Time-Dependent Deep Image Prior for Dynamic MRI

Jaejun Yoo, Kyong Hwan Jin, Harshit Gupta, Jérôme Yerly, Matthias Stuber, and Michael Unser, *Fellow, IEEE*

Abstract— We propose a novel unsupervised deep-learning-based algorithm for dynamic magnetic resonance imaging (MRI) reconstruction. Dynamic MRI requires rapid data acquisition for the study of moving organs such as the heart. We introduce a generalized version of the deep-image-prior approach, which optimizes the weights of a reconstruction network to fit a sequence of sparsely acquired dynamic MRI measurements. Our method needs neither prior training nor additional data. In particular, for cardiac images, it does not require the marking of heartbeats or the reordering of spokes. The key ingredients of our method are threefold: 1) a fixed low-dimensional manifold that encodes the temporal variations of images; 2) a network that maps the manifold into a more expressive latent space; and 3) a convolutional neural network that generates a dynamic series of MRI images from the latent variables and that favors their consistency with the measurements in *k*-space. Our method outperforms the state-of-the-art methods quantitatively and qualitatively in both retrospective and real fetal cardiac datasets. To the best of our knowledge, this is the first unsupervised deep-learning-based method that can reconstruct the continuous variation of dynamic MRI sequences with high spatial resolution.

Index Terms— Accelerated MRI, unsupervised learning.

I. INTRODUCTION

THE aim of dynamic magnetic resonance imaging (MRI) is to capture the dynamics associated with moving organs,

Jaejun Yoo, Harshit Gupta, and Michael Unser are with the Biomedical Imaging Group, EPFL, 1015 Lausanne, Switzerland (e-mail: jaejun.yoo88@gmail.com; harshit.gupta.cor@gmail.com; michael.unser@epfl.ch).

Kyong Hwan Jin was with the Biomedical Imaging Group, EPFL, 1015 Lausanne Switzerland. He is now with the Department of Information and Communication Engineering, Daegu Gyeongbuk Institute of Science and Technology, Daegu 42988, South Korea (e-mail: kyong.jin@dgist.ac.kr).

Jérôme Yerly and Matthias Stuber are with the Department of Radiology, University Hospital (CHUV), 1011 Lausanne, Switzerland and University of Lausanne (UNIL), 1015 Lausanne, Switzerland, and also with the Center for Biomedical Imaging (CIBM), Switzerland (e-mail: yerlyj.mri@gmail.com; matthias.stuber@chuv.ch).

which requires a fast imaging process. A typical approach is to accelerate data acquisition by a partial sampling of the *k*-space. The resulting partial loss of data must then be compensated to maintain the image quality. Several methods have addressed this by exploiting spatial or temporal redundancies, including parallel MRI [1]–[4], k-t acceleration methods [5]–[7], compressed sensing (CS) MRI [8]–[16], low-rank methods [17]–[19], manifold-learning methods [20]–[25], and many others. In the specific case of cardiac applications, the current state-of-the-art methods further improve the reconstruction by exploiting the fact that the heart motion is approximately cyclic. They typically use electrocardiograms or self-gating techniques [26], [27]. However, all of these methods are limited by constraints over the signal-to-noise ratio (SNR), restrictions in the coil design, hand-picked priors, multiple processing steps, or inefficient algorithms in their deployment of the standard convex-optimization techniques.

More recently, inspired by the development of deep-learning techniques in various imaging modalities [28]–[31], supervised-learning approaches have been applied to the fast and accurate reconstruction of partially sampled MRI [32]–[40]. These methods, however, heavily depend on a training dataset, especially on ground-truth data (*i.e.*, fully sampled measurements), which are typically unavailable for dynamic MRI. Unlike the direct deep-learning approaches, the model-based deep-learning framework of [41] formulates the image recovery as an optimization scheme. By unrolling an iterative algorithm, it minimizes a cost function that combines data consistency and a deep-learned prior. Because the learned prior incorporates patient-specific noise patterns into the algorithm, this approach successfully recovers images with fast reconstruction and acceptable quality. However, it still requires ground-truth data to train the denoising network.

A. Contribution

In this paper, we propose an unsupervised learning framework in which a generative network is optimized to reconstruct a sequence of golden-angle radial lines in *k*-space, also called spokes. Inspired by deep image priors (DIP) [42], we use the architecture of a convolutional neural networks (CNN) as an implicit prior to constrain the search space of the optimization problem. To learn the temporal dependencies of the dynamic measurements, we impose a one-dimensional manifold parameterized by time. Using this explicit cue, the network learns to encode the temporal variations of the sequential images into the spatial closeness of the samples on the imposed
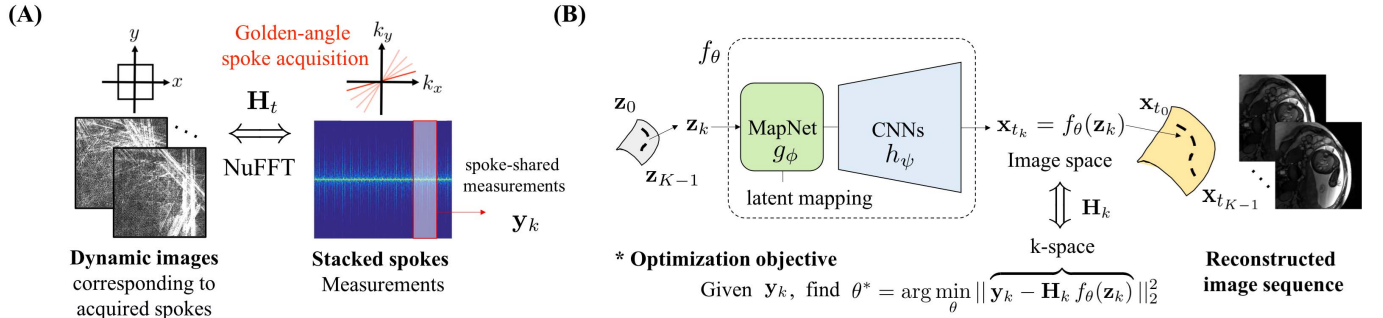
Fig. 1.   Overview of our framework. (A) Schematic illustration of the dynamic MRI data-acquisition procedure. We use a nonuniform fast Fourier transform with a golden-angle scheme and spoke sharing. (B) Proposed framework based on latent mapping and deep image priors. The block labeled g_ϕ is a mapping network (MapNet) of fully connected layers and the block labeled h_ψ denotes the generative CNNs.

manifold. This simple temporal coupling enables our model to outperform the other CS algorithms [15], [16], [26], [27] without bells and whistles—note that our approach is purely unsupervised and optimized in an end-to-end manner. We further improve the reconstruction by introducing a mapping network (MapNet) that brings more flexibility to our latent space [43], [44]. MapNet consists of a few fully connected layers with nonlinear activations that learns to map the fixed manifold into a more expressive latent space. This allows the subsequent generative network to adapt its input to a given dataset, thereby improving image quality (Figure 1).

In short, our generative model takes the latent variables from MapNet and reconstructs dynamic images by exploiting its powerful structural prior. With the extensive analyses in Section IV and experimental results in Section V, we show that the manifold design and MapNet are both essential to achieve good reconstructions.

B. Related Work

1) Unsupervised Learning: Starting from the seminal work of DIP [42], there have been several studies that applied unsupervised learning to medical imaging, such as MRI [45] and positron emission tomography [46], albeit both cases address the reconstruction of static images. The authors of [47] used DIP for video compression. Their work is the closest work to ours but, unlike our goal (the reconstruction of an image sequence), theirs is to find compact codes for the representation of video frames. To find such codes, they optimize both the network weights and latent variables. Without any constraint on the latent space, however, the latent codes may diverge to an arbitrary space. To prevent this, they imposed either low-rank or similarity constraints on the latent sequence. As it turns out, this optimization not only requires additional effort to tune hyper-parameters but also entails a singular-value decomposition at each iteration, which severely increases the computational burden. By contrast, our solution imposes an explicit manifold and lets a mapping network adapt the manifold to data, which makes the training much simpler and easier. In addition, their forward model is an identity operator, while ours is an MR measurement operator with severe undersampling.

2) Manifold Learning: Recently, several manifold-learning methods have been proposed for dynamic MRI reconstruc-

tion [20]–[23]. The main assumption of these approaches is that each image frame can be modeled as a point on some smooth, low-dimensional manifold in a high-dimensional space [48]. Because the knowledge of the manifold is essential for these approaches, most methods require navigator signals that are used to estimate the manifold structure. Specifically, a series of radial spokes with the same orientations are needed periodically, which results in a large overhead. In addition, without dedicated setups, these approaches cannot be readily applied to the golden-angle sequences implemented on several scanners. In this paper, we do not consider such additional acquisitions. Thus, these manifold-learning methods cannot be applied to our case. In [25], the authors proposed a two-step approach for the application of the manifold approach to the navigator-free setup. They first estimated the manifold using the low-frequency part of the k-space measurements, which is used in model optimization. However, when the low-frequency part is not densely sampled (e.g., only five radial spokes per frame), the manifold estimation becomes inaccurate, which severely affects the quality of the final reconstruction. By contrast, our method does not require navigator signals and, in fact, is agnostic to sampling patterns since the latent manifold is captured by the network in a data-adaptive way.

II. METHODS

We first briefly recapitulate the content of deep image prior (Section II-A) as well as the physics of dynamic MRI (Section II-B). Then, we describe our method based on DIP with a mapping network and on the learning of the underlying latent manifolds (Section II-C).

A. Deep Image Prior

The deep image prior (DIP) [42] is a recent approach that has been proposed to solve static linear inverse problems, such as image denoising, inpainting, and superresolution. DIP has been found to capture advanced image statistics in a purely unsupervised way with neither pre-training nor external training data. Taking a random but fixed latent variable $\mathbf{z} \in \mathbb{R}^L$ as input, DIP optimizes the parameters $\boldsymbol{\theta}$ of an untrained neural network $f_{\boldsymbol{\theta}}$ to produce an output $f_{\boldsymbol{\theta}}(\mathbf{z})$ that is consistent with the measurement $\mathbf{y} \in \mathbb{R}^M$. The problem being solved is

formalized as

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{H}(f_{\boldsymbol{\theta}}(\mathbf{z}))\|_2^2, \tag{1}$$

where $\mathbf{H} \in \mathbb{R}^{M \times N}$ is a forward model. For example, in the superresolution problem, $\mathbf{y}$ is a noisy, low-resolution image and $\mathbf{H}$ is a downsampling operator. The output of the optimized network $\mathbf{x}^* = f_{\boldsymbol{\theta}^*}(\mathbf{z})$ then yields a reconstructed image of surprisingly good quality. This has been ascribed to the implicit representation bias of the CNN architecture, which favors a natural-looking output image over a noisy unstructured one. In this paper, we extend the concept of DIP to solve the more challenging problem of dynamic MRI reconstruction.

B. Dynamic MRI

We use a radial 2D MRI acquisition scheme where the instrumentation is such that it physically records a temporal sequence of radial lines of the Fourier transform of a fixed slice (image) of a 3D volumetric object. The underlying 2D image is represented by a vector $\mathbf{x} \in \mathbb{C}^N$, where N is the number of pixels. At a given time point t, the vector of k-space measurements $\mathbf{y}(t) \in \mathbb{C}^{M_0}$ consists of the uniform samples of the 2D Fourier transform of the image taken along a radial line at some orientation $\vartheta = \vartheta(t)$. Because of the central-slice theorem, these measurements can also be interpreted as the 1D Fourier transform of the Radon transform of the image at angle ϑ. By repeating this process with a sufficiently dense sequence of angles $\vartheta_k \in [0, \pi)$, and assuming the images to be static, one obtains a complete dataset from which a high-quality (static) image can be reconstructed using standard tomographic techniques. Now, the difficulty with dynamic imaging is that the underlying image is not static but varies through time, which calls for a more sophisticated reconstruction procedure.

1) Forward Model: The measurement process that relates the image at time t and the k-space measurements with angle $\vartheta = \vartheta(t)$ is linear and formally described by the relation

$$\mathbf{y}(t) = \mathbf{H}(\vartheta)\mathbf{x}(t), \tag{2}$$

where $\mathbf{H}(\vartheta)$ is the $(M_0 \times N)$ system matrix that represents the combined effect of taking the 2D Fourier transform of $\mathbf{x}$ and resampling along a radial line with direction ϑ. The type of measurement provided by (2) is referred to as an angular *spoke*. In practice, we acquire a series of K spokes taken at regularly spaced time point $t_k = t_0 + k\Delta t$, $k = 0, \ldots, (K-1)$ with step size Δt. The spoke orientations follow the golden-angle strategy

$$\vartheta_k = \vartheta_0 + \omega_0 k \Delta t, \tag{3}$$

where ϑ_k gives the orientation of a spoke at time $t_k = t_0 + k\Delta t$, with ω_0 its angular velocity. The golden-angle specificity is the irrationality condition $(\omega_0 \Delta t / \pi) \notin \mathbb{Q}$, which is approximated by setting $(\omega_0 \Delta t) \approx 111.25°$ [15]. Then, our task is to reconstruct the image sequence $\{\mathbf{x}(t_k)\}_{k=0}^{K-1}$ from the measurement sequence $\{\mathbf{y}(t_k)\}_{k=0}^{K-1}$.

2) Spoke-Sharing: The ambitious goal of accelerated dynamic MRI is to reconstruct $\{\mathbf{x}(t_k)\}_{k=0}^{K-1}$—or, even better, $\mathbf{x}(t)$ for $t \in [t_0, T_{K-1}]$—from the finite set of measurements $\{\mathbf{y}(t_k)\}$. However, a single orientation per frame does not provide enough information to recover the corresponding instantaneous two-dimensional image $\mathbf{x}(t_k)$. To overcome this issue, we assume that the changes are slow over some small number of neighboring spokes (n_s), so that $\mathbf{x}(t) \approx \mathbf{x}(t_k)$ for all $t \in \mathbb{T}_k = [t_k - n_s \Delta t/2, t_k + n_s \Delta t/2)$. The sharing parameter $n_s \in 2\mathbb{N}+1$ is the number of radial lines used for the reconstruction of one frame; it controls the temporal resolution.

To further describe this pooling process, we introduce the augmented measurement vector $\mathbf{y}_k = (\mathbf{y}(t_m))_{m=k-(n_s-1)/2}^{k+(n_s-1)/2}$ of size $M = (n_s \times M_0)$. Correspondingly, we define the column-wise concatenated system matrix $\mathbf{H}_k = (\mathbf{H}(\vartheta_k))_{m=k-(n_s-1)/2}^{k+(n_s-1)/2}$, whose time dependence is indicated by the index k. This results in the forward imaging model

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}(t_k), \tag{4}$$

where the matrix $\mathbf{H}_k \in \mathbb{C}^{M \times N}$ encodes the (pseudo-simultaneous) acquisition of n_s spokes at time t_k. The underlying strategy is called spoke sharing. Because of the irrationality condition of the golden-angle approach, no direction will ever be measured twice. While the imaging model (4) is more favorable than (2) because of the augmented number of measurements, the problem is still ill-posed because $M = n_s M_0$ remains smaller than N (the number of unknowns). The common practice, therefore, is to introduce an appropriate regularizer. In this paper, we propose to constrain the solution by applying a deep image prior that is shared among all frames.

C. Proposed Framework

To address the dynamic MRI reconstruction problem, we propose a new framework, time-dependent deep image prior (TD-DIP). We first modify the original DIP so that it takes a sequence of input and output pairs (Figure 1). More specifically, we optimize an untrained neural network $f_{\boldsymbol{\theta}}$ to map a sequence of inputs $\{\mathbf{z}_k\}_{k=0}^{K-1}$ to the spoke-shared measurements $\{\mathbf{y}_k\}_{k=0}^{K-1}$, thereby reconstructing the sequence of images $\{\mathbf{x}(t_k)\}_{k=0}^{K-1}$ by searching for

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \frac{1}{K} \sum_{k=0}^{K-1} \|\mathbf{y}_k - \mathbf{H}_k f_{\boldsymbol{\theta}}(\mathbf{z}_k)\|^2, \tag{5}$$

leading to $\mathbf{x}^*(t_k) = f_{\boldsymbol{\theta}^*}(\mathbf{z}_k)$. Note that the optimization is done in the measurement domain. This enforces the image sequence to be consistent with the measurements, while the modified DIP scheme regularizes the reconstructed images.

1) Manifold Design: To fully exploit the characteristics of dynamic MRI, the underlying model must be able to effectively encode the temporal variations of the measurements while preserving the structure of the individual frames. To this end, we propose to design a manifold $\mathcal{Z}$, thereby effectively injecting a specific prior into the network. For example, an ordered sequence $\{\mathbf{z}_k\}$ from a straight-line manifold will guide the network to associate spatial closeness of input

---

Algorithm 1 TD-DIP for Dynamic MRI. We Use Adam Optimizer [49] With $n_{\text{iter}} = 10{,}000$ and $B = 1$

Input: Set of measurements $\{\mathbf{y}_k\}_{k=0}^{K-1}$, number of iterations n_{iter}, batch size B, and number of cycles p.
1) Select a manifold $\mathcal{Z}$.
2) Sample $\{\mathbf{z}_k\}_{k=0}^{K-1}$ from $\mathcal{Z}$.
3) Optimize $\boldsymbol{\theta}$.
   for n_{iter} iterations **do**
   - Randomly sample a batch $\{k_0, \ldots, k_{B-1}\}$ of size B from $\{0, \ldots, K-1\}$.
   - Compute the batch loss of (6).
   - Update $\boldsymbol{\theta}$ with gradient $\nabla_{\boldsymbol{\theta}} L_B(\boldsymbol{\theta})$.
   end for
4) Reconstruct images $\left\{(h \circ g)_{\boldsymbol{\theta}^*}(\mathbf{z}_k)\right\}_{k=0}^{K-1}$.

---

variables with temporal closeness of images. This encourages the network to reconstruct an image sequence with temporally similar attributes. For a nearly periodic signal such as the cardiac motion, we can encode the expected behavior by letting the manifold take the structure of a three-dimensional helix.

2) Mapping Network (MapNet): Although a careful choice of temporally meaningful manifolds typically results in an excellent performance, the fact that the design is hand-crafted may also sometimes limit the performance of the network [43]. To add flexibility to our model and to exploit the rich representation power of the network, we introduce a mapping network (MapNet) refer to as g_ϕ. In our design, MapNet involves a few fully connected layers with nonlinearities. It learns to map a fixed manifold into the more expressive latent space $\mathcal{W} = g_\phi(\mathcal{Z})$. More specifically, our model $f_{\boldsymbol{\theta}}$ now has a hierarchical architecture that consists of the MapNet g_ϕ followed by the CNN h_ψ so that $f_{\boldsymbol{\theta}} = h_\psi \circ g_\phi$ and $\boldsymbol{\theta} = \{\phi, \psi\}$ (Figure 1 (B)). This leads us to replace (5) by

$$L_K(\boldsymbol{\theta}) = \frac{1}{K} \sum_{k=0}^{K-1} \|\mathbf{y}_k - \mathbf{H}_k(h \circ g)_{\boldsymbol{\theta}}(\mathbf{z}_k)\|^2. \tag{6}$$

The role of g_ϕ is to appropriately warp the input manifold to facilitate h_ψ in its reconstruction of the true dynamics. Overall, the insertion of g_ϕ provides better flexibility to our model and lets us efficiently exploit the representation power of neural networks, resulting in a good reconstruction.

3) Final Algorithm: Our optimization scheme is given in Algorithm 1. We minimize the loss function (6) using standard gradient-descent methods [49] for n_{iter} iterations. At each iteration, instead of (6), a batch loss $L_B(\boldsymbol{\theta})$ is updated where a batch $\{k_0, \ldots, k_{B-1}\}$ of size B is randomly sampled from the index set $\{0, \ldots, K-1\}$. The corresponding input variables $\{\mathbf{z}_{k_b}\}_{b=0}^{B-1}$ are fed to the network and its parameters are updated using the gradient with respect to $\boldsymbol{\theta}$.

III. EXPERIMENTS

A. Datasets

All experimental datasets are breath-hold MR images. We assume a twofold upsampling of measurements for every dataset. Therefore, the size of the reconstructed fields of view is half that of the first dimension of the measurements.

1) Retrospective Dataset: A cardiac cine dataset was acquired using a 3T whole-body MRI scanner (Siemens; Tim Trio) equipped with a 32-element cardiac coil array. The acquisition sequence was bSSFP and prospective cardiac gating was used. The imaging parameters were as follows: FOV = (300×300) mm^2, acquisition matrix size = (128×128), TE/TR = $1.37/2.7$ ms, receiver bandwidth = 1184 Hz/pixel, and flip angle = $40°$. The number of frames was 23 and the temporal resolution was 43.2 ms. The resulting fully sampled Cartesian trajectories are used as ground truth. To retrospectively simulate the radial sampling, we implemented the forward model using the golden-angle strategy. Sinograms are obtained as shown in Figure 1. The number of spokes for reconstructing a frame is $n_s = 13$. For a single-cycle simulation, the dimension of the sinograms is $(K \times n_s \times M_{\boldsymbol{\omega}} \times C) = (23 \times 13 \times 256 \times 32)$. For a multicycle simulation, we acquire $p = 13$ cycles, which results in $K = 13 \cdot 23 = 299$ frames.

2) Fetal Cardiac Dataset: Fetal cardiac MRI data were acquired on a 1.5 T clinical MR scanner (MAGNETOM Aera, Siemens AG, Healthcare Sector, Erlangen, Germany) with an 18-channel body array coil and a 32-channel spine coil for signal reception. We used an untriggered continuous 2D bSSFP sequence that was modified to acquire radial readouts with a golden-angle trajectory [27]. The acquisition parameters were: FOV = (260×260) mm^2, acquisition matrix size = (256×256) pixels, slice thickness = 4.0 mm, TE/TR = $1.99/4.1$ ms, RF excitation angle = $70°$, radial readouts = 1400, acquisition time = 6.7 s, and bandwidth = 1028 Hz/pixel. The number of shared spokes for reconstructing a frame is $n_s = 5$.

B. Baseline Methods

We apply three baseline methods.
1) **NUFFT** is a nonuniform, zero-filled fast Fourier transform.[1]
2) **GRASP** [15] is a golden-angle radial sparse parallel MRI algorithm, which extends the idea of k-t SPARSE-SENSE [12] to volumetric golden-angle radial acquisitions. Here, the spoke-sharing strategy is not applied.
3) **Reordering Method (RD)** [26], [27] is a three-step algorithm. RD first reconstructs real-time images of limited image quality and uses these images to reorder or self-gate the measurements, which in turn are used for the final reconstruction with k-t SPARSE-SENSE [12]. In the retrospective experiment, where we know the phase indices, we use the exact order of frames for self-gating.

C. Estimation of Cardiac Cycles

For the processing of the fetal cardiac dataset, RD and our algorithm both require a rough estimate of the number of cardiac cycles seen over the whole duration of a sequence of data acquisition. It can be typically obtained from k-space. Simple techniques to estimate the cardiac cycles from

---

[1] https://github.com/marchdf/python-nufft

TABLE I
ARCHITECTURE OF OUR GENERATIVE CONVOLUTIONAL NETWORK (h_ψ). CONV.: CONVOLUTION;
BN: BATCH NORMALIZATION; NN INTERP.: NEAREST-NEIGHBOR INTERPOLATION

Operation Layer	Number of Filters	Size of Each Filter (XYC)	Strides (XY)	Zero Padding (XY)	Size of Output Image (XYC)
Input of h_ψ ($L = 64$)					$8 \times 8 \times 1$
Conv+BN+ReLU	128	$3 \times 3 \times 1$	1×1	1×1	$8 \times 8 \times 128$
Conv+BN+ReLU	128	$3 \times 3 \times 128$	1×1	1×1	$8 \times 8 \times 128$
NN interp.			2×2		$16 \times 16 \times 128$
$2\times$(Conv+BN+ReLU)	128	$3 \times 3 \times 128$	1×1	1×1	$16 \times 16 \times 128$
NN interp.			2×2		$32 \times 32 \times 128$
$2\times$(Conv+BN+ReLU)	128	$3 \times 3 \times 128$	1×1	1×1	$32 \times 32 \times 128$
NN interp.			2×2		$64 \times 64 \times 128$
$2\times$(Conv+BN+ReLU)	128	$3 \times 3 \times 128$	1×1	1×1	$64 \times 64 \times 128$
NN interp.			2×2		$128 \times 128 \times 128$
$2\times$(Conv+BN+ReLU)	128	$3 \times 3 \times 128$	1×1	1×1	$128 \times 128 \times 128$
Conv.	2	$3 \times 3 \times 128$	1×1	1×1	$128 \times 128 \times 2$

radial data have been previously reported by [26], [50], [51]. Radial acquisition schemes sample the center of k-space at every readout, which supports the extraction of physiological motion signals. The central k-space coefficient of a radial readout (*i.e.*, the echo peak) corresponds to the complex sum of the transverse magnetization across the entire image volume. In the presence of moving structures such as a beating heart, changes in the overall transverse magnetization due to motion will induce a modulation of the consecutive echo peaks. (Trajectory imperfections and eddy currents can also modulate echo peaks, but their frequency responses differ from the physiological motion frequencies and, thus, can be filtered out.) The resulting signal can then be used to estimate the number of cardiac cycles and to inform the manifold network. For our fetal cardiac dataset, we find that the time-course has approximately 13 periods so that we finally set $p = 13$.

D. Evaluation Metric

We use the regressed SNR and structural similarity index (SSIM) [52] as quantitative metrics, where higher scores correspond to better reconstructions. With the oracle $\mathbf{x}$ and the reconstructed image $\mathbf{x}^*$, RSNR and SSIM are given by

$$\text{RSNR} = \max_{a,b\in\mathbb{R}} 20 \log \frac{\|\mathbf{x}\|_2}{\|\mathbf{x} - a\,\mathbf{x}^* + b\|_2},$$
$$\text{SSIM} = \frac{(2\mu_{\mathbf{x}^*}\mu_{\mathbf{x}} + c_1)(2\sigma_{\mathbf{x}^*\mathbf{x}} + c_2)}{(\mu_{\mathbf{x}^*}^2 + \mu_{\mathbf{x}}^2 + c_1)(\sigma_{\mathbf{x}^*}^2 + \sigma_{\mathbf{x}}^2 + c_2)}. \quad (7)$$

Here, $\mu_{\mathbf{x}^*}$ and $\mu_{\mathbf{x}}$ are the expectations, $\sigma_{\mathbf{x}^*}^2$ and $\sigma_{\mathbf{x}}^2$ are the variances, and $\sigma_{\mathbf{x}^*\mathbf{x}}^2$ is the covariance of $\mathbf{x}^*$ and $\mathbf{x}$, respectively. In addition, c_1 and c_2 are stabilization parameters chosen as $c_1 = (0.01\xi)^2$ and $c_1 = (0.03\xi)^2$ with ξ being the dynamic range of the pixel intensity.

E. Implementation Details

We use an Intel i7-7820X (3.60GHz) CPU and an NVIDIA Titan X (Pascal) GPU. Pytorch 1.0.0 on Python 3.6 is used to implement our generative model.[2] The network is optimized until $n_{\text{iter}} = 10,000$ with $B = 1$ using the Adam optimizer [49] with default settings and a learning rate of 10^{-3}.

F. Architectures

The mapping network g_ϕ comprises two consecutive fully connected layers of 512 hidden dimensions with ReLUs in between. It outputs an $L = 64$-dimensional latent vector which is reshaped to (8×8) for the subsequent generative network h_ψ (Table I). The generative network consists of convolutional layers, batch normalization layers, ReLUs, and nearest-neighbor interpolations. We apply zero-padding before convolution to let the size of the output mirror that of the input. At the last layer, ReLU is not used. The output has two channels because MRI images take complex values.

IV. DESIGN OF THE LATENT SPACE

In this section, we analyze the individual components of our model and compare the performance with baselines. We first demonstrate the simplest setup that reconstructs a single heart cycle. We then move on to a more complicated dataset that has multiple heart cycles.

A. Straight-Line Manifold for a Single Heart Cycle

A straight-line manifold can help the network to encode the temporal variations of images. To implement it, we first sample $\mathbf{z}_0$, $\mathbf{z}_{K-1} \sim \mathcal{U}(\mathbb{R}^L)$. Then, the intermediate $\mathbf{z}_k$ are obtained by linear interpolation. This yields a straight-line manifold that simply joins the end points as

$$\mathbf{z}_k = (1 - \alpha_k)\,\mathbf{z}_0 + \alpha_k\,\mathbf{z}_{K-1}, \quad (8)$$

where $\alpha_k = k/(K - 1)$.

Although simple, this configuration already outperforms the other baseline methods and successfully reconstructs the dynamics for a single cycle dataset (Table II).

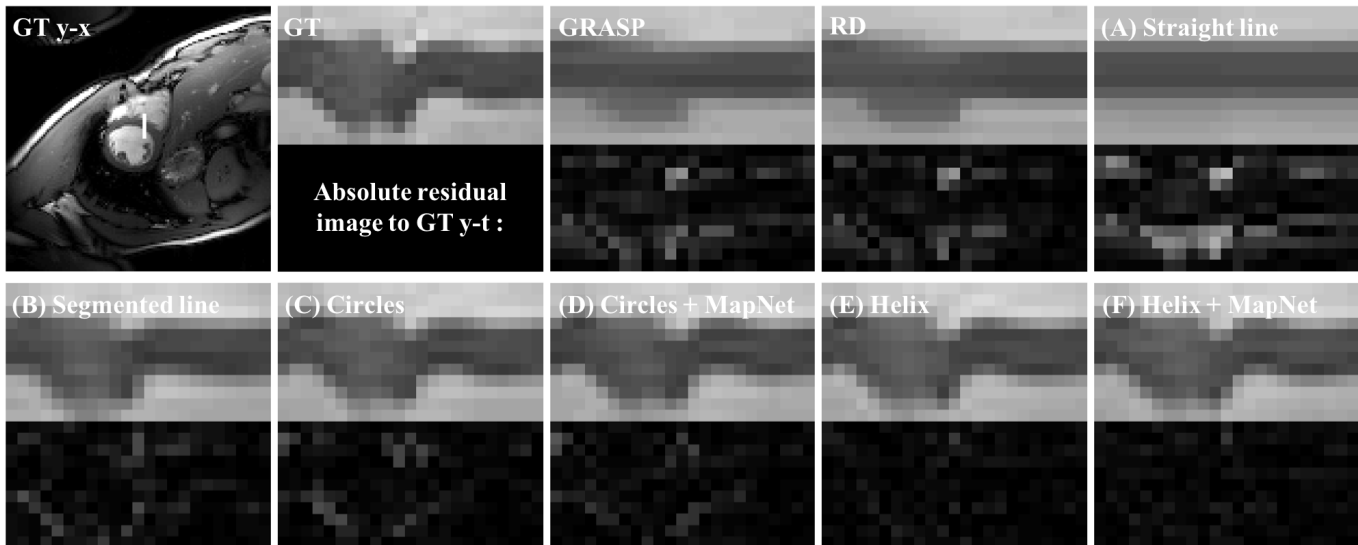[2]Our code is available at https://github.com/jaejun-yoo/TDDIP

Fig. 2. Visual comparison of reconstructed (y-t) images using the baseline methods and ours with each configuration in Table III. The reconstructed images from fully sampled Cartesian trajectories are used as a ground truth. A white line at the heart region indicates the cross section that is visualized. Here, for simulating RD [26], [27], we reorder the spokes of each frame from 13 periods resulting in 169 spokes per frame. For a better comparison, the absolute residual images to the ground truth are provided in the lower panels.

TABLE II

PERFORMANCE ON THE RETROSPECTIVE
DATASET FOR A SINGLE HEART CYCLE

Method	RSNR (dB)	SSIM
NUFFT	8.4	0.582
GRASP [15]	24.1	0.980
Straight line ($L = 64$)	**26.5**	**0.985**

B. Manifolds for Multiple Heart Cycles

In practice, the measurements generally span several heart cycles. To better exploit the fact that the cardiac movement has a nearly periodic behavior, it is of interest to explore more sophisticated manifolds.

- **Segmented Line**. We first sample $p+1 = 14$ landmarks $\{\mathbf{z}^{(\tau)}\}_{\tau \in [0...13]} \sim \mathcal{U}(\mathbb{R}^L)$, where p is the number of cardiac periods. We generate a set of equispaced intermediate $\mathbf{z}_k$ of each segment by a linear combination of $\mathbf{z}^\tau$ and $\mathbf{z}^{(\tau+1)}$, $\forall \tau \in [0 \dots 12]$.

- **Circles**. Let $\mathbf{z}_k = (z_1^{(k)}, z_2^{(k)}, \mathbf{z}_{\text{slack}}) \in \mathbb{R}^L$ and $\mathbf{z}_{\text{slack}} \sim \mathcal{U}(\mathbb{R}^{L-2})$. The first two coordinates $(z_1^{(k)}, z_2^{(k)})$ are points from a unit circle with p cycles. The slack coordinates do not depend on k. Thus, we have that

$$\mathbf{z}_k = \left(\cos(\frac{2\pi\ p\ k}{(K-1)}), \sin(\frac{2\pi\ p\ k}{(K-1)}), \mathbf{z}_{\text{slack}} \right). \quad (9)$$

- **Helix.** Similar to "Circles", the first two coordinates of $\mathbf{z}_k$ are points from a unit circle with p cycles. The slack coordinates $\mathbf{z}_{\text{slack}} \sim \mathcal{U}(\mathbb{R}^{L-2})$ are now scaled by $\frac{k}{(K-1)}$, leading to

$$\mathbf{z}_k = \left(\cos(\frac{2\pi\ p\ k}{(K-1)}), \sin(\frac{2\pi\ p\ k}{(K-1)}), \frac{k\ \mathbf{z}_{\text{slack}}}{(K-1)} \right). \quad (10)$$

1) Effect of the Manifolds: In Figure 2, we show the reconstructed (y-t) images of the cross section that is denoted by

TABLE III

PERFORMANCE ON THE RETROSPECTIVE DATASET FOR MULTIPLE
HEART CYCLES. AVERAGED RSNR AND SSIM OVER FIVE RUNS
OF EACH CNN LATENT SPACE DESIGN ARE SHOWN

Method	RSNR (dB)	SSIM
NUFFT	8.5	0.579
GRASP [15]	24.2	0.982
Reordered method (RD) [26], [27]	25.0	0.986
Straight line ($L = 64$)	20.6	0.973
Segmented line ($L = 64$)	25.9	0.987
Circles ($L = 64$)	27.1	0.990
Circles ($L = 2$) + MapNet ($L = 64$)	27.5	0.991
Helix ($L = 64$)	27.6	0.990
Helix ($L = 3$) + **MapNet** ($L = 64$)	**28.1**	**0.991**

a white line in the GT (y-x) image.[3] When we use a straight-line manifold, the network fails to capture the heart movement and outputs the same static image over all frames. This is natural since most of the pixels are static and the dynamic parts are localized in a small area. Thus, the network easily finds a local minimum that corresponds to an image that remains constant over all frames. However, as soon as we switch to "periodic-like" manifold designs, the network starts to reconstruct the movement (Table III). For example, when we use a line with 13 segments as input, the performance is better than the RD that uses the same information. Using circles with 13 repetitions as input manifold, we improve even further. However, the helix input manifold gives the best performance among the others without MapNet because the heartbeat is nearly periodic.

2) Effect of the Mapping Network: In addition to the choice of its manifold, our method has another design choice: its mapping network. By introducing MapNet, the network can adapt its input manifold to a given dataset, which allows us to further improve the reconstruction (Table III). This can be

---

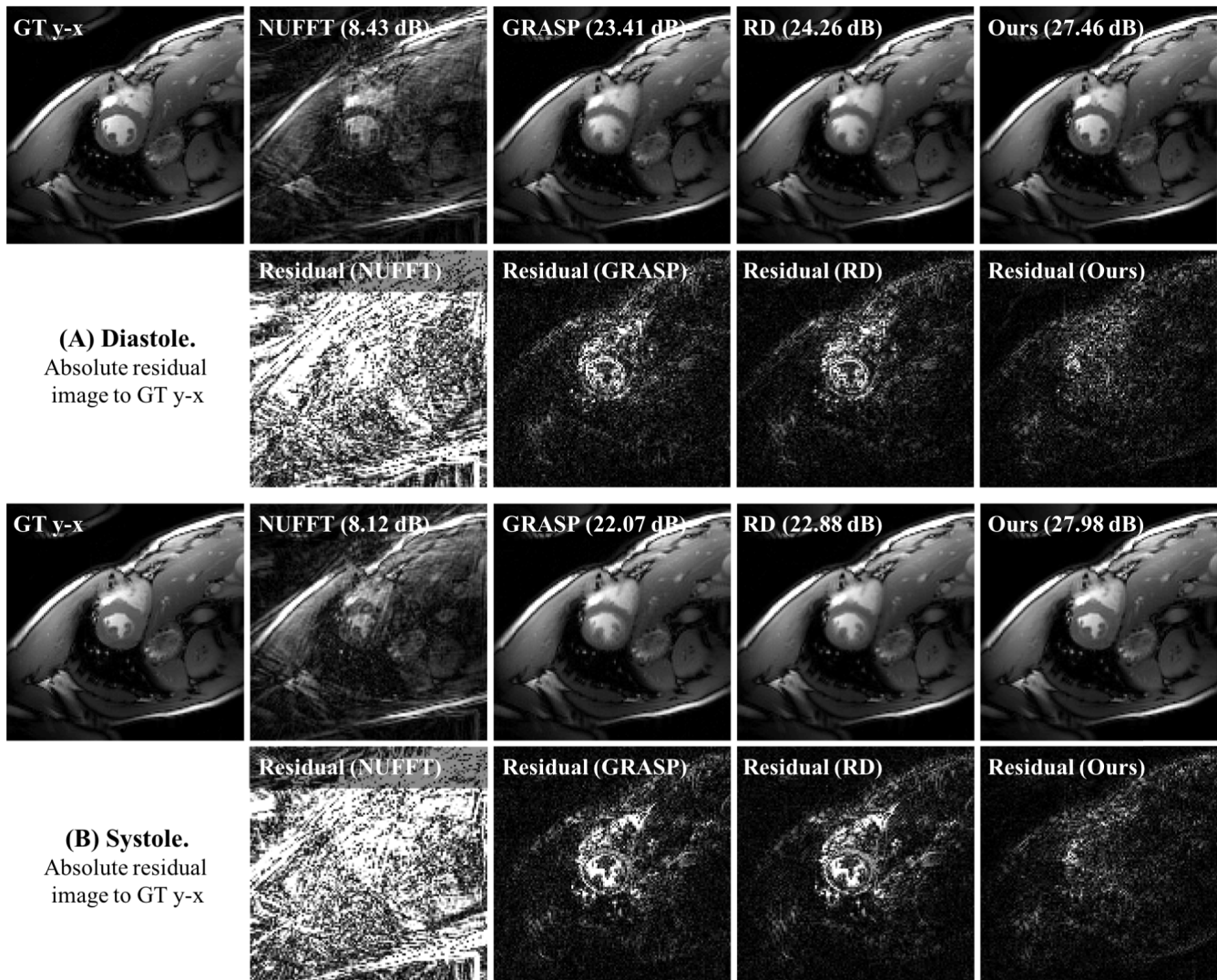[3]For display purposes, we show only one cycle of our cross section.

Fig. 3. Visual comparison of reconstructed (y-x) images using NUFFT, GRASP [15], RD [26], [27], and our method at both (A) diastole and (B) systole. The absolute residual image to the ground truth is also given.

clearly seen in the t-SNE visualization of the mapped latent space (Figure 7), which we discuss in Section VI.

In summary, our analysis shows that both a careful design of the manifold and the use of a mapping network are necessary to achieve the best performance. Based on these, from now on, we use 'Helix+MapNet' as our default setup.

V. RESULTS

We first show results on the retrospective dataset, where the desired behaviors of the reconstruction methods are well-defined. We then illustrate on the fetal cardiac dataset that the observations extend well to a real scenario.

A. Retrospective Dataset: Multiple Heart Cycles

The benefits of our method are evident in both the (y-t) view (Figure 2) and (y-x) view (Figure 3) of each frame. In Figure 2, both GRASP and RD reconstruct the movement of the heart. RD shows a better performance than GRASP, which was expected because it takes advantage of the period information that is estimated while reordering the frames. However, as can be seen in the residuals, GRASP and RD show significant errors in the reconstruction of the dynamics. In the (y-x) view of Figure 3, we report the RSNR scores of each method for the corresponding frames. Here, GRASP shows blurring artifacts, while the residual image reveals errors around the wall of the heart in both the GRASP and RD reconstructions. By contrast, our method gives better results with fewer artifacts. This difference is more prominent at systole (Figure 2 (B)), a heart phase when the cardiac motions are captured better by our model than the others, quantitatively and qualitatively. The systolic phase captured by GRASP and RD is too flat, causing the severe errors concentrated near the heart area. Especially, the RSNR value of RD drops from 24.26 dB to 22.88 dB, while our method works similarly at both phases (27.46 dB at diastole and 27.98 dB at systole). This is clearly seen in the residual images. Compared to the other baselines, our method provides the closest reconstruction to the ground truth with very small errors. Here, for better visualization, we increased the magnitude of the residual images ten times. We see likewise in Figure 4 that our method provides a stable reconstruction quality over the entire cardiac cycle.
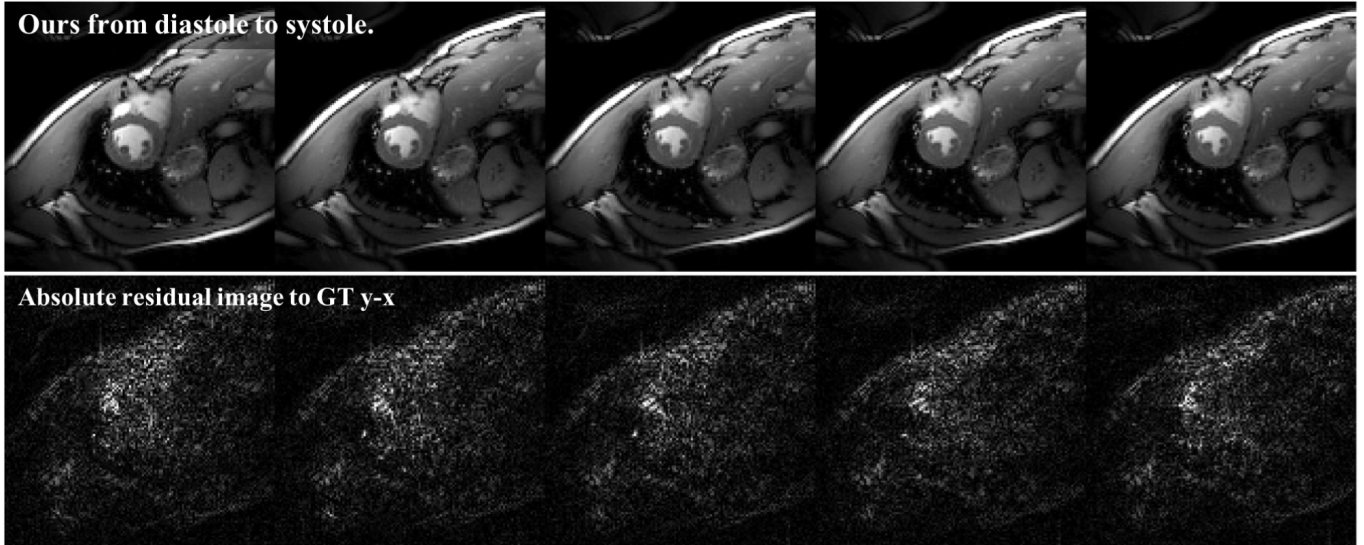
Fig. 4. Entire cardiac cycle reconstructed by our method and the absolute residual images to the ground truth.

B. Fetal Cardiac Dataset

Having demonstrated the superior behavior of our method on the retrospective dataset, we now assess our model on real data. In the absence of ground truth, we shall take the static image that is generated from all spokes as pseudo-gold standard (Figure 5 first column)—note that it is of high quality only in the regions that are not moving.

Like in the retrospective experiments, GRASP and RD are able to reconstruct cardiac movements at both systole and diastole. RD gives much better reconstructions, especially in the dynamic region (Figures 5 (A) and (A")). However, RD shows a spurious artifact at the edge area (Figure 5 (B)) and fails to find the detailed structures of the static background (Figure 5 (C)). By contrast, our method produces better-resolved features in dynamic areas. Specifically, it preserves well the hyperintense dot-like structures (leftmost and middle arrows in Figures 5 (A) and (A")) as well as the line-like structure (the rightmost arrow in Figures 5 (A) and (A")). In addition, our method successfully reconstructs the static areas. For example, it recovers even a small dot-like feature (leftmost black arrow in Figure 5 (B)), which RD fails to recover. It does not suffer from artifacts at the edges (white arrows in Figure 5 (B)) and recovers the low-intensity background areas as well (Figure 5 (C)).

In Figure 6, it is apparent that NUFFT completely fails in capturing the fetal cardiac beats. The GRASP reconstruction is less noisy but still far from satisfactory. RD fares better; unfortunately, its reordering process can lead it to superpose in the same frame spokes that belong to different phases of the cardiac cycle. By contrast, our method reconstructs each frame with data from just a few neighboring spokes ($n_s = 5$),[4] thus avoiding the mingling of different cycles. The reconstructed systolic phase captures the true motion of the heart better. The cross section from our method is similar to that of RD but

the motion may look a little smoother in our case (Figure 6, top row). This is expected because RD uses retrospectively gated data, while our method uses real-time data (*i.e.*, ours uses much fewer data per image). In addition, this is also partly due to our result having more frames for a single cycle than RD. Specifically, unlike our reconstruction of entire image frames, RD reconstructs a single cardiac cycle since it uses multiple Fourier slices of different time frames to recover a single image. In addition to the absence of ground truth, this complicates the direct comparison between the results of RD and our method.
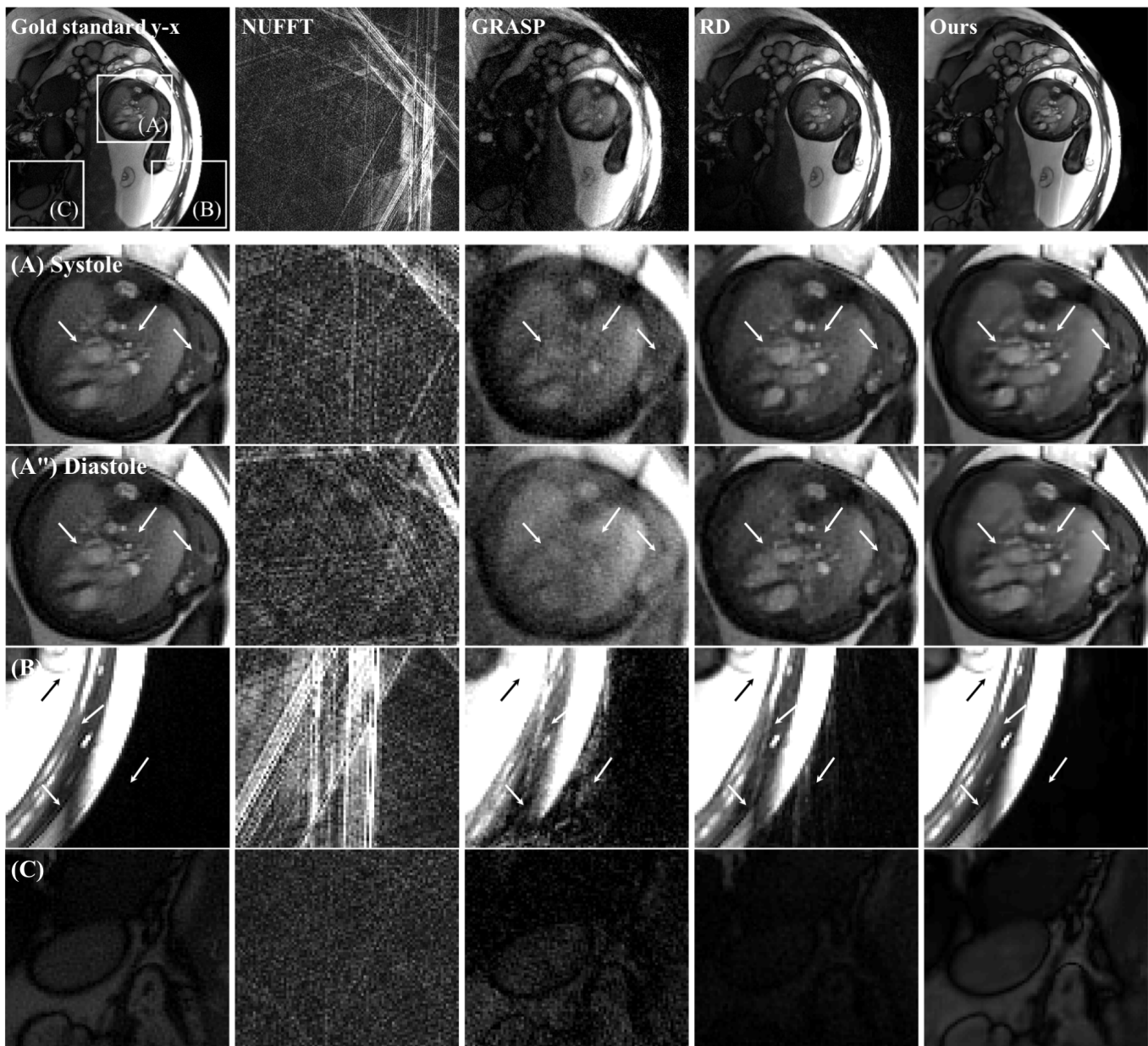
VI. DISCUSSION

A. t-SNE Visualization of the Latent and Image Spaces

To assess the extent of structural change as a function of time, we used t-stochastic neighborhood embeddings (t-SNE) [53]. t-SNE is a nonlinear dimension-reduction technique that is widely used for the mapping of high-dimensional data such as images or network features into a low-dimensional space of two or three dimensions. Specifically, the mapping is such that similar points are grouped together and dissimilar points are modeled by distant points with high probability.

In Figure 7 (A), we first show the t-SNE result of the original manifold when the variables are generated according to (10) with $L = 64$ and $p = 13$. Unsurprisingly, this recovers a helix with 13 cycles. In Figure 7 (B), we show the t-SNE result of the 64-dimensional latent variables $g_{\phi*}$, where $\mathbf{z}_k$ is generated according to (10) with $L = 3$ and $p = 13$. It suggests that the latent variables still lie on a helical manifold with 13 periods, while the height of the helix is shortened compared to the original input in Figure 7 (A)—the first and the last cycles become closer, but they are not entirely collapsed to a single circle. This implies that MapNet successfully encoded both the similarity and the dissimilarity among different cardiac cycles.

Finally, in Figure 7 (C), we display the t-SNE result of the (256×256)-dimensional reconstructed images. It again shows a helical geometry with 13 local folds, each of which corresponding to a single cycle of the cardiac motion.

---

[4]We have also tried to reconstruct images with fewer spokes ($n_s = 3$) and even without sharing the spokes ($n_s = 1$). The quality of reconstructed images degrades as the number of shared spokes decreases, but our method converges and reconstructs the movements (results not shown).

Fig. 5. Visual comparison of reconstructed fetal hearts: (y-x) images at both systole and diastole. The top row is the results of NUFFT, GRASP [15], RD [26], [27], and our method (Helix + MapNet). The four bottom rows show magnified views, where A and A" are the heart region at systole and diastole, respectively.

The irregular shape and length of each fold implies different movements and rates of the reconstructed cardiac cycles, respectively. This reveals that the near-periodicity of the data is well represented by the network. As shown in Section IV, we ascribe this to the introduction of MapNet that can warp the given manifold in a data-adaptive fashion, while retaining the prior information that we inject when we impose the geometry of the manifold.

B. Benefits of Our Approach

1) Continuous Dynamic Reconstruction: One major benefit of our approach is that it lets us reconstruct temporally continuous dynamic images. We showed that the network f_{θ^*} successfully captures the underlying nonlinear dynamics of the image manifold, and the input variable $\mathbf{z}_k$ lets us reconstruct the image at the corresponding timestamp (Figure 7). Because our method represents images as a learned parametric function f_{θ^*}, we can recover nontrivial intra-frame images by navigating between two consecutive input variables, which would not be possible with other standard interpolation methods such as temporal bilinear interpolation. Note that this is a unique benefit of our method that the other algorithms cannot provide.

2) Memory Savings: In the methods based on compressed sensing (CS), the gradient updates of the iterative optimization process necessitate memory that is large enough to hold the target reconstruction volume. For example, the reconstruction of 5,000 frames with spatial size (256×256) would need one to handle data of size $(256 \times 256 \times 5,000)$, which demands for over a gigabyte of memory. Our approach, by contrast, requires much less memory. It optimizes the neural network
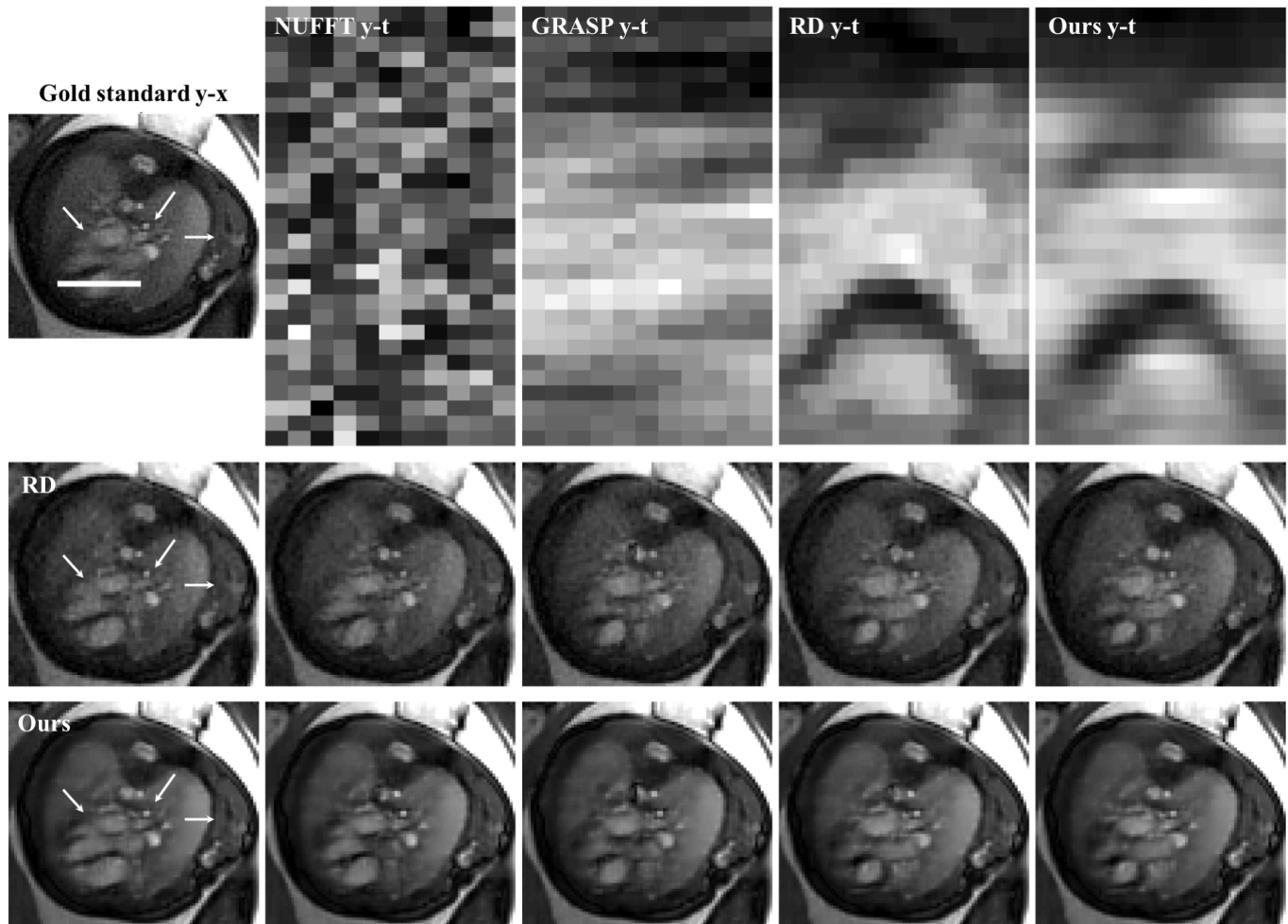
Fig. 6.   Visual comparison of reconstructed fetal hearts: (y-t) images. A white line in the gold standard indicates the cross section that is visualized. The top row contains the results of NUFFT, GRASP [15], RD [26], [27], and our method (Helix + MapNet). The two bottom rows provide a comparison between RD and our method from diastole to systole.
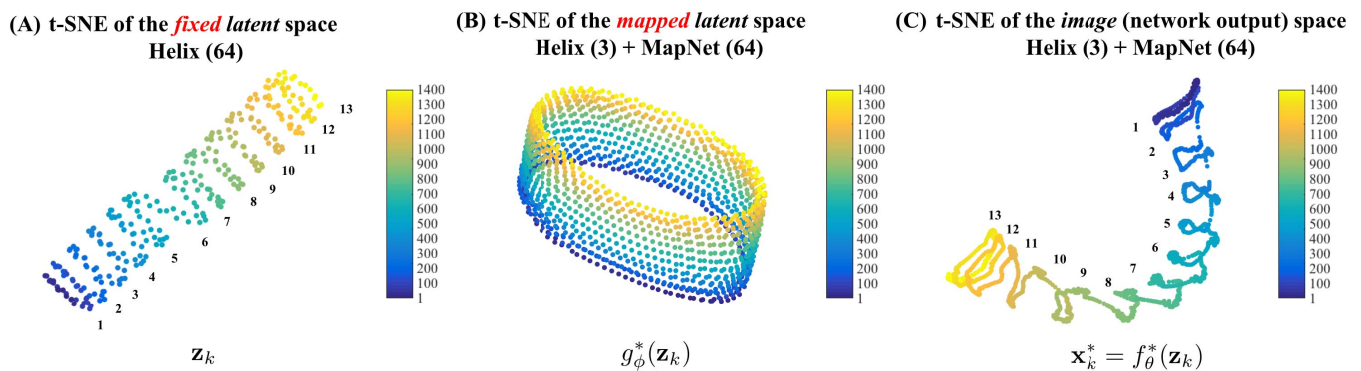

Fig. 7.   Visualization of three-dimensional t-SNE embeddings. (A) the fixed latent space of a helix in 64-dimension ('Helix'); (B) the mapped latent space by the mapping network in 64-dimension ('Helix + MapNet'); and (C) its corresponding image space in (256×256) dimensions. Here, the temporal index is color-coded (1,400 frames). There are approximately 13 cycles of heart motion, which are also clearly seen in the embedded helix of the reconstructed images.

using batches, which requires the simultaneous handling of only those frames that correspond to the batch size. In short, the fact that our proposed approach handles few 2D images whereas CS handles a 2D+t extended sequence leads to

substantial savings, particularly for golden-angle dynamic MRI with many frames. In our approach, we only store a 2D generative model; for example, its memory demands for the spatial size (256×256) are about half-a-dozen

megabytes. This cost is negligible compared to that of the CS approach.

3) Easier Optimization With Fewer Hyperparameters: Because our model is fully automated, it leads to a simpler optimization task with fewer hyperparameters than the conventional methods. For instance, once the model is set, k-t SENSE requires three additional, interdependent hyperparameters whose optimal values are found only after some substantial grid-search effort. In contrast, our approach has only two hyperparameters (an initial learning rate and a number of iterations) that are easy to set and do not affect much the final performance as long as the model converges.

C. Limitations and Future Work

1) Faster Forward Model: Regarding the execution time, the major bottleneck of our method is the slow forward model. It depends on the NUFFT package which, in its current implementation, does not benefit from a GPU and is a major cause for slowdown. Indeed, NUFFT takes 47 % of the entire running time of our algorithm per each iteration; the average processing time for 100 repetitions is 6.55 s for back and forth NUFFTs, and 3.08 s for the remaining parts. With a more efficient, GPU-friendly implementation such as [54], our algorithm could be substantially accelerated. One obvious way to achieve this goal is to migrate to a lower-level programming language than Python. Another way out is to use the Cartesian sampling pattern. Since our framework is not restricted to the radial sampling pattern, this will remove the bottleneck while maintaining all the benefits of using our method.

2) Extension to Free-Breathing Setting: Although breath-held cine MRI is widely used, it is often challenging for children and patients with heart failure or respiratory complications. One of the standard alternatives is an ungated real-time imaging in free-breathing mode using nonuniform acquisitions at the expense of lower spatial and/or temporal resolution [55], [56]. Several manifold-learning methods [20], [24], [25] have been proposed to handle this more difficult scenario. It would be interesting to extend our generative framework to this free-breathing setting in the spirit of what has been accomplished in [57] recently.

3) Extension to Multi-Slice Imaging: In this paper, our main focus is mainly on reconstructing 2D+t with a single slice measurement dataset. Similar to what we did for encoding temporal variations, we believe that our technique can be extended to multi-slice data by designing a manifold that encodes the spatial variations across different slices. Another easy extension would be to use a 3D CNNs architecture. However, from an experiment point of view, we are still far from performing 3D imaging with high temporal resolution.

4) Different Architectural Variations: Because the network architecture that we explored in the paper is never exhaustive, there can be many architectural variations that would bring more improvement than the currently reported results. For example, we believe that adopting the advanced generative architectures in computer vision area such as style-based models [43] or good initialization techniques of the network parameters would further improve the performance easily.

VII. CONCLUSION

In this paper, we proposed an unsupervised deep-learning-based algorithm for the reconstruction of dynamic magnetic resonance images (MRI). It provides high spatial resolution with access to the subframe—or even continuous—temporal control of dynamic images. By designing a one-dimensional manifold, combined with a mapping network, our generative network model fully exploits the representation power of the networks as well as their structural priors. Our study showed that the proposed method successfully reconstructs dynamic MRI in an end-to-end manner and outperforms the state-of-the-art CS approaches by 3.1 dB. To the best of our knowledge, this is the first unsupervised-learning approach in accelerated dynamic MRI.

ACKNOWLEDGMENT

REFERENCES

[1] M. A. Griswold *et al.*, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magn. Reson. Med., Off. J. Int. Soc. Magn. Reson. Med.*, vol. 47, no. 6, pp. 1202–1210, Jun. 2002.

[2] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magn. Reson. Med., Off. J. Int. Soc. Magn. Reson. Med.*, vol. 42, no. 5, pp. 952–962, Nov. 1999.

[3] P. Kellman, F. H. Epstein, and E. R. McVeigh, "Adaptive sensitivity encoding incorporating temporal filtering (TSENSE)," *Magn. Reson. Med., Off. J. Int. Soc. Magn. Reson. Med.*, vol. 45, no. 5, pp. 846–852, May 2001.

[4] F. A. Breuer, P. Kellman, M. A. Griswold, and P. M. Jakob, "Dynamic autocalibrated parallel imaging using temporal GRAPPA (TGRAPPA)," *Magn. Reson. Med., Off. J. Int. Soc. Magn. Reson. Med.*, vol. 53, no. 4, pp. 981–985, Apr. 2005.

[5] J. Tsao, P. Boesiger, and K. P. Pruessmann, "$k - t$ BLAST and k-t SENSE: Dynamic MRI with high frame rate exploiting spatiotemporal correlations," *Magn. Reson. Med., Off. J. Int. Soc. Magn. Reson. Med.*, vol. 50, no. 5, pp. 1031–1042, Nov. 2003.

[6] F. Huang, J. Akao, S. Vijayakumar, G. R. Duensing, and M. Limkeman, "*k*-t GRAPPA: A *k*-space implementation for dynamic MRI with high reduction factor," *Magn. Reson. Med., Off. J. Int. Soc. Magn. Reson. Med.*, vol. 54, no. 5, pp. 1172–1184, Nov. 2005.

[7] D. Xu, K. F. King, and Z.-P. Liang, "Improving $k - t$ SENSE by adaptive regularization," *Magn. Reson. Med., Off. J. Int. Soc. Magn. Reson. Med.*, vol. 57, no. 5, pp. 918–930, May 2007.

[8] H. Jung, J. C. Ye, and E. Y. Kim, "Improved $k - t$ BLAST and $k - t$ SENSE using FOCUSS," *Phys. Med. Biol.*, vol. 52, no. 11, p. 3201, 2007.

[9] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed sensing in dynamic MRI," *Magn. Reson. Med., Off. J. Int. Soc. Magn. Reson. Med.*, vol. 59, no. 2, pp. 365–373, Feb. 2008.

[10] J. Ji and T. Lang, "Dynamic MRI with compressed sensing imaging using temporal correlations," in *Proc. 5th IEEE Int. Symp. Biomed. Imag., From Nano Macro*, May 2008, pp. 1613–1616.

[11] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye, "$k - t$ FOCUSS: A general compressed sensing framework for high resolution dynamic MRI," *Magn. Reson. Med., Off. J. Int. Soc. Magn. Reson. Med.*, vol. 61, no. 1, pp. 103–116, Jan. 2009.

[12] R. Otazo, D. Kim, L. Axel, and D. K. Sodickson, "Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI," *Magn. Reson. Med.*, vol. 64, no. 3, pp. 767–776, 2010.

[13] Y. Wang and L. Ying, "Compressed sensing dynamic cardiac cine MRI using learned spatiotemporal dictionary," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 4, pp. 1109–1120, Apr. 2014.

[14] L. Feng *et al.*, "Highly accelerated real-time cardiac cine MRI using $k − t$ SPARSE-SENSE," *Magn. Reson. Med.*, vol. 70, no. 1, pp. 64–74, 2013.

[15] L. Feng *et al.*, "Golden-angle radial sparse parallel MRI: Combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI," *Magn. Reson. Med.*, vol. 72, no. 3, pp. 707–717, Sep. 2014.

[16] L. Feng, L. Axel, H. Chandarana, K. T. Block, D. K. Sodickson, and R. Otazo, "XD-GRASP: Golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing," *Magn. Reson. Med.*, vol. 75, no. 2, pp. 775–788, Feb. 2016.

[17] S. G. Lingala, Y. Hu, E. DiBella, and M. Jacob, "Accelerated dynamic MRI exploiting sparsity and low-rank structure: K-t SLR," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1042–1054, May 2011.

[18] R. Otazo, E. Candès, and D. K. Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components," *Magn. Reson. Med.*, vol. 73, no. 3, pp. 1125–1136, Mar. 2015.

[19] S. Ravishankar, B. E. Moore, R. R. Nadakuditi, and J. A. Fessler, "Low-rank and adaptive sparse signal (LASSI) models for highly accelerated dynamic imaging," *IEEE Trans. Med. Imag.*, vol. 36, no. 5, pp. 1116–1128, May 2017.

[20] S. Poddar and M. Jacob, "Dynamic MRI using SmooThness regularization on manifolds (SToRM)," *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 1106–1115, Apr. 2016.

[21] U. Nakarmi, Y. Wang, J. Lyu, D. Liang, and L. Ying, "A kernel-based low-rank (KLR) model for low-dimensional manifold recovery in highly accelerated dynamic MRI," *IEEE Trans. Med. Imag.*, vol. 36, no. 11, pp. 2297–2307, Nov. 2017.

[22] U. Nakarmi, K. Slavakis, J. Lyu, and L. Ying, "M-MRI: A manifold-based framework to highly accelerated dynamic magnetic resonance imaging," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 19–22.

[23] U. Nakarmi, K. Slavakis, and L. Ying, "MLS: Joint manifold-learning and sparsity-aware framework for highly accelerated dynamic magnetic resonance imaging," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 1213–1216.

[24] S. Poddar, Y. Q. Mohsin, D. Ansah, B. Thattaliyath, R. Ashwath, and M. Jacob, "Manifold recovery using kernel low-rank regularization: Application to dynamic imaging," *IEEE Trans. Comput. Imag.*, vol. 5, no. 3, pp. 478–491, Sep. 2019.

[25] A. H. Ahmed, R. Zhou, Y. Yang, P. Nagpal, M. Salerno, and M. Jacob, "Free-breathing and ungated cardiac cine using navigator-less spiral SToRM," 2019, *arXiv:1901.05542*. [Online]. Available: http://arxiv.org/abs/1901.05542

[26] J. Yerly *et al.*, "Coronary endothelial function assessment using self-gated cardiac cine MRI and $k − t$ sparse SENSE," *Magn. Reson. Med.*, vol. 76, no. 5, pp. 1443–1454, Nov. 2016.

[27] J. Chaptinel *et al.*, "Fetal cardiac cine magnetic resonance imaging *in utero*," *Sci. Rep.*, vol. 7, no. 1, pp. 1–10, Nov. 2017.

[28] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. McCann, and M. Unser, "CNN-based projected gradient descent for consistent CT image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1440–1453, May 2018.

[29] E. Kang, W. Chang, J. Yoo, and J. C. Ye, "Deep convolutional framelet denosing for low-dose CT via wavelet residual network," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1358–1369, Jun. 2018.

[30] J. Yoo, A. Wahab, and J. C. Ye, "A mathematical framework for deep learning in elastic source imaging," *SIAM J. Appl. Math.*, vol. 78, no. 5, pp. 2791–2818, Jan. 2018.

[31] J. Yoo *et al.*, "Deep learning diffuse optical tomography," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 877–887, Apr. 2020.

[32] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-Net for compressive sensing MRI," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2016, pp. 10–18.

[33] S. Wang *et al.*, "Accelerating magnetic resonance imaging via deep learning," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 514–517.

[34] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, Sep. 2017.

[35] K. C. Tezcan, C. F. Baumgartner, R. Luechinger, K. P. Pruessmann, and E. Konukoglu, "MR image reconstruction using deep density priors," *IEEE Trans. Med. Imag.*, vol. 38, no. 7, pp. 1633–1642, Jul. 2019.

[36] K. Hammernik *et al.*, "Learning a variational network for reconstruction of accelerated MRI data," *Magn. Reson. Med.*, vol. 79, no. 6, pp. 3055–3071, Jun. 2018.

[37] Y. Han, J. Yoo, H. H. Kim, H. J. Shin, K. Sung, and J. C. Ye, "Deep learning with domain adaptation for accelerated projection-reconstruction MR," *Magn. Reson. Med.*, vol. 80, no. 3, pp. 1189–1205, Feb. 2018.

[38] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for dynamic MR image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 491–503, Feb. 2018.

[39] A. Hauptmann, S. Arridge, F. Lucka, V. Muthurangu, and J. A. Steeden, "Real-time cardiovascular MR with spatio-temporal artifact suppression using deep learning–proof of concept in congenital heart disease," *Magn. Reson. Med.*, vol. 81, no. 2, pp. 1143–1156, Feb. 2019.

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

[41] S. Biswas, H. K. Aggarwal, and M. Jacob, "Dynamic MRI using model-based deep learning and SToRM priors: MoDL-SToRM," *Magn. Reson. Med.*, vol. 82, no. 1, pp. 485–494, Jul. 2019.

[42] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.

[43] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4401–4410.

[44] Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "Stargan v2: Diverse image synthesis for multiple domains," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 8188–8197.

[45] A. Yazdanpanah, O. Afacan, and S. Warfield, "Non-learning based deep parallel MRI reconstruction (NLDpMRI)," *Proc. SPIE*, vol. 10949, Feb. 2019, Art. no. 1094904.

[46] K. Gong, C. Catana, J. Qi, and Q. Li, "PET image reconstruction using deep image prior," *IEEE Trans. Med. Imag.*, vol. 38, no. 7, pp. 1655–1665, Jul. 2019.

[47] R. Hyder and M. S. Asif, "Generative models for low-rank video representation and reconstruction," 2019, *arXiv:1902.11132*. [Online]. Available: http://arxiv.org/abs/1902.11132

[48] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Nov. 2006.

[49] D. Kingma and J. Ba, "ADAM: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Diego CA, USA, May 2015, pp. 1–15.

[50] L. Di Sopra, D. Piccini, S. Coppo, M. Stuber, and J. Yerly, "An automated approach to fully self-gated free-running cardiac and respiratory motion-resolved 5D whole-heart MRI," *Magn. Reson. Med.*, vol. 82, no. 6, pp. 2118–2132, Jul. 2019.

[51] A. C. Larson, R. D. White, G. Laub, E. R. McVeigh, D. Li, and O. P. Simonetti, "Self-gated cardiac cine MRI," *Magn. Reson. Med.*, vol. 51, no. 1, pp. 93–102, Jan. 2004.

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

[53] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[54] T. Lu, T. Marin, Y. Zhuo, Y.-F. Chen, and C. Ma, "Nonuniform fast Fourier transform on TPUs," in *Proc. IEEE 18th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2021, pp. 19–22.

[55] S. Uribe *et al.*, "Whole-heart cine MRI using real-time respiratory self-gating," *Magn. Reson. Med.*, vol. 57, no. 3, pp. 606–613, 2007.

[56] X. Feng, M. Salerno, C. M. Kramer, and C. H. Meyer, "Non-Cartesian balanced steady-state free precession pulse sequences for real-time cardiac MRI," *Magn. Reson. Med.*, vol. 75, no. 4, pp. 1546–1555, Apr. 2016.

[57] Q. Zou, A. H. Ahmed, P. Nagpal, S. Kruger, and M. Jacob, "Dynamic imaging using a deep generative SToRM (Gen-SToRM) model," *IEEE Trans. Med. Imag.*, early access, Mar. 15, 2021, doi: 10.1109/TMI.2021.3065948.