Semi-Supervised Learning of Mutually Accelerated MRI Synthesis without Fully-Sampled Ground Truths

Mahmut Yurt, Salman Ul Hassan Dar, Muzaffer Özbey, Berk Tınaz, Kader Karlı Oğuz, Tolga Çukur

Abstract—Learning-based synthetic multi-contrast MRI commonly involves deep models trained using high-quality images of source and target contrasts, regardless of whether source and target domain samples are paired or unpaired. This results in undesirable reliance on fully-sampled acquisitions of all MRI contrasts, which might prove impractical due to limitations on scan costs and time. Here, we propose a novel semi-supervised deep generative model that instead learns to recover high-quality target images directly from accelerated acquisitions of source and target contrasts. To achieve this, the proposed model introduces novel multi-coil tensor losses in image, k-space and adversarial domains. These selective losses are based only on acquired k-space samples, and randomized sampling masks are used across subjects to capture relationships among acquired and non-acquired k-space regions. Comprehensive experiments on multi-contrast neuroimaging datasets demonstrate that our semi-supervised approach yields equivalent performance to gold-standard fully-supervised models, while outperforming a cascaded approach that learns to synthesize based on reconstructions of undersampled data. Therefore, the proposed approach holds great promise to improve the feasibility and utility of accelerated MRI acquisitions mutually undersampled across both contrast sets and k-space.

Index Terms—magnetic resonance imaging (MRI), accelerated MRI, image synthesis, semi-supervised

I. INTRODUCTION

MRI is a clinical powerhouse in neuroimaging due to its noninvasiveness and excellent soft-tissue contrast. Its unique ability to image the same anatomy under a diverse set of tissue contrasts empowers it to accumulate complementary diagnostic information within a single exam session [1], [2]. However, prolonged scans and increased costs associated with multi-contrast protocols often limit the diversity and quality of MRI exams [3], [4]. A promising solution against this limitation is synthesis of missing or unacceptably low-quality images within the protocol from available high-quality images [5]. Multi-contrast MRI synthesis methods can enhance radiological assessments as well as image analysis tasks such as registration, segmentation, or detection [6]–[8].

In recent years, there has been emerging interest in learning-based MRI synthesis based on deep neural networks, given their state-of-the-art performance in other computer vision [9]–[13] and medical imaging tasks [14]–[16]. An earlier group of studies proposed deep models with convolutional neural networks (CNNs) to learn nonlinear latent representations that mediate conversion from source to target images [17]–[21]. These studies typically involved encoder-decoder architectures, where the encoder embeds hierarchical image features onto a latent space that is later used by the decoder to recover the target image [17]–[21]. For improved capture of structural details, a second group has proposed deep architectures based on conditional generative adversarial networks (GAN) [22]–[26], where the generator that performs the source-to-target mapping benefits from the game-theoretic interplay with the discriminator [9]. Pioneering studies have exploited pixel- or feature-wise correspondence between source-target images in an adversarial setup [22], [32], [33]. Later studies have proposed unified models capable of multiple types of contrast conversion [23], [27], [28], or multi-tasking frameworks [23], [27], [35] to reduce computational complexity. These previous studies have collectively highlighted the immense potential of learning-based synthesis in multi-contrast MRI. That said, both CNN and GAN models are canonically trained in a fully-supervised setup based on pixel-wise, adversarial or perceptual losses between synthesized and ground truth target images. Supervised models require large datasets of high-quality images from Nyquist-sampled source and target acquisitions, paired within subjects [17], [22], [23]. Yet, compilation of paired, high-quality datasets might prove impractical due to scan time and cost considerations [3], [4]. As such, there is a dire need for methods with lower reliance on supervision to improve practicability of learning-based MRI synthesis.

Recent efforts to lower supervision requirements in MRI synthesis have predominantly focused on model training in the absence of paired images across subjects. For unpaired training, a successful approach has been to replace pixel-wise losses in GAN models with cycle-consistency, shape-consistency or mutual information losses [22], [27], [37]–[40]. Similar to supervised models, unpaired models that unify...
multiple contrast conversion tasks have also been introduced to reduce computational complexity [13], [38]. As an alternative, [41], [42] have proposed a hybrid method where the model is trained on a composite dataset with both paired and unpaired samples. These previous methods have increased the capacity of synthesis models to learn from unpaired data, but they still leverage high-quality MR images reconstructed from fully-sampled k-space acquisitions. While training of MRI reconstruction models from undersampled data has received recent interest [43], [44], to the best of our knowledge, no prior study has considered learning of MRI synthesis models from undersampled source or target acquisitions.

Here, we propose a novel semi-supervised deep generative model for multi-contrast MRI synthesis, namely ssGAN, to avoid reliance on fully-sampled k-space acquisitions. The proposed model is trained directly on undersampled acquisitions, and it generates high-quality target images given undersampled multi-coil source acquisitions. To do this, ssGAN introduces novel multi-coil tensor losses in image, k-space and adversarial domains. These selective losses are based only on acquired k-space samples, and randomized sampling masks are used across subjects to capture relationships among acquired and non-acquired k-space regions. Comprehensive experiments performed on brain MRI clearly demonstrate that ssGAN achieves equivalent performance to gold-standard acquisitions of source and target contrasts, and it synthesizes multi-coil target images directly from undersampled multi-coil source acquisitions. To do this, ssGAN introduces novel selective loss functions expressed based on acquired k-space samples in target contrast acquisitions. Meanwhile, ssGAN outperforms a cascade-model that first reconstructs undersampled acquisitions on multi-coil source acquisitions. To do this, ssGAN introduces novel multi-coil tensor losses in image, k-space and adversarial domains, selectively expressed based on acquired k-space samples in target contrast acquisitions.

The proposed method substantially lowers data requirements in MRI synthesis by enabling model training and inference from undersampled acquisitions.

**Contributions:**

- To the best of our knowledge, this is the first semi-supervised learning method for multi-contrast MRI synthesis that performs model training on undersampled source and target acquisitions.
- The proposed method synthesizes target images directly from undersampled multi-coil source acquisitions.
- The proposed method introduces novel multi-coil tensor losses in image, k-space and adversarial domains, selectively expressed based on acquired k-space samples in target contrast acquisitions.
- The proposed method substantially lowers data requirements in MRI synthesis by enabling model training and inference from undersampled acquisitions.

## II. METHODS

In this section, we first overview basics of generative adversarial networks, and the foundation of the proposed architecture for semi-supervised multi-contrast MRI synthesis. We then describe in detail the datasets and experiments conducted to evaluate the proposed methodology.

### A. Generative Adversarial Networks

Generative adversarial networks (GANs) [9] are deep generative models comprising a pair of competing subnetworks: a generator (\(G\)) and a discriminator (\(D\)). \(G\) aims to map a random noise vector \(z\) to a sample resembling a target domain distribution, whereas \(D\) aims to distinguish between real and fake samples of the target domain [9]. These two subnetworks are alternately trained via an adversarial loss function, formulated as follows:

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

(1)

where \(E\) denotes expectation, and \(y\) is an arbitrary real sample in the target domain. Upon convergence, \(G\) is expected to generate realistic target domain samples that \(D\) cannot tell apart from the real ones [9]. While the initial GAN models generated target samples from a random noise vector, later studies have demonstrated success in image-to-image translation with conditional GAN (cGAN) models that additionally receive as input a source domain image \(x\) [10]. The adversarial loss function is therefore modified by conditioning \(G\) on \(x\):

\[
L_{cGAN} = -E_{x,y}[(D(y - 1)^2] - E_x[D(G(x))^2]
\]

When spatially aligned source-target images are available, a pixel-wise loss can be further included [12]:

\[
L_{cGAN} = -E_{x,y}[(D(y - 1)^2] - E_x[D(G(x))^2] + E_{x,y}||y - G(x)||_1
\]

(3)

Several studies have demonstrated variants of cGAN models on multi-contrast MRI that synthesize target contrast images from source contrast images of the same underlying anatomy [22], [23], [25]–[29], [31]–[33], [35], [36]. These models typically learn the source-to-target mapping in a fully-supervised setup. A comprehensive training set is needed containing high-quality source and target images reconstructed from fully-sampled k-space acquisitions (\(x_1, y_1\)), where \(x_1\) is an arbitrary source, \(y_1\) is an arbitrary target image in the training set, and \(I\) denotes the the sampling mask for Nyquist-sampled acquisitions. These fully-supervised models have demonstrated state-of-the-art performance for synthetic multi-contrast MRI. However, they are limited due to reliance on fully-sampled acquisitions that might prove impractical. Therefore, there is a critical need for methods that can directly learn from undersampled MRI data.

### B. Semi-Supervised Generative Adversarial Networks

Here, we propose a novel semi-supervised GAN model, namely ssGAN, to mitigate the dependency of MRI synthesis models on supervised training with Nyquist-sampled source and target acquisitions. ssGAN is trained on undersampled acquisitions of source and target contrasts, and it synthesizes multi-coil target images directly from undersampled multi-coil acquisitions of the source contrast. To do this, ssGAN introduces novel selective loss functions expressed based on only the acquired subset of k-space samples in the target contrast (Fig. 1). Details regarding the optimization objectives of ssGAN are provided in the remainder of this section.
ssGAN receives as input Fourier reconstructions of either fully-sampled or undersampled acquisitions of the source contrast, and learns to synthesize high-quality images of the target contrast. The generator $G$ in ssGAN produces target contrast images via a forward mapping:

$$G(X^n_\Lambda) = \hat{y}, \text{ with } X^n_\Lambda = \{x^n_1, \ldots, x^n_n\}$$  \hspace{1cm} (4)

where $X^n_\Lambda$ denotes multi-coil source contrast images acquired with a k-space sampling mask $\Lambda$, $n$ denotes the number of receive coils with sensitivity maps $\hat{C}_m$ computed via ESPIRIT [48], and $\hat{y}$ denotes the synthesized coil-combined target contrast image. Note that ssGAN considers that only undersampled acquisitions of the target contrast are available, where $Y^n_\Omega = \{y^n_1, \ldots, y^n_m\}$ denotes Fourier reconstructions of multi-target contrast acquisitions collected with a sampling mask $\Omega$ and $m$ receive coils of true coil sensitivities $C^n_Y$.

As no high-quality reference for the target contrast image is assumed, ssGAN expresses novel selective loss functions based on only the acquired subset of k-space samples. To do this, the synthesized coil-combined image is first projected onto individual coils as follows:

$$\hat{Y}^m = P(\hat{y}, \hat{C}_Y^m) = \hat{y} \cdot \hat{C}_Y^m$$  \hspace{1cm} (5)

where $\hat{Y}^m$ denotes the synthesized multi-coil target contrast images, $\hat{C}_Y^m$ denotes estimated coil sensitivity maps computed via ESPIRIT [48], and $P$ is the operator that performs the coil projection in the image domain as dot product takes vectors and outputs a scalar, element-wise multiplication between the input image and coil sensitivity maps. The multi-coil target image projections are then subjected to the binary sampling mask in Fourier domain:

$$\hat{Y}_\Omega = F^{-1}(\hat{Y}^m_\Omega)$$  \hspace{1cm} (6)
where \( \mathcal{F} \) denotes the forward and \( \mathcal{F}^{-1} \) denotes the inverse Fourier transform, \( M \) is the operator that performs binary masking in k-space to with a given sampling mask. In Eq. \( (6) \) \( kY_m^n \) and \( \hat{Y}_m^n \) denote undersampled multi-coil data respectively in k-space and image domain for the synthesized target contrast image. The selective loss function in ssGAN is then defined between undersampled synthesized and undersampled ground truth data for the target contrast, based on three loss components: multi-coil tensor image loss, multi-coil tensor k-space loss, and multi-coil tensor adversarial loss. Each loss term is described below.

1) **Multi-Coil Tensor Image Loss**: The first component of the selective loss function is a multi-coil tensor image loss defined based on undersampled multi-coil data in image domain, between synthesized and ground truth target images:

\[
L_i = E_X \mathcal{X} m \| |Y_m^n - \hat{Y}_m^n|_1 \tag{7}
\]

where \( Y_m^n \) denotes the multi-coil ground truth target images from acquired acquisitions, and \( \hat{Y}_m^n \) denotes the undersampled target images generated by ssGAN.

2) **Multi-Coil Tensor k-space Loss**: The quality of the synthesized images in ssGAN is further enhanced via a multi-coil tensor k-space loss expressed between the Fourier-domain data of the synthesized and ground truth images.

\[
L_k = E_X \mathcal{X} m \| |h(\mathcal{F}(Y_m^n)/\beta) - h(\mathcal{F}(\hat{Y}_m^n)/\beta)|_1 \|_1 \tag{8}
\]

where \( h \) is a \( \tanh \) function with a normalization constant \( \beta \) to provide a comparable signal intensities across k-space, and \( \mathcal{F}(Y_m^n) - \mathcal{F}(\hat{Y}_m^n) \) stand for k-space data of the ground truth and synthesized multi-coil images, respectively.

3) **Multi-Coil Tensor Adversarial Loss**: The level of realism in the synthesized images is advanced via a multi-coil adversarial loss function evaluated between image-domain data of the synthesized and ground truth multi-coil images:

\[
L_a = -E_{\hat{Y}_m^n} [(D(Y_m^n) - 1)^2] - E_X \mathcal{X} m [D(\hat{Y}_m^n)^2] \tag{9}
\]

where \( D \) denotes the discriminator that distinguishes between undersampled ground truth and synthesized images.

The final selective loss function for ssGAN is constructed as a weighted combination of the three multi-coil tensor loss terms described as \( L_{ssGAN} = \lambda_k L_k + \lambda_i L_i + \lambda_a L_a \), where \( \lambda_k \), \( \lambda_i \), and \( \lambda_a \) denote the relative weighting of the tensor k-space, image, and adversarial losses. Note that the selective loss function in along with randomization of the k-space sampling masks across training subjects enables ssGAN to effectively capture complex relationships between acquired and non-acquired k-space coefficients. In turn, ssGAN can successfully recover high-quality target images without requiring Nyquist-sampled acquisitions of the target contrast.

### C. Datasets

The proposed ssGAN model was demonstrated on the public IXI dataset (https://brain-development.org/ixi-dataset/) containing multi-contrast single-coil magnitude brain images and an in-house dataset of multi-contrast multi-coil complex brain images.

1) **The IXI Dataset**: \( T_1 \)- and \( T_2 \)-weighted single-coil magnitude brain MR images of 94 subjects were used, where 64 were reserved for training, 10 for validation, and 20 for testing. See supplementary materials for scan parameters. Since multi-contrast images within subjects were spatially unaligned, \( T_2 \)-weighted images were registered onto \( T_1 \)-weighted images via FSL [49] prior to experiments. Registration was performed using an affine transformation based on mutual information. For demonstrations, brain images for individual cross-sections were retrospectively undersampled in two-dimensions to yield acceleration ratios \( R = [2 : 1 : 10] \), via uniform random sampling with a \( 10 \times 10 \) central fully-sampled k-space region.

2) **In vivo Brain Dataset**: \( T_2 \)-and PD-weighted multi-coil complex images of 10 subjects were used, where 7 were reserved for training, 1 for validation, and 2 for testing. See supplementary materials for scan parameters. Because there was negligible interscan motion, no spatial registration was performed. Data were collected on a 3T Siemens Magnetom scanner using a 32-channel receive-only head coil at Bilkent University, Ankara, Turkey. Imaging protocols were approved by the local ethics committee at Bilkent University, and all participants provided written informed consent. To lower computational complexity, geometric-decomposition coil compression was performed to reduce the number of coils from 32 to 5 [50]. For demonstrations, brain images for individual cross-sections were retrospectively undersampled in two-dimensions to yield acceleration ratios \( R = [2 : 1 : 4] \), via uniform random sampling with a \( 16 \times 16 \) central fully-sampled k-space region.

### D. Implementation Details

The architecture of the generator and discriminator in ssGAN were adopted from a previous state-of-the-art study that demonstrated success in multi-contrast MRI synthesis [22]. The generator contained an encoder of 3 convolutional layers, a residual network of 9 ResNet blocks, and a decoder of 3 convolutional layers in series. The discriminator contained a convolutional neural network of 5 convolutional layers in series. An unlearned coil-combination block was placed at the input of the generator, so the generator recovered real-imaginary parts of the target image given real-imaginary parts of the coil-combined source image. The coil-combined target image was backprojected onto individual coils, and the complex target images from each coil were sequentially fed to an unconditional patch discriminator. The generator and the discriminator were alternately trained for 100 epochs with a batch size of 1 using the ADAM optimizer with first and second gradient moments of \( \beta_1 = 0.5 \) and \( \beta_2 = 0.999 \). The learning rate of the optimizer was set to 0.0002 in the first 50 epochs and was linearly decayed to 0 in the last 50 epochs. Cross-validation was used to select the relative weighting of the selective loss function components (\( \lambda_i \), \( \lambda_k \), ...)
loss in the validation set. The set of parameters ($\lambda_i = 100$, $\lambda_k = 3000$, $\lambda_a = 1$, $\beta = 5000$) that yielded near-optimal performance in both datasets were used in all experiments. Implementations were run on NVIDIA 1080 Ti and 2080 Ti GPUs in Python 2.7 using PyTorch. Code will be available at https://github.com/icon-lab/mrirecon.

E. Competing Methods

The proposed semi-supervised ssGAN model was comparatively demonstrated for multi-contrast MRI synthesis against several state-of-the-art methods.

1) pix2pix [12] (fully-sampled source, fully-sampled target): The fully-supervised pix2pix model based on paired, Nyquist-sampled source-target acquisitions sets a gold-standard for synthesis performance. pix2pix learns a mapping between coil-combined, magnitude source-target images, so it is geared for single-coil MRI synthesis. Here, pix2pix was trained with single-coil nonselective variants of tensor losses in ssGAN. The generator-discriminator architectures were taken from [22], and hyperparameters were optimized via cross-validation.

2) CycleGAN [11] (fully-sampled source, fully-sampled target): The CycleGAN model based on unpaired albeit fully-sampled source-target acquisitions is another gold-standard reference. CycleGAN also learns the contrast mapping between coil-combined magnitude MR images, making it suitable for single-coil synthesis. The network architecture, hyperparameters, and loss functions (single-coil nonselective adversarial and cycle-consistency) in [22] were adopted.

3) fsGAN (undersampled source, fully-sampled target): The fsGAN model was constructed as a gold-standard supervised baseline in cases where target acquisitions were Nyquist sampled but source acquisitions were undersampled. fsGAN learns to map Fourier reconstructions of undersampled, multi-coil source acquisitions onto target images. Here it was trained using nonselective variants of tensor image, k-space and adversarial loss functions in ssGAN. The network architecture was matched to ssGAN. Hyperparameters were selected via cross-validation and identical to ssGAN.

4) CasGAN (undersampled source, undersampled target): CasGAN is a cascaded method that sequentially performs reconstruction and synthesis to cope with mutually accelerated source-target acquisitions. Here compressed-sensing reconstructions were first performed to recover source-target images from undersampled acquisitions. A fully-supervised synthesis model was then learned based on these reconstructions. CS reconstructions were implemented using SparseMRI [45] for single-coil data, and L1-SPIRiT [47] for multi-coil data (https://people.eecs.berkeley.edu/mlustig/Software.html). Hyperparameters were selected via cross-validation. In SparseMRI, the number of iterations was 4, weight for total variation regularization was 0.0001 and weight for wavelet-domain L1 regularization was 0.0001. In SPIRiT, the kernel size was 5 × 5, weight for wavelet-domain L1-regularization was 0.1, weight for Tikhonov regularization during kernel estimation was 0.001, number of iterations was 10 for PD-weighted images and 20 for T2-weighted images. The synthesis model had identical architecture, loss functions and hyperparameters to fsGAN.

F. Experiments

1) Reliability against deficiencies in training data: Robustness against deficiencies in the quality and amount of training data was examined on the IXI dataset. Multiple independent ssGAN models were trained for $T_1 \rightarrow T_2$ synthesis while varying the training dataset. Variations were introduced by altering the acceleration ratio of target contrast acquisitions across $R_{\text{target}} = [2 : 1 : 10]$, whereas ssGAN was trained with $R_{\text{target}} \in [2 : 1 : 10]$. ssGAN-k with $k = R_{\text{target}}$. All models were trained with $R_{\text{source}} = 1$. Performance metrics for CycleGAN at $n_{T} = 8$ remain outside the display windows (see Fig. S1 for a broader display that also shows this model).

Fig. 2: Reliability of ssGAN against training data deficiencies. Evaluations were performed for $n_{T} = [8 : 8 : 64]$. For each $n_{T}$, pix2pix and CycleGAN were trained with $R_{\text{target}} = 1$, whereas ssGAN was trained with $R_{\text{target}} \in [2 : 1 : 10]$. ssGAN-k with $k = R_{\text{target}}$. All models were trained with $R_{\text{source}} = 1$. Performance metrics for CycleGAN at $n_{T} = 8$ remain outside the display windows (see Fig. S1 for a broader display that also shows this model).

F. Experiments

1) Reliability against deficiencies in training data: Robustness against deficiencies in the quality and amount of training data was examined on the IXI dataset. Multiple independent ssGAN models were trained for $T_1 \rightarrow T_2$ synthesis while varying the training dataset. Variations were introduced by altering the acceleration ratio of target contrast acquisitions across $R_{\text{target}} = [2 : 1 : 10]$, and by altering the number of training subjects across $n_{T} = [8 : 8 : 64]$. As gold-standard baselines, independent pix2pix and CycleGAN models were trained for the same variations in $n_{T}$ (albeit with $R_{\text{target}} = 1$). Additional experiments were conducted on $T_2 \rightarrow T_1$ synthesis, where ssGAN models with $R_{\text{target}} \in \{2, 3, 4\}$ were compared against pix2pix and CycleGAN. Fully-sampled source acquisitions $R_{\text{source}} = 1$ were assumed for all experiments.
2) Single-coil synthesis: Experiments were conducted on brain images from the IXI dataset to demonstrate synthesis performance on single-coil data. Demonstrations were performed on the $T_1 \rightarrow T_2$ and $T_2 \rightarrow T_1$ synthesis tasks with ssGAN, fsGAN and CasGAN methods. Independent ssGAN and CasGAN models were trained for different target accelerations: ssGAN-$k$ and CasGAN-$k$ trained with $R_{target} = k$, where $k \in \{2, 3, 4\}$. ssGAN and all competing methods were separately trained for $R_{source} = \{2, 3, 4\}$. A radiological evaluation was conducted on $T_2 \rightarrow PD$ and PD $\rightarrow T_2$ synthesis tasks with ssGAN, fsGAN and CasGAN. Independent ssGAN and CasGAN models were trained for different target accelerations: ssGAN-$k$ and CasGAN-$k$ trained with $R_{target} = k$, where $k \in \{2, 3, 4\}$. All competing methods were separately trained for $R_{source} = \{2, 3, 4\}$. A radiological evaluation was conducted on $T_2 \rightarrow PD$ and PD $\rightarrow T_2$ synthesis tasks with $R_{source} = 2, 4$. Opinion scores of an expert radiologist with more than 25 years of experience were considered. The quality of the synthesized images was rated based on similarity to reference images from fully-sampled acquisitions, on a five-point scale (0: unacceptable, 1: poor, 2: limited, 3: moderate, 4: good, 5: perfect match). For each synthesis task, radiological evaluations were performed on 5 different cross-sections randomly taken from each subject.

3) Multi-coil synthesis: Experiments were conducted on brain images from the in vivo dataset to demonstrate synthesis performance on multi-coil data. Demonstrations were performed on the $T_2 \rightarrow PD$ and PD $\rightarrow T_2$ synthesis tasks with ssGAN, fsGAN and CasGAN. Independent ssGAN and CasGAN models were trained for different target accelerations: ssGAN-$k$ and CasGAN-$k$ trained with $R_{target} = k$, where $k \in \{2, 3, 4\}$. ssGAN and all competing methods were separately trained for $R_{source} = \{2, 3, 4\}$. A radiological evaluation was conducted on $T_2 \rightarrow PD$ and PD $\rightarrow T_2$ synthesis tasks with $R_{source} = 2, 4$. Opinion scores of an expert radiologist with more than 25 years of experience were considered. The quality of the synthesized images was rated based on similarity to reference images from fully-sampled acquisitions, on a five-point scale (0: unacceptable, 1: poor, 2: limited, 3: moderate, 4: good, 5: perfect match). For each synthesis task, radiological evaluations were performed on 5 different cross-sections randomly taken from each subject.

4) Ablation studies: Experiments were conducted to individually examine the effects of the tensor image, k-space and adversarial loss functions on synthesis quality. Demonstrations were performed on IXI for $T_1 \rightarrow T_2$ and $T_2 \rightarrow T_1$ synthesis tasks. Four independent ssGAN models were trained: ssGAN (with all loss functions, ssGAN(w/o image) without the image loss, ssGAN(w/o k-space) without the k-space loss, and ssGAN(w/o adv) without the adversarial loss. Acceleration rates of $R_{source} = 4$ and $R_{target} = 4$ were assumed.

Synthesis performance was evaluated using peak signal-to-noise ratio (PSNR), structural similarity (SSIM) and mean-squared error (MSE) metrics (all MSE reports reflect measured $MSE \times 100$). Metrics were measured on coil-combined magnitude images derived from synthesized and reference target contrasts. The reference image was based on Fourier reconstructions of fully-sampled target acquisitions. In Tables, summary statistics of quantitative metrics were provided as mean ± std across test subjects. Significance of PSNR, SSIM, MSE, and radiological opinions scores was assessed via Kruskal Wallis H-test ($p < 0.05$) to collectively compare ssGAN models vs pix2pix, ssGAN models vs CycleGAN, and ssGAN models vs fsGAN, and via Wilcoxon signed-rank test ($p < 0.05$) to individually compare ssGAN-2 vs CasGAN-2, ssGAN-3 vs CasGAN-3, and ssGAN-4 vs CasGAN-4.

III. RESULTS

A. Robustness of semi-supervised learning against deficiencies in training data

We first performed comprehensive experiments to examine the reliability of ssGAN against deficiencies in training data.

Fig. 3: ssGAN was demonstrated on IXI for $T_2 \rightarrow T_1$ mapping against pix2pix and CycleGAN with ($R_{source} = 1$). Synthesized images from ssGAN-2, pix2pix, and CycleGAN are displayed together with the reference (i.e., target) and source images in the first row. The corresponding error maps for the synthesized images are displayed in the second row (see colorbar). For comparison with ssGAN-3,4, see Fig. S2.

Models were learned for $T_1 \rightarrow T_2$ mapping in the IXI dataset, while acceleration ratio of target acquisitions ranged in $R_{target} = [2 : 1 : 10]$ and number of training subjects ranged in $n_T = [8 : 8 : 64]$. As baselines, gold-standard pix2pix and CycleGAN models were also trained for matching $n_T$ but with $R_{target} = 1$. All models were input high-quality source images reconstructed from Nyquist-sampled acquisitions ($R_{source} = 1$). Synthesis performance in terms of PSNR, SSIM, and MSE is displayed in Fig. 2 as a function of $R_{target}$ and $n_T$. The reported measurements indicate that synthesis quality of ssGAN is on par with the gold-standard pix2pix model ($p > 0.05$), where ssGAN performance is within $[-0.51, 0.13]$ dB PSNR, $[-0.70, 0.02]$ % SSIM, and $[0.019, -0.004]$ MSE windows of pix2pix. Meanwhile, ssGAN outperforms CycleGAN with 1.18 dB higher PSNR, 1.16 % SSIM, and 0.081 lower MSE ($p < 0.05$). Importantly, ssGAN models trained with varying $R_{target}$ yield highly similar performance, where ssGAN achieves near-optimal synthesis quality while undersampling target acquisitions up to 10-fold.

Measurements reported in Fig. 2 also demonstrate that increasing $n_T$ improves performance of all competing methods. Comparing $n_T = 16$ against $n_T = 32$, average improvements in (PSNR, SSIM, MSE) are $(0.68 \text{dB}, 0.57 \%, -0.025)$ for ssGAN, $(0.58 \text{dB}, 0.47 \%, -0.020)$ for pix2pix, and $(0.32 \text{dB}, 0.24 \%, -0.012)$ for CycleGAN. Note that ssGAN improves scan efficiency by accelerating target acquisitions, so in principle training data from a larger group of subjects can be collected at high acceleration rates for training ssGAN, compared to fully-supervised or unpaired models. For instance, given a total, active scan time of 126 min, Nyquist-sampled k-space data for $T_1$ and $T_2$-weighted images can be collected in 16 subjects for pix2pix and CycleGAN. In the same duration, a protocol with undersampled target acquisitions ($R_{target} = 10$) can be performed in 32 subjects for ssGAN, resulting in performance benefits of $(0.34 \text{dB}, 0.04 \%, -0.012)$ over pix2pix and $(0.74 \text{dB}, 0.48 \%, -0.031)$ over CycleGAN. Therefore, ssGAN enables elevated diversity in the training set to improve accuracy and practicality of learning-based MRI synthesis.

![Image](image.png)
corresponding error maps for the synthesized images are displayed in the second row. Representative results from two different subjects are displayed, along with the source and reference images.

Table 1: Image Quality for Mutually Accelerated Single-Coil MRI Synthesis

| Method | T1 → T2 | T2 → T1 |
|--------|---------|---------|
| PSNR | SSIM | MSE | PSNR | SSIM | MSE |
| ssGAN | 26.18 | 92.58 | 0.259 | 27.01 | 93.54 | 0.215 |
| fsGAN | 25.51 | 91.61 | 0.302 | 26.81 | 93.1 | 0.227 |
| CasGAN | 26.78 | 92.24 | 0.279 | 27.63 | 93.86 | 0.225 |

We then extended the demonstrations of ssGAN against pix2pix and CycleGAN by comparisons on the T2 → T1 synthesis task in the IXI dataset with a fixed number of training subjects nT = 32 (used hereafter in all evaluations in IXI). Measurements of synthesis quality are reported in Table S1. The reported measurements reveal that ssGAN models maintain near-optimal synthesis quality par with pix2pix, and on average they outperform CycleGAN with (0.41 dB, 0.34%, -0.015) improvement in (PSNR, SSIM, MSE). Representative results displayed in Fig. 3 corroborate the quantitative findings by showing that ssGAN offers a similar level of accuracy in tissue depiction to pix2pix, while it synthesizes higher quality images compared to CycleGAN that suffers from elevated errors.

B. Single-coil image synthesis in mutually accelerated multi-contrast MRI

To examine the synthesis performance of ssGAN in mutually accelerated MRI, we conducted experiments on IXI where both source and target acquisitions were undersampled.
Single-coil image synthesis was considered with $T_1 \rightarrow T_2$ and $T_2 \rightarrow T_1$ recovery tasks. ssGAN was compared against a gold-standard supervised model (fsGAN) that was trained on undersampled source acquisitions but Nyquist-sampled target acquisitions, and against a sequential model (CasGAN) that first reconstructed undersampled acquisitions, and then trained a synthesis model on the reconstructed source and target images. The target acceleration ratio varied in $R_{\text{target}} = \{2, 3, 4\}$ for ssGAN and CasGAN resulting in three separate models for each method: ssGAN-k and CasGAN-k with $k = R_{\text{target}}$. Meanwhile, the acceleration ratio for the source acquisitions varied in $R_{\text{source}} = \{2, 3, 4\}$ for all methods.

Quantitative metrics for synthesis performance are listed in Table I for varying $R_{\text{source}}$ and $R_{\text{target}}$ values. Overall, ssGAN models at moderate acceleration factors for the target acquisition yield near-optimal performance on par with the gold-standard fsGAN model ($p > 0.05$). On average across $R_{\text{target}}$, ssGAN outperforms CasGAN by 0.92 dB in PSNR, 3.18 % in SSIM, and $-0.130$ in MSE ($p < 0.05$, except for PD $\rightarrow$ T2 with $R_{\text{source}} = 4$). Incremental steps from $R_{\text{target}} = 1$ to $R_{\text{target}} = 4$ result in an average performance loss of 0.12 dB PSNR, 0.30 % SSIM, and 0.010 MSE for ssGAN, and 0.82 dB PSNR, 2.25 % SSIM, and 0.07 MSE for CasGAN. Similar to single-coil results, this finding demonstrates the utility of the selective loss function in ssGAN to cope with moderately undersampled target acquisitions.

For further validation of the quantitative assessments, radiological evaluations were performed for ssGAN-2, fsGAN and CasGAN-2 on $T_2 \rightarrow$ PD and PD $\rightarrow$ T2 synthesis tasks. Representative synthetic images are displayed in Fig. 5 and Fig. S3, whereas results of radiological evaluation are shown in Fig. 6. ssGAN images are visually similar to fsGAN, whereas they manifest superior synthesis quality compared to CasGAN. In terms of opinion score, ssGAN maintains a high-level of synthesis quality on par with ssGAN ($p > 0.05$, except for $R_{\text{source}} = 2$), and on average a modest score difference of 0.325 is observed. In contrast, ssGAN yields superior performance to CasGAN with an average improvement of 1.075 in opinion score across tasks ($p < 0.05$).

### D. Ablation Studies

Ablation experiments were conducted to demonstrate the contribution of individual loss components in ssGAN. Independent ssGAN models were trained while the loss components were selectively ablated (see Experiments). Models were learned for $T_1 \rightarrow T_2$ and $T_2 \rightarrow T_1$ mappings in the IXI dataset. The effects of image and k-space losses were evaluated using PSNR, SSIM and MSE metrics, whereas the effect of adversarial loss was assessed using Frechlet Inception Distance (FID) scores and visual inspection as common in literature [51]. Quantitative metrics listed in Table S2 indicate that the selective image and k-space losses serve to improve synthesis quality in both $T_1$ and $T_2$ recovery tasks. Meanwhile,

| $R_{\text{source}} = 2$ | $R_{\text{source}} = 3$ | $R_{\text{source}} = 4$ |
|------------------------|------------------------|------------------------|
| $T_1 \rightarrow$ PD   | $T_2 \rightarrow$ T1   | PSNR SSIM MSE PSNR SSIM MSE PSNR SSIM MSE |
| ssGAN                  | 25.04 ± 0.338          | 24.77 ± 0.345          |
| CasGAN-2               | 25.14 ± 0.315          | 24.65 ± 0.356          |
| CasGAN-3               | 25.2 ± 0.317           | 24.61 ± 0.358          |
| CasGAN-4               | 25.01 ± 0.325          | 24.16 ± 0.398          |
| ssGAN                  | 24.77 ± 0.374          | 24.87 ± 0.386          |
| CasGAN-2               | 24.77 ± 0.374          | 24.87 ± 0.386          |
| CasGAN-3               | 25.44 ± 0.406          | 24.2 ± 0.444           |
| CasGAN-4               | 22.68 ± 0.579          | 22.83 ± 0.597          |
| ssGAN                  | 22.68 ± 0.579          | 22.83 ± 0.597          |
| CasGAN-2               | 22.68 ± 0.579          | 22.83 ± 0.597          |
| CasGAN-3               | 22.68 ± 0.579          | 22.83 ± 0.597          |
| CasGAN-4               | 22.68 ± 0.579          | 22.83 ± 0.597          |

### C. Multi-coil image synthesis in mutually accelerated multi-contrast MRI

Next, we conducted experiments on the in vivo brain dataset to demonstrate multi-coil MRI synthesis with the proposed ssGAN model. Multi-coil image synthesis was considered for $T_2 \rightarrow$ PD and PD $\rightarrow$ T2 recovery tasks. As in single-coil synthesis, ssGAN was compared against fsGAN and CasGAN. $R_{\text{target}} = \{2, 3, 4\}$ and $R_{\text{source}} = \{2, 3, 4\}$ were considered.
the selective adversarial loss component increases the realism of synthetic images with decreased FID scores.

IV. Discussion

Here, we introduced a novel semi-supervised deep generative model for image synthesis in multi-contrast MRI that is mutually accelerated across both contrast sets and k-space. As opposed to supervised models [17]–[23], ssGAN learns to synthesize high-quality target-contrast images in the absence of training sets composed of costly acquisitions of Nyquist-sampled source and target contrasts. ssGAN achieves synthesis quality on par with gold-standard supervised models for a broad range of acceleration ratios. This performance leap is mediated by selective loss functions in image, k-space, and adversarial domains. Unlike prior synthesis methods, ssGAN processes multi-coil complex MRI data and learns to synthesize directly from undersampled source acquisitions. Therefore, ssGAN holds great promise in advancing the practicality and utility of multi-contrast MRI synthesis.

Comprehensive experiments were conducted on single- and multi-coil neuroimaging datasets to demonstrate the proposed approach. Our experiments indicate that ssGAN achieves equivalent performance to gold-standard fully-supervised models, pix2pix [11] with fully-sampled source acquisitions, and fgGAN with undersampled source acquisitions. Importantly, ssGAN maintains near-optimal synthesis for acceleration factors up to 10 for target acquisitions given the same amount of training subjects, significantly reducing the data requirements for model training. Furthermore, ssGAN outperforms an alternative weakly-supervised synthesis model CasGAN based on a sequential hybrid of compressed-sensing reconstructions followed by supervised synthesis. Compared to CasGAN, ssGAN enables end-to-end learning of target contrast recovery, alleviating potential propagation of errors across the pipeline and reducing computational complexity.

Here, we demonstrated ssGAN with uniform-density random undersampling in k-space. An alternative would be to collect low-resolution source-target images by Nyquist-sampling within a central k-space region to achieve similar acceleration. The problem would then be transformed into a superresolution task on coil-combined images [52], [53]. However, this superresolution task is challenging when both source and target images are low resolution, so external priors might be required to enable recovery of high-spatial-frequency information. Another alternative would be to perform variable-density sampling where central k-space is more densely sampled [45]. Variable-density schemes trade-off high-frequency coverage in return for increased signal-to-noise ratio in acquired data. As such, they might improve aggregate performance metrics (e.g., PSNR, MSE) that are dominated by low-spatial-frequency errors, particularly for higher acceleration factors used in ssGAN and CasGAN. Note, however, that uniform-density sampling improves peripheral k-space coverage to expand high-spatial-frequency information, and in turn contributes to recovery of detailed tissue structure.

In the current study, we demonstrated the proposed method on one-to-one synthesis tasks with a single source and a single target contrast. In multi-contrast protocols, multiple source and/or target contrasts might be available, one might additionally be interested in many-to-one and many-to-many synthesis tasks [23], [27]. In such cases, a many-to-many variant of ssGAN can be constructed by concatenating the multitude of source and target contrasts as separate input and output channels, respectively [23]. The selective loss function along with the k-space masking and coil projection operators can then be defined separately for each target contrast.

The ssGAN implementation considered here leverages a multi-coil tensor loss between undersampled versions of the synthesized and reference target images. This image-domain loss implicitly assumes that the source and target-contrast acquisitions are spatially registered. The datasets examined here were either aligned or a registration step was performed during preprocessing. If an end-to-end alternative is desired that can cope with misaligned source-target acquisitions, deep network-based registration models can be cascaded to the input of ssGAN for spatial registration [54]. It remains important future work to explore the extent of improvements in synthesis performance with integrated registration and synthesis.

The semi-supervised learning framework that ssGAN leverages undersampled albeit paired acquisitions of source and target contrasts from the same set of subjects. Our results suggest that successful ssGAN models can be trained even with relatively modest size datasets. However, more complex models including order of magnitude higher number of parameters such as 3D architectures might require substantial datasets for reliable training. In such cases, a variant of ssGAN that permits training on a hybrid of paired and unpaired images or directly on unpaired images would be valuable. To do this, the cycle-consistent counterpart of the selective loss function in ssGAN can be devised [22], [27], [37], [41], [42].

In summary, here we proposed a semi-supervised learning framework based on generative adversarial networks that can recover high-quality target images without demanding Nyquist-sampled ground truth. While the superior data-efficiency of ssGAN was primarily demonstrated for within-modality contrast conversion in the brain, it can also be adopted to other anatomies, other recovery tasks including multi-parametric MRI synthesis, or cross-modality mappings between MRI and other imaging modalities [39], [41], [55].
[41] C.-B. Jin, H. Kim, M. Liu, W. Jung, S. Joo, E. Park, Y. S. Ahn, I. H. Han, J. I. Lee, and X. Cui, “Deep CT to MR synthesis using paired and unpaired data,” Sensors, vol. 19, no. 10, 2019.

[42] Z. Wang, Y. Lin, K. -T. T. Cheng, and X. Yang, “Semi-supervised mp-MRI data synthesis with StitchLayer and auxiliary distance maximization,” Medical Image Analysis, vol. 59, p. 101565, 2020.

[43] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Ügurbil, and M. Akçaçay, “Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data,” Magnetic Resonance in Medicine, vol. 84, no. 6, pp. 3172–3191, 2020.

[44] E. K. Cole, J. M. Pauly, S. S. Vasanawala, and F. Ong, “Unsupervised MRI reconstruction with generative adversarial networks,” arXiv:2008.13065, 2020.

[45] M. Lustig, D. Donoho, and J. M. Pauly, “Sparse MRI: The application of compressed sensing for rapid MR imaging,” Magnetic Resonance in Medicine, vol. 58, no. 6, pp. 1182–1195, 2007.

[46] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, “Compressed sensing MRI,” IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 72–82, 2008.

[47] M. Lustig and J. M. Pauly, “SPIRiT: iterative self-consistent parallel imaging reconstruction from arbitrary k-space,” Magnetic Resonance in Medicine, vol. 64, no. 2, pp. 457–471, 2010.

[48] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig, “ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA,” Magnetic Resonance in Medicine, vol. 71, no. 3, pp. 990–1001, 2014.

[49] S. Smith, M. Jenkinson, M. Woolrich, C. Beckmann, T. Behrens, H. Johansen-Berg, P. Bannister, M. Luca, I. Drobnjak, D. Flitney, R. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, M. Brady, and P. Matthews, “Advances in functional and structural MR image analysis and implementation as FSL,” NeuroImage, vol. 23 Suppl 1, pp. S208–19, 2004.

[50] T. Zhang, J. M. Pauly, S. S. Vasanawala, and M. Lustig, “Coil compression for accelerated imaging with Cartesian sampling,” Magnetic Resonance in Medicine, vol. 69, no. 2, pp. 571–582, 2013.

[51] T. DeVries, A. Romero, L. Pineda, G. W. Taylor, and M. Drozdzal, “On the evaluation of conditional GANs,” arXiv:1907.08175, 2019.

[52] C. Zhao, B. E. Dewey, D. L. Pham, P. A. Calabresi, D. S. Reich, and J. L. Prince, “Smore: A self-supervised anti-aliasing and super-resolution algorithm for mri using deep learning,” IEEE Transactions on Medical Imaging, vol. 40, no. 3, pp. 805–817, 2021.

[53] Q. Lyu, H. Shan, C. Steber, C. Helis, C. Whitlow, M. Chan, and G. Wang, “Multi-contrast super-resolution MRI through a progressive network,” IEEE Transactions on Medical Imaging, vol. 39, no. 9, pp. 2738–2749, 2020.

[54] X. Yang, R. Kwitt, and M. Niethammer, “Quicksilver: Fast predictive image registration – a deep learning approach,” NeuroImage, vol. 158, pp. 378–396, 2017.

[55] B. Dewey, C. Zhao, J. Reinhold, A. Carass, K. Fitzgerald, E. Sotirchos, S. Saidha, J. Oh, D. Pham, P. Calabresi, P. van zijl, and J. Prince, “DeepHarmony: a deep learning approach to contrast harmonization across scanner changes,” Magnetic Resonance Imaging, vol. 64, 2019.
**Supp. Text**

Acquisition Parameters of the IXI Dataset

- **T1-weighted images**: TR = 9.81 ms, TE = 4.603 ms, flip angle = 8°, matrix size = 256 × 256 × 150, spatial resolution = 0.94 × 0.94 × 1.2 mm³, acquisition time = 4:42.
- **T2-weighted images**: TR = 8178.34 ms, TE = 100 ms, flip angle = 90°, matrix size = 256 × 256 × 150, spatial resolution = 0.94 × 0.94 × 1.2 mm³, acquisition time = 3:11.

Acquisition Parameters of the In Vivo Brain Dataset

- **T2-weighted images**: 3D spin-echo sequence, TR = 1000 ms, TE = 118 ms, flip angle = 90°, imaging matrix = 256 × 192 × 88, spatial resolution = 1 × 1 × 2 mm³, acquisition time = 17:39.
- **PD-weighted images**: 3D spin-echo sequence, TR = 750 ms, TE = 12 ms, flip angle = 90°, imaging matrix = 256 × 192 × 88, spatial resolution = 1 × 1 × 2 mm³, acquisition time = 13:14.
Fig. S1: Reliability of ssGAN against training data deficiencies. Evaluations were performed for \( n_T = [8 : 8 : 64] \). For each \( n_T \), pix2pix and CycleGAN were trained with \( R_{\text{target}} = 1 \), whereas ssGAN was trained with \( R_{\text{target}} \in [2 : 1 : 10] \), ssGAN-\( k \) with \( k = R_{\text{target}} \). All models were trained with \( R_{\text{source}} = 1 \).
### TABLE S1: Quality of Synthesis in the IXI Dataset for Nyquist-Sampled Source Acquisitions

| Method    | PSNR ($T_1 \rightarrow T_2$) | SSIM ($T_1 \rightarrow T_2$) | MSE ($T_1 \rightarrow T_2$) | PSNR ($T_2 \rightarrow T_1$) | SSIM ($T_2 \rightarrow T_1$) | MSE ($T_2 \rightarrow T_1$) |
|-----------|-------------------------------|------------------------------|-----------------------------|-------------------------------|------------------------------|-----------------------------|
| pix2pix   | 28.57                         | 95.33                        | 0.15                        | 28.62                         | 95.95                        | 0.154                       |
|           | $\pm 1.39$                    | $\pm 1.35$                  | $\pm 0.048$                | $\pm 1.45$                    | $\pm 1.38$                  | $\pm 0.057$                |
| CycleGAN  | 27.91                         | 94.66                        | 0.178                       | 28.05                         | 95.4                         | 0.175                       |
|           | $\pm 1.52$                    | $\pm 1.51$                  | $\pm 0.062$                | $\pm 1.41$                    | $\pm 1.52$                  | $\pm 0.062$                |
| ssGAN-2   | 28.63                         | 95.25                        | 0.148                       | 28.52                         | 95.79                        | 0.158                       |
|           | $\pm 1.42$                    | $\pm 1.33$                  | $\pm 0.049$                | $\pm 1.61$                    | $\pm 1.46$                  | $\pm 0.061$                |
| ssGAN-3   | 28.56                         | 95.24                        | 0.15                        | 28.44                         | 95.72                        | 0.16                        |
|           | $\pm 1.39$                    | $\pm 1.32$                  | $\pm 0.048$                | $\pm 1.55$                    | $\pm 1.41$                  | $\pm 0.059$                |
| ssGAN-4   | 28.6                          | 95.21                        | 0.149                       | 28.42                         | 95.7                         | 0.162                       |
|           | $\pm 1.43$                    | $\pm 1.34$                  | $\pm 0.05$                 | $\pm 1.56$                    | $\pm 1.42$                  | $\pm 0.061$                |
Fig. S2: ssGAN was demonstrated on IXI for $T_2 \rightarrow T_1$ mapping against pix2pix and CycleGAN with $(R_{\text{source}} = 1)$. Synthesized images from ssGAN-2, ssGAN-3, ssGAN-4, pix2pix, and CycleGAN are displayed together with the reference and source images in the first row. The corresponding error maps for the synthesized images are displayed in the second row.
Fig. S3: The proposed ssGAN model was demonstrated on the in vivo brain dataset for multi-coil complex PD $\rightarrow$ $T_2$ synthesis task with source contrast acquisitions undersampled by $R_{\text{target}} = 2$. Representative results from two different subjects are shown in separate rows. Synthesized coil-combined images from fsGAN, ssGAN, and CasGAN are displayed along with the source image and the reference image.
**TABLE S2**: Effects of Tensor Losses on Synthesis Quality

|         | T1 to T2 | T2 to T1 |
|---------|----------|----------|
|         | PSNR     | SSIM     | MSE   | PSNR | SSIM | MSE |
| ssGAN   | 25.25    | 90.75    | 0.32  | 26.47 | 92.27 | 0.24 |
| ssGAN (w/o image) | 25.12    | 90.36    | 0.328 | 26.12 | 91.78 | 0.264 |
| ssGAN (w/o k-space) | 24.96    | 90.30    | 0.341 | 26.41 | 92.19 | 0.246 |
|         | T1 to T2 | T2 to T1 |
|         | FID      | FID      |
| ssGAN   | 22.89    | 16.80    |
| ssGAN (w/o adv) | 23.13    | 24.35    |