 unpaired 3T and 7T MR images, respectively, we trained a 2-dimensional (2D) CycleGAN model, evaluating its performance on a paired dataset of 22 participants scanned at 3T and 7T. Independent testing on 22 distinct participants affirmed the model's proficiency in accurately predicting various tissue types, encompassing cerebral spinal fluid, gray matter, and white matter. Our approach provides a reliable and efficient methodology for synthesizing 7T images, achieving a median Dice of 6.82%, 7.63%, and 4.85% for Cerebral Spinal Fluid (CSF), Gray Matter (GM), and White Matter (WM), respectively, in the testing dataset, thereby significantly aiding in harmonizing heterogeneous datasets. Furthermore, it delineates the potential of GANs in amplifying the contrast-to-noise ratio (CNR) from 3T, potentially enhancing the diagnostic capability of the images. While acknowledging the risk of model overfitting, our research underscores a promising progression towards harnessing the benefits of 7T MR systems in research investigations while preserving compatibility with existent 3T MR data. This work was previously presented at the ISMRM 2021 conference (Diniz, Helmet, Santini, Aizenstein, & Ibrahim, 2021).

KEYWORDS

7 Tesla Magnetic Resonance Imaging, 3 Tesla Magnetic Resonance Imaging, Cross-Modality Image Translation, Generative Adversarial Networks, CycleGAN, Deep Learning, Neuroimaging

1 | INTRODUCTION

Magnetic Resonance Imaging (MRI) technology has offered high definition and noninvasive insights into the intricacies of human brain structure and function (Hagmann et al., 2006). This advancement has considerably improved our understanding of brain morphological and functional changes throughout life and in various neurodegenerative conditions (Ashburner & Friston, 2000). In addition, it has spearheaded the discovery of potential biomarkers for numerous neurological disorders (Ross et al., 2012), thereby opening the doors to early detection and therapeutic intervention.

The field of MRI has been shifting from conventional/clinical 3 Tesla (3T) MRI systems towards high-field strength 7 Tesla (7T) MRI systems (Uğurbil, 2014). The benefits of 7T MRI include augmented signal-to-noise ratio (SNR), enhanced visualization of brain structures due to reduced voxel size, and an improved contrast-to-noise ratio (CNR) resulting from more distinguishable tissue relaxation times (Uğurbil, 2014). Despite these advantages, increased costs, restricted availability, increased susceptibility artifacts, a higher specific absorption rate (SAR), and radiofrequency (RF) inhomogeneity may complicate image acquisition and interpretation (van Osch & Webb, 2014).

The field has seen significant advances in the design and application of RF coils that are instrumental in overcoming the challenges of 7T MRI, chiefly ensuring high-quality and high-resolution acquisition

(Santini, Wood, et al., 2021). However, transitioning longitudinal patient studies from the widely used 3T systems to the more advanced 7T systems presents substantial hurdles. The primary reason for this is the innate differences in the images generated by the two systems, especially the contrast variations among the soft tissues. These may affect direct image comparisons and complicate longitudinal analyses (Obusez et al., 2018).

Other factors, such as variability in image quality and systematic differences between varying field strengths, can complicate statistical analyses of aggregate data (Fortin et al., 2018). These difficulties underscore the need for efficient spatial adaptive data normalization methodologies capable of handling variances across different field strengths; they are paramount in strengthening our understanding of the neuroanatomical shifts and progressions associated with normal aging and pathological processes.

Artificial intelligence (AI) and machine learning technologies have significantly shaped numerous sectors, including healthcare (Bohr & Memarzadeh, 2020). In particular, medical imaging is being improved by the potential of deep learning methods, which extends their reach into the complex realm of neuroimaging (Shen, Wu, & Suk, 2017). Specifically, applying Convolutional Neural Networks (CNNs) in image classification tasks has improved diagnostic precision and efficiency (Litjens et al., 2017). Building on this momentum, Generative Adversarial Networks (GANs) and their variants have emerged as powerful tools for medical image analysis, including tasks such as image synthesis, segmentation, and data augmentation (Nie et al., 2018).

Introduced by Goodfellow et al., (2014), GANs comprise a unique model structure employing two deep neural networks: the generator and the discriminator. They engage in an adversarial process, aiming to generate synthetic data resembling the original distribution. These networks have demonstrated proficiency in modeling complex data distributions (Arjovsky, Chintala, & Bottou, 2017; Salimans et al., 2016), which is beneficial for various image transformation tasks.

CycleGANs, an extension of GANs, provide a solution to unpaired image-to-image translation tasks (Zhu, Park, Isola, & Efros, 2017), including photo enhancement (Chen & Koltun, 2017), style transfer (Gatys, Ecker, & Bethge, 2016), and image synthesis (T.-C. Wang et al., 2018). In particular, within the realm of medical imaging, its applications have been transformative. CycleGANs have also shown potential in tasks such as lesion synthesis (Guerrero et al., 2018). These networks can enrich rare disease datasets by generating synthetic lesion images, facilitating better disease detection and diagnosis. In addition, they have been used in tasks such as organ segmentation by transforming organ-specific images into simplified representations that facilitate the segmentation process (Cai, Zhang, Cui, Zheng, & Yang, 2019).

One significant application pertains to the translation of images in different modalities, such as converting Computed Tomography (CT) images to MRI images and vice versa (Wolterink, Leiner, Viergever, & Išgum, 2017). In practical scenarios, patients may not undergo CT and MRI scans due to cost, radiation exposure, or other considerations. By training in unpaired CT and MRI datasets, CycleGAN has demonstrated

the capability to generate synthetic but anatomically accurate MR images from CT scans and, conversely, to produce CT-like images from MR scans.

Cross-modal image translation, defined by the machine learning and imaging community, consists of converting images from one modality to another, retaining the crucial content while modifying the style to resemble the target modality (Zhu et al., 2017). 3T and 7T, despite employing the identical principle of nuclear magnetic resonance, exhibit divergent operational mechanics due to distinct field strengths. This differentiation manifests itself in significant differences in image quality, resolution, CNR, susceptibility effects, and spectral separation (Vaughan et al., 2001). In this regard, 3T MRI can be considered one modality and 7T MRI another. The conversion process from 3T to 7T through GANs fulfills this criterion, as it retains the intrinsic information (content) from the 3T images and translates the image properties (style) to align with that of 7T MRI.

This study uses a dataset of unpaired 3T and 7T MR images to train a CycleGAN-based model capable of translating 3T MRI data into synthetic 7T data. This approach aims to address the problem of spatial adaptive MR data harmonization, allowing the combined analysis of cross-sectional and longitudinal data from 3T and 7T MRI systems.

2 | MATERIALS & METHODS

2.1 | Model Architecture

The landscape of machine learning has been significantly shaped by the emergence of GANs, as conceptualized by Goodfellow et al., (2014). This novel model structure, built on game-theoretic principles, employs two deep neural networks in strategic interaction: the generator and the discriminator. The generator aims to generate synthetic data replicating the actual data distribution, while the discriminator strives to discern the synthesized data from the real data. This adversarial dance evolves over iterative training sessions, yielding refined performance from both networks and improved proficiency in modeling complex data distributions (Arjovsky et al., 2017; Salimans et al., 2016).

Taking a step further, CycleGANs provide a robust solution to unpaired image-to-image translation, a challenging problem in computer vision. Zhu et al., (2017) introduced CycleGAN's architecture, comprising two GANs containing a generator and a discriminator. The generator creates compelling images of the target domain, deceiving the discriminator, which aims to differentiate real images from the translated ones within the target domain. The design of CycleGAN is such that each generator maps an input image from a source domain to a target domain and vice versa to ensure cyclical consistency (Zhu et al., 2017).

A feature that distinguishes CycleGAN from traditional GANs is its cycle consistency loss. This loss function aims to ensure that the cycle of translating an image from one domain to another and back again reconstructs the original image. Let \mathcal{X} denote the domain of images, and \mathcal{Y} denote another domain of images. We can represent an image from

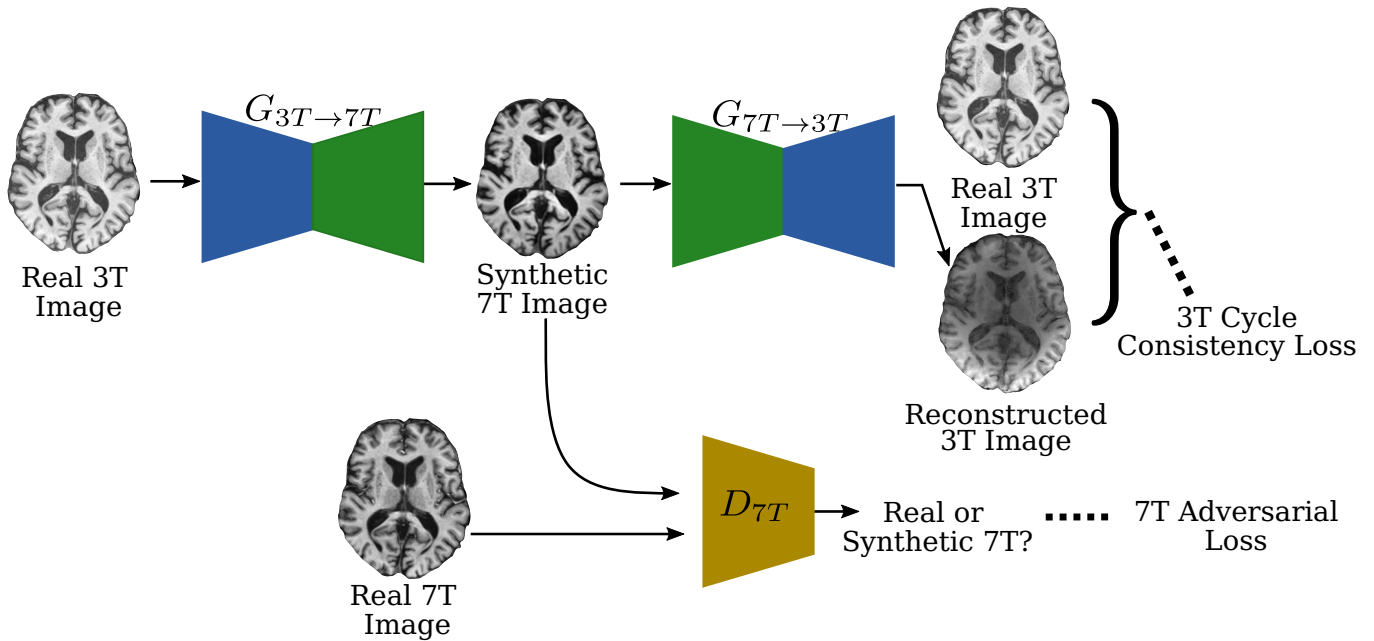


FIGURE 1 The cyclical consistency of the CycleGAN. This figure visually demonstrates the 3T cycle consistency. The 7T loop is conceptually symmetric. The generators are marked as performing the mappings $G_{3T \rightarrow 7T} : 3T \rightarrow 7T$ and $G_{7T \rightarrow 3T} : 7T \rightarrow 3T$. The Real 3T image is fed to the $G_{3T \rightarrow 7T}$ (top left), producing the synthetic 7T image (top middle). The D_{7T} discriminator differentiates between the Real and Synthetic 7T images and strives to minimize 7T adversarial loss (bottom left). The synthetic 7T image is fed into the $G_{7T \rightarrow 3T}$ generator, which produces the reconstructed 3T image. The 3T cycle consistency loss ensures the 3T image can be reconstructed from the synthetic 7T image (top right).

\mathcal{X} as $x \in \mathcal{X}$, and an image from \mathcal{Y} as $y \in \mathcal{Y}$. Let us construct image datasets from each domain: $\mathcal{D}_{\mathcal{X}} = \{x_1, x_2, \dots, x_m\}$ consisting of M images from \mathcal{X} , and $\mathcal{D}_{\mathcal{Y}} = \{y_1, y_2, \dots, y_n\}$ consisting of N images from \mathcal{Y} . Now, let the mappings $G : \mathcal{X} \rightarrow \mathcal{Y}$ and $F : \mathcal{Y} \rightarrow \mathcal{X}$ be the generators between the two domains. Then, the cycle consistency loss is:

$$L_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim \mathcal{D}_{\mathcal{X}}} [\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim \mathcal{D}_{\mathcal{Y}}} [\|G(F(y)) - y\|_1]. \quad (1)$$

$\mathbb{E}_{x \sim \mathcal{D}_{\mathcal{X}}}$ and $\mathbb{E}_{y \sim \mathcal{D}_{\mathcal{Y}}}$ are the expectations taken over the datasets $\mathcal{D}_{\mathcal{X}}$ and $\mathcal{D}_{\mathcal{Y}}$, respectively. The terms $\|F(G(x)) - x\|_1$ and $\|G(F(y)) - y\|_1$ measure the absolute difference between the original image and the image that has been translated to the other domain and back, i.e., $F(G(x))$ means G translates an image $x \in \mathcal{D}_{\mathcal{X}}$ into the domain \mathcal{Y} and then translated to \mathcal{X} by F . Similarly, $G(F(y))$ takes an image $y \in \mathcal{D}_{\mathcal{Y}}$, translates it into domain \mathcal{X} by F and then converts it back to \mathcal{Y} by G . The cycle consistency loss should be small if the model performs well, indicating that the original and the cycled-back image are nearly identical.

Our study uses a CycleGAN to deal with unpaired 3T and 7T MRI data. Figure 1 shows a schematic representation of the architecture and functioning of the CycleGAN model. Our implementation integrates two generator networks based on the U-Net architecture (Ronneberger, Fischer, & Brox, 2015), with residual blocks inserted between the encoding and decoding stages (He, Zhang, Ren, & Sun, 2016).

U-Net is a CNN architecture renowned for its utility in biomedical image segmentation. It is named after its U-shaped structure, which consists of a contracting path to capture context and a symmetric expanding

path for precise localization of features. The expanding pathway uses transposed convolutions, with skip connections that transfer feature maps from the contracting path to recover spatial information.

Residual blocks, proposed in the ResNet architecture by He et al., (2016) resolve training difficulties encountered with deep neural networks. These blocks consist of several convolutions, after which the input is added to the output, creating a residual connection. This configuration allows the network to learn residual functions concerning the layer's inputs, facilitating the training of deeper models.

The model also uses instance normalization and stride-2 convolutions. Instance normalization is often used in style transfer tasks (Ulyanov, Vedaldi, & Lempitsky, 2016). It calculates the mean and variance for each instance separately, leading to the relative scaling of the activations. This technique is beneficial for disentangling content and style in image generation tasks. Stride-2 convolutions, frequently utilized in CNNs, help reduce the spatial dimensions of feature maps and allow the model to learn more abstract representations in deeper layers (Goodfellow, Bengio, & Courville, 2016).

Our CycleGAN model uses 70×70 PatchGANs for the discriminator networks. PatchGANs, as described by Isola et al., (2017), classify whether each patch in an image is real or fake instead of assessing the picture as a whole. This patch-level classification enables the generation of sharper and more contextually accurate images in tasks such as image-to-image translation.

2.2 | Dataset

We leveraged an extensive collection of unpaired and paired 3T and 7T MR images from diverse sources to ensure a rich and comprehensive dataset. The 3T MRI data were primarily sourced from the Human Connectome Project (HCP), a globally recognized brain imaging dataset that contains high definition multimodal brain imaging data from healthy adults (Sotiropoulos et al., 2013). Specifically, structural 3T images were sourced from a subset of 1112 scans in the HCP's 1200 subjects data release. The images were acquired using a Siemens 3T Connectome Skyra and a standard 32-channel Siemens receive head coil at Washington University in St. Louis. The scans were conducted using a T1-weighted (T1w) Magnetization Prepared Rapid Gradient Echo (MPRAGE) protocol with a repetition time (TR) of 2400 ms, echo time (TE) of 2.14 ms, flip angle (FA) of 8°, field of view (FOV) of 224 mm × 224 mm, and isotropic resolution of 0.7 mm, lasting approximately 8 minutes.

The inclusion criteria for the 3T HCP data required participants to be between 22 and 35 years old with no significant history of psychiatric disorder, substance abuse, neurological or cardiovascular disease and the ability to provide valid informed consent. Participants were required to have a Mini Mental Status Exam score above 28.

Exclusion criteria included multiple nonprovoked seizures or a diagnosis of epilepsy, genetic disorders such as cystic fibrosis, the use of prescription medications for migraines in the past 12 months, and conditions such as multiple sclerosis, cerebral palsy, brain tumor, stroke, sickle cell disease, thyroid hormone treatment in the past 12 months, current treatment for diabetes (excluding gestational or diet-controlled diabetes), head injury, and premature birth. Other exclusion criteria were the presence of unsafe metals or devices in the body (such as cardiac pacemakers, cochlear implants, aneurysm clips), current or historical use of chemotherapy or immunomodulatory agents that could affect the brain, pregnancy, and moderate to severe claustrophobia.

The 7T dataset was sourced from the 7TBRP dataset, acquired at the RF Research Facility at the University of Pittsburgh, a specialized research center for high-field imaging. The dataset included 490 subjects' structural 7T MR images. The scans were acquired using a Siemens Magnetom 7T whole-body MR scanner and an in-house developed first generation of the Tic-Tac-Toe head radio-frequency coil system (Santini et al., 2018; Santini, Wood, et al., 2021; Krishnamurthy et al., 2019; Santini, Koo, et al., 2021). The scans were performed using a T1w MPRAGE protocol with a TR of 3000 ms, TE of 2.17 ms, bandwidth of 391 Hz/Px, GRAPPA reconstruction with acceleration factor $R = 2$, and isotropic resolution of 0.75 mm, lasting approximately 5 minutes.

The exclusion criteria for the 7TBRP dataset included pregnancy or lactation, acute medical problems that could result in neurocognitive or brain dysfunction, including diabetes mellitus, coronary artery disease, and causes cerebral vasculities, such as peripheral vascular disease. Other exclusion criteria were contraindications to MRI scanning, such as electronic implants, magnetically activated implants, tattoos above the shoulders, or brain implants. Data acquisition in all studies was performed following the protocol approved by local institutional review

boards. All participants were older than 18 years and were able to provide their written informed consent prior to participation.

A set of paired (acquired in the same participants) 3T-7T data from the 7TBRP dataset was used to compare the synthetic 7T data from 3T and the real 7T data. The data set consisted of MR images from 22 subjects. The study-specific inclusion and exclusion criteria for the paired data were similar to those for the unpaired 7T datasets.

2.3 | Data Preprocessing

The first stage of pre-processing, bias field correction, was performed using FMRIB's Automated Segmentation Tool (FAST) part of FSL (Y. Zhang, Brady, & Smith, 2001). This technique mitigates the intensity inhomogeneity observed within MR images due to receive coil profile and patient positioning differences (Vovk, Pernus, & Likar, 2007). If not correctly trained, deep learning models can learn from these inhomogeneities and become biased, leading to sub-optimal performance. Bias field correction can rectify this, enabling the model to concentrate on the salient features of the image (Arnold et al., 2001). This process results in enhanced feature extraction, improved segmentation, and registration results, which further enhances the effectiveness of subsequent processing and training phases (Tustison et al., 2010).

Following bias field correction, we performed spatial normalization using the Statistical Parametric Mapping (SPM) toolbox's coregister function, co-registering the images via normalized mutual information to the Montreal Neurological Institute 152 standard (MNI152) space (Ashburner, 2007; Penny, Friston, Ashburner, Kiebel, & Nichols, 2011). Normalization facilitates the comparison and integration of data between subjects, as each voxel location corresponds to the same anatomical structure across all images (Evans, Janke, Collins, & Baillet, 2012). This alignment minimizes the need for the neural network to learn these invariances, potentially improving performance (Cao et al., 2017).

In the next stage, we transform the image intensities to a standard normal range of $[-1, 1]$ using SPM's image calculator function. This transformation minimizes intersubject variability, promoting improved image data consistency (Nyul & Udupa, 1999). Furthermore, intensity normalization allows the model to focus more on structural or semantic differences between images than on absolute pixel intensities (Goodfellow et al., 2016). This is critical during the early stages of training as it prevents the gradients from vanishing or exploding, resulting in a more stable optimization process and faster convergence (LeCun, Bottou, Orr, & Müller, 2002). Furthermore, intensity normalization can also make the learning process less dependent on the specific units used. This is especially important when combining data from multiple sources with different measurement units or scales (Bishop & Nasrabadi, 2006).

Finally, we skull-stripped the images using the Brain Extraction Tool (BET) within FSL, removing non-brain tissue from the MRI images and thereby reducing dimensionality, which can improve the efficiency of subsequent model training (Iglesias & Sabuncu, 2015; Smith, 2002). For computational efficiency, we transformed 3D images into 2D by

extracting axial slices using FSL's `fslslice` function, which is less computationally intensive while preserving critical anatomical information.

2.4 | Model Training

Training a deep learning model requires orchestrating various strategies and methods to optimize performance. In the context of our CycleGAN model implemented within the TensorFlow framework (Abadi et al., 2016), we had to consider various elements such as computational hardware, optimization algorithms, weight initialization, regularization techniques, and learning rate schedules.

We used NVIDIA Titan RTX GPUs (24GB of memory) for training and adopted an iterative training strategy; the discriminators were updated four times for each update of the generators within the training loop. This prevented the generators from overpowering the discriminators, which is critical to supporting the dynamics of the adversarial process (Goodfellow et al., 2014).

We used early stopping as a regularization technique to control the model's capacity and prevent overfitting. We stopped training when no significant improvement in the performance of the validation set was observed, *i.e.*, the validation loss did not decrease by 0.1% over five consecutive epochs (Prechelt, 2002). Furthermore, L1/L2 regularization was employed to encourage the model to learn distributed and sparse representations, which can increase its generalizability (Ng, 2004).

We used the Adam optimizer (Kingma & Ba, 2014) for the optimization algorithm. Its adaptive learning rate and efficient computation make it a superior choice for training deep learning models, especially for achieving faster and more effective convergence. The Xavier weight initialization method was used for all convolutional layers (Glorot & Bengio, 2010). In addition, all bias terms were initialized to zero. Proper weight initialization can significantly improve the convergence rate and prevent problems such as vanishing or exploding gradients, making the optimization process more stable.

Finally, we incorporated learning rate annealing into the training process to fine-tune the speed at which the model learns. The learning rate was set at 10^{-4} for the first 100 epochs and was linearly decreased to zero. By gradually reducing the learning rate, the model can make significant updates to learn the gross structure of the data in the early stages and then fine-tune its weights in later stages, allowing for smoother and more effective convergence (Bengio, 2012).

2.5 | Model Evaluation

The traditional approach to assessing GANs, primarily focusing on discriminator loss, needs an explicit measure of sample quality. This work thus evaluates the quality of training using a pre-trained deep neural network to embed the samples, inspired by Heusel et al., (2017) methodology. We adopted a ResNet-50 architecture pre-trained on ImageNet (He et al., 2016), given its effectiveness in similarity searching in the embedding space for various forms of imaging data, including MRI, even

without fine-tuning (Yosinski, Clune, Bengio, & Lipson, 2014). In this context, the quality of the samples is evaluated on the basis of their realism and "closeness" to the target domain. We calculated the Fréchet ResNet Distance (FRD) and the Cosine ResNet Distance (CRD) between each sample in the ResNet-50 embedding space to facilitate a comparison between 3T and 7T MRI features.

Let $\mathcal{D}_{\mathcal{X}'}$ = $G(\mathcal{D}_{\mathcal{Y}})$ denote the dataset of synthetic samples and $\text{ResNet}(\mathcal{D}_{\mathcal{X}})$ and $\text{ResNet}(\mathcal{D}_{\mathcal{X}'})$ denote the ResNet-50 embeddings of real and synthetic samples, respectively. The FRD, calculated as the Fréchet Distance (Dowson & Landau, 1982), or, equivalently, the Wasserstein-2 distance (Vallender, 1974) between the $\text{ResNet}(\mathcal{D}_{\mathcal{X}})$ and $\text{ResNet}(\mathcal{D}_{\mathcal{X}'})$ embeddings is given by:

$$\text{FRD}(\mathcal{N}(\mu_{\mathcal{X}}, \Sigma_{\mathcal{X}}), \mathcal{N}(\mu_{\mathcal{X}'}, \Sigma_{\mathcal{X}'}))^2 = \|\mu_{\mathcal{X}} - \mu_{\mathcal{X}'}\|_2^2 + \text{Tr}(M), \quad (2)$$

where

$$M = \left(\Sigma_{\mathcal{X}} + \Sigma_{\mathcal{X}'} - 2 \left(\Sigma_{\mathcal{X}}^{\frac{1}{2}} \cdot \Sigma_{\mathcal{X}'} \cdot \Sigma_{\mathcal{X}}^{\frac{1}{2}} \right)^{\frac{1}{2}} \right),$$

$\mathcal{N}(\mu_{\mathcal{X}}, \Sigma_{\mathcal{X}})$ and $\mathcal{N}(\mu_{\mathcal{X}'}, \Sigma_{\mathcal{X}'})$ are the multivariate Gaussian distributions over $\text{ResNet}(\mathcal{D}_{\mathcal{X}})$ and $\text{ResNet}(\mathcal{D}_{\mathcal{X}'})$, respectively, $\|\cdot\|_2$ represents the Euclidean norm, and Tr the trace.

Likewise, the CRD is the average cosine distance between each real $z \in \text{ResNet}(\mathcal{D}_{\mathcal{X}})$ and synthetic $z' \in \text{ResNet}(\mathcal{D}_{\mathcal{X}'})$:

$$\text{CRD} = 1 - \frac{1}{M \cdot N} \sum_{i=1}^M \sum_{j=1}^N \frac{z_i \cdot z'_j}{\|z_i\|_2 \cdot \|z'_j\|_2}, \quad (3)$$

where z_i represents the i -th feature embedding in $\text{ResNet}(\mathcal{D}_{\mathcal{X}})$ and z'_j represents the j -th feature in $\text{ResNet}(\mathcal{D}_{\mathcal{X}'})$. M and N are the total real and synthetic samples, respectively. This expression calculates the cosine similarity between each pair of feature embeddings, normalizes it by their Euclidean norms, and then averages the distances over all pairs.

Finally, we measured the sample quality over the synthetically generated 7T MR images. The accuracy of the generated synthetic 7T MR images was evaluated by cross-referencing these images with their respective real 7T ground truth counterparts. The images were segmented into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) using FSL's FAST segmentation tool. We then used two key and widely used evaluation metrics in medical imaging; the Dice Coefficient (Dice) and the Percentual Area Difference (PAD). These metrics were indispensable for evaluating the model's performance and provided imperative insights for future model enhancement.

The Dice, introduced by Dice, (1945), measures the overlap between the predicted and the ground-truth segmentation. Mathematically, for an expected pixel set \mathcal{A} and the ground truth pixel set \mathcal{B} , the Dice is:

$$\text{Dice} = 2 \cdot \frac{|\mathcal{A} \cap \mathcal{B}|}{(|\mathcal{A}| + |\mathcal{B}|)}, \quad (4)$$

where $|\cdot|$ represents the cardinality of a set. The range of Dice ranges from 0 (indicating no overlap) to 1 (indicating a perfect overlap), making it an excellent indicator of the congruence between predicted and real segmentation (Zou et al., 2004).

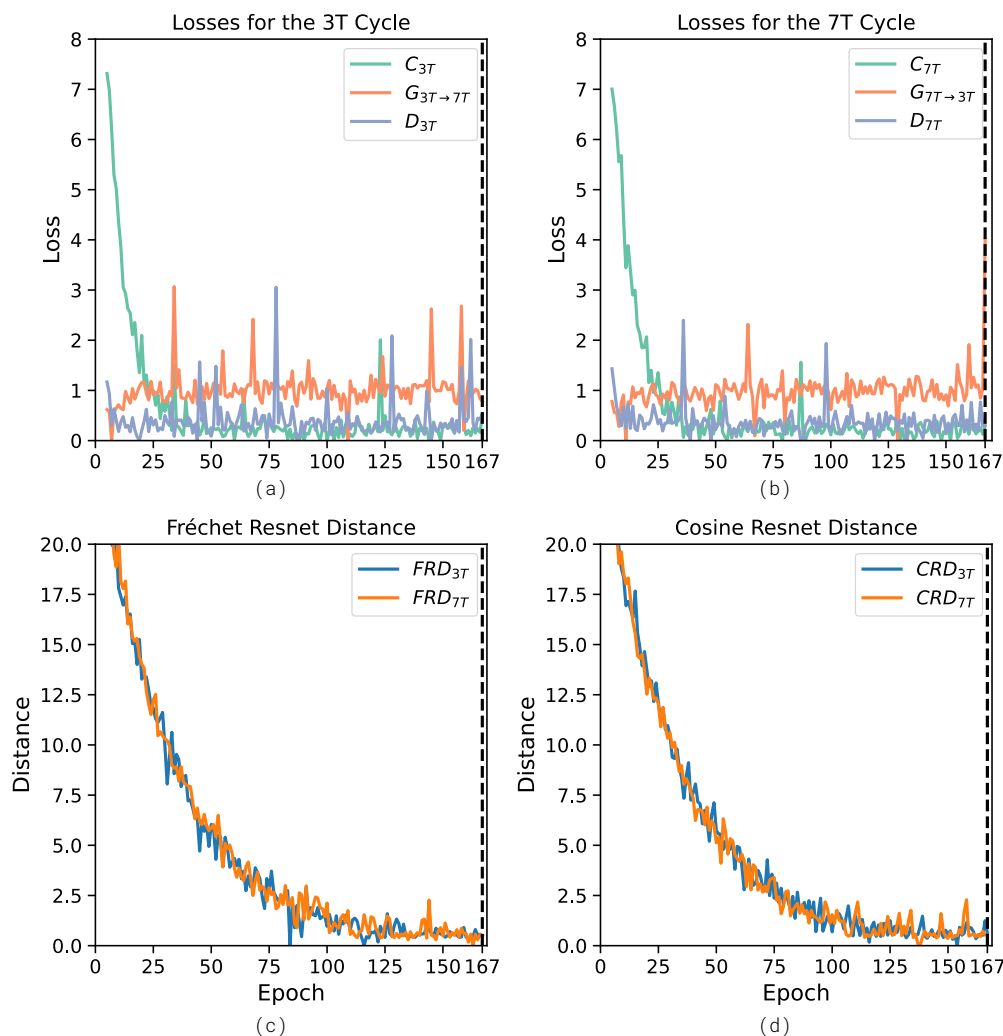


FIGURE 2 Progression of CycleGAN training dynamics and quality metrics for 3T to 7T MRI translation. (a) Generator, Discriminator and Cycle Consistency Losses for the 3T loop, (b) Generator, Discriminator, and Cycle Consistency Losses for the 7T loop, (c) Fréchet ResNet Distance (FRD), (d) Cosine Resnet Distance (CRD). The graphs illustrate the evolution of loss functions and quality metrics over training epochs. The vertical black dotted line mark epoch 167, where we halted training since we did not observe an improvement bigger than 0.1% in the 3T cycle consistency loss for five consecutive epochs.

Conversely, the PAD measures the discrepancy in the size between the predicted and the ground-truth segmentation. Specifically, it evaluates the model's accuracy in estimating the size of the segmented objects. For an expected area A and the ground truth segmentation area B , the PAD is computed as:

$$\text{PAD} = 100\% \cdot \frac{|A - B|}{B}, \quad (5)$$

where $|\cdot|$ is the absolute value. The PAD is inversely proportional to the performance of the segmentation, with smaller values indicating a superior performance (Huang, Sun, Tseng, Li, & Qian, 2019).

3 | RESULTS

The performance of the model was assessed using multiple parameters, including generators, discriminators, and cycle consistency losses in the validation dataset; FRD and CRD embedding quality measures; and a detailed analysis of the prediction for various brain tissue types and Dice and PAD sample quality metrics from independent test data.

3.1 | Training Performance

We assessed the model's performance by first analyzing the generator and discriminator losses in the validation dataset. The generator and the

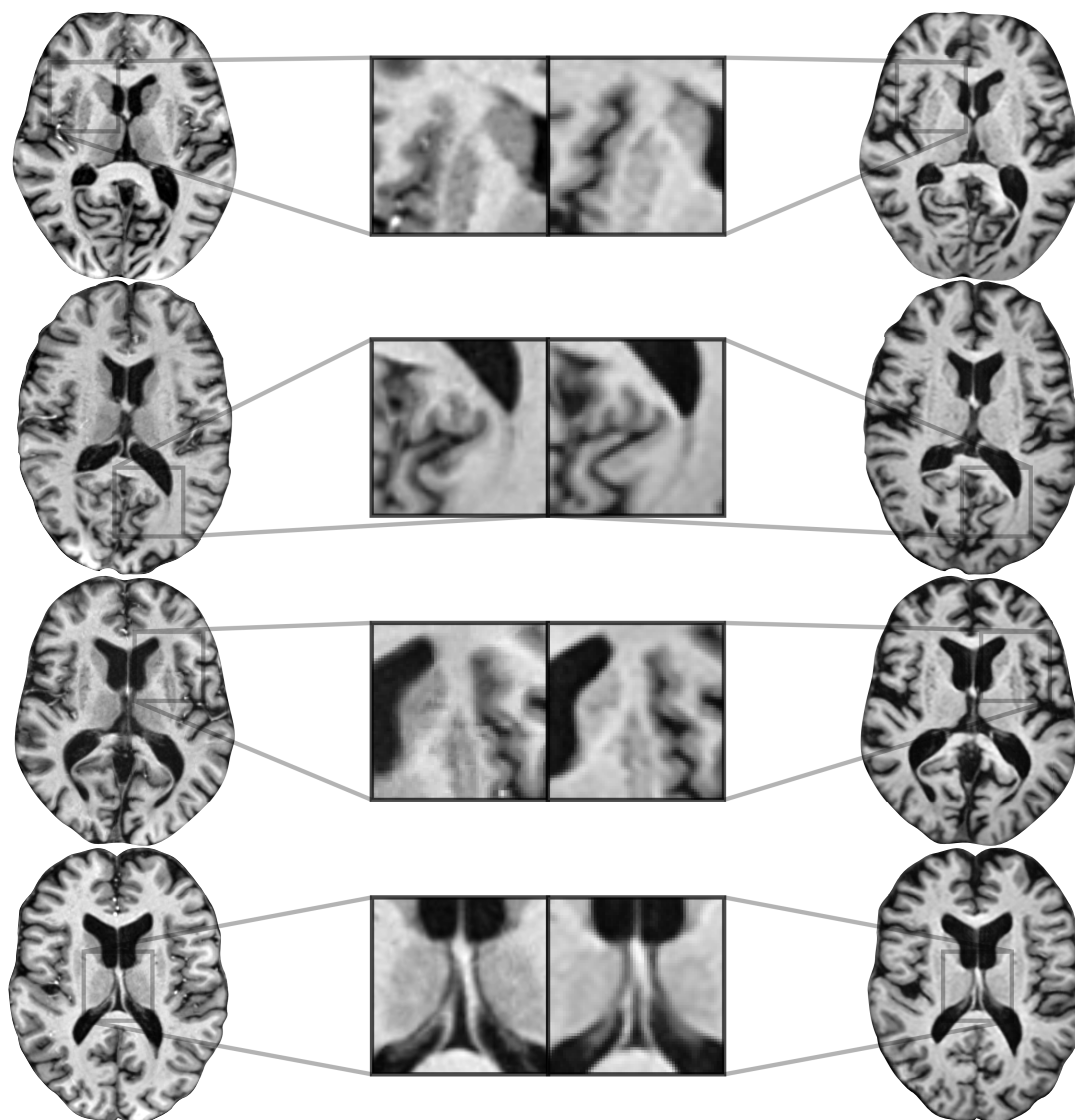


FIGURE 3 Examples of real 7T MR images (left column) and the corresponding synthetic 7T images (right column) generated by our model.

discriminator essentially participate in a minimax game, where the generator seeks to create images that the discriminator cannot distinguish from the real 7T images. In contrast, the discriminator aims to accurately classify real images from synthetic ones (Goodfellow et al., 2014).

Figures 2(a) and 2(b) show that the tug-of-war between generators and discriminators resulted in the generator losses gradually decreasing, demonstrating the generators' increasing proficiency in creating images that the discriminators struggle to differentiate from real MRI images. In tandem, the discriminators' losses also decreased, indicative of their improving ability to distinguish between real and generated images. At minima, the convergence of the losses substantiates the efficacy of the model in producing synthetic MRI-like images from real data.

However, this is evidence of improvement only during the early stages of training. Beyond that, since the generator and discriminator losses plateaued after the 14th epoch, just the cycle-consistency losses suggest that the output continued to improve. Hence, to quantitatively

assess the progression of the CycleGAN, we apply the metrics FRD and CRD described in Methods 2.5 to the data generated at each training epoch. This strategy allows us to examine the training dynamics of the CycleGAN. As shown in Figures 2(c) and 2(d), FRD and CRD though slightly noisy, stop decreasing beyond the 125th epoch. In general, Figure 2 shows that the embedded MRI captures more relevant features in the data, with the measures of transformation quality showing a consistent and smooth decrease throughout the training process.

3.2 | Visual Representation of Results

Further, an independent test dataset evaluated the generalization ability of the model, which involved using the model to translate 3T images from new subjects, unseen during training, into synthetic 7T images.

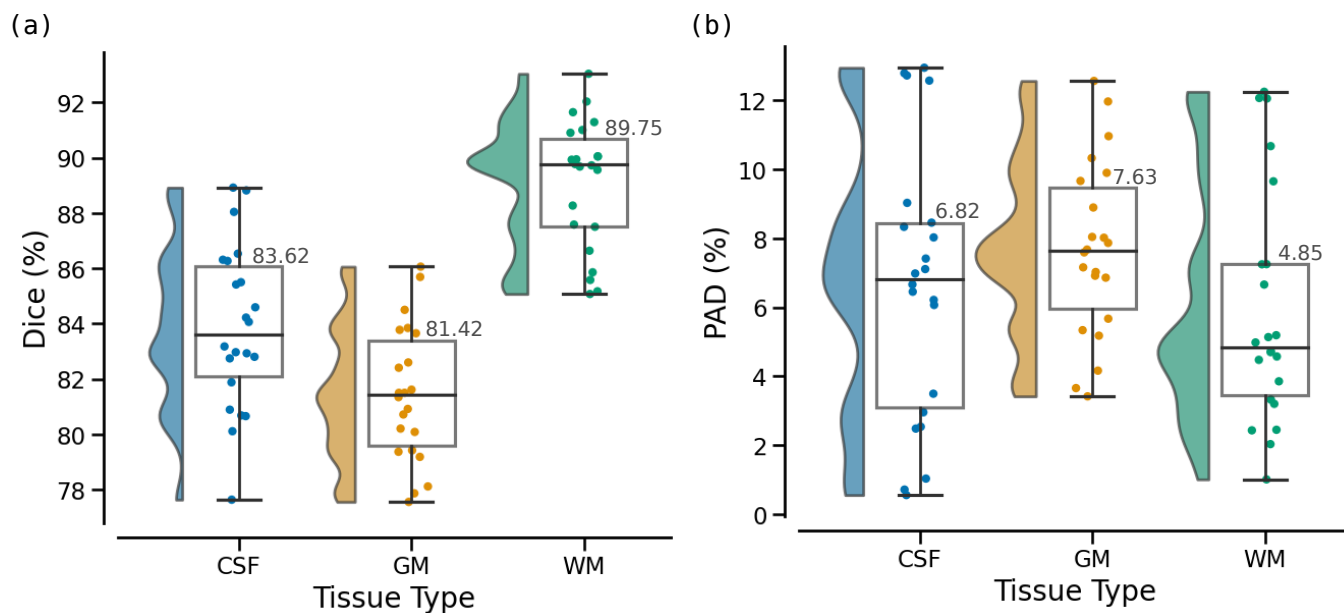


FIGURE 4 Group-level segmentation performance for the validation (22 subjects) data set across different tissue types—Cerebral Spinal Fluid (CSF), Gray Matter (GM), and White Matter (WM). We plot single patient data overlaid on group-level whisker box plots (center, median; box, 25th to 75th percentiles; whiskers, 1.5 interquartile range) and the smoothed data distribution. The median Dice for CSF, GM and WM are 83.62%, 81.42%, and 89.75%, respectively. The median percentual area difference (PAD) for CSF, GM, and WM are 6.82%, 7.63%, and 4.85%, respectively.

Figure 3 shows examples of real and synthetic 7T MRI image pairs generated by our model. The synthetic images manifest convincing resolution and contrast properties that mimic the attributes of real 7T images, demonstrating the model's success in generating high-fidelity 7T-like images from 3T data.

3.3 | Tissue Type Specific Prediction

Diving deeper, we analyzed the performance of the model in generating specific brain tissue types—CSF, GM, and WM—by computing the Dice and the PAD. The results, visualized using box plots (Figure 4), revealed that the model achieved a median Dice of 83.62% for CSF, 81.42% for GM and 89.75% for WM. Concurrently, the median PAD was 6.82% for CSF, 7.63% for GM, and 4.85% for WM. These figures demonstrate consistent performance across all tissue types, corroborating the model's robustness in translating detailed tissue-specific characteristics from 3T to 7T.

4 | DISCUSSION

The present study sought to contribute with a novel approach to spatial adaptive MR data normalization between the 3T and 7T MR modalities, an increasing challenge in medical imaging. Using CycleGAN, an unsupervised generative adversarial network, our model demonstrated promising performance in generating clear 7T-like MR images from 3T inputs, evidenced by the high Dice and PAD (Taha & Hanbury, 2015;

Zou et al., 2004). These scores affirm the model's competence in maintaining the original morphological features of the brain and preserving the clinical relevance of the images, which has a significant implication in the advancement of neurological research (Moeskops et al., 2016; Ronneberger et al., 2015).

The successful performance of the model, especially in tasks that require global contrast properties, supports the hypothesis that our approach enables a seamless transition to 7T MR systems without jeopardizing the quality of previously obtained 3T data (Ashburner & Friston, 2000; Johansen-Berg & Behrens, 2006). This holds promise for facilitating consistent and robust data analysis, mitigating potential bias and loss of statistical power associated with data from difference MRI field strengths in longitudinal studies (Schaer et al., 2008; Little & Rubin, 2019).

Previous research efforts have probed into creating 7T images using advanced deep-learning methodologies (Hou et al., 2016; Qin et al., 2019). However, our model stands out due to its unsupervised learning aspect and its ability to function without resorting to information in the frequency domain (Klosowski & Frahm, 2017; Huang et al., 2019). The effectiveness of the CycleGAN model in minimizing image variance underlines its potential to tackle the challenge of cross-modality image translation between different MRI field strengths (Adriany et al., 2008; Van der Velden et al., 2015).

Our study offers a promising approach to cross-modality MR image translation, allowing the efficient utilization of 3T and 7T MRI technologies in longitudinal studies examining brain health and pathology (Uğurbil, 2014; Keuken & Forstmann, 2015). The strategy we propose

constitutes a progress towards exploiting the capabilities of cutting-edge neuroimaging technologies while preserving the value of existing imaging data. This spatial adaptive normalization tactic helps bridge the technological gap, potentially accelerating the identification and validation of imaging biomarkers for neurological conditions (Bourgeat et al., 2015; W. Zhang et al., 2015).

Our research has made a significant step forward in advancing the application of CycleGAN to transform 3T to 7T MR data. However, the success of the translation task is highly dependent on various factors, including the quality and nature of the input data, the model architecture, and the training procedure. Despite our progress, it is essential to consider the following factors as potential limitations that can impact the results.

One of the primary limitations concerns the 2D nature of the implemented model. Adopting a 2D CycleGAN model was primarily due to computational considerations; however, the inherent lack of a three-dimensional context by independently processing each slice results in slice-to-slice inconsistencies (Çiçek, Abdulkadir, Lienkamp, Brox, & Ronneberger, 2016; Milletari, Navab, & Ahmadi, 2016). Furthermore, patch-based image processing can result in loss of local spatial information, which is critical in medical image segmentation (Tajbakhsh et al., 2020). These inconsistencies may impact the quality and fidelity of the synthesized 7T images. As Yushkevich et al., (2006) have demonstrated, 3D analysis can significantly enhance image interpretation and extract critical information from complex structures. Although implementing 3D CycleGAN architecture would alleviate this concern, the extensive computational demand often hinders its deployment.

The second potential pitfall is associated with an intrinsic property of CycleGANs, namely the non-bijective or many-to-many mapping between domains (Zhu et al., 2017). This can lead to struggles with controlling the mode of output (Ghosh, Kulharia, Nambodiri, Torr, & Dokania, 2018) and non-functional transformations, which introduce potentially unrealistic information, or "hallucinations," into the synthetic images (Almahairi, Rajeshwar, Sordani, Bachman, & Courville, 2018). These limitations and the difficulty of maintaining the semantic consistency of anatomical structures in synthesized images (Chartsias, Joyce, Giuffrida, & Tsiftaris, 2018) hinder their diagnostic applicability (G. Wang et al., 2020). Consequently, future research should focus on devising strategies to rectify this challenge, such as regularizing the model with a perceptual loss (Johnson, Alahi, & Fei-Fei, 2016) or improving training procedures to facilitate a better generalization (Roth et al., 2020).

Variations in image quality across different MRI scanners and imaging protocols can also influence our model's performance (Jovicich et al., 2006; Kruggel, Turner, Muftuler, & Alzheimer's Disease Neuroimaging, 2010). This issue is frequently encountered in multi-center studies and can generate images that do not accurately represent real-world variance, reducing its practical use. Therefore, future research is warranted to assess the model's resilience against such variations and verify its performance on diverse datasets, including those featuring pathological changes (van Opbroek, Ikram, Vernooij, & de Bruijne, 2015).

The question of data dependence and the potential for error in translation must also be mentioned. The translation's success relies heavily on the quality of the input data, meaning that errors during preprocessing, like inaccurate skull-stripping or normalization, could adversely impact the output (Iglesias & Sabuncu, 2015). Similarly, the translation process could inadvertently introduce or exaggerate noise or artifacts not present in authentic 7T images, leading to misleading interpretations (Chen & Koltun, 2017).

The performance evaluation of segmentation algorithms plays a significant role in medical imaging. It determines the efficacy and reliability of these algorithms in clinical practice. We employed the Dice and PAD scores. Despite its widespread usage, the Dice score has a few limitations. One major drawback is that it needs to account for the geometric correspondence between the predicted and ground truth segments (Taha & Hanbury, 2015). It does not consider how well the predicted segment's shape, location, and orientation match the ground truth. In some scenarios, this information is critical, mainly when the size and position of the segment are clinically significant (e.g., tumor segmentation) (Crum, Camara, & Hill, 2006).

Conversely, the PAD evaluates the size estimation accuracy of segmentation results. While PAD effectively measures size discrepancies, it does not provide information on the spatial overlap between the predicted and the ground truth segmentation. Thus, two segmentations with the same area but in different locations would yield a zero PAD, failing to account for the spatial mismatch (Isensee, Kickingereder, Wick, Bendszus, & Maier-Hein, 2017). This limitation is critical in situations where the accurate localization of the object of interest is essential.

4.1 | Conclusion and future work

Our study underscores the transformative potential of blending advanced machine learning techniques with clinical applications, particularly within neuroimaging (Obermeyer & Emanuel, 2016; Esteva et al., 2019). Our results provide a powerful testimony to the potential of CycleGAN in handling unpaired data, thus offering a compelling solution to MR translation across different field strengths.

However, despite its merits, the CycleGAN model has limitations, and these challenges delineate promising avenues for future enhancement. Our study used generators based on the U-Net architecture (Ronneberger et al., 2015), which have delivered excellent results in medical image segmentation tasks. However, U-Net-based models primarily exploit local dependencies in the data, potentially neglecting long-range relationships between pixels. In contrast, with the advent of Vision Transformers (ViTs), there is potential to enhance our model's performance further (Dosovitskiy et al., 2020). ViTs can capture these intricate, long-range dependencies, which could be highly advantageous for tasks involving complex morphological structures such as brain MR images (T.-C. Wang et al., 2018). This, in turn, could improve the segmentation performance, particularly for brain images that exhibit non-local relationships between anatomical structures (Petit et al., 2021).

ACKNOWLEDGMENTS

The National Institutes of Health supported this work under award numbers: R01MH111265, R01AG063525, R01MH076079, and RF1AG025516. This research was partly supported by the University of Pittsburgh Center for Research Computing, RRID:SCR_022735, through the resources provided. Specifically, this work used the H2P cluster, supported by NSF award number OAC-2117681. The Human Connectome Project, WU-Minn Consortium partly provided 3T data (Principal Investigators: David Van Essen and Kamil Uğurbil; 1U54MH091657) funded by NIH and Washington University. CNPq partially supported Eduardo Diniz under the award: Ph.D. Scholarship Ciência sem Fronteiras, Brazil (#210150/2014-9).

DATA AVAILABILITY

The trained model and code are available upon request.

FINANCIAL DISCLOSURE

None reported.

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Devin, M. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems [Journal Article]. arXiv preprint. arXiv:1603.04467.
- Adriany, G., Van de Moortele, P.-F., Ritter, J., Moeller, S., Auerbach, E. J., Akgün, C., ... Uğurbil, K. (2008). A geometrically adjustable 16-channel transmit/receive transmission line array for improved rf efficiency and parallel imaging performance at 7 tesla [Journal Article]. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 59(3), 590–597.
- Almahairi, A., Rajeshwar, S., Sordani, A., Bachman, P., & Courville, A. (2018). Augmented cyclegan: Learning many-to-many mappings from unpaired data [Conference Proceedings]. In *International conference on machine learning* (pp. 195–204). PMLR.
- Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein generative adversarial networks [Conference Proceedings]. In *International conference on machine learning* (pp. 214–223). PMLR.
- Arnold, J. B., Liow, J. S., Schaper, K. A., Stern, J. J., Sled, J. G., Shattuck, D. W., ... Rottenberg, D. A. (2001). Qualitative and quantitative evaluation of six algorithms for correcting intensity nonuniformity effects [Journal Article]. *Neuroimage*, 13(5), 931–43. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/11304088> doi: 10.1006/nimg.2001.0756
- Ashburner, J. (2007). A fast diffeomorphic image registration algorithm [Journal Article]. *Neuroimage*, 38(1), 95–113. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/17761438> doi: 10.1016/j.neuroimage.2007.07.007
- Ashburner, J., & Friston, K. J. (2000). Voxel-based morphometry—the methods [Journal Article]. *Neuroimage*, 11(6 Pt 1), 805–21. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/10860804> doi: 10.1006/nimg.2000.0582
- Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures [Book Section]. In *Neural networks: Tricks of the trade: Second edition* (pp. 437–478). Springer.
- Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4) [Book]. Springer.
- Bohr, A., & Memarzadeh, K. (2020). *Artificial intelligence in healthcare* [Book]. Academic Press.
- Bourgeat, P., Villemagne, V. L., Dore, V., Brown, B., Macaulay, S. L., Martins, R., ... Group, A. R. (2015). Comparison of mr-less pib suvr quantification methods [Journal Article]. *Neurobiol Aging*, 36 Suppl 1, S159–66. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/25257985> doi: 10.1016/j.neurobiolaging.2014.04.033
- Cai, J., Zhang, Z., Cui, L., Zheng, Y., & Yang, L. (2019). Towards cross-modal organ translation and segmentation: A cycle- and shape-consistent generative adversarial network [Journal Article]. *Med Image Anal*, 52, 174–184. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/30594770> doi: 10.1016/j.media.2018.12.002
- Cao, X., Yang, J., Zhang, J., Nie, D., Kim, M., Wang, Q., & Shen, D. (2017). Deformable image registration based on similarity-steered cnn regression [Conference Proceedings]. In *Medical image computing and computer assisted intervention—miccai 2017: 20th international conference, quebec city, qc, canada, september 11–13, 2017, proceedings, part i 20* (pp. 300–308). Springer.
- Çiçek, O., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3d u-net: learning dense volumetric segmentation from sparse annotation [Conference Proceedings]. In *Medical image computing and computer-assisted intervention—miccai 2016: 19th international conference, athens, greece, october 17–21, 2016, proceedings, part ii 19* (pp. 424–432). Springer.
- Chartsias, A., Joyce, T., Giuffrida, M. V., & Tsaftaris, S. A. (2018). Multimodal mr synthesis via modality-invariant latent representation [Journal Article]. *IEEE Trans Med Imaging*, 37(3), 803–814. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/29053447> doi: 10.1109/TMI.2017.2764326
- Chen, Q., & Koltun, V. (2017). Photographic image synthesis with cascaded refinement networks [Conference Proceedings]. In *Proceedings of the IEEE international conference on computer vision* (pp. 1511–1520).
- Crum, W. R., Camara, O., & Hill, D. L. (2006). Generalized overlap measures for evaluation and validation in medical image analysis [Journal Article]. *IEEE Trans Med Imaging*, 25(11), 1451–61. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/17117774> doi: 10.1109/TMI.2006.880587
- Dice, L. R. (1945). Measures of the amount of ecologic association

- between species [Journal Article]. *Ecology*, 26(3), 297–302.
- Diniz, E., Helmet, K., Santini, T., Aizenstein, H., & Ibrahim, T. (2021). Image-to-image translation of 3t to 7t mri using generative adversarial networks: A step towards longitudinal harmonization [Conference Proceedings]. In *International society for magnetic resonance in medicine annual meeting*.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... Gelly, S. (2020). *An image is worth 16x16 words: Transformers for image recognition at scale* [Journal Article]. arXiv preprint. arXiv:2010.11929.
- Dowson, D., & Landau, B. (1982). The fréchet distance between multivariate normal distributions [Journal Article]. *Journal of multivariate analysis*, 12(3), 450–455.
- Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... Dean, J. (2019). A guide to deep learning in healthcare [Journal Article]. *Nat Med*, 25(1), 24–29. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/30617335> doi: 10.1038/s41591-018-0316-z
- Evans, A. C., Janke, A. L., Collins, D. L., & Baillet, S. (2012). Brain templates and atlases [Journal Article]. *Neuroimage*, 62(2), 911–22. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/22248580> doi: 10.1016/j.neuroimage.2012.01.024
- Fortin, J. P., Cullen, N., Sheline, Y. I., Taylor, W. D., Aselcioglu, I., Cook, P. A., ... Shinohara, R. T. (2018). Harmonization of cortical thickness measurements across scanners and sites [Journal Article]. *Neuroimage*, 167, 104–120. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/29155184> doi: 10.1016/j.neuroimage.2017.11.024
- Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks [Conference Proceedings]. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 2414–2423).
- Ghosh, A., Kulharia, V., Nambodiri, V. P., Torr, P. H., & Dokania, P. K. (2018). Multi-agent diverse generative adversarial networks [Conference Proceedings]. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 8513–8521).
- Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks [Conference Proceedings]. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249–256). JMLR Workshop and Conference Proceedings.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning* [Book]. MIT press.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative adversarial nets [Journal Article]. *Advances in neural information processing systems*, 27.
- Guerrero, R., Qin, C., Oktay, O., Bowles, C., Chen, L., Joules, R., ... Rueckert, D. (2018). White matter hyperintensity and stroke lesion segmentation and differentiation using convolutional neural networks [Journal Article]. *Neuroimage Clin*, 17, 918–934. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/29527496> doi: 10.1016/j.nicl.2017.12.022
- Hagmann, P., Jonasson, L., Maeder, P., Thiran, J. P., Wedeen, V. J., & Meuli, R. (2006). Understanding diffusion mr imaging techniques: from scalar diffusion-weighted imaging to diffusion tensor imaging and beyond [Journal Article]. *Radiographics*, 26(Suppl 1), S205–23. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/17050517> doi: 10.1148/rg.26si065510
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition [Conference Proceedings]. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 770–778).
- Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium [Journal Article]. *Advances in neural information processing systems*, 30.
- Hou, L., Samaras, D., Kurc, T. M., Gao, Y., Davis, J. E., & Saltz, J. H. (2016). Patch-based convolutional neural network for whole slide tissue image classification [Conference Proceedings]. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 2424–2433).
- Huang, X., Sun, W., Tseng, T. B., Li, C., & Qian, W. (2019). Fast and fully-automated detection and segmentation of pulmonary nodules in thoracic ct scans using deep convolutional neural networks [Journal Article]. *Comput Med Imaging Graph*, 74, 25–36. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/30954678> doi: 10.1016/j.compmedimag.2019.02.003
- Iglesias, J. E., & Sabuncu, M. R. (2015). Multi-atlas segmentation of biomedical images: A survey [Journal Article]. *Med Image Anal*, 24(1), 205–219. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/26201875> doi: 10.1016/j.media.2015.06.012
- Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., & Maier-Hein, K. H. (2017). Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge [Conference Proceedings]. In *Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries* (pp. 287–297). Springer. Revised Selected Papers 3.
- Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks [Conference Proceedings]. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 1125–1134).
- Johansen-Berg, H., & Behrens, T. E. (2006). Just pretty pictures? what diffusion tractography can add in clinical neuroscience [Journal Article]. *Current opinion in neurology*, 19(4), 379.
- Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution [Conference Proceedings]. In *Computer vision-eccv 2016: 14th european conference, amsterdam, the netherlands, october 11-14, 2016, proceedings, part ii 14* (pp. 694–711). Springer.
- Jovicich, J., Czanner, S., Greve, D., Haley, E., van der Kouwe, A., Gollub, R., ... Dale, A. (2006). Reliability in multi-site structural mri studies:

- effects of gradient non-linearity correction on phantom and human data [Journal Article]. *Neuroimage*, 30(2), 436–43. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/16300968> doi: 10.1016/j.neuroimage.2005.09.046
- Keuken, M. C., & Forstmann, B. U. (2015). A probabilistic atlas of the basal ganglia using 7 t mri [Journal Article]. *Data Brief*, 4, 577–82. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/26322322> doi: 10.1016/j.dib.2015.07.028
- Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization* [Journal Article]. arXiv preprint. arXiv:1412.6980.
- Klosowski, J., & Frahm, J. (2017). Image denoising for real-time mri [Journal Article]. *Magn Reson Med*, 77(3), 1340–1352. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/27079944> doi: 10.1002/mrm.26205
- Krishnamurthy, N., Santini, T., Wood, S., Kim, J., Zhao, T., Aizenstein, H. J., & Ibrahim, T. S. (2019). Computational and experimental evaluation of the tic-tac-toe rf coil for 7 tesla mri [Journal Article]. *PLoS One*, 14(1), e0209663. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/30629618> doi: 10.1371/journal.pone.0209663
- Kruggel, F., Turner, J., Muftuler, L. T., & Alzheimer's Disease Neuroimaging, I. (2010). Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the adni cohort [Journal Article]. *Neuroimage*, 49(3), 2123–33. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/19913626> doi: 10.1016/j.neuroimage.2009.11.006
- LeCun, Y., Bottou, L., Orr, G. B., & Müller, K.-R. (2002). Efficient backprop [Book Section]. In *Neural networks: Tricks of the trade* (pp. 9–50). Springer.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciampi, F., Ghafoorian, M., ... Sanchez, C. I. (2017). A survey on deep learning in medical image analysis [Journal Article]. *Med Image Anal*, 42, 60–88. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/28778026> doi: 10.1016/j.media.2017.07.005
- Little, R. J., & Rubin, D. B. (2019). *Statistical analysis with missing data* (Vol. 793) [Book]. John Wiley & Sons.
- Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation [Conference Proceedings]. In *2016 fourth international conference on 3d vision (3dv)* (pp. 565–571). IEEE.
- Moeskops, P., Viergever, M. A., Mendrik, A. M., de Vries, L. S., Benders, M. J., & Išgum, I. (2016). Automatic segmentation of mr brain images with a convolutional neural network [Journal Article]. *IEEE Trans Med Imaging*, 35(5), 1252–1261. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/27046893> doi: 10.1109/TMI.2016.2548501
- Ng, A. Y. (2004). Feature selection, l_1 vs. l_2 regularization, and rotational invariance [Conference Proceedings]. In *Proceedings of the twenty-first international conference on machine learning* (p. 78).
- Nie, D., Trullo, R., Lian, J., Wang, L., Petitjean, C., Ruan, S., ... Shen, D. (2018). Medical image synthesis with deep convolutional adversarial networks [Journal Article]. *IEEE Trans Biomed Eng*, 65(12), 2720–2730. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/29993445> doi: 10.1109/TBME.2018.2814538
- Nyul, L. G., & Udupa, J. K. (1999). On standardizing the mr image intensity scale [Journal Article]. *Magn Reson Med*, 42(6), 1072–81. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/10571928> doi: 10.1002/(sici)1522-2594(199912)42:6<1072::aid-mrm11>3.0.co;2-m
- Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future – big data, machine learning, and clinical medicine [Journal Article]. *N Engl J Med*, 375(13), 1216–9. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/27682033> doi: 10.1056/NEJMp1606181
- Obusez, E. C., Lowe, M., Oh, S.-H., Wang, I., Bullen, J., Ruggieri, P., ... others (2018). 7t mr of intracranial pathology: preliminary observations and comparisons to 3t and 1.5 t. *Neuroimage*, 168, 459–476.
- Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical parametric mapping: the analysis of functional brain images* [Book]. Elsevier.
- Petit, O., Thome, N., Rambour, C., Themyr, L., Collins, T., & Soler, L. (2021). U-net transformer: Self and cross attention for medical image segmentation [Conference Proceedings]. In *Machine learning in medical imaging: 12th international workshop, mlmi 2021, held in conjunction with miccai 2021, strasbourg, france, september 27, 2021, proceedings 12* (pp. 267–276). Springer.
- Prechelt, L. (2002). Early stopping—but when? [Book Section]. In *Neural networks: Tricks of the trade* (pp. 55–69). Springer.
- Qin, C., Schlemper, J., Caballero, J., Price, A. N., Hajnal, J. V., & Rueckert, D. (2019). Convolutional recurrent neural networks for dynamic mr image reconstruction [Journal Article]. *IEEE Trans Med Imaging*, 38(1), 280–290. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/30080145> doi: 10.1109/TMI.2018.2863670
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation [Conference Proceedings]. In *Medical image computing and computer-assisted intervention—miccai 2015: 18th international conference, munich, germany, october 5–9, 2015, proceedings, part iii 18* (pp. 234–241). Springer.
- Ross, D. E., Ochs, A. L., Seabaugh, J. M., Demark, M. F., Shrader, C. R., Marwitz, J. H., & Havranek, M. D. (2012). Progressive brain atrophy in patients with chronic neuropsychiatric symptoms after mild traumatic brain injury: a preliminary study [Journal Article]. *Brain Inj*, 26(12), 1500–9. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/22721509> doi: 10.3109/02699052.2012.694570
- Roth, K., Milbich, T., Sinha, S., Gupta, P., Ommer, B., & Cohen, J. P. (2020). Revisiting training strategies and generalization performance in deep metric learning [Conference Proceedings]. In *International conference on machine learning* (pp. 8242–8252). PMLR.
- Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training gans [Journal

- Article]. *Advances in neural information processing systems*, 29.
- Santini, T., Koo, M., Farhat, N., Campos, V. P., Alkhateeb, S., Vieira, M. A. C., ... Ibrahim, T. S. (2021). Analysis of hippocampal subfields in sickle cell disease using ultrahigh field mri [Journal Article]. *Neuroimage Clin*, 30, 102655. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/34215139> doi: 10.1016/j.nicl.2021.102655
- Santini, T., Wood, S., Krishnamurthy, N., Martins, T., Aizenstein, H. J., & Ibrahim, T. S. (2021). Improved 7 tesla transmit field homogeneity with reduced electromagnetic power deposition using coupled tic tac toe antennas [Journal Article]. *Sci Rep*, 11(1), 3370. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/33564013> doi: 10.1038/s41598-020-79807-9
- Santini, T., Zhao, Y., Wood, S., Krishnamurthy, N., Kim, J., Farhat, N., ... Ibrahim, T. S. (2018). In-vivo and numerical analysis of the eigenmodes produced by a multi-level tic-tac-toe head transmit array for 7 tesla mri [Journal Article]. *PLoS One*, 13(11), e0206127. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/30481187> doi: 10.1371/journal.pone.0206127
- Schaer, M., Cuadra, M. B., Tamarit, L., Lazeyras, F., Eliez, S., & Thiran, J. P. (2008). A surface-based approach to quantify local cortical gyrification [Journal Article]. *IEEE Trans Med Imaging*, 27(2), 161-70. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/18334438> doi: 10.1109/TMI.2007.903576
- Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis [Journal Article]. *Annu Rev Biomed Eng*, 19, 221-248. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/28301734> doi: 10.1146/annurev-bioeng-071516-044442
- Smith, S. M. (2002). Fast robust automated brain extraction [Journal Article]. *Hum Brain Mapp*, 17(3), 143-55. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/12391568> doi: 10.1002/hbm.10062
- Sotiropoulos, S. N., Jbabdi, S., Xu, J., Andersson, J. L., Moeller, S., Auerbach, E. J., ... Consortium, W. U.-M. H. (2013). Advances in diffusion mri acquisition and processing in the human connectome project [Journal Article]. *Neuroimage*, 80, 125-43. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/23702418> doi: 10.1016/j.neuroimage.2013.05.057
- Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool [Journal Article]. *BMC Med Imaging*, 15(1), 29. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/26263899> doi: 10.1186/s12880-015-0068-x
- Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J. N., Wu, Z., & Ding, X. (2020). Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation [Journal Article]. *Med Image Anal*, 63, 101693. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/32289663> doi: 10.1016/j.media.2020.101693
- Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4itk: improved n3 bias correction [Journal Article]. *IEEE Trans Med Imaging*, 29(6), 1310-20. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/20378467> doi: 10.1109/TMI.2010.2046908
- Ulyanov, D., Vedaldi, A., & Lempitsky, V. (2016). Instance normalization: The missing ingredient for fast stylization [Journal Article]. arXiv preprint. arXiv:1607.08022.
- Uğurbil, K. (2014). Magnetic resonance imaging at ultrahigh fields [Journal Article]. *IEEE transactions on biomedical engineering*, 61(5), 1364-1379.
- Vallender, S. (1974). Calculation of the wasserstein distance between probability distributions on the line [Journal Article]. *Theory of Probability & Its Applications*, 18(4), 784-786.
- Van der Velden, T. A., Italiaander, M., van der Kemp, W. J., Raaijmakers, A. J., Schmitz, A. T., Luijten, P. R., ... Klomp, D. W. (2015). Radiofrequency configuration to facilitate bilateral breast 31p mr spectroscopic imaging and high-resolution mri at 7 tesla [Journal Article]. *Magnetic resonance in medicine*, 74(6), 1803-1810.
- van Opbroek, A., Ikram, M. A., Vernooij, M. W., & de Bruijne, M. (2015). Transfer learning improves supervised image segmentation across imaging protocols [Journal Article]. *IEEE Trans Med Imaging*, 34(5), 1018-30. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/25376036> doi: 10.1109/TMI.2014.2366792
- van Osch, M. J., & Webb, A. G. (2014). Safety of ultra-high field mri: What are the specific risks? [Journal Article]. *Current Radiology Reports*, 2, 1-8.
- Vaughan, J. T., Garwood, M., Collins, C. M., Liu, W., DelaBarre, L., Adriany, G., ... Ugurbil, K. (2001). 7t vs. 4t: Rf power, homogeneity, and signal-to-noise comparison in head images [Journal Article]. *Magn Reson Med*, 46(1), 24-30. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/11443707> doi: 10.1002/mrm.1156
- Vovk, U., Pernus, F., & Likar, B. (2007). A review of methods for correction of intensity inhomogeneity in mri [Journal Article]. *IEEE Trans Med Imaging*, 26(3), 405-21. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/17354645> doi: 10.1109/TMI.2006.891486
- Wang, G., Gong, E., Banerjee, S., Martin, D., Tong, E., Choi, J., ... Zaharchuk, G. (2020). Synthesize high-quality multi-contrast magnetic resonance imaging from multi-echo acquisition using multi-task deep generative model [Journal Article]. *IEEE Trans Med Imaging*, 39(10), 3089-3099. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/32286966> doi: 10.1109/TMI.2020.2987026
- Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., & Catanzaro, B. (2018). High-resolution image synthesis and semantic manipulation with conditional gans [Conference Proceedings]. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 8798-8807).
- Wolterink, J. M., Leiner, T., Viergever, M. A., & Isgum, I. (2017). Generative adversarial networks for noise reduction in low-dose ct [Journal Article]. *IEEE Trans Med Imaging*, 36(12), 2536-2545. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/28574346> doi: 10.1109/TMI.2017.2708987

- Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? [Journal Article]. *Advances in neural information processing systems*, 27.
- Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., & Gerig, G. (2006). User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability [Journal Article]. *Neuroimage*, 31(3), 1116–1128.
- Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., & Shen, D. (2015). Deep convolutional neural networks for multi-modality isointense infant brain image segmentation [Journal Article]. *Neuroimage*, 108, 214–24. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/25562829> doi: 10.1016/j.neuroimage.2014.12.061
- Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm [Journal Article]. *IEEE transactions on medical imaging*, 20(1), 45–57.
- Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks [Conference Proceedings]. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223–2232).
- Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M., Kaus, M. R., Haker, S. J., ... Kikinis, R. (2004). Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports [Journal Article]. *Academic radiology*, 11(2), 178–189.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.