# Fast Unsupervised MRI Reconstruction Without Fully-Sampled Ground Truth Data Using Generative Adversarial Networks

Elizabeth K. Cole[1], Frank Ong [1], Shreyas S. Vasanawala[2], and John M. Pauly[1]

[1]Department of Electrical Engineering, Stanford University
[2]Department of Radiology, Stanford University
{*ekcole, franko, vasanawala, pauly*}*@stanford.edu*

## Abstract

*Most deep learning (DL) magnetic resonance imaging (MRI) reconstruction approaches rely on supervised training algorithms, which require access to high-quality, fully-sampled ground truth datasets. In MRI, acquiring fully-sampled data is time-consuming, expensive, and, in some cases, impossible due to limitations on data acquisition speed. We present a DL framework for MRI reconstruction that does not require any fully-sampled data using unsupervised generative adversarial networks. We test our proposed method on 2D knee MRI data and 2D+time abdominal dynamic contrast enhanced (DCE) MRI data. In the DCE-MRI dataset, as is the case with many dynamic MRI sequences, ground truth was not possible to acquire and therefore, supervised DL reconstruction was not feasible. We show that our unsupervised method produces reconstructions which are better than compressed sensing in terms of image metrics and the recovery of anatomical structure, with faster inference time. In contrast to most deep learning reconstruction techniques, which are supervised, this method does not need any fully-sampled data. With the proposed method, accelerated imaging and accurate reconstruction can be performed in applications in cases where fully-sampled datasets are difficult to obtain or unavailable.*

## 1. Introduction

Magnetic resonance imaging (MRI) is a Fourier imaging modality which enables comprehensive evaluation of soft-tissue anatomy and physiology. Unfortunately, MRI scans are intrinsically long because data is acquired sequentially in the Fourier domain - in contrast with traditional camera imaging, where all pixels are acquired simultaneously. The acquisition time can be reduced by acquiring less data through sampling below Nyquist rate in the Fourier domain, otherwise known as k-space. However, acquisition acceleration results in poor image quality when using Fourier inversion for reconstruction. Many iterative reconstruction methods, such as parallel imaging (PI) [15, 37] and compressed sensing (CS) [29]. Recently, deep learning (DL) methods [17, 7, 32, 52, 10, 8, 1, 46, 11, 9] have proven to be more powerful than traditional methods, providing more robustness for different anatomy and sequences, higher quality, and faster reconstruction speed.

Most DL techniques for MRI reconstruction must be supervised, involving a large number of high-quality, fully-sampled acquisitions for training. In MRI, collecting fully-sampled data is time-consuming, expensive, and, in some cases, impossible due to physical limitations on data acquisition rate. For example, dynamic contrast-enhanced (DCE) imaging is a type of MRI exam in which data is acquired during intravenous contrast injection. The goal is to resolve tissue uptake of a contrast agent; however, often the data acquisition rate is slower than the contrast dynamics. To achieve clinically acceptable temporal resolutions, spatial resolution and/or signal-to-noise ratio (SNR) must be traded off, impacting downstream interpretation and ultimately limiting the utility of DCE-MRI in clinical decision-making. DL has the potential to enable high-value rapid DCE-MRI scans, however; without high-quality ground truth, the vast majority of existing supervised DL-based reconstruction methods cannot be used for DCE-MRI and other types of dynamic MRI acquisitions such as real-time, cardiac cine, and phase-contrast imaging.

There are a couple current ways to address this problem. First, parallel imaging-compressed sensing (PI-CS) reconstructions can be used as the ground truth for a weakly supervised DL framework [8]. However, the image quality of PI-CS is potentially an upper bound for the image quality of the DL model. Another way is to reformulate

DL training by creating a hybrid supervised-unsupervised model [48, 57, 23, 45, 51, 6, 50, 44, 35, 18].

One such hybrid supervised-unsupervised work by Sim *et al*. [44] used unpaired datasets to train Cycle Generative Adversarial Network (GAN) for inverse problems. This was a step forward towards MRI reconstruction for applications where no fully-sampled data is available. However, the remaining limitation was that this work still used fully-sampled data from the same type of acquisition as the intended reconstruction task. Therefore, the network had a strong prior from this ground truth data related to the target distribution. This method is unproven to work in applications where no fully-sampled data from the same type of acquisition is available.

Another hybrid approach, Noise2Noise [23] and similar variants [18, 26, 27] successfully trained a model for noise removal based only on noisy training data. One limitation of this method is that it requires at least two independent prospectively undersampled scans of the same underlying image, which is clinically impractical due to long scan times, patient motion between and throughout scans, as well as changing dynamics from scan to scan. Alternatively, this method can use two differently retrospectively undersampled acquisitions, which defeats the original purpose of training without clean data. Therefore, this precludes the use of existing formulations of Noise2Noise for cases where no fully-sampled data is available.

A promising direction to address MRI reconstruction without using ground truth is deploying GANs [14, 36]. GANs have proven very useful in creating visually appealing natural images [56], modeling underlying data distributions [14, 38], and constructing generative models for MRI reconstruction [32, 52, 31]. AmbientGAN [4] learns a generative model from underdetermined linear systems. The authors attempted to train a generative model on only partial observations with tasks in image inpainting, denoising and deconvolution. They showed that despite never training the discriminator network on fully observed data, their generator network was still able to produce samples which recovered the underlying data distribution of small-scale simulated datasets, such as MNIST [22] and celebA [28].

The objective of AmbientGAN is similar to the problem we are trying to solve in unsupervised MRI reconstruction, where we only have undersampled k-space measurements for our training set. One notable difference, elaborated below, is that our reconstruction model is a conditional GAN where we directly learn a mapping from the input undersampled k-space to the output reconstructed image. In contrast, AmbientGAN learns a network which only takes latent codes as inputs.

In this work, we propose a fully unsupervised generative method for learned MRI reconstruction that only uses undersampled datasets and never uses any fully-sampled data, with inference and high-quality reconstruction using unrolled networks. This enables efficient DL reconstruction when it is impossible or difficult to obtain fully-sampled data. To evaluate this method, we first implement the method on retrospectively undersampled datasets and compare the results to CS and a semi-supervised DL model using standard image quality metrics. Then, we evaluate on prospectively undersampled DCE datasets, where a ground truth was not available; thus, we can only compare our method against CS.

## 2. Methods

We consider the standard multi-channel Fourier acquisition model for MRI, which can be written as:

$$y = Ax + \epsilon. \tag{1}$$

where $y$ is the measured k-space data, $A$ is the imaging model, $x$ is the latent true image, and $\epsilon$ is the additive complex Gaussian noise. The imaging model $A$ consists of data subsampling, a Fourier transform, and signal modulation by coil sensitivity maps. We consider acquisitions with randomized sampling masks; thus, A is a random matrix drawn from a known distribution $p_A$, where $p_A$ is the Poisson-disc variable density sampling [5] distribution. The overall k-space data distribution is denoted as $p_y$.

In GANs [14], a pair of neural networks are jointly trained. The generator network tries to map samples from a low-dimensional distribution that is easy to sample from (such as Gaussian noise) to samples from a high-dimensional space. Meanwhile, the discriminator network tries to differentiate between generated and real samples. To jointly train the networks, a min-max game is employed, where the loss function of the generator is based on the output of the discriminator. Convergence of a GAN is signified by obtaining equilibrium between the generator and discriminator. Many GAN loss objectives exist, including those of deep convolutional GAN (DCGAN) [38], least-squares GAN (LSGAN) [30], Wasserstein GAN (WGAN) [2], and WGAN with gradient penalty (WGAN-GP) [16].

In a standard DL setting where ground truth is available, one could formulate a semi-supervised GAN for MRI reconstruction by using fully-sampled reconstructions as the real images [43]. Figure 1a illustrates the overall framework of training a semi-supervised GAN when fully-sampled datasets are available. However, when we only have access to training datasets where $A$ is underdetermined due to data subsampling, we don't have ground truth to use as those real images.

AmbientGAN attempts to solve this problem of lack of clean training data by training the discriminator to differentiate between a real measurement from a simulated measurement of a generated image. The original AmbientGAN

objective is:

$$\min_G \max_D \mathbb{E}_{y\sim p_y}\left[q\left(D(y)\right)\right] + \mathbb{E}_{z\sim p_z, A\sim p_A}\left[q\left(1-D\left(A(G(z))\right)\right)\right] \quad (2)$$

where $G$ is the generator network, $D$ is the discriminator network, and $y$ denotes the observed subsampled measurements drawn from the distribution $p_y$. The vector $z$ denotes a random latent vector of a distribution $p_z$ that is easy to sample from, such as IID Gaussian noise. $A$ denotes the inverse model of a subsampling mask randomly chosen from a set of subsampling masks. $p_A$ is the Poisson-disc variable density sampling distribution. $q(\cdot)$ denotes the quality function used to define the objective. For DCGAN [50], $q(t) = t$. The generated images are given by $G(z)$ and generated measurements are given by $A(G(z))$.

Note that in AmbientGAN, the input vector $z$ is mapped from a distribution of random latent vectors to a higher dimensional image space. However, for reconstruction, we want a generative mapping from k-space to an image space, not from a latent distribution to an image space. Therefore, we propose to use a conditional GAN, where the input is actually subsampled k-space data $y$ from the distribution $p_y$, not the noise vector $z$. Again, $y$ denotes the observed subsampled measurements drawn from the distribution $p_y$. In our setting, we attempt to reconstruct a set of input subsampled k-space data $y^{input}$. We use a separate training dataset as the real measurements that are fed into the discriminator, $y^{train}$. Both $y^{input}$ and $y^{train}$ are drawn from the same distribution of subsampled k-space $y$, and are disjoint sets. $G(y^{input})$ is the resulting generated high-resolution image from the input undersampled k-space data.

Concretely, our overall proposed objective is:

$$\min_G \max_D \mathbb{E}_{y\sim p_y}\left[q\left(D\left(y\right)\right)\right] + \mathbb{E}_{y\sim p_y, A\sim p_A}\left[q(1-D\left(A(G\left(y\right))\right))\right]. \quad (3)$$

where the discriminator is trained to distinguish between real measurements $y^{train}$ and generated measurements $A(G(y^{input}))$.

In this work, we propose to use the adversarial loss of WGAN-GP because WGAN-GP has been shown to potentially have the best convergence [16]. Later, we compare our results of WGAN-GP to that of LSGAN. The gradient penalty loss for WGAN-GP is:

$$\lambda \max_D \mathbb{E}_{y\sim p_y}\left[\left(\left\|\nabla_y D\left(y\right)\right\|_2 - 1\right)^2\right. \quad (4)$$

where we set $\lambda = 10$ [16].

A subtle difference between the AmbientGAN setting and our setting is that all our variables are in the complex domain instead of real. However, this can be dealt with by splitting the real and imaginary components into two channels, or using custom complex-valued building blocks [9].

One potential concern with this objective is that the generator may directly output the zero-filled reconstruction of $y$ because both generator and discriminator take undersampled k-space measurements as inputs. The key to addressing this problem is that the simulated imaging model $A$ in the objective is independently drawn with respect to the generator inputs $y$. Therefore, the simulated measurements $A(G(y))$ are unlikely to be sampled at the same k-space locations as the input $y$. The discriminator would likely enforce the generator to fill in missing k-space measurements, because otherwise the discriminator can easily classify the generated data as fake.

Our proposed unsupervised framework is shown in Figure 1b. The input to the generator network is an undersampled complex-valued k-space data and the output is a reconstructed two-dimensional complex-valued image. An imaging matrix comprised of coil sensitivity maps, an FFT and a randomized undersampling mask (drawn independently from the input k-space measurements) is applied to the generated image to simulate the imaging process. The discriminator takes simulated and observed measurements as inputs and tries to differentiate between them.

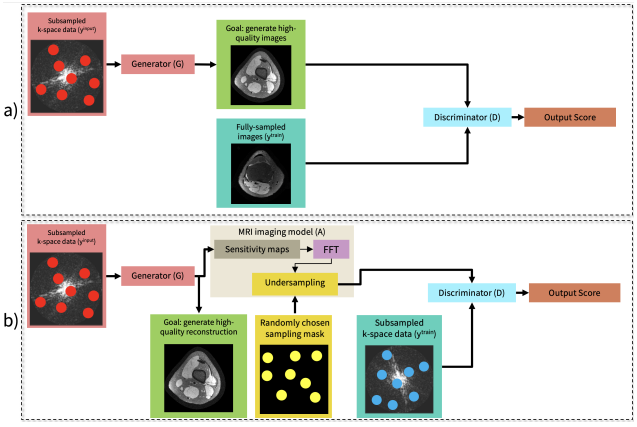


Figure 1. Existing semi-supervised training (a) and our proposed unsupervised training (b). (a) Framework overview in a semi-supervised setting with a conditional GAN when fully-sampled datasets are available. (b) Our proposed framework overview in an unsupervised setting. The input to the generator network is an undersampled complex-valued k-space data and the output is a reconstructed two-dimensional complex-valued image. Next, an imaging matrix comprised of coil sensitivity maps, an FFT and a randomized undersampling mask (drawn independently from the input k-space measurements) is applied to the generated image to simulate the imaging process. The discriminator takes simulated and observed measurements as inputs and tries to differentiate between them.

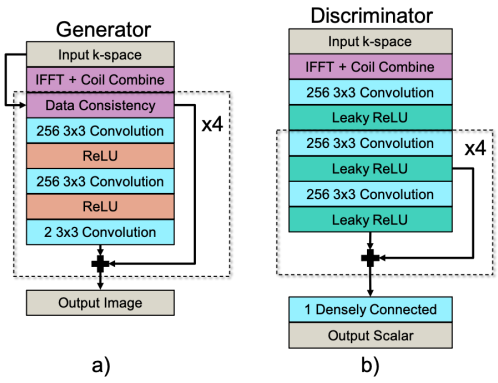


Figure 2. Network architectures. (a) The generator architecture, which is an unrolled network based on the Iterative Shrinkage-Thresholding Algorithm and includes data consistency. The generator is trained in both k-space and image domain. (b) The discriminator architecture, which uses leaky ReLU to backpropagate small negative gradients into the generator. The discriminator is trained only in image domain.

### 2.1. Network Details

The WGAN-GP [16] quality and penalty functions are used in our proposed training objective. We also compare to the results of using the LSGAN [30] objective functions. An unrolled network [10] based on the Iterative Shrinkage-Thresholding Algorithm (ISTA) [3] is used as the generator architecture, shown in Figure 2a. The unrolled network is a common data-driven approach to reconstruction which also incorporates known MR physics [33, 8, 24, 51, 42, 48, 25, 9].

To solve Eq. 1 for an estimate of $\hat{x}$, the problem can be formulated as a standard algorithm for inverse problems as:

$$\hat{x} = argmin_x \left\| Ax - y \right\|_2^2 + R\left(x\right) \tag{5}$$

where $x$ is the reconstructed set of images, $A$ is the imaging model, $y$ is the measured data in k-space, and $R$ is some regularization function. To solve Eq. 5, the unrolled network repeats two different blocks: a data consistency block and a de-noising block. The update block of the unrolled network is used to enforce consistency with the measured data samples. This block, also known as the data consistency block, makes sure that the final reconstructed image is consistent with the measured data to minimize the chance of hallucination. The gradient for the least-squares component in is computed for the $m-$th image estimate $x^m$:

$$\nabla^m = A^H A x^m - A^H y, \tag{6}$$

where the adjoint of $A$ is denoted as $A^H$. The gradient $\nabla^m$ from Eq. 6 is used to update the current image estimate as:

$$x^{m+1} = x^m + t\nabla^m. \tag{7}$$

The step size is denoted by $t$. We initialize the step size $t$ to -2 and learn a different step size, $t^m$, for each iteration of the unrolled model. The rest of the network uses convolutional layers to de-noise the image estimate. These two blocks are then repeated as iterations.

The discriminator architecture, based on a simple convolutional neural network with residual structure, is shown in Figure 2b. Residual connections help to enforce the original structure of the input image, and accelerate training convergence. Leaky ReLU is used as the final activation function instead of ReLU (which is more commonly used in CNNs) in order for the discriminator to be able to backpropagate small negative gradients into the generator.

The generator and discriminator of both GANs were each trained with 2D convolutional layers and 256 feature maps. The generator had 4 residual blocks and 5 iterations of the unrolled network. Parameters were not shared across iterations of the generator. The number of feature maps, unrolled iterations, and residual blocks were chosen for each model to maximize the computational capacity of the network, and thus maximize the reconstruction quality. Many authors have shown that as the size of a network increases, so does its performance [34]. Cheng *et al.* specifically showed that for unrolled network architectures, as the number of iterations increased, the reconstructions' PSNR and SSIM increased, while the NRMSE decreased [8]. Therefore, we chose the biggest model that would fit on our GPUs. We used the performance of our unsupervised GAN on the knee dataset to obtain an estimate of how long the GAN needed to train before convergence was reached.

All networks were trained with a batch size of one and optimized with the Adam optimizer [21]. In all networks, the generator and discriminator were trained for one iteration per training step. Networks were trained using an NVIDIA Titan Xp graphics card. The proposed methods were implemented in Python using Tensorflow. In the spirit of reproducible research, we will provide a software package in Tensorflow to reproduce the results described in this article.

### 2.2. Dataset Details

Two datasets were obtained with Institutional Review Board (IRB) approval and subject informed consent. The first dataset was a set of fully sampled 3T knee images acquired using 8 channel coil arrays and a 3D FSE CUBE sequence with proton density weighting including fat saturation [20]. Fifteen subjects were used for training and 3 subjects were used for testing. Each subject had a complex-valued volume of size 320x320x256 that was split across the readout (X) dimension into slices of size 320x256. Because a fully-sampled ground truth exists for this scenario, we can quantitatively validate our results using standard image quality metrics. We created undersampled images

by applying pseudo-random Poisson-disc variable-density sampling masks to the fully-sampled k-space. Acceleration factors of the sampling masks ranged from 2 to 14. Although we initially use fully-sampled acquisitions to create sub-sampled acquisitions, it is critical to note that the generator and discriminator are never trained with fully-sampled acquisitions.

The second dataset consisted of DCE acquisitions of the abdomen, with a fat-suppressed free-breathing SPGR acquisition [54]. 886 subjects were used for training and 50 subjects were used for testing. It was impossible to obtain fully-sampled data for DCE because the dynamics of the intravenously injected contrast were faster than could be captured at full sampling by the imaging hardware. Each scan acquired a complex-valued volume of size 192x180x80x18, which was split into images of size 180x80. The raw data was compressed from 32 channels to 6 virtual channels using a singular-value-decomposition-based compression algorithm [55]. Data were fully sampled in the $k_x$ direction and subsampled in the $k_y$ and $k_z$ directions. Images were prospectively subsampled with an average acceleration factor of 5.15.

## 2.3. Experiments

### 2.3.1 Retrospectively Undersampled Knee Experiments

First, we trained two GANs, one semi-supervised, and one using our unsupervised method, on the set of knee scans. We did this to quantitatively evaluate the reconstruction performance gap between a traditional semi-supervised GAN and our proposed unsupervised method. We also compared our proposed unsupervised method to CS with $L_1$-wavelet regularization, another reconstruction method which requires no fully-sampled data, and which is routinely used in our clinical practice. For each knee scan, we used a fully-sampled calibration region of $20\times20$ in the center of k-space. To compute image quality, we evaluated average normalized root-mean-square error (NRMSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM) [53] between the reconstructed image and the fully-sampled ground truth on test datasets.

In order to evaluate the two different methods which do not use ground truth, our proposed unsupervised GAN and CS, across different acceleration regimes, we evaluated the reconstruction performance on the set of knee scans of the unsupervised GAN as a function of the acceleration factor of the datasets. We compared this to the reconstruction performance of CS. In both experiments, separate training, validation, and test set were used. The hyperparameters of the training and network architectures were tuned based on the results of the validation set. The hyperparameters of the CS reconstruction were determined based on the L1 difference between the output and the ground truth of the knee validation set, as well as on visual quality for both datasets. The Berkeley Advanced Reconstruction Toolbox (BART) [47] was used to estimate sensitivity maps, generate undersampling masks, and perform a CS reconstruction of these datasets for comparison purposes. Coil sensitivity maps was generated using ESPIRiT [49].

### 2.3.2 Prospectively Undersampled DCE Experiments

Second, we trained our unsupervised GAN on the set of abdominal DCE scans. Because DCE must be undersampled for adequate temporal resolution, we have no ground truth to quantitatively assess reconstruction performance. Instead, we compare to a CS [29] reconstruction that is used in our routine clinical practice, and qualitatively evaluate the sharpness of the vessels and other anatomical structures in the generated images. Both the networks and CS reconstructed the DCE data one phase at a time. We calculated the inference times per slice of CS and our proposed unsupervised method to understand the difference in reconstruction times and clinical implications.

# 3. Results

## 3.1. Retrospectively Undersampled Knee Results

Figure 3 displays a comparison of PSNR, SSIM, and NRMSE between reconstructions from CS with $L_1$-wavelet regularization, our proposed unsupervised GAN, and a standard semi-supervised GAN on our test dataset. The proposed unsupervised GAN had superior PSNR, NRMSE,

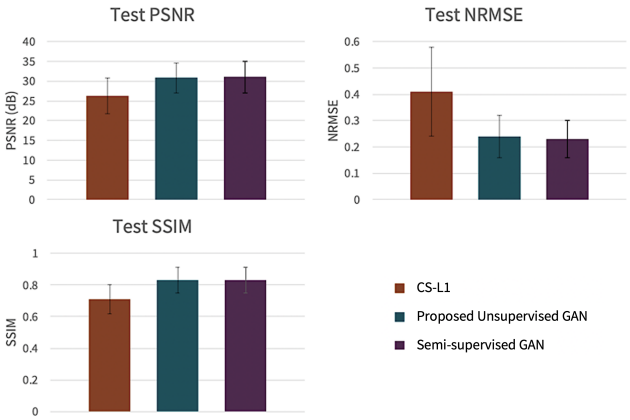


Figure 3. Image metrics calculated on test datasets with reconstructions from CS with L1-wavelet regularization, the proposed unsupervised GAN, and the semi-supervised GAN. The error bars on each series represent the standard deviation of each image metric. The error bars of CS are much larger than the error bars of the unsupervised and semi-supervised GANs. This suggests that the unsupervised generative method is more consistent over each test example than CS.

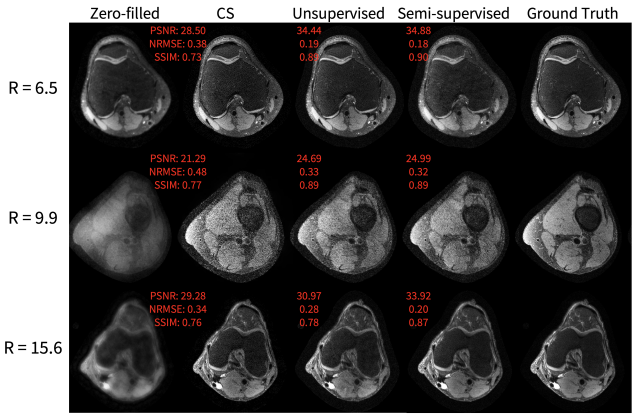


Figure 4. Knee application representative results, showing reconstructions, from left to right: zero-filled, compressed-sensing, our proposed unsupervised, semi-supervised, and fully-sampled ground truth. The acceleration factors of the input image are 6.5, 9.9, and 15.6, from top to bottom. The quantitative metrics that are plotted next to the images are for the slice that is shown. In all rows, the unsupervised GAN has superior PSNR, NRMSE, and SSIM compared to CS. In the first row, the unsupervised GAN has metrics that are notably worse than the semi-supervised GAN. In the middle row and last rows, the unsupervised GAN has metrics that come close to the performance of the semi-supervised GAN.

and SSIM compared to the CS reconstruction. Additionally, the proposed unsupervised GAN only had 0.78% worse PSNR, 4.17% worse NRMSE, and equal SSIM compared to the semi-supervised GAN. The error bars on each series represent the standard deviation of each image metric. The error bars of CS are much larger than the error bars of the unsupervised and semi-supervised GANs. This suggests that the unsupervised generative method is more consistent over each test example than CS.

Representative results in the knee scenario are shown in Figure 4. The columns represent, from left to right, reconstructions from zero-filled, compressed-sensing, our proposed unsupervised, semi-supervised, and fully-sampled ground truth. The acceleration factors of the input images are 6.5, 9.9, and 15.6, from top to bottom. In all rows, the unsupervised GAN has superior PSNR, NRMSE, and SSIM compared to CS. In the first row, the unsupervised GAN has metrics that are worse than the semi-supervised GAN. In the middle and last rows, the unsupervised GAN has metrics that come relatively close to the performance of the semi-supervised GAN. In the unsupervised GAN, the generator markedly improves the image quality by recovering vessels and structures that were not visible before but uses no ground truth data in the training.

The results of the reconstruction performance on the set of knee scans of the unsupervised GAN as a function of the acceleration factor of the training datasets is shown in Figure 5. This figure shows three different image metrics:

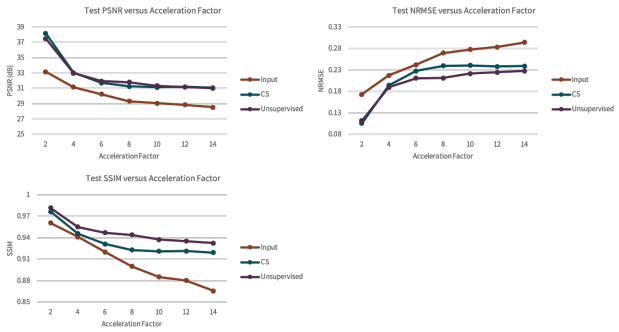


Figure 5. The results of the reconstruction performance on the set of knee scans of the unsupervised GAN as a function of the acceleration factor of the training datasets. The y-axis represents PSNR, NRMSE, or SSIM, depending on the plot. The x-axis represents the acceleration factor of the datasets.

PSNR, NRMSE, and SSIM. The gap between PSNR of CS and the unsupervised model is negligible over all of the acceleration factors. The gap between NRMSE of CS and the unsupervised model is negligible at first for low accelerations but becomes significant at an acceleration of 6 and beyond. The gap between SSIM of CS and the unsupervised model is negligible at first for an acceleration of 2 but becomes significant at an acceleration of 4 and beyond. This behavior is consistent with what we would expect because performance of CS typically drops off at an acceleration of around 4-6 [41]

### 3.2. Prospectively Undersampled DCE Results

Representative DCE results are shown in Figure 6. From left to right, Figure 6 shows reconstructions from input zero-filled, WGAN-GP (our proposed method), CS with L1-wavelet regularization, and LSGAN. Our proposed method, trained under WGAN-GP, greatly improves the input image quality by recovering sharpness and adding more structure. Additionally, the proposed method produces a sharper reconstruction in ROIs such as kidneys and liver vessels compared to CS and LSGAN. For example, in the first row, the kidneys of our WGAN-GP are sharper than that of the input, CS, and LSGAN. In the middle row, liver vessels are more visible in WGAN-GP. In the final row, details of the pelvis are sharpest in WGAN-GP.

We compared the average inference time per two-dimensional DCE slice between CS with L1-wavelet regularization (CS-L1) and our unsupervised GAN. On average, CS-L1 had an average inference time of 1.39 seconds per slice, with a standard deviation of 0.23 seconds. On average, our proposed unsupervised method had an average inference time of 0.23 seconds per slice, with a standard deviation of 0.23 seconds. The inference time of our unsupervised method is approximately 7 times faster than CS-L1.

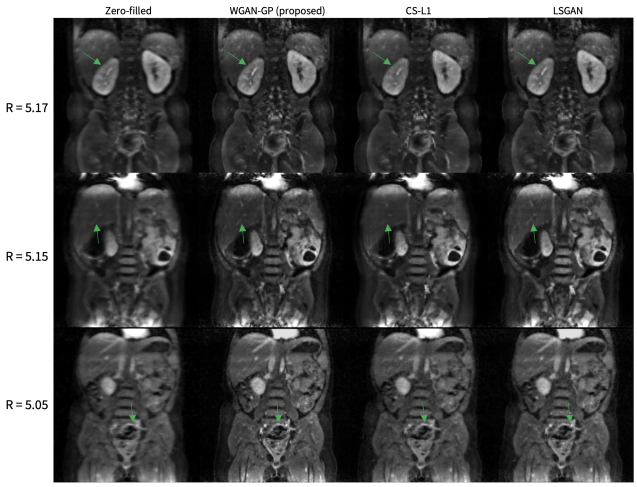


Figure 6. 2D DCE application representative results, under two different adversarial objectives, namely WGAN-GP and LSGAN, as well as zero-filled and CS with L1-wavelet regularization. Our proposed method, trained under WGAN-GP, greatly improves the input image quality by recovering sharpness and adding more structure, as shown by the green arrows. Additionally, the proposed method produces a sharper reconstruction in ROIs such as kidneys and liver vessels compared to CS and LSGAN. For example, in the first row, the kidneys of the unsupervised GAN are visibly much sharper than that of the input and CS. In the middle row, liver vessels are more visible in WGAN-GP. In the final row, details of the pelvis are sharpest in WGAN-GP.

# 4. Discussion

## 4.1. Summary

Our unsupervised generative method allows training high-quality image reconstruction DL models without ever using any fully-sampled data. This method differs from the current paradigm of supervised and hybrid unsupervised-supervised MRI reconstruction by removing the use of any fully-sampled data during training. We demonstrate our method on DCE, a clinically important application where fully-sampled data is impossible to acquire.

In the knee application, reconstructions from our proposed unsupervised method achieve superior SSIM, PSNR, and NRMSE compared to a CS reconstruction. The semi-supervised GAN was slightly superior in comparison to our proposed unsupervised GAN, with better PSNR, better NRMSE, and equal SSIM. This is expected because the semi-supervised GAN has access to fully-sampled data, which gives the network a stronger prior.

As the acceleration factor of the training knee datasets is increased, the reconstruction performance of both our proposed unsupervised method and CS, measured by PSNR, NRMSE, and SSIM, decreases. This trend is as one would expect because as the acceleration factor of the training dataset increases, the GAN has a smaller range of sampled k-space to learn from.

In the DCE application, generated images of our proposed unsupervised method are sharper than those reconstructed by CS. This could be because our method learned a better model of the underlying data distribution of fully-sampled k-space. Data consistency of the unrolled [10] generative network could also play a role in greater sharpness. Additionally, inference time of our method is 7 times faster than CS.

In our experimental setup, the respective sizes of the two datasets are quite different (18 versus 886 subjects). We could expect that the performance of the proposed method may be improved by using a larger dataset for the knee experiments. However, it was easier to tune the hyperparameters of the network for the knee dataset because we have the fully-sampled data to compute model performance metrics on. The DCE reconstruction is a harder problem because of the greater variance in anatomy and contrast of this dataset compared to the knee dataset. Therefore, it potentially requires training with more samples. Additionally, we cannot compute any quantitative metrics on the DCE dataset due to lack of fully-sampled data. Therefore, it is more difficult to optimally tune the hyperparameters of the network without ground truth data to quantitively compare performance results.

Through our results, we have demonstrated that when fully-sampled data is available, such as in the retrospectively undersampled knee application, semi-supervised training should still be used for best reconstruction quality. However, in the situations where fully-sampled data is not available for training a reconstruction model, our unsupervised method can still produce reconstructions which are comparable to a semi-supervised counterpart and better, as well as faster, than CS. Although in this work, we demonstrated an application to DCE, a common acquisition type across a wide range of oncologic imaging indications, these concepts can potentially be leveraged for many other acquisitions where fully-sampled data is difficult or impossible to acquire, such as in cardiac imaging, including volumetric cine and 4D flow, as well as in neurologic imaging, such as DTI and fMRI.

## 4.2. Advantages

The main advantage of this method over existing DL reconstruction methods is the elimination of fully-sampled data. Another benefit is that other additional datasets from the same anatomy or sequence are not needed to use as ground-truth, as in some other works on unpaired training [44, 24]. Additionally, our proposed method produces better quality reconstruction compared to baseline CS methods.

Not only does our method produce higher quality recon-

structions compared to CS, but our model's inference time also is much faster (approximately 7 times faster) than the inference time of CS. Faster reconstruction time is a huge advantage of our method compared to traditional CS-based methods. This has implications in higher patient throughput, faster results for radiologists, and enabling real-time imaging.

This method is extremely generalizable, and could be easily extended to other GAN loss objectives such as WGAN [2] or DCGAN [38] and other network architectures such as variational networks [17], U-Nets [39], or hybrid-domain networks [11]. This technique can be applied to many different dimensionalities and applications, and could thus be demonstrated for 2D slices, 2D slices plus a time dimension, 3D volumes, and 4D datasets. Additionally, this method could also be useful for high noise environments where the acquisition of high SNR data is difficult. Other adverse situations where obtaining ground truth data is precluded include real-time imaging (due to motion) and arterial spin labeling (due to low SNR). Further applications where it is hard to fully sample include time-resolved MR angiography, cardiac cine, low contrast agent imaging, EPI-based sequences, diffusion tensor imaging, and fMRI. Outside of MRI, this method can potentially have applications in other areas where obtaining fully-sampled data is difficult or impossible, such as dynamic PET [13] or computed tomography (CT) [12].

### 4.3. Limitations

One limitation of this method is that because our framework uses a GAN, the training of the generator and discriminator must be balanced, so that they don't become unstable. This can potentially be done by tuning the number of iterations the discriminator and generator are trained per training step to balance both networks. Also, the model sizes and parameters could be optimized for each dataset, although this was not the main focus of this work.

Because no fully-sampled data existed for our DCE dataset, it was difficult to quantitatively validate the DCE experimental results. In the future, we plan to validate these results against clinically relevant measures. For renal DCE, this could include glomerular filtration rate (GFR), which is measured by blood draws.

In the future, image quality could potentially be improved by adding some kind of perceptual loss to the loss function of the generator, such as a total variation loss [6, 40] of the generated image, a feature reconstruction loss between the generated and real images [19], or a style reconstruction loss between the generated and real images [19].

## 5. Conclusion

In this work, we propose an unsupervised GAN framework for image reconstruction without ever using ground truth. We show that our proposed method outperforms existing traditional methods such as CS, while being 7 times faster. Our proposed method has NRMSE, PSNR, and SSIM values which come close to the performance of a semi-supervised GAN. In contrast to most deep learning reconstruction techniques, which are supervised, this method does not need any fully-sampled data. With our proposed method, accelerated imaging and accurate reconstruction can be performed in MRI applications where fully-sampled datasets are difficult or impossible to obtain, such as dynamic imaging.

## References


[1] Hemant K. Aggarwal, Merry P. Mani, and Mathews Jacob. MoDL: Model-Based Deep Learning Architecture for Inverse Problems. *IEEE Transactions on Medical Imaging*, 2019. 1

[2] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *34th International Conference on Machine Learning, ICML 2017*, 2017. 2, 8

[3] Amir Beck and Marc Teboulle. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 3 2009. 4

[4] Ashish Bora, Eric Price, and Alexandros G. Dimakis. Ambientgan: Generative models from lossy measurements. In *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 2018. 2

[5] Robert Bridson. Fast poisson disk sampling in arbitrary dimensions. In *ACM SIGGRAPH 2007 Sketches, SIGGRAPH'07*, 2007. 2

[6] Feiyu Chen, Joseph Y Cheng, John M Pauly, and Shreyas S Vasanawala. Semi-Supervised Learning for Reconstructing Under-Sampled MR Scans. In *ISMRM*, 2019. 2, 8

[7] Feiyu Chen, Valentina Taviani, Itzik Malkiel, Joseph Y. Cheng, Jonathan I. Tamir, Jamil Shaikh, Stephanie T. Chang, Christopher J. Hardy, John M. Pauly, and Shreyas S. Vasanawala. Variable-Density Single-Shot Fast Spin-Echo MRI with Deep Learning Reconstruction by Using Variational Networks. *Radiology*, 289(2):366–373, 11 2018. 1

[8] Joseph Y. Cheng, Feiyu Chen, Marcus T. Alley, John M. Pauly, and Shreyas S. Vasanawala. Highly Scalable Image Reconstruction using Deep Neural Networks with Bandpass Filtering. *arXiv*, 5 2018. 1, 4

[9] Elizabeth Cole, Joseph Cheng, John Pauly, and Shreyas Vasanawala. Analysis of deep complex-valued convolutional neural networks for MRI reconstruction and phase-focused applications. *Magnetic Resonance in Medicine*, 86(2):1093–1109, 8 2021. 1, 3, 4

[10] Steven Diamond, Vincent Sitzmann, Felix Heide, and Gordon Wetzstein. Unrolled optimization with deep priors, 5 2017. 1, 4, 7

[11] Taejoon Eo, Yohan Jun, Taeseong Kim, Jinseong Jang, Ho Joon Lee, and Dosik Hwang. KIKI-net: Cross-domain Convolutional Neural Networks for Reconstructing Undersampled Magnetic Resonance Images. *Magnetic Resonance in Medicine*, 2018. 1, 8

[12] Isabel O. Gallegos, Srivathsan Koundinyan, April Suknot, Edward S. Jimenez, Kyle R. Thompson, and Ryan N. Goodner. Unsupervised learning methods to perform material identification tasks on spectral computed tomography data. In Gary P. Grim, H. Bradford Barber, Lars R. Furenlid, and Jeffrey A. Koch, editors, *Radiation Detectors in Medicine, Industry, and National Security XIX*, page 15. SPIE, 9 2018. 8

[13] Kuang Gong, Ciprian Catana, Jinyi Qi, and Quanzheng Li. Direct patlak reconstruction from dynamic PET using unsupervised deep learning. In Samuel Matej and Scott D. Metzler, editors, *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, page 100. SPIE, 5 2019. 8

[14] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, 2014. 2

[15] Mark A. Griswold, Peter M. Jakob, Robin M. Heidemann, Mathias Nittka, Vladimir Jellus, Jianmin Wang, Berthold Kiefer, and Axel Haase. Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, 47(6):1202–1210, 2002. 1

[16] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved Training of Wasserstein GANs. *Advances in Neural Information Processing Systems*, 3 2017. 2, 3, 4

[17] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P. Recht, Daniel K. Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, 6 2018. 1, 8

[18] Peizhou Huang, Chaoyi Zhang, Hongyu Li, Kumar Gaire, Ruiying Liu, Xiaoliang Zhang, Xiaojuan Li, and Leslie Ying. Deep MRI Reconstruction without Ground Truth for Training. *arXiv*. 2

[19] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. *Eccv*, 3 2016. 8

[20] M. Lustig M. T. Alley M. Uecker P. Virtue P. Lai K. Epperson, A. M. Sawyer, , and S. S. Vasanawala. Creation of Fully Sampled MR Data Repository for Compressed Sensing of the Knee. In *SMRT 22nd Annual Meeting*, Salt Lake City, Utah, USA, 2013. 4

[21] Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015. 4

[22] LeCun Yann, Cortes Corinna, and Burges Christopher. THE MNIST DATABASE of handwritten digits. *The Courant Institute of Mathematical Sciences*, 1998. 2

[23] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2Noise: Learning image restoration without clean data. *35th International Conference on Machine Learning, ICML 2018*, 7:4620–4631, 2018. 2

[24] Ke Lei, Morteza Mardani, John M. Pauly, and Shreyas S. Vasawanala. Wasserstein GANs for MR Imaging: from Paired to Unpaired Training, 10 2019. 4, 7

[25] Dong Liang, Jing Cheng, Ziwen Ke, and Leslie Ying. Deep Magnetic Resonance Image Reconstruction: Inverse Problems Meet Neural Networks. *IEEE Signal Processing Magazine*, 2020. 4

[26] Jiaming Liu, Cihat Eldeniz, Yu Sun, Weijie Gan, Sihao Chen, Hongyu An, and Ulugbek S. Kamilov1. RED-N2N: Image reconstruction for MRI using deep CNN priors trained without ground truth. In *ISMRM*, 2020. 2

[27] Jiaming Liu, Yu Sun, Cihat Eldeniz, Weijie Gan, Hongyu An, and Ulugbek S. Kamilov. RARE: Image Reconstruction Using Deep Priors Learned Without Groundtruth. *IEEE Journal of Selected Topics in Signal Processing*, 14(6):1088–1099, 10 2020. 2

[28] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Large-scale Celebfaces Attributes (celeba) Dataset. 2018. 2

[29] Michael Lustig, David Donoho, and John M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 12 2007. 1, 5

[30] Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, Zhen Wang, and Stephen Paul Smolley. Least Squares Generative Adversarial Networks. 11 2016. 2, 4

[31] Morteza Mardani, Enhao Gong, Joseph Y. Cheng, Shreyas Vasanawala, Greg Zaharchuk, Marcus Alley, Neil Thakur, Song Han, William Dally, John M. Pauly, and Lei Xing. Deep Generative Adversarial Networks for Compressed Sensing Automates MRI, 5 2017. 2

[32] Morteza Mardani, Enhao Gong, Joseph Y. Cheng, Shreyas S. Vasanawala, Greg Zaharchuk, Lei Xing, and John M. Pauly. Deep generative adversarial neural networks for compressive sensing MRI. *IEEE Transactions on Medical Imaging*, 38(1):167–179, 1 2019. 1, 2

[33] Morteza Mardani, Qingyun Sun, Shreyas Vasawanala, Vardan Papyan, Hatef Monajemi, John Pauly, and David Donoho. Neural Proximal Gradient Descent for Compressive Imaging. *Advances in Neural Information Processing Systems*, 6 2018. 4

[34] Preetum Nakkiran, Gal Kaplun, Yamini Bansal, Tristan Yang, Boaz Barak, and Ilya Sutskever. Deep Double Descent: Where Bigger Models and More Data Hurt. *arXiv*, 12 2019. 4

[35] Gyutaek Oh, Byeongsu Sim, Hyung Jin Chung, Leonard Sunwoo, and Jong Chul Ye. Unpaired Deep Learning for Accelerated MRI Using Optimal Transport Driven CycleGAN. *IEEE Transactions on Computational Imaging*, 2020. 2

[36] Arthur Pajot, Emmanuel De Bézenac, and Patrick Gallinari. Unsupervised adversarial image reconstruction. In *7th International Conference on Learning Representations, ICLR 2019*, 2019. 2

[37] Klaas P. Pruessmann, Markus Weiger, Markus B. Scheidegger, and Peter Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 1999. 1

[38] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. In *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, 2016. 2, 8

[39] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015. 8

[40] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 11 1992. 8

[41] Christopher M. Sandino, Joseph Y. Cheng, Feiyu Chen, Morteza Mardani, John M. Pauly, and Shreyas S. Vasanawala. Compressed Sensing: From Research to Clinical Practice With Deep Neural Networks: Shortening Scan Times for Magnetic Resonance Imaging. *IEEE Signal Processing Magazine*, 37(1):117–127, 1 2020. 6

[42] Christopher M Sandino, Peng Lai, Shreyas S Vasanawala, and Joseph Y Cheng. Accelerating Cardiac Cine MRI Using a Deep Learning-Based ESPIRiT Reconstruction. *arXiv:1911.05845 [cs, eess]*, 5 2020. 4

[43] Priyanka Shende, Mahesh Pawar, and Sandeep Kakde. A brief review on: MRI images reconstruction using GAN. In *Proceedings of the 2019 IEEE International Conference on Communication and Signal Processing, ICCSP 2019*, 2019. 2

[44] Byeongsu Sim, Gyutaek Oh, and Jong Chul Ye. Optimal Transport Structure of CycleGAN for Unsupervised Learning for Inverse Problems. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8644–8647. IEEE, 5 2020. 2, 7

[45] Shakarim Soltanayev and Young Chun. Training deep learning based denoisers without ground truth data. In *NIPS*, 2018. 2

[46] Roberto Souza, R Marc Lebel, Richard Frayne, and Rfrayne@ucalgary Ca. A Hybrid, Dual Domain, Cascade of Convolutional Neural Networks for Magnetic Resonance Image Reconstruction. In *Proceedings of Machine Learning Research 102:437–446*, 2019. 1

[47] Jonathan I. Tamir, Frank Ong, Joseph Y. Cheng, Martin Uecker, and Michael Lustig. Generalized Magnetic Resonance Image Reconstruction using The Berkeley Advanced Reconstruction Toolbox. *Proceedings of the ISMRM 2016 Data Sampling and Image Reconstruction Workshop*, 2016. 5

[48] Jonathan I Tamir, Stella X Yu, and Michael Lustig. Unsupervised Deep Basis Pursuit: Learning inverse problems without ground-truth data. In *International Society for Magnetic Resonance in Medicine*, Montreal, Quebec, Canada, 2019. 2, 4

[49] Martin Uecker, Peng Lai, Mark J. Murphy, Patrick Virtue, Michael Elad, John M. Pauly, Shreyas S. Vasanawala, and Michael Lustig. ESPIRiT - An eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. *Magnetic Resonance in Medicine*, 71(3):990–1001, 3 2014. 5

[50] Zhirong Wu, Yuanjun Xiong, Stella X. Yu, and Dahua Lin. Unsupervised Feature Learning via Non-parametric Instance Discrimination. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3733–3742. IEEE, 6 2018. 2, 3

[51] Burhaneddin Yaman, Seyed Amir Hossein Hosseini, Steen Moeller, Jutta Ellermann, Kamil Ugurbil, and Mehmet Akcakaya. Self-Supervised Physics-Based Deep Learning MRI Reconstruction Without Fully-Sampled Data. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 921–925. IEEE, 4 2020. 2, 4

[52] Guang Yang, Simiao Yu, Hao Dong, Greg Slabaugh, Pier Luigi Dragotti, Xujiong Ye, Fangde Liu, Simon Arridge, Jennifer Keegan, Yike Guo, and David Firmin. DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1310–1321, 6 2018. 1, 2

[53] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 5

[54] Tao Zhang, Joseph Y. Cheng, Aaron G. Potnick, Richard A. Barth, Marcus T. Alley, Martin Uecker, Michael Lustig, John M. Pauly, and Shreyas S. Vasanawala. Fast pediatric 3D free-breathing abdominal dynamic contrast enhanced MRI with high spatiotemporal resolution. *Journal of Magnetic Resonance Imaging*, 41(2):460–473, 2 2015. 5

[55] Tao Zhang, John M. Pauly, Shreyas S. Vasanawala, and Michael Lustig. Coil compression for accelerated imaging with Cartesian sampling. *Magnetic Resonance in Medicine*, 2013. 5

[56] Jun-yan Zhu, Philipp Krähenbühl, Eli Shechtman, and Alexei Efros. Generative Visual Manipulation on the Natural. *European Conference on Computer Vision*, 2016. 2

[57] Magauiya Zhussip, Shakarim Soltanayev, and Se Young Chun. Training Deep Learning Based Image Denoisers From Undersampled Measurements Without Ground Truth and Without Image Prior. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10247–10256. IEEE, 6 2019. 2