Image Synthesis in Multi-Contrast MRI With Conditional Generative Adversarial Networks

Salman UH. Dar, *Student Member, IEEE*, Mahmut Yurt, Levent Karacan, Aykut Erdem,
Erkut Erdem, and Tolga Çukur, *Senior Member, IEEE*

Abstract—Acquiring images of the same anatomy with multiple different contrasts increases the diversity of diagnostic information available in an MR exam. Yet, the scan time limitations may prohibit the acquisition of certain contrasts, and some contrasts may be corrupted by noise and artifacts. In such cases, the ability to synthesize unacquired or corrupted contrasts can improve diagnostic utility. For multi-contrast synthesis, the current methods learn a nonlinear intensity transformation between the source and target images, either via nonlinear regression or deterministic neural networks. These methods can, in turn, suffer from the loss of structural details in synthesized images. Here, in this paper, we propose a new approach for multi-contrast MRI synthesis based on conditional generative adversarial networks. The proposed approach preserves intermediate-to-high frequency details via an adversarial loss, and it offers enhanced synthesis performance via pixel-wise and perceptual losses for registered multi-contrast images and a cycle-consistency loss for unregistered images. Information from neighboring cross-sections are utilized to further improve synthesis quality. Demonstrations on T_1 - and T_2- weighted images from healthy subjects and patients clearly indicate the superior performance of the proposed approach compared to the previous state-of-the-art methods. Our synthesis approach can help improve the quality and versatility of the multi-contrast MRI exams without the need for prolonged or repeated examinations.

Index Terms—Generative adversarial network, image synthesis, multi-contrast MRI, pixel-wise loss, cycle-consistency loss.

I. INTRODUCTION

MAGNETIC resonance imaging (MRI) is pervasively used in clinical applications due to the diversity of contrasts it can capture in soft tissues. Tailored MRI pulse sequences enable the generation of distinct contrasts while imaging the same anatomy. For instance, T_1-weighted brain images clearly delineate gray and white matter tissues, whereas T_2-weighted images delineate fluid from cortical tissue. In turn, multi-contrast images acquired in the same subject increase the diagnostic information available in clinical and research studies. However, it may not be possible to collect a full array of contrasts given considerations related to the cost of prolonged exams and uncooperative patients, particularly in pediatric and elderly populations [1]. In such cases, acquisition of contrasts with relatively shorter scan times might be preferred. Even then a subset of the acquired contrasts can be corrupted by excessive noise or artifacts that prohibit subsequent diagnostic use [2]. Moreover, cohort studies often show significant heterogeneity in terms of imaging protocol and the specific contrasts that they acquire [3]. Thus, the ability to synthesize missing or corrupted contrasts from other successfully acquired contrasts has potential value for enhancing multi-contrast MRI by increasing availability of diagnostically-relevant images, and improving analysis tasks such as registration and segmentation [4].

Cross-domain synthesis of medical images has recently been gaining popularity in medical imaging. Given a subject's image x in X (source domain), the aim is to accurately estimate the respective image of the same subject y in Y (target domain). Two main synthesis approaches are registration-based [5]–[7] and intensity-transformation-based methods [8]–[24]. Registration-based methods start by generating an atlas based on a co-registered set of images, x_1 and y_1, respectively acquired in X and Y [5]. These methods further make the assumption that within-domain images from separate subjects are related to each other through a geometric warp. For synthesizing y_2 from x_2, the warp that transforms x_1 to x_2 is estimated, and this warp is then applied

on y_1. Since they only rely on geometric transformations, registration-based methods that rely on a single atlas can suffer from across-subject differences in underlying morphology [23]. For example, inconsistent pathology across a test subject and the atlas can cause failure. Multi-atlas registration in conjunction with intensity fusion can alleviate this limitation, and has been successfully used in synthesizing CT from MR images [6], [7]. Nevertheless, within-domain registration accuracy might still be limited even in normal subjects [23].

An alternative is to use intensity-based methods that do not rely on a strict geometric relationship among different subjects' anatomies [8]–[24]. One powerful approach for multi-contrast MRI is based on the compressed sensing framework, where each patch in the source image x_2 is expressed as a sparse linear combination of patches in the atlas image x_1 [10], [22]. The learned sparse combinations are then applied to estimate patches in y_2 from patches in y_1. To improve matching of patches across domains, generative models were also proposed that use multi-scale patches and tissue segmentation labels [16], [18]. Instead of focusing on linear models, recent studies aimed to learn more general non-linear mappings that express individual voxels in y_1 in terms of patches in x_1, and then predict y_2 from x_2 based on these mappings. Nonlinear mappings are learned on training data via techniques such as nonlinear regression [8], [9], [23] or location-sensitive neural networks [19]. An important example is Replica that performs random forest regression on multiresolution image patches [23]. Replica demonstrates great promise in multi-contrast MR image synthesis. However, dictionary construction at different spatial scales is independent, and the predictions from separate random forest trees are averaged during synthesis. These may lead to loss of detailed structural information and suboptimal synthesis performance.

Recently an end-to-end framework for MRI image synthesis has been proposed, Multimodal, based on deep neural networks [21]. Multimodal trains a neural network that receives as input images in multiple source contrasts and predicts the image in the target contrast. This method performs multiresolution dictionary construction and image synthesis in a unified framework, and it was demonstrated to yield higher synthesis quality compared to non-network-based approaches even when only a subset of the source contrasts is available. That said, Multimodal assumes the availability of spatially-registered multi-contrast images. In addition, Multimodal uses mean absolute error loss functions that can perform poorly in capturing errors towards higher spatial frequencies [25]–[27].

Here we propose a novel approach for image synthesis in multi-contrast MRI based on generative adversarial network (GAN) architectures. Adversarial loss functions have recently been demonstrated for various medical imaging applications with reliable capture of high-frequency texture information [28]–[48]. In the domain of cross-modality image synthesis, important applications include CT to PET synthesis [29], [40], MR to CT synthesis [28], [33], [38], [42], [48], CT to MR synthesis [36], and retinal vessel map to image synthesis [35], [41]. Inspired by this success, here we introduce conditional GAN models for synthesizing images of
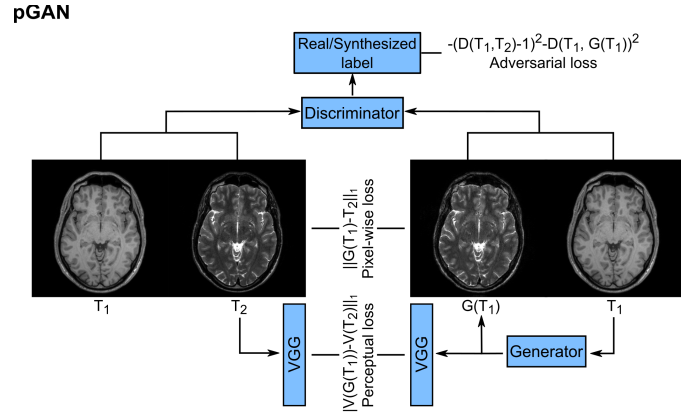
pGAN

Fig. 1. The pGAN method is based on a conditional adversarial network with a generator G, a pre-trained VGG16 network V, and a discriminator D. Given an input image in a source contrast (e.g., T_1-weighted), G learns to generate the image of the same anatomy in a target contrast (e.g., T_2-weighted). Meanwhile, D learns to discriminate between synthetic (e.g., T_1–$G(T_1)$) and real (e.g., T_1–T_2) pairs of multi-contrast images. Both subnetworks are trained simultaneously, where G aims to minimize a pixel-wise, a perceptual and an adversarial loss function, and D tries to maximize the adversarial loss function.

distinct contrasts from a single modality, with demonstrations on multi-contrast brain MRI in normal subjects and glioma patients. For improved accuracy, the proposed method also leverages correlated information across neighboring cross-sections within a volume. Two implementations are provided for use when multi-contrast images are spatially registered (pGAN) and when they are unregistered (cGAN). For the first scenario, we train pGAN with pixel-wise loss and perceptual loss between the synthesized and true images (Fig. 1) [25], [49]. For the second scenario, we train cGAN after replacing the pixel-wise loss with a cycle loss that enforces the ability to reconstruct back the source image from the synthesized target image (Fig. 2) [50]. Extensive evaluations are presented on multi-contrast MRI images (T_1- and T_2-weighted) from healthy normals and glioma patients. The proposed approach yields visually and quantitatively enhanced accuracy in multi-contrast MRI synthesis compared to state-of-the-art methods (Replica and Multimodal) [21], [23].

II. METHODS

A. Image Synthesis via Adversarial Networks

Generative adversarial networks are neural-network architectures that consist of two sub-networks; G, a generator and D, a discriminator. G learns a mapping from a latent variable z (typically random noise) to an image y in a target domain, and D learns to discriminate the generated image $G(z)$ from the real image y [51]. During training of a GAN, both G and D are learned simultaneously, with G aiming to generate images that are indistinguishable from the real images, and D aiming to tell apart generated and real images. To do this, the following adversarial loss function (L_{GAN}) can be used:

$$L_{GAN}(G, D) = E_y[\log D(y)] + E_z[\log(1 - D(G(z)))], \quad (1)$$

where E denotes expected value. G tries to minimize and D tries to maximize the adversarial loss that improves
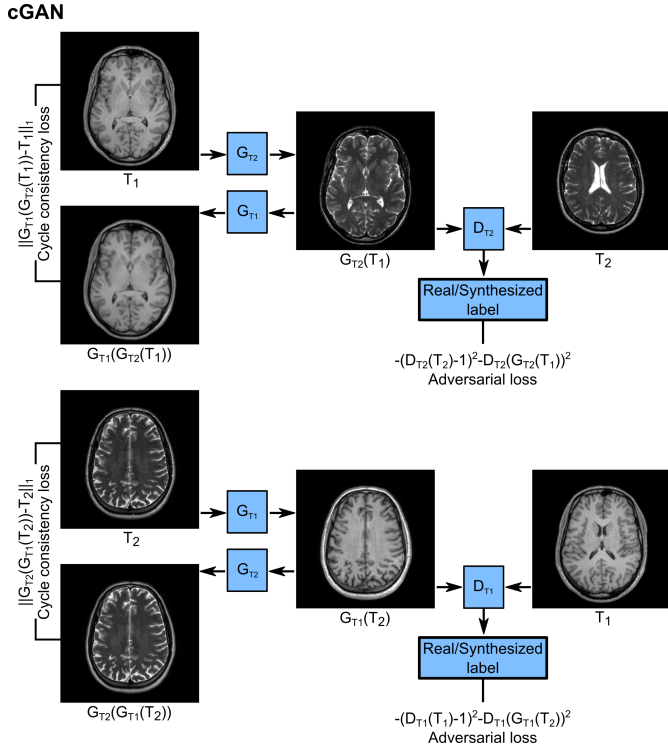
cGAN

Fig. 2. The cGAN method is based on a conditional adversarial network with two generators (G_{T1}, G_{T2}) and two discriminators (D_{T1}, D_{T2}). Given a T_1-weighted image, G_{T2} learns to generate the respective T_2-weighted image of the same anatomy that is indiscriminable from real T_2-weighted images of other anatomies, whereas D_{T2} learns to discriminate between synthetic and real T_2-weighted images. Similarly, G_{T1} learns to generate realistic a T_1-weighted image of an anatomy given the respective T_2-weighted image, whereas D_{T1} learns to discriminate between synthetic and real T_1-weighted images. Since the discriminators do not compare target images of the same anatomy, a pixel-wise loss cannot be used. Instead, a cycle-consistency loss is utilized to ensure that the trained generators enable reliable recovery of the source image from the generated target image.

modeling high-spatial-frequency information [26]. Both G and D are trained simultaneously. Upon convergence, G is capable of producing realistic counterfeit images that D cannot recognize [51]. To further stabilize the training process, the negative log-likelihood cost for adversarial loss in (1) can be replaced by a squared loss [52]:

$$L_{GAN}(D, G) = -E_y[(D(y) - 1)^2] - E_z[D(G(z))^2] \quad (2)$$

Recent studies in computer vision have demonstrated that GANs are very effective in image-to-image translation tasks [49], [50]. Image-to-image translation concerns transformations between different representations of the same underlying visual scene [49]. These transformations can be used to convert an image between separate domains, e.g., generating semantic segmentation maps from images, colored images from sketches, or maps from aerial photos [49], [53], [54]. Traditional GANs learn to generate samples of images from noise. However, in image-to-image translation, the synthesized image has statistical dependence on the source image. To better capture this dependency, conditional GANs can be employed that receive the source image as an additional input [55]. The resulting network can then be trained based on the following

adversarial loss function:

$$L_{condGAN}(D, G) = -E_{x,y}[(D(x, y) - 1)^2]$$
$$-E_{x,z}[D(x, G(x, z))^2], \quad (3)$$

where x denotes the source image.

An analogous problem to image-to-image translation tasks in computer vision exists in MR imaging where the same anatomy is acquired under multiple different tissue contrasts (e.g., T_1- and T_2-weighted images). Inspired by the recent success of adversarial networks, here we employed conditional GANs to synthesize MR images of a target contrast given as input an alternate contrast. For a comprehensive solution, we considered two distinct scenarios for multi-contrast MR image synthesis. First, we assumed that the images of the source and target contrasts are perfectly registered. For this scenario, we propose pGAN that incorporates a pixel-wise loss into the objective function as inspired by the pix2pix architecture [49]:

$$L_{L1}(G) = E_{x,y,z}[\|y - G(x, z)\|_1], \quad (4)$$

where L_{L1} is the pixel-wise L1 loss function. Since the generator G was observed to ignore the latent variable in pGAN, the latent variable was removed from the model.

Recent studies suggest that incorporation of a perceptual loss during network training can yield visually more realistic results in computer vision tasks. Unlike loss functions based on pixel-wise differences, perceptual loss relies on differences in higher feature representations that are often extracted from networks pre-trained for more generic tasks [25]. A commonly used network is VGG-net trained on the ImageNet [56] dataset for object classification. Here, following [25], we extracted feature maps right before the second max-pooling operation of VGG16 pre-trained on ImageNet. The resulting loss function can be written as:

$$L_{Perc}(G) = E_{x,y}[\|V(y) - V(G(x))\|_1], \quad (5)$$

where V is the set of feature maps extracted from VGG16.

To synthesize each cross-section y from x we also leveraged correlated information across neighboring cross-sections by conditioning the networks not only on x but also on the neighboring cross-sections of x. By incorporating the neighboring cross-sections (3), (4) and (5) become:

$$L_{condGAN-k}(D, G) = -E_{\mathbf{x}_k,y}[(D(\mathbf{x}_k, y) - 1)^2]$$
$$-E_{\mathbf{x}_k}[D(\mathbf{x}_k, G(\mathbf{x}_k))^2], \quad (6)$$
$$L_{L1-k}(G) = E_{\mathbf{x}_k,y}[\|y - G(\mathbf{x}_k)\|_1], \quad (7)$$
$$L_{Perc-k}(G) = E_{\mathbf{x}_k,y}[\|V(y) - V(G(\mathbf{x}_k))\|_1], \quad (8)$$

where $\mathbf{x}_k = [x_{-\lfloor\frac{k}{2}\rfloor}, \ldots, x_{-2}, x_{-1}, x, x_{+1}, x_{+2}, \ldots, x_{+\lfloor\frac{k}{2}\rfloor}]$ is a vector consisting of k consecutive cross-sections ranging from $-\lfloor\frac{k}{2}\rfloor$ to $\lfloor\frac{k}{2}\rfloor$, with the cross section x in the middle, and $L_{condGAN-k}$ and L_{L1-k} are the corresponding adversarial and pixel-wise loss functions. This yields the following aggregate loss function:

$$L_{pGAN} = L_{condGAN-k}(D, G) + \lambda L_{L1-k}(G)$$
$$+\lambda_{perc} L_{perc-k}(G), \quad (9)$$

where L_{pGAN} is the complete loss function, λ controls the relative weighing of the pixel-wise loss and λ_{perc} controls the relative weighing of the perceptual loss.

In the second scenario, we did not assume any explicit registration between the images of the source and target contrasts. In this case, the pixel-wise and perceptual losses cannot be leveraged since images of different contrasts are not necessarily spatially aligned. To limit the number of potential solutions for the synthesized image, here we proposed cGAN that incorporates a cycle-consistency loss as inspired by the cycleGAN architecture [50]. The cGAN method consists of two generators (G_x, G_y) and two discriminators (D_x, D_y). G_y tries to generate $G_y(x)$ that looks similar to y and D_y tries to distinguish $G_y(x)$ from the images y. On the other hand, G_x tries to generate $G_x(y)$ that looks similar to x and D_x tries to distinguish $G_x(y)$ from the images x. This architecture incorporates an additional loss to ensure that the input and target images are consistent with each other, called the cycle consistency loss L_{cycle}:

$$L_{cycle}(G_x, G_y) = E_x[\|x - G_x(G_y(x))\|_1]$$
$$+ E_y[\|y - G_y(G_x(y))\|_1]. \quad (10)$$

This loss function enforces that property that after projecting the source images onto the target domain, the source image can be re-synthesized with minimal loss from the projection. Lastly, by incorporating the neighboring cross-sections, the cycle consistency and adversarial loss functions become:

$$L_{cycle-k}(G_x, G_y) = E_{\mathbf{x}_k}[\|\mathbf{x}_k - G_x(G_y(\mathbf{x}_k))\|_1]$$
$$+ E_{\mathbf{y}_k}[\|\mathbf{y}_k - G_y(G_x(\mathbf{y}_k))\|_1]. \quad (11)$$
$$L_{GAN-k}(D_y, G_y) = -E_{\mathbf{y}_k}[(D_y(\mathbf{y}_k) - 1)^2]$$
$$- E_{\mathbf{x}_k}[D_y(G_y(\mathbf{x}_k))^2] \quad (12)$$

This yields the following aggregate loss function for training:

$$L_{cGAN}(D_x, D_y, G_x, G_y)$$
$$= L_{GAN-k}(D_x, G_x) + L_{GAN-k}(D_y, G_y)$$
$$+ \lambda_{cycle} L_{cycle-k}(G_x, G_y). \quad (13)$$

where L_{cGAN} is the complete loss function, and λ_{cycle} controls the relative weighing of the cycle consistency loss.

While training both pGAN and cGAN, we made a minor modification in the adversarial loss function. As implemented in [50], the generator was trained to minimize $E_{\mathbf{x}_k}[(D(\mathbf{x}_k, G(\mathbf{x}_k)) - 1)^2]$ instead of $-E_{\mathbf{x}_k}[(D(\mathbf{x}_k, G(\mathbf{x}_k)))^2]$.

B. MRI Datasets

For registered images, we trained both pGAN and cGAN models. For unregistered images, we only trained cGAN models. The experiments were performed on three separate datasets: the MIDAS dataset [57], the IXI dataset (http://brain-development.org/ixi-dataset/) and the BRATS dataset (https://sites.google.com/site/braintumorsegmentation/home/brats2015). MIDAS and IXI datasets contained data from healthy subjects, whereas the BRATS dataset contained data from patients with structural abnormality (i.e., brain tumor). For each dataset, subjects were sequentially selected in the

order that they were shared on the public databases. Subjects with images containing severe motion-artifacts across the volume were excluded from selection. The selected set of subjects were then sequentially split into training, validation and testing sets. Protocol information for each dataset is described below.

1) MIDAS Dataset: T_1- and T_2-weighted images from 66 subjects were analyzed, where 48 subjects were used for training, 5 were used for validation and 13 were used for testing. From each subject, approximately 75 axial cross sections that contained brain tissue and that were free of major artifacts were manually selected. T_1-weighted images: 3D gradient-echo FLASH sequence, TR=14ms, TE=7.7ms, flip angle=25°, matrix size=256x176, 1 mm isotropic resolution, axial orientation. T_2-weighted images: 2D spin-echo sequence, TR=7730ms, TE=80ms, flip angle=90°, matrix size=256×192, 1 mm isotropic resolution, axial orientation.

2) IXI Dataset: T_1- and T_2-weighted images from 40 subjects were analyzed, where 25 subjects were used for training, 5 were used for validation and 10 were used for testing. When T_1-weighted images were registered onto T_2-weighted images, nearly 90 axial cross sections per subject that contained brain tissue and that were free of major artifacts were selected. When T_2-weighted images were registered onto T_1-weighted images, nearly 110 cross sections were selected. In this case due to poor registration quality we had to remove a test subject. T_1-weighted images: TR=9.813ms, TE=4.603ms, flip angle=8°, volume size $= 256 \times 256 \times 150$, voxel dimensions $= 0.94mm \times 0.94mm \times 1.2mm$, sagittal orientation. T_2-weighted images: TR=8178ms, TE=100ms, flip angle=90°, volume size $= 256 \times 256 \times 150$, voxel dimensions $= 0.94 \times 0.94 \times 1.2$ mm^3, axial orientation.

3) BRATS Dataset: T_1- and T_2-weighted images from 41 low-grade glioma patients with visible lesions were analyzed, where 24 subjects were used for training, 2 were used for validation and 15 were used for testing. From each subject, approximately 100 axial cross sections that contained brain tissue and that were free of major artifacts were manually selected. Different scanning protocols were employed on separate sites.

Note that each dataset comprises a different number of cross-sections per subject, and we only retained cross-sections that contained brain tissue and that were free of major artifacts. As such, we varied the number of subjects across datasets to balance the total number of images used, resulting in approximately 4000–5000 images per dataset.

Control analyses were performed to rule out biases due to the specific selection or number of subjects. To do this, we performed model comparisons using an identical number of subjects (40) within each dataset. This selection included nonoverlapping training, validation and testing sets, such that 25 subjects were used for training, 5 for validation and 10 for testing. In IXI, we sequentially selected a completely independent set of subjects from those reported in the main analyses. This selection was then sequentially split into training/validation/testing sets via a 4-fold cross-validation procedure. Since the number of subjects available was smaller in MIDAS and BRATS, we performed 4-fold cross-validation

by randomly sampling nonoverlapping training, validation and testing sets in each fold. No overlap was allowed among testing sets across separate folds, or among the training, testing and validation sets within each fold.

4) Data Normalization: To prevent suboptimal model training and bias in quantitative assessments, datasets were normalized to ensure comparable ranges of voxel intensities across subjects. The multi-contrast MRI images in the IXI and MIDAS datasets were acquired using a single scan protocol. Therefore, for each contrast, voxel intensity was normalized within each subject to a scale of [0 1] via division by the maximum intensity within the brain volume. The protocol variability in the BRATS dataset was observed to cause large deviations in image intensity and contrast across subjects. Thus, for normalization, the mean intensity across the brain volume was normalized to 1 within individual subjects. To attain an intensity scale in [0 1], three standard deviations above the mean intensity of voxels pooled across subjects was then mapped to 1.

C. Image Registration

For the first scenario, multi-contrast images from a given subject were assumed to be registered. Note that the images contained in the MIDAS and IXI datasets are unregistered. Thus, the T_1- and T_2-weighted images in these datasets were registered prior to network training. In the MIDAS dataset, the voxel dimensions for T_1- and T_2-weighted images were identical, so a rigid transformation based on a mutual information cost function was observed to yield high quality registration. In the IXI dataset, however, voxel dimensions for T_1- and T_2-weighted images were quite distinct. For improved registration accuracy, we therefore used an affine transformation with higher degrees of freedom based on a mutual information cost in this case. No registration was needed for the BRATS dataset that was already registered. No registration was performed for the second scenario. All registrations were implemented in FSL [58], [59].

D. Network Training

Since we consider two different scenarios for multi-contrast MR image synthesis, network training procedures were distinct. In the first scenario, we assumed perfect alignment between the source and target images, and we then used pGAN to learn the mapping from the source to the target contrast. In a first variant of pGAN (k=1), the input image was a single cross-section of the source contrast, and the target was the respective cross-section of the desired contrast. Note that neighboring cross sections in MR images are expected to show significant correlation. Thus, we reasoned that additional information from adjacent cross-sections in the source contrast should improve synthesis. To do this, a second variant of pGAN was implemented where multiple consecutive cross-sections (k=3, 5, 7) of the source contrast were given as input, with the target corresponding to desired contrast at the central cross-section.

For the pGAN network, we adopted the generator architecture from [25], and the discriminator architecture from [50]

(see Supp. Methods for details). Tuning hyperparameters in deep neural networks, especially in complex models such as GANs, can be computationally intensive [60], [61]. Thus, it is quite common in deep learning research to perform one-fold cross-validation [30], [35] or even directly adopt hyperparameter selection from published work [24], [28], [29], [38], [48], [62]. For computational efficiency, here we selected the optimum weightings of loss functions and number of epochs by performing one-fold cross-validation. We partitioned the datasets into training, validation and test sets, each set containing images from distinct subjects. Multiple models were trained for varying number of epochs (in the range [100 200]) and relative weighting of the loss functions (λ in the set $\{10,100,150\}$, and λ_{perc} in the set $\{10,100,150\}$). Parameters were selected based on the validation set, and performance was then assessed on the test set. Among the datasets here, IXI contains the highest-quality images with visibly lower noise and artifact levels compared to MIDAS and visibly sharper images compared to BRATS. To prevent overfitting to noise, artifacts or blurry images, we therefore performed cross-validation of GAN models on IXI, and used the selected parameters in the remaining datasets. Weightings of both pixel-wise and perceptual loss were selected as 100 and the number of epochs was set to 100 (the benefits of perceptual loss on synthesis performance are demonstrated in MIDAS and IXI; Supp. Table IV). Remaining hyperparameters were adopted from [50], where the Adam optimizer was used with a minibatch size of 1 [63]. In the first 50 epochs, the learning rates for the generator and discriminator were 0.0002. In the last 50 epochs, the learning rate was linearly decayed from 0.0002 to 0. During each iteration the discriminator loss function was halved to slow down the learning process of the discriminator. Decay rates for the first and second moments of gradient estimates were set as $\beta1= 0.5$ and $\beta2=0.999$, respectively. Instance normalization was applied [64]. All weights were initialized using normal distribution with 0 mean and 0.02 std.

In the second scenario, we did not assume any alignment between the source and target images, and so we used cGAN to learn the mapping between unregistered source and target images (cGAN$_{unreg}$). Similar to pGAN, two variants of cGAN were considered that worked on a single cross-section (k=1) and on multiple consecutive cross-sections. Because training of cGAN brings substantial computational burden compared to pGAN, we only examined k=3 for cGAN. This latter cGAN variant was implemented with multiple consecutive cross-sections of the source contrast. Although cGAN does not assume alignment between the source and target domains, we wanted to examine the effects of loss functions used in cGAN and pGAN. For comparison purposes, we also trained separate cGAN networks on registered multi-contrast data (cGAN$_{reg}$). The cross-validation procedures, and the architectures of the generator and discriminator were identical to those for pGAN. Multiple models were trained for varying number of epochs (in the range [100 200]), and λ_{cycle} in the set $\{10,100,150\}$). Model parameters were selected based on performance on the validation set, and model performance was then assessed on the test set. The relative weighting of the

cycle consistency loss function was selected as $\lambda_{cycle} = 100$, and the model was trained for 200 epochs. In the first 100 epochs, the learning rate for both networks were set to 0.0002, and in the remaining 100 epochs, the learning rate was linearly decayed from 0.0002 to 0. During each iteration the discriminator loss function was divided by 2 to slow down the learning process of the discriminator.

E. Competing Methods

To demonstrate the proposed approach, two state-of-the-art methods for MRI image synthesis were implemented. The first method was Replica that estimates a nonlinear mapping from image patches in the source contrast onto individual voxels in the target contrast [23]. Replica extracts image features at different spatial scales, and then performs a multi-resolution analysis via random forests. The learned nonlinear mapping is then applied on test images. Code posted by the authors of the Replica method was used to train the models, based on the procedures/parameters described in [23].

The second method was Multimodal that uses an end-to-end neural network to estimate the target image given the source image as input. A neural-network implementation implicitly performs multi-resolution feature extraction and synthesis based on these features. Trained networks can then be applied on test images. Code posted by the authors of the Multimodal method was used to train the models, based on procedures/parameters described in [21].

The proposed approach and the competing methods were compared on the same training and test data. Since the proposed models were implemented for unimodal mapping between two separate contrasts, Replica and Multimodal implementations were also performed with only two contrasts.

F. Experiments

1) Comparison of GAN-Based Models: Here we first questioned whether the direction of registration between multi-contrast images affects the quality of synthesis. In particular, we generated multiple registered datasets from T_1- and T_2-weighted images. In the first set, T_2-weighted images were registered onto T_1-weighted images (yielding $T_{2\#}$). In the second set, T_1-weighted images were registered onto T_2-weighted images (yielding $T_{1\#}$). In addition to the direction of registration, we also considered the two possible directions of synthesis (T_2 from T_1; T_1 from T_2).

For MIDAS and IXI, the above-mentioned considerations led to four distinct cases: a) $T_1 \rightarrow T_{2\#}$, b) $T_{1\#} \rightarrow T_2$, c) $T_2 \rightarrow T_{1\#}$, d) $T_{2\#} \rightarrow T_1$. Here, T_1 and T_2 are unregistered images, $T_{1\#}$ and $T_{2\#}$ are registered images, and $\rightarrow$ corresponds to the direction of synthesis. For each case, pGAN and cGAN were trained based on two variants, one receiving a single cross-section, the other receiving multiple (3, 5 and 7) consecutive cross-sections as input. This resulted in a total of 32 pGAN and 12 cGAN models. Note that the single-cross section cGAN contains generators for both contrasts, and trains a model that can synthesize in both directions. For the multi cross-section cGAN, however, a separate model was trained for synthesis direction. For BRATS, no registration

was needed, and this resulted in only two distinct cases for consideration: a) $T_1 \rightarrow T_2$ and d) $T_2 \rightarrow T_1$. A single variant of pGAN (k=3) and cGAN (k=1) was considered.

2) Comparison to State-of-the-Art Methods: To investigate how well the proposed methods perform with respect to state-of-the-art approaches, we compared the pGAN and cGAN models with Replica and Multimodal. Models were compared using the same training, and testing sets, and these sets comprised images from different groups of subjects. The synthesized images were compared with the true target images as reference. Both the synthesized and the reference images were normalized to a maximum intensity of 1. To assess the synthesis quality, we measured the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [65] metrics between the synthesized image and the reference.

3) Spectral Density Analysis: While PSNR and SSIM serve as common measures to evaluate overall quality, they primarily capture characteristics dominated by lower spatial frequencies. To examine synthesis quality across a broader range of frequencies, we used a spectral density similarity (SDS) metric. The rationale for SDS is similar to that for the error spectral plots demonstrated in [66], where error distribution is analyzed across spatial frequencies. To compute SDS, synthesized and reference images were transformed into k-space, and separated into four separate frequency bands: low (0–25%), intermediate (25–50%), high-intermediate (50–75%), and high (75–100% of the maximum spatial frequency in k-space). Within each band, SDS was taken as the Pearson's correlation between vectors of magnitude k-space samples of the synthesized and reference images. To avoid bias from background noise, we masked out background regions to zero before calculating the quality measures.

4) Generalizability: To examine the generalizability of the proposed methods, we trained pGAN, cGAN, Replica and Multimodal on the IXI dataset and tested the trained models on the MIDAS dataset. The following cases were examined: $T_1 \rightarrow T_{2\#}$, $T_{1\#} \rightarrow T_2$, $T_2 \rightarrow T_{1\#}$, and $T_{2\#} \rightarrow T_1$. During testing, ten sample images were synthesized for a given source image, and the results were averaged to mitigate nuisance variability in individual samples. When T_1-weighted images were registered onto T_2-weighted images, within-cross-section voxel dimensions were isotropic for both datasets and no extra pre-processing step was needed. However, when T_2-weighted images were registered, voxel dimensions were anisotropic for IXI yet isotropic for MIDAS. To avoid spatial mismatch, voxel dimensions were matched via trilinear interpolation. Because a mismatch of voxel thickness in the cross-sectional dimension can deteriorate synthesis performance, single cross-section models were considered.

5) Reliability Against Noise: To examine the reliability of synthesis against image noise, we trained pGAN and Multi-modal on noisy images. The IXI dataset was selected since it contains high-quality images with relatively low noise levels. Two separate sets of noisy images were then generated by adding Rician noise to the source and target contrast images respectively. The noise level was fixed within subjects and randomly varied across subjects by changing the Rician shape parameter in [0 0.2]. For noise-added target images,

TABLE I
QUALITY OF SYNTHESIS IN THE MIDAS DATASET
SINGLE CROSS-SECTION MODELS

	$cGAN_{unreg}$		$cGAN_{reg}$		pGAN	
	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
$T_1 \to T_{2\#}$	0.829 ±0.017	23.66 ±0.632	0.895 ±0.014	26.56 ±0.432	**0.920** ±0.014	**28.79** ±0.580
$T_{1\#} \to T_2$	0.823 ±0.021	23.85 ±0.420	0.854 ±0.024	25.47 ±0.556	**0.876** ±0.028	**27.07** ±0.618
$T_2 \to T_{1\#}$	0.826 ±0.015	23.20 ±0.503	0.892 ±0.017	26.53 ±1.169	**0.912** ±0.017	**27.81** ±1.424
$T_{2\#} \to T_1$	0.821 ±0.021	22.56 ±1.008	0.863 ±0.022	26.15 ±0.974	**0.883** ±0.023	**27.31** ±0.983

$T_{1\#}$ is registered onto the respective T_2 image; and $T_{2\#}$ is registered onto the respective T_1 image; and $\to$ indicates the direction of synthesis. PSNR and SSIM measurements are reported as mean±std across test images. Boldface marks the model with the highest performance.

TABLE II
QUALITY OF SYNTHESIS IN THE MIDAS DATASET
MULTI CROSS-SECTION MODELS (K = 3)

	$cGAN_{unreg}$		$cGAN_{reg}$		pGAN	
	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
$T_1 \to T_{2\#}$	0.829 ±0.016	23.65 ±0.650	0.895 ±0.014	26.62 ±0.489	**0.926** ±0.014	**29.34** ±0.592
$T_{1\#} \to T_2$	0.797 ±0.027	23.37 ±0.604	0.862 ±0.022	25.83 ±0.384	**0.883** ±0.027	**27.49** ±0.643
$T_2 \to T_{1\#}$	0.824 ±0.015	24.00 ±0.628	0.900 ±0.017	27.04 ±1.238	**0.920** ±0.016	**28.16** ±1.303
$T_{2\#} \to T_1$	0.805 ±0.021	23.55 ±0.782	0.864 ±0.022	26.44 ±0.871	**0.887** ±0.023	**27.42** ±1.127

Boldface marks the model with the highest performance.

background masking was performed prior to training and no perceptual loss was used in pGAN to prevent overfitting to noise. Separate models were trained using noise-added source and original target images, and using original source and noise-added target images.

Statistical significance of differences among methods was assessed with nonparametric Wilcoxon signed-rank tests across test subjects. Neural network training and evaluation was performed on NVIDIA Titan X Pascal and Xp GPUs. Implementation of pGAN and cGAN was carried out in Python using the Pytorch framework [67]. Code for replicating the pGAN and cGAN models will be available on http://github.com/icon-lab/mrirecon. Replica was based on a MATLAB implementation, and a Keras implementation [68] of Multimodal with the Theano backend [69] was used.

III. RESULTS

A. Comparison of GAN-Based Models

We first evaluated the proposed models on T_1- and T_2-weighted images from the MIDAS and IXI datasets. We considered two cases for T_2 synthesis (a. $T_1 \to T_{2\#}$, b. $T_{1\#} \to T_2$, where # denotes the registered image), and two cases for T_1 synthesis (c. $T_2 \to T_{1\#}$, d. $T_{2\#} \to T_1$). Table I lists PSNR and SSIM for pGAN, $cGAN_{reg}$ trained on registered data, and $cGAN_{unreg}$ trained on unregistered data in the MIDAS dataset. We find that pGAN outperforms $cGAN_{unreg}$ and $cGAN_{reg}$ in all cases (p<0.05). Representative results for $T_1 \to T_{2\#}$ are displayed in Fig. 3a and $T_{2\#} \to T_1$ are displayed in Supp. Fig. Ia, respectively. pGAN yields higher synthesis quality compared to $cGAN_{reg}$. Although $cGAN_{unreg}$ was trained on unregistered images, it can faithfully capture fine-grained structure in the synthesized contrast. Overall, both pGAN and cGAN yield synthetic images of remarkable visual similarity to the reference. Supp. Tables II and III (k=1) lists PSNR and SSIM across test images for T_2 and T_1 synthesis with both directions of registration in the IXI dataset. Note that there is substantial mismatch between the voxel dimensions of the source and target contrasts in the IXI dataset, so $cGAN_{unreg}$ must map between the spatial sampling grids of the source and the target. Since this yielded suboptimal performance,

measurements for $cGAN_{unreg}$ are not reported. Overall, similar to the MIDAS dataset, we observed that pGAN outperforms the competing methods (p<0.05). On average, across the two datasets, pGAN achieves 1.42dB higher PSNR and 1.92% higher SSIM compared to cGAN. These improvements can be attributed to pixel-wise and perceptual losses compared to cycle-consistency loss on paired images.

In MR images, neighboring voxels can show structural correlations, so we reasoned that synthesis quality can be improved by pooling information across cross sections. To examine this issue, we trained multi cross-section pGAN (k = 3, 5, 7), $cGAN_{reg}$ and $cGAN_{unreg}$ models (k = 3; see Methods) on the MIDAS and IXI datasets. PSNR and SSIM measurements for pGAN are listed in Supp. Table II, and those for cGAN are listed in Supp. Table III. For pGAN, multi cross-section models yield enhanced synthesis quality in all cases. Overall, k=3 offers optimal or near-optimal performance while maintaining relatively low model complexity, so k=3 was considered thereafter for pGAN. The results are more variable for cGAN, with the multi-cross section model yielding a modest improvement only in some cases. To minimize model complexity, k=1 was considered for cGAN.

Table II compares PSNR and SSIM of multi cross-section pGAN and cGAN models for T_2 and T_1 synthesis in the MIDAS dataset. Representative results for $T_1 \to T_{2\#}$ are shown in Fig. 3b and $T_{2\#} \to T_1$ are shown in Supp. Fig. Ib. Among multi cross-section models, pGAN outperforms alternatives in PSNR and SSIM (p<0.05), except for SSIM in $T_{2\#} \to T_1$. Moreover, compared to the single cross-section pGAN, the multi cross-section pGAN improves PSNR and SSIM values. These measurements are also affirmed by improvements in visual quality for the multi cross-section model in Fig. 3 and Supp. Fig. I. In contrast, the benefits are less clear for cGAN. Note that, unlike pGAN that works on paired images, the discriminators in cGAN work on unpaired images from the source and target domains. In turn, this can render incorporation of correlated information across cross sections less effective. Supp. Tables II and III compare PSNR and SSIM of multi cross-section pGAN and cGAN models for T_2 and T_1 synthesis in the IXI dataset. The multi cross-section pGAN outperforms $cGAN_{reg}$ in all cases (p<0.05). Moreover, the multi cross-section pGAN outperforms the single cross-section pGAN in all cases (p<0.05), except in $T_1 \to T_{2\#}$. On average, across the two datasets, multi cross-section
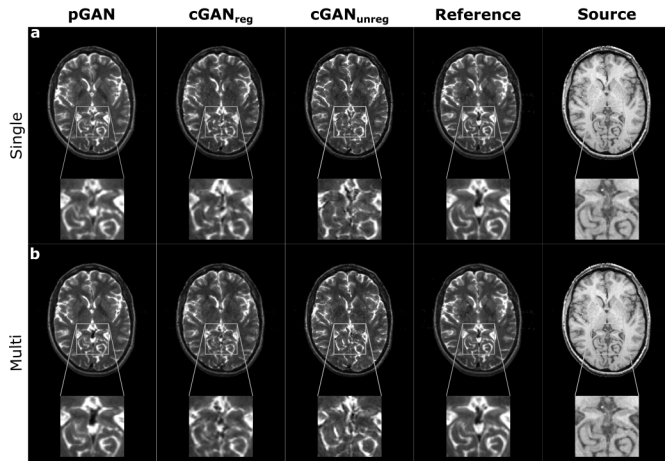
Fig. 3. The proposed approach was demonstrated for synthesis of T_2-weighted images from T_1-weighted images in the MIDAS dataset. Synthesis was performed with pGAN, cGAN trained on registered images (cGAN$_{reg}$), and cGAN trained on unregistered images (cGAN$_{unreg}$). For pGAN and cGAN$_{reg}$, training was performed using T_2-weighted images registered onto T_1-weighted images ($T_1 \rightarrow T_{2\#}$). Synthesis results for (a) the single cross-section, and (b) multi cross-section models are shown along with the true target image (reference) and the source image (source). Zoomed-in portions of the images are also displayed. While both pGAN and cGAN yield synthetic images of striking visual similarity to the reference, pGAN is the top performer. Synthesis quality is improved as information across neighboring cross sections is incorporated, particularly for the pGAN method.

pGAN achieves 0.63dB higher PSNR and 0.89% higher SSIM compared to single cross-section pGAN.

B. Comparison to State-of-the-Art Methods

Next, we demonstrated the proposed methods against two state-of-the-art techniques for multi-contrast MRI synthesis, Replica and Multimodal. We trained pGAN, cGAN$_{reg}$, Replica, and Multimodal on T_1- and T_2-weighted brain images in the MIDAS and IXI datasets. Note that Replica performs ensemble averaging across random forest trees and Multimodal uses mean-squared error measures that can lead to overemphasis of low frequency information. In contrast, conditional GANs use loss functions that can more effectively capture details in the intermediate to high spatial frequency range. Thus, pGAN should synthesize sharper and more realistic images as compared to the competing methods. Table III lists PSNR and SSIM for pGAN, Replica and Multimodal (cGAN$_{reg}$ listed in Supp. Table I) in the MIDAS dataset. Overall, pGAN outperforms the competing methods in all examined cases (p<0.05), except for SSIM in T_2 synthesis, where pGAN and Multimodal perform similarly. The proposed method is superior in depiction of detailed tissue structure as visible in Supp. Fig. II (for comparisons in coronal and sagittal cross-sections see Supp. Figs. IV, V). Table IV lists PSNR and SSIM across test images synthesized via pGAN, Replica and Multimodal (cGAN$_{reg}$ listed in Supp. Table I) for the IXI dataset. Overall, pGAN outperforms the competing methods in all examined cases (p<0.05). The proposed method is superior in depiction of detailed tissue structure as visible in Fig. 4 and Supp. Fig. III (see also Supp. Figs. IV, V).

TABLE III
A- QUALITY OF SYNTHESIS IN THE MIDAS DATASET

	pGAN		Replica		Multimodal	
	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
$T_1 \rightarrow T_{2\#}$	**0.926** ±0.014	**29.34** ±0.592	0.877 ±0.027	26.18 ±0.638	0.924 ±0.012	28.33 ±0.501
$T_{1\#} \rightarrow T_2$	0.883 ±0.027	**27.49** ±0.643	0.838 ±0.039	25.27 ±0.468	**0.889** ±**0.020**	26.73 ±0.461
$T_2 \rightarrow T_{1\#}$	**0.920** ±0.016	**28.16** ±**1.303**	0.840 ±0.028	20.00 ±1.207	0.886 ±0.022	22.13 ±1.325
$T_{2\#} \rightarrow T_1$	**0.887** ±**0.023**	**27.42** ±**1.127**	0.827 ±0.031	20.29 ±1.066	0.872 ±0.020	23.08 ±1.280

Boldface marks the model with the highest performance.
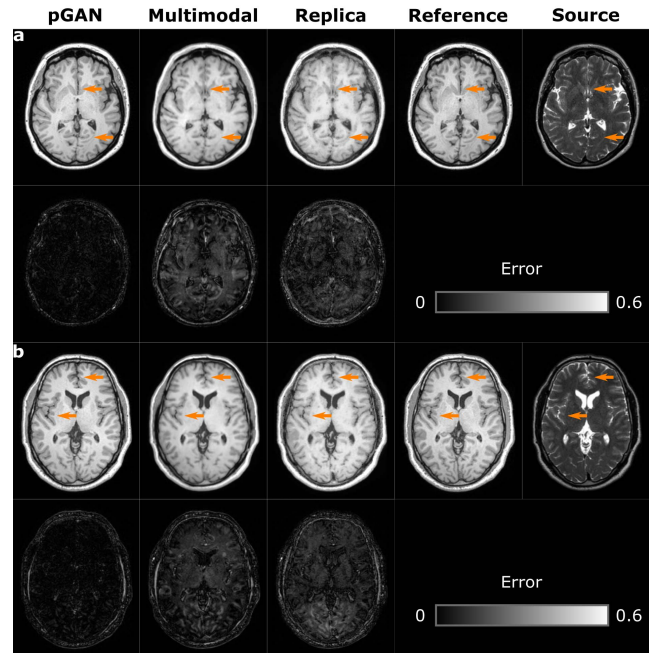

Fig. 4. The proposed approach was demonstrated for synthesis of T_1-weighted images from T_2-weighted images in the IXI dataset. $T_2 \rightarrow T_{1\#}$ and $T_{2\#} \rightarrow T_1$ synthesis were performed with pGAN, Multimodal and Replica. Synthesis results for (a) $T_2 \rightarrow T_{1\#}$, and (b) $T_{2\#} \rightarrow T_1$ along with their corresponding error maps are shown along with the true target image (reference) and the source image (source). The proposed method outperforms competing methods in terms of synthesis quality. Regions that are inaccurately synthesized by the competing methods are reliably depicted by pGAN (marked with arrows). The use of adversarial loss enables improved accuracy in synthesis of intermediate-spatial-frequency texture in T_2-weighted images compared to Multimodal and Replica that show some degree of blurring.

Following assessments on datasets comprising healthy subjects, we demonstrated the performance of the proposed methods on patients with pathology. To do this, we trained and tested pGAN, cGAN$_{reg}$, Replica, and Multimodal on T_1- and T_2-weighted brain images from the BRATS dataset. Similar to the previous evaluations, here we expected that the proposed method would synthesize more realistic images with improved preservation of fine-grained tissue structure. Table V lists PSNR and SSIM across test images synthesized via pGAN, Replica and Multimodal (cGAN$_{reg}$ listed in Supp. Table I; for measurements on background-removed images in MIDAS, IXI and BRATS see Supp. Table V). Overall, pGAN is the
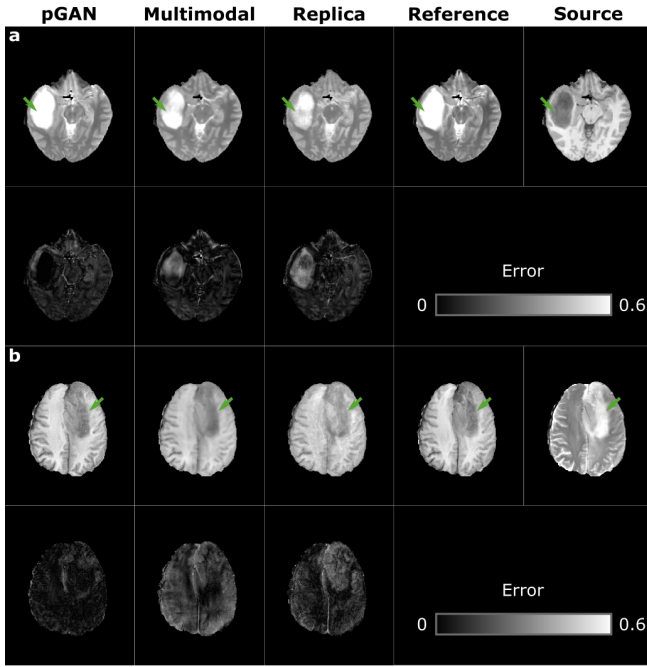
Fig. 5. The proposed approach was demonstrated on glioma patients for synthesis of T_2-weighted images from T_1-weighted images, and T_2-weighted images from T_1-weighted images in the BRATS dataset. Synthesis results for (a) $T_1 \rightarrow T_2$, and (b) $T_1 \rightarrow T_2$ along with their corresponding error maps are shown along with the true target image (reference) and the source image (source). Regions of inaccurate synthesis with Replica and Multimodal are observed near pathologies (marked with arrows). Meanwhile, the pGAN method enables reliable synthesis with visibly improved depiction of intermediate spatial frequency information.

TABLE IV
QUALITY OF SYNTHESIS IN THE IXI DATASET

	pGAN		Replica		Multimodal	
	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
$T_1 \rightarrow T_{2\#}$	**0.948** ±**0.014**	**29.77** ±**1.568**	0.912 ±0.028	25.40 ±2.084	0.936 ±0.015	27.72 ±0.910
$T_{1\#} \rightarrow T_2$	**0.917** ±**0.012**	**27.89** ±**0.887**	0.863 ±0.023	24.08 ±1.427	0.898 ±0.014	26.11 ±0.769
$T_2 \rightarrow T_{1\#}$	**0.926** ±**0.013**	**27.27** ±**0.960**	0.865 ±0.013	20.46 ±0.921	0.895 ±0.015	22.61 ±1.105
$T_{2\#} \rightarrow T_1$	**0.953** ±**0.012**	**29.55** ±**1.423**	0.887 ±0.033	21.82 ±1.600	0.936 ±0.017	25.91 ±1.689

Boldface marks the model with the highest performance.

top performing method in all cases ($p < 0.05$), except for SSIM in $T_1 \rightarrow T_2$ where pGAN and Multimodal perform similarly. Moreover, cGAN performs favorably in PSNR over competing methods. Representative images for T_2 and T_1 synthesis are displayed in Fig. 5 (see also Supp. Figs. IV, V). It is observed that regions near pathologies are inaccurately synthesized by Replica and Multimodal. Meanwhile, the pGAN method enables reliable synthesis with visibly improved depiction of structural details. Across the datasets, pGAN outperforms the state-of-the-art methods by 2.85dB PSNR and 1.23% SSIM.

Next, we performed additional control analyses via 4-fold cross validation to rule out potential biases due to subject selection. Supp. Tables IX–XI list PSNR and SSIM across test images synthesized via pGAN and Multimodal separately for

TABLE V
QUALITY OF SYNTHESIS IN THE BRATS DATASET

	pGAN		Replica		Multimodal	
	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
$T_1 \rightarrow T_2$	**0.946** ±**0.009**	**27.19** ±**1.456**	0.924 ±0.014	24.64 ±1.615	0.939 ±0.011	25.09 ±1.013
$T_2 \rightarrow T_1$	**0.940** ±**0.009**	**25.80** ±**1.867**	0.917 ±0.007	24.49 ±1.230	0.935 ±0.010	23.78 ±2.080

Boldface marks the model with the highest performance.
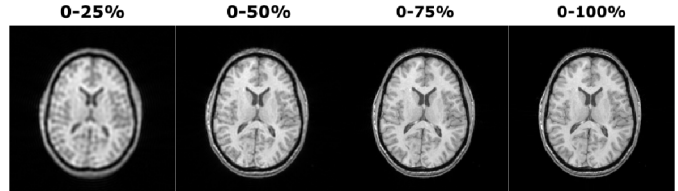

Fig. 6. The T_1-weighted image of a sample cross-section from the MIDAS dataset was processed with an ideal filter in k-space. The filter was broadened sequentially to include higher frequencies (0-25%, 0-50%, 0-75%, 0-100% of the maximum spatial frequency). The filtered images respectively show the contribution of low, intermediate, high-intermediate and high frequency bands. The bulk shape and contrast of the imaged object is captured in the low frequency band, whereas the fine structural details such as edges are captured in the intermediate and partly high-intermediate frequency bands. There is no apparent contribution from the high frequency band.

all 4 folds. We find that there is minimal variability in pGAN performance across folds. Across the datasets, pGAN variability is merely 0.70% in PSNR and 0.37% in SSIM, compared to Multimodal variability of 2.26% in PSNR and 0.46% in SSIM. The results of these control analyses are also highly consistent with those in the original set of subjects reported in Supp. Table I. We find that there is minimal variability in pGAN performance between the main and control analyses. Across the datasets, pGAN variability is 1.42% in PSNR and 0.73% in SSIM, compared to Multimodal variability of 2.98% in PSNR and 0.97% in SSIM.

C. Spectral Density Analysis

To corroborate visual observations regarding improved depiction of structural details, we measured spectral density similarity (SDS) between synthesized and reference images across low, intermediate, high-intermediate and high spatial frequencies (see Methods). Fig. 6 shows filtered versions of a T_1-weighted image in the MIDAS dataset, where the filter is broadened sequentially to include higher frequencies so as to visualize the contribution of individual bands. Intermediate and high-intermediate frequencies primarily correspond to edges and other structural details in MR images, so we expected pGAN to outperform competing methods in these bands. Fig. 7 shows representative synthesis results in the image and spatial frequency (k-space) domains. Supp. Table VI lists SDS across the test images synthesized via pGAN, cGAN$_{reg}$, Replica and Multimodal in the all datasets. In the MIDAS dataset, pGAN outperforms the competing methods at low and intermediate frequencies ($p < 0.05$), except in T_1 synthesis where it performs similarly to Multimodal.
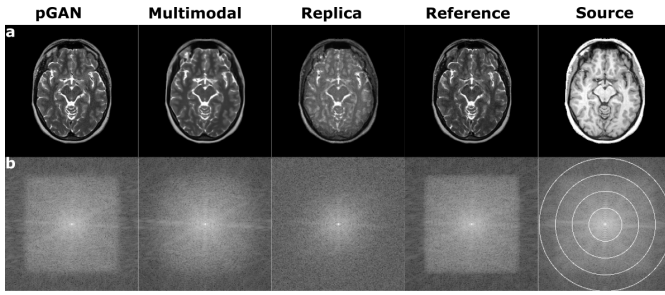
Fig. 7. Synthesis results are shown for a sample cross section from the IXI dataset along with the true target (reference) and the source image (source). Images shown in (a) the spatial domain (b) the spatial-frequency (k-space) domain. White circular boundaries in the k-space representation of the source delineate the boundaries of the low, intermediate, high-intermediate and high frequency bands. The pGAN method more accurately synthesizes the target image as evidenced by the better match in energy distribution across k-space.

In the IXI dataset, pGAN yields superior performance to competing methods in all frequency bands ($p < 0.05$). In the BRATS dataset, pGAN achieves higher SDS than the competing methods at low, intermediate and high-intermediate frequencies in T_2 synthesis and at low frequencies in T_1 synthesis ($p < 0.05$). Across the datasets, pGAN outperforms the state-of-the-art methods by 0.056 at low, 0.061 at intermediate and 0.030 at high-intermediate frequencies.

D. Generalizability

Next, we examined synthesis methods in terms of their generalization performance. Supp. Table VII lists SSIM and PSNR for pGAN, $cGAN_{reg}$, Replica and Multimodal trained on the IXI dataset and tested on the MIDAS dataset. Overall, the proposed methods are the top performers. In $T_1 \rightarrow T_{2\#}$, Multimodal is the leading performer with 1.9% higher SSIM SSIM ($p < 0.05$) than pGAN. In $T_{1\#} \rightarrow T_2$, pGAN outperforms competing methods in PSNR ($p < 0.05$). In $T_2 \rightarrow T_{1\#}$, pGAN is again the leading performer with 1.9% higher SSIM ($p < 0.05$) than Multimodal. In $T_{2\#} \rightarrow T_1$, $cGAN_{reg}$ is the leading performer with 1.22dB higher PSNR ($p < 0.05$) SSIM than pGAN. We also assessed the level of performance degradation between within-dataset synthesis (trained and tested on MIDAS) and across-dataset synthesis (trained on IXI, tested on MIDAS). Overall, pGAN and Multimodal show similar degradation levels. While pGAN is the top performer in terms of SSIM, cGAN yields a modest advantage in PSNR. On average, percentage degradation is 20.83% in PSNR and 11.70% in SSIM for pGAN, 22.22% in PSNR and 10.12% in SSIM for Multimodal, 15.85% in PSNR and 12.85% in SSIM for $cGAN_{reg}$, and 11.40% in PSNR and 14.51% in SSIM for Replica. Note that percentage degradation in PSNR is inherently limited for Replica, which yields low PSNR for within-dataset synthesis.

E. Reliability Against Noise

Lastly, we examined reliability of synthesis against noise (Supp. Fig. VI). Supp. Table VIII list SSIM and PSNR for pGAN and Multimodal trained on noise-added source and target images from IXI, respectively. For noisy source images, pGAN outperforms Multimodal in all examined cases ($p < 0.05$) except for SSIM in $T_1 \rightarrow T_{2\#}$. On average, pGAN achieves 1.74dB higher PSNR and 2.20% higher SSIM than Multimodal. For noisy target images, pGAN is the top performer in PSNR in $T_{1\#} \rightarrow T_2$, $T_2 \rightarrow T_{1\#}$ ($p < 0.05$) and performs similarly to Multimodal in the remaining cases. On average, pGAN improves PSNR by 0.61dB. (Note, however, that for noisy target images, reference-based quality measurements are biased by noise particularly towards higher frequency bands; see Supp. Fig. VII.) Naturally, synthesis performance is lowered in the presence of noise. We assessed the performance degradation when the models were trained on noise-added images as compared to when the models were trained on original images. Overall, pGAN and Multimodal show similar performance degradation with noise. For noisy source images, degradation is 5.27% in PSNR and 2.17% in SSIM for pGAN, and 3.77% in PSNR, 2.66% in SSIM for Multimodal. For noisy target images, degradation is 16.70% in PSNR and 12.91% in SSIM for pGAN, and 15.19% in PSNR, 10.06% in SSIM for Multimodal.

IV. DISCUSSION

A multi-contrast MRI synthesis approach based on conditional GANs was demonstrated against state-of-the-art methods in three publicly available brain MRI datasets. The proposed pGAN method uses adversarial loss functions and correlated structure across neighboring cross-sections for improved synthesis. While many previous methods require registered multi-contrast images for training, a cGAN method was presented that uses cycle-consistency loss for learning to synthesize from unregistered images. Comprehensive evaluations were performed for two distinct scenarios where training images were registered and unregistered. Overall, both proposed methods yield synthetic images of remarkable visual similarity to reference images, and pGAN visually and quantitatively improves synthesis quality compared to state-of-the-art methods [21], [23]. These promising results warrant future studies on broad clinical populations to fully examine diagnostic quality of synthesized images in pathological cases.

Several previous studies proposed the use of neural networks for multi-contrast MRI synthesis tasks [13], [19]–[21], [24]. A recent method, Multimodal, was demonstrated to yield higher quality compared to conventional methods in brain MRI datasets [21]. Unlike conventional neural networks, the GAN architectures proposed here are generative networks that learn the conditional probability distribution of the target contrast given the source contrast. The incorporation of adversarial loss as opposed to typical squared or absolute error loss leads to enhanced capture of detailed texture information about the target contrast, thereby enabling higher synthesis quality.

While our synthesis approach was primarily demonstrated for multi-contrast brain MRI here, architectures similar to pGAN and cGAN have been proposed in other medical image synthesis applications such as cross-modality synthesis or data augmentation [28], [29], [33]–[36], [38]–[42], [48]. The discussions below highlight key differences between the current study and previous work:

(1) [29], [40], [42], [48] proposed conditional GANs for cross-modality synthesis applications. One important proposed application is CT to PET synthesis [29], [40]. For instance, [29] fused the output of GANs and convolutional networks to enhance tumor detection performance from synthesized images; and [40] demonstrated competitive tumor detection results from synthesized versus real images. Another important application is MR to CT synthesis [42], [48]. In [42] and [48], patch-based GANs were used for locally-aware synthesis, and contextual information was incorporated by training an ensemble of GAN models recurrently. Our approach differs in the following aspects: (i) Rather than cross-modality image synthesis, we focus on within-modality synthesis in multi-contrast MRI. MRI provides excellent delineation among soft tissues in the brain and elsewhere, with the diversity of contrasts that it can capture [70]. Therefore, synthesizing a specific MRI contrast given another poses a different set of challenges than performing MR-CT or CT-PET synthesis where CT/PET shows relatively limited contrast among soft tissues [71]. (ii) We demonstrate multi-cross section models to leverage correlated information across neighboring cross-sections within a volume. (iii) We demonstrate pGAN based on both pixel-wise and perceptual losses to enhance synthesis quality.

(2) Architectures similar to cGAN with cycle-consistency loss were recently proposed to address the scarcity of paired training data in MR-CT synthesis tasks [28], [33], [36], [38], [39]. [33] also utilized a gradient-consistency loss to enhance the segmentation performance on CT images synthesized from MR data. Reference [36] performed data-augmentation for enhanced segmentation performance using MR images synthesized from CT data. Reference [39] coupled synthesis and segmentation networks to perform improved segmentation on synthesized CT images using MR labels. Our work differs in the following aspects: (i) As aforementioned, we consider within-modality synthesis as opposed to cross-modality synthesis. (ii) We consider paired image synthesis with cGAN to comparatively evaluate its performance against two state-of-the-art methods (Replica and Multimodal) for paired image synthesis.

(3) An architecture resembling pGAN was proposed for synthesizing retinal images acquired with fundus photography given tabular structural annotations [41]. Similar to pGAN, this previous study incorporated a perceptual loss to improve synthesis quality. Our work differs in the following aspects: (i) Synthesis of vascular fundus images in the retina given annotations is a distinct task than synthesis of a target MR contrast given another source MR contrast in the brain. Unlike the relatively focused delineation between vascular structures and background in retinal images, in our case, there are multiple distinct types of brain tissues that appear at divergent signal levels in separate MR contrasts [71]. (ii) We demonstrate multi-cross section models to leverage correlated information across neighboring cross-sections within an MRI volume.

(4) A recent study suggested the use of multiple cross-sections during MR-to-CT synthesis [72]. In comparison to [72], our approach is different in that: (i) We incorporate an adversarial loss function to better preserve

intermediate-to-high frequency details in the synthesized images. (ii) We perform task- and model-specific optimization of the number of cross-section considering both computational complexity and performance. (iii) As aforementioned, we consider within-modality synthesis as opposed to cross-modality synthesis.

Few recent studies have independently proposed GAN models for multi-contrast MRI synthesis [62], [73], [74]. Perhaps, the closest to our approach are [62] and [73] where conditional GANs with pixel-wise loss were used for improved segmentation based on synthesized FLAIR, T_1- and T_2-weighted images. Our work differs from these studies in the following aspects: (i) We demonstrate improved multi-contrast MRI synthesis via cycle-consistency loss to cope with un-registered images. (ii) We demonstrate improved multi-contrast synthesis performance via the inclusion of a perceptual loss to pGAN. (iii) We demonstrate multiple cross-section models to leverage correlated information across neighboring cross-sections within multi-contrast MRI volumes. (iv) We quantitatively demonstrate that conditional GANs better preserve detailed tissue structure in synthesized multi-contrast images compared to conventional methods [21], [23].

The proposed approach might be further improved by considering several lines of development. Here we presented multi-contrast MRI results while considering two potential directions for image registration ($T_1 \rightarrow T_{2\#}$ and $T_{1\#} \rightarrow T_2$ for T_2 synthesis). We observed that the proposed methods yielded high-quality synthesis regardless of the registration direction. Comparisons between the two directions based on reference-based metrics are not informative because the references are inevitably distinct (e.g., $T_{2\#}$ versus T_2), so determining the optimal direction is challenging. Yet, with substantial mismatch between the voxel sizes in the source and target contrasts, the cGAN method learns to interpolate between the spatial sampling grids of the source and the target. To alleviate performance loss, a simple solution is to resample each contrast separately to match the voxel dimensions. Alternatively, the spatial transformation between the source and target images can first be estimated via multi-modal registration [75]. The estimated transformation can then be cascaded to the output of cGAN. A gradient cycle consistency loss can also be incorporated to prevent the network from learning the spatial transformation between the source and the target [33]. Another cause for performance loss arises when MR images for a given contrast are corrupted by higher levels of noise than typical. Our analyses on noise-added images imply a certain degree of reliability against moderate noise in T_1- or T_2-weighted images. However, an additional denoising network could be incorporated to earlier layers in GAN models when source images have higher noise, and to later layers when target images have elevated noise [76].

Synthesis accuracy can also be improved by generalizing the current approach to predict the target based on multiple source contrasts. In principle, both pGAN and cGAN can receive as input multiple source contrasts in addition to multiple cross sections as demonstrated here. In turn, this generalization can offer improved performance when a subset of the source contrast is unavailable. The performance of conditional

GAN architectures in the face of missing inputs warrants further investigation. Alternatively, an initial fusion step can be incorporated that combines multi-contrast source images in the form of a single fused image fed as input to the GAN [77].

Our analyses on noise-added images indicate that, for target contrasts that are inherently noisier, a downweighing of perceptual loss might be necessary. The proposed models include a hyperparameter for adjusting the relative weighing of the perceptual loss against other loss terms. Thus, a cross-validation procedure can be performed for the specific set of source-target contrasts at hand to optimize model parameters. It remains important future work to assess the optimal weighing of perceptual loss as a function of noise level for specific contrasts. Alternatively, denoising can be included as a preprocessing step to improve reliability against noise. Note that such denoising has recently been proposed for learning-based sampling pattern optimization in MRI [78].

An important concern regarding neural-network based methods is the availability of large datasets for successful training. The cGAN method facilitates network training by permitting the use of unregistered and unpaired multi-contrast datasets. While here we performed training on paired images for unbiased comparison, cGAN permits the use of unpaired images from distinct sets of subjects. As such, it can facilitate compilation of large datasets that would be required for improved performance via deeper networks. Yet, further performance improvements may be viable by training networks based on a mixture of paired and unpaired training data [15].

Recently, cross-modality synthesis with GANs was leveraged as a pre-processing step to enhance various medical imaging tasks such as segmentation, classification or tumor detection [29], [33], [36], [39], [40], [79], [80]. For instance, [29] fused the output of GANs and convolutional networks to enhance tumor detection from synthesized PET images, and [40] demonstrated competitive detection performance with real versus synthesized PET images. [33] trained GANs based on cycle-consistency loss to enhance segmentation performance from synthesized CT images. Reference [36] showed that incorporating synthesized MR images with the real ones can improve the performance of a segmentation network [39]. GANs also showed enhanced performance in liver lesion classification in synthetic CT [79], and chest pathology classification in synthetic X-ray images [80]. These previous reports suggest that the multi-contrast MRI synthesis methods proposed here might also improve similar post-processing tasks. It remains future work to assess to what extent improvements in synthesis quality translate to tasks such as segmentation or detection.

V. CONCLUSION

We proposed a new multi-contrast MRI synthesis method based on conditional generative adversarial networks. Unlike most conventional methods, the proposed method performs end-to-end training of GANs that synthesize the target contrast given images of the source contrast. The use of adversarial loss functions improves accuracy in synthesis of detailed structural information in the target contrast. Synthesis performance is further improved by incorporating pixel-wise and perceptual losses in the case of registered images, and a cycle-consistency loss for unregistered images. Finally, the proposed method leverages information across neighboring cross-sections within each volume to increase accuracy of synthesis. The proposed method outperformed state-of-the-art synthesis methods in multi-contrast brain MRI datasets from healthy subjects and glioma patients. Given the prohibitive costs of prolonged exams due to repeated acquisitions, only a subset contrasts might be collected with adequate quality, particularly in pediatric and elderly patients and in large cohorts [1], [3]. Multi-contrast MRI synthesis might be helpful in those worst-case situations by offering a substitute for highly-corrupted or even unavailable contrasts. Therefore, our GAN-based approach holds great promise for improving the diagnostic information available in clinical multi-contrast MRI.

REFERENCES

[1] B. B. Thukral, "Problems and preferences in pediatric imaging," *Indian J. Radiol. Imag.*, vol. 25, no. 4, pp. 359–364, Oct. 2015.

[2] K. Krupa and M. Bekiesińska-Figatowska, "Artifacts in magnetic resonance imaging," *Polish J. Radiol.*, vol. 80, pp. 93–106, Feb. 2015.

[3] C. M. Stonnington et al., "Interpreting scan data acquired from multiple scanners: A study with Alzheimer's disease," *NeuroImage*, vol. 39, no. 3, pp. 1180–1185, Feb. 2008.

[4] J. E. Iglesias, E. Konukoglu, D. Zikic, B. Glocker, K. van Leemput, and B. Fischl, "Is synthesizing MRI contrast useful for inter-modality analysis?" in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2013, pp. 631–638.

[5] M. I. Miller, G. E. Christensen, Y. Amit, and U. Grenander, "Mathematical textbook of deformable neuroanatomies," *Proc. Natl. Acad. Sci. USA.*, vol. 90, no. 24, pp. 11944–11948, Dec. 1993.

[6] N. Burgos et al., "Attenuation correction synthesis for hybrid PET-MR scanners: Application to brain studies," *IEEE Trans. Med. Imag.*, vol. 33, no. 12, pp. 2332–2341, Dec. 2014.

[7] J. Lee, A. Carass, A. Jog, C. Zhao, and J. L. Prince, "Multi-atlas-based CT synthesis from conventional MRI with patch-based refinement for MRI-based radiotherapy planning," *Proc. SPIE*, vol. 10133, Feb. 2017, Art. no. 101331I.

[8] A. Jog, A. Carass, S. Roy, D. L. Pham, and J. L. Prince, "MR image synthesis by contrast learning on neighborhood ensembles," *Med. Image Anal.*, vol. 24, no. 1, pp. 63–76, Aug. 2015.

[9] A. Jog, S. Roy, A. Carass, and J. L. Prince, "Magnetic resonance image synthesis through patch regression," in *Proc. IEEE Int. Symp. Biomed. Imaging*, Apr. 2013, pp. 350–353.

[10] S. Roy, A. Carass, and J. Prince, "A compressed sensing approach for MR tissue contrast synthesis," in *Proc. Biennial Int. Conf. Inf. Process. Med. Imaging*, 2011, pp. 371–383.

[11] S. Roy, A. Jog, A. Carass, and J. L. Prince, "Atlas based intensity transformation of brain MR images," in *Proc. Int. Workshop Multimodal Brain Image Anal.*, 2013, pp. 51–62.

[12] Y. Huang, L. Shao, and A. F. Frangi, "Simultaneous super-resolution and cross-modality synthesis of 3D medical images using weakly-supervised joint convolutional sparse coding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 5787–5796.

[13] V. Sevetlidis, M. V. Giuffrida, and S. A. Tsaftaris, "Whole image synthesis using a deep encoder-decoder Network," in *Proc. Int. Workshop Simul. Synth. Med. Imaging*, 2016, pp. 127–137.

[14] R. Vemulapalli, H. van Nguyen, and S. K. Zhou, "Unsupervised cross-modal synthesis of subject-specific scans," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 630–638.

[15] Y. Huang, L. Shao, and A. F. Frangi, "Cross-modality image synthesis via weakly coupled and geometry co-regularized joint dictionary learning," *IEEE Trans. Med. Imaging*, vol. 37, no. 3, pp. 815–827, Mar. 2018.

[16] D. H. Ye, D. Zikic, B. Glocker, A. Criminisi, and E. Konukoglu, "Modality propagation: Coherent synthesis of subject-specific scans with data-driven regularization," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2013, pp. 606–613.

[17] S. Roy, A. Carass, N. Shiee, D. L. Pham, and J. L. Prince, "MR contrast synthesis for lesion segmentation," in *Proc. IEEE Int. Symp. Biomed. Imaging*, Apr. 2010, pp. 932–935.

[18] N. Cordier, H. Delingette, M. Lê, and N. Ayache, "Extended modality propagation: Image synthesis of pathological cases," *IEEE Trans. Med. Imaging*, vol. 35, no. 12, pp. 2598–2608, Dec. 2016.

[19] H. van Nguyen, K. Zhou, and R. Vemulapalli, "Cross-domain synthesis of medical images using efficient location-sensitive deep Network," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 677–684.

[20] T. Joyce, A. Chartsias, and S. A. Tsaftaris, "Robust multi-modal MR image synthesis," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2017, pp. 347–355.

[21] A. Chartsias, T. Joyce, M. V. Giuffrida, and S. A. Tsaftaris, "Multimodal MR synthesis via modality-invariant latent representation," *IEEE Trans. Med. Imaging*, vol. 37, no. 3, pp. 803–814, Mar. 2018.

[22] S. Roy, A. Carass, and J. L. Prince, "Magnetic resonance image example-based contrast synthesis," *IEEE Trans. Med. Imaging*, vol. 32, no. 12, pp. 2348–2363, Dec. 2013.

[23] A. Jog, A. Carass, S. Roy, D. L. Pham, and J. L. Prince, "Random forest regression for magnetic resonance image synthesis," *Med. Image Anal.*, vol. 35, pp. 475–488, Jan. 2017.

[24] C. Zhao, A. Carass, J. Lee, Y. He, and J. L. Prince, "Whole brain segmentation and labeling from CT using synthetic MR images," in *Machine Learning in Medical Imaging*. Cham, Switzerland: Springer, 2017, pp. 291–298.

[25] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision—ECCV*. Cham, Switzerland: Springer, Sep. 2016, pp. 694–711.

[26] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial Network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Aug. 2017, pp. 105–114.

[27] A. Dosovitskiy and T. Brox, "Generating images with perceptual similarity metrics based on deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 658–666.

[28] J. M. Wolterink, A. M. Dinkla, M. H. F. Savenije, P. R. Seevinck, C. A. T. van den Berg, and I. Isgum, "Deep MR to CT synthesis using unpaired data," in *Proc. Int. Workshop Simul. Synth. Med. Imaging*, 2017, pp. 14–23.

[29] A. Ben-Cohen, E. Klang, S. P. Raskin, M. M. Amitai, and H. Greenspan, "Virtual PET images from CT data using deep convolutional Networks: Initial results," in *Proc. Int. Workshop Simul. Synth. Med. Imaging*, 2017, pp. 49–57.

[30] F. Mahmood, R. Chen, and N. J. Durr, "Unsupervised reverse domain adaptation for synthetic medical images via adversarial training," *IEEE Trans. Med. Imaging*, vol. 37, no. 12, pp. 2572–2581, Dec. 2018.

[31] H. Huang, P. S. Yu, and C. Wang. (2018). "An introduction to image synthesis with generative adversarial nets." [Online]. Available: https://arxiv.org/abs/1803.04469

[32] Y. Hu *et al.*, "Freehand ultrasound image simulation with spatially-conditioned generative adversarial Networks," in *Proc. Int. Workshop Reconstruction Anal. Moving Body Organs*, 2017, pp. 105–115.

[33] Y. Hiasa *et al.*, "Cross-modality image synthesis from unpaired data using CycleGAN," in *Proc. Int. Workshop Simul. Synth. Med. Imaging*, Sep. 2018, pp. 31–41.

[34] J. T. Guibas, T. S. Virdi, and P. S. Li. (2017). "Synthetic medical images from dual generative adversarial networks." [Online]. Available: https://arxiv.org/abs/1709.01872

[35] P. Costa *et al.*, "End-to-end adversarial retinal image synthesis," *IEEE Trans. Med. Imag.*, vol. 37, no. 3, pp. 781–791, Mar. 2018.

[36] A. Chartsias, T. Joyce, R. Dharmakumar, and S. A. Tsaftaris, "Adversarial image synthesis for unpaired multi-modal cardiac data," in *Proc. Int. Workshop Simul. Synth. Med. Imaging*, Sep. 2017, pp. 3–13.

[37] F. Calimeri, A. Marzullo, C. Stamile, and G. Terracina, "Biomedical data augmentation using generative adversarial neural Networks," in *Proc. Int. Conf. Artif. Neural Netw.*, Oct. 2017, pp. 626–634.

[38] J. M. Wolterink, A. M. Dinkla, M. H. F. Savenije, P. R. Seevinck, and C. A. T. van den Berg, "MR-to-CT synthesis using cycle-consistent generative adversarial networks," in *Proc. Neural Inf. Process. Syst. (NIPS)*, Long Beach, CA, USA, 2017.

[39] Y. Huo, Z. Xu, S. Bao, A. Assad, R. G. Abramson, and B. A. Landman, "Adversarial synthesis learning enables segmentation without target modality ground truth," in *Proc. IEEE 15th Int. Symp. Biomed. Imaging*, Apr. 2018, pp. 1217–1220.

[40] L. Bi, J. Kim, A. Kumar, D. Feng, and M. Fulham, "Synthesis of positron emission tomography (PET) images via multi-channel generative adversarial networks (GANs)," in *Proc. Int. Workshop Reconstruction Anal. Moving Body Organs*, Sep. 2017, pp. 43–51.

[41] H. Zhao, H. Li, S. Maurer-Stroh, and L. Cheng, "Synthesizing retinal and neuronal images with generative adversarial nets," *Med. Image Anal.*, vol. 49, pp. 14–26, Jul. 2018.

[42] D. Nie, R. Trullo, C. Petitjean, S. Ruan, and D. Shen, "Medical image synthesis with deep convolutional adversarial Networks," in *Proc. Med. Image Comput. Comput.-Assist. Intervent*, May 2017, pp. 417–425.

[43] M. Mardani *et al.*, "Deep generative adversarial neural Networks for compressive sensing MRI," *IEEE Trans. Med. Imaging*, vol. 38, no. 1, pp. 167–179, Jan. 2019.

[44] T. M. Quan, T. Nguyen-Duc, and W.-K. Jeong, "Compressed sensing MRI reconstruction with cyclic loss in generative adversarial Networks," *IEEE Trans. Med. Imaging*, vol. 37, no. 6, pp. 1488–1497, Aug. 2018.

[45] G. Yang *et al.*, "DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1310–1321, Jun. 2018.

[46] O. Shitrit *et al.*, "Accelerated magnetic resonance imaging by adversarial neural Network," in *Proc. Int. Workshop Deep Learn. Med. Image Anal.*, Sep. 2017, pp. 30–38.

[47] Y. Wang *et al.*, "3D conditional generative adversarial Networks for high-quality PET image estimation at low dose," *NeuroImage*, vol. 174, pp. 550–562, Jul. 2018.

[48] D. Nie *et al.*, "Medical image synthesis with deep convolutional adversarial Networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 12, pp. 2720–2730, Dec. 2018.

[49] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial Networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1125–1134.

[50] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial Networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2223–2232.

[51] I. J. Goodfellow *et al.*, "Generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[52] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial Networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2813–2821.

[53] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional Networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Sep. 2015, pp. 3431–3440.

[54] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu, "Sketch2Photo: Internet image montage," *ACM Trans. Graph.*, vol. 28, no. 5, p. 124, Dec. 2009.

[55] M. Mirza and S. Osindero. (2014). "Conditional generative adversarial nets." [Online]. Available: https://arxiv.org/abs/1411.1784

[56] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[57] E. Bullitt *et al.*, "Vessel tortuosity and brain tumor malignancy: A blinded study[1]," *Acad. Radiol.*, vol. 12, no. 10, pp. 1232–1240, Oct. 2005.

[58] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, "Improved optimization for the robust and accurate linear registration and motion correction of brain images," *NeuroImage*, vol. 17, no. 2, pp. 825–841, Oct. 2002.

[59] M. Jenkinson and S. Smith, "A global optimisation method for robust affine registration of brain images," *Med. Image Anal.*, vol. 5, no. 2, pp. 143–156, Jun. 2001.

[60] P. Murugan. (2017). "Hyperparameters optimization in deep convolutional neural Network / Bayesian approach with Gaussian process prior." [Online]. Available: https://arxiv.org/abs/1712.07233

[61] T. Hinz, N. Navarro-Guerrero, S. Magg, and S. Wermter, "Speeding up the hyperparameter optimization of deep convolutional neural networks," *Int. J. Comput. Intell. Appl.*, vol. 17, no. 2, Jun. 2018, Art. no. 1850008.

[62] B. Yu, L. Zhou, L. Wang, J. Fripp, and P. Bourgeat, "3D cGAN based cross-modality MR image synthesis for brain tumor segmentation," in *Proc. IEEE 15th Int. Symp. Biomed. Imaging*, Apr. 2018, pp. 626–630.

[63] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, Aug. 2015, pp. 12–24.

[64] D. Ulyanov, A. Vedaldi, and V. Lempitsky. (2016). "Instance normalization: The missing ingredient for fast stylization." [Online]. Available: https://arxiv.org/abs/1607.08022

[65] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[66] T. H. Kim and J. P. Haldar, "The Fourier radial error spectrum plot: A more nuanced quantitative evaluation of image reconstruction quality," in *Proc. IEEE 15th Int. Symp. Biomed. Imaging*, Apr. 2018, pp. 61–64.

[67] A. Paszke *et al.*, "Automatic differentiation in PyTorch," in *Proc. Neural Inf. Process. Syst. (NIPS)*, Long Beach, CA, USA, 2017.

[68] F. Chollet, *Keras*. San Francisco, CA, USA: GitHub, 2015.

[69] T. D. Team *et al.*. (2016). "Theano: A Python framework for fast computation of mathematical expressions." [Online]. Available: https://arxiv.org/abs/1605.02688

[70] M. A. Bernstein, K. F. King, and X. J. Zhou, *Handbook of MRI Pulse Sequences*. New York, NY, USA: Academic, 2004.

[71] D. G. Nishimura, *Principles of Magnetic Resonance Imaging*. Stanford, CA, USA: Stanford Univ., 1996.

[72] L. Xiang, Q. Wang, D. Nie, Y. Qiao, and D. Shen, "Deep embedding convolutional neural network for synthesizing CT image from T1-Weighted MR image," *Med. Image Anal.*, vol. 47, pp. 31–44, Jul. 2018.

[73] Q. Yang, N. Li, Z. Zhao, X. Fan, E. I.-C. Chang, and Y. Xu. (2018). "MRI image-to-image translation for cross-modality image registration and segmentation." [Online]. Available: https://arxiv.org/abs/1801.06940

[74] C. Han *et al.*, "GAN-based synthetic brain MR image generation," in *Proc. IEEE 15th Int. Symp. Biomed. Imaging*, Apr. 2018, pp. 734–738.

[75] X. Yang, R. Kwitt, M. Styner, and M. Niethammer, "Quicksilver: Fast predictive image registration—A deep learning approach," *Neuroimage*, vol. 158, pp. 378–396, Sep. 2017.

[76] D. Jiang, W. Dou, L. Vosters, X. Xu, Y. Sun, and T. Tan, "Denoising of 3D magnetic resonance images with multi-channel residual learning of convolutional neural Network," *Jpn. J. Radiol.*, vol. 36, no. 9, pp. 566–574, Sep. 2018.

[77] S. Qi *et al.*, "Multimodal fusion with reference: Searching for joint neuromarkers of working memory deficits in schizophrenia," *IEEE Trans. Med. Imaging*, vol. 37, no. 1, pp. 93–105, Jan. 2018.

[78] B. Gözcü *et al.*, "Learning-based compressive MRI," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1394–1406, Jun. 2018.

[79] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018.

[80] H. Salehinejad, S. Valaee, T. Dowdell, E. Colak, and J. Barfett, "Generalization of deep neural networks for chest pathology classification in X-Rays using generative adversarial Networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2018, pp. 990–994.