

# Dynamic Imaging Using Deep Bi-Linear Unsupervised Representation (DEBLUR)

Abdul Haseeb Ahmed<sup>ID</sup>, Member, IEEE, Qing Zou<sup>ID</sup>, Member, IEEE, Prashant Nagpal<sup>ID</sup>, and Mathews Jacob<sup>ID</sup>, Member, IEEE

**Abstract**— Bilinear models such as low-rank and dictionary methods, which decompose dynamic data to spatial and temporal factor matrices are powerful and memory-efficient tools for the recovery of dynamic MRI data. Current bilinear methods rely on sparsity and energy compaction priors on the factor matrices to regularize the recovery. Motivated by deep image prior, we introduce a novel bilinear model, whose factor matrices are generated using convolutional neural networks (CNNs). The CNN parameters, and equivalently the factors, are learned from the undersampled data of the specific subject. Unlike current unrolled deep learning methods that require the storage of all the time frames in the dataset, the proposed approach only requires the storage of the factors or compressed representation; this approach allows the direct use of this scheme to large-scale dynamic applications, including free breathing cardiac MRI considered in this work. To reduce the run time and to improve performance, we initialize the CNN parameters using existing factor methods. We use sparsity regularization of the network parameters to minimize the overfitting of the network to measurement noise. Our experiments on free-breathing and ungated cardiac cine data acquired using a navigated golden-angle gradient-echo radial sequence show the ability of our method to provide reduced spatial blurring as compared to classical bilinear methods as well as a recent unsupervised deep-learning approach.

**Index Terms**— Bilinear model, cardiac MRI, dynamic imaging, image reconstruction, unsupervised learning.

## I. INTRODUCTION

DYNAMIC MRI (DMRI) plays an important role in clinical applications such as cardiac cine MRI, which is commonly used by clinicians for the anatomical and functional assessment of organs. The clinical practice is to acquire the

Manuscript received 21 December 2021; revised 19 February 2022; accepted 12 April 2022. Date of publication 18 April 2022; date of current version 30 September 2022. This work was supported in part by NIH under Grant R01EB019961, Grant R01 AG067078, and Grant 1R01EB031169-01, and in part by the MRI instrument under Grant 1S10OD025025-01. (Corresponding author: Abdul Haseeb Ahmed.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board (IRB) at the University of Iowa approved the acquisition of the data, and written consent was obtained from the subject.

Abdul Haseeb Ahmed, Qing Zou, and Mathews Jacob are with the Department of Electrical and Computer Engineering, The University of Iowa, Iowa City, IA 52242 USA (e-mail: abdul-ahmed@uiowa.edu; zou-qing@uiowa.edu; mathews-jacob@uiowa.edu).

Prashant Nagpal is with the Department of Radiology, University of Wisconsin-Madison, Madison, WI, USA (email: pnagpal@wisc.edu).

Digital Object Identifier 10.1109/TMI.2022.3168559

cine data using breath holding to achieve good spatial and temporal resolution. However, it is difficult for many subjects, including children, patients with myocardial infarction, and chronic obstructive pulmonary disease patients, to hold their breath [1]. In addition, multiple breath-holds prolong the scan time, adversely impacting patient comfort and compliance.

Several computational approaches have been introduced to reduce the breath-held duration in cardiac cine and to enable free-breathing imaging. Early approaches relied on carefully designed signal models to exploit the structure of the data in  $x$ - $f$  space [2]–[4], sparsity [5], or binned the data to different phases [6] which facilitates the signal recovery from undersampled measurements. In recent years, bilinear models [7]–[10], which represent the signal as the product of spatial and temporal factors, have emerged as powerful alternatives for the recovery of large scale data. The factor-based framework has been combined with other priors, including low-rank and sparsity [7], [9], low-rank +sparse [11], blind compressed sensing that learns a dictionary from the data [12], and motion-compensation [13], [14]. Recently, non-linear manifold models that rely on kernel low-rank relation have been shown to outperform the subspace-based models in the context of free-breathing and ungated cardiac MRI [15]–[17]. A major benefit of the factor-based methods is the significantly reduced memory demand of these algorithms, in addition to the good reconstructions they offer. Specifically, the spatial and temporal factors are significantly smaller in dimension than the dynamic dataset, which facilitates the recovery of large 3D volumes [10], [18]. The distinction between the methods can be considered as the specific priors applied on the spatial and temporal factors, including energy priors in the low-rank setting [9], unit column norm and sparsity priors in the blind dictionary learning setting [12], and kernel priors in the manifold setting [15], [16].

Model-based deep unrolled learning methods [20]–[22], which interleave convolutional neural network (CNN) blocks with data-consistency enforcing optimization modules, have been shown as the promising alternatives for classical methods in image recovery. A challenge with these schemes is their high memory demand during training, which results from the recurrent use of CNN blocks. In particular, the intermediate results at each unroll need to be stored for back-propagation. Several unrolled architectures were introduced for the recovery of breath-held ECG-gated cardiac cine MRI with factor representations [21], [23]–[26]. The gated approach allows the binning of data from multiple cardiac cycles to a few

cardiac phase images (e.g. 20-30), which makes it suitable for memory hungry unrolled networks. The availability of pre-acquired fully-sampled data from multiple subjects, which is readily available in the cine setting, makes it possible to train the above algorithms in a supervised learning mode. Unfortunately, the memory demand and lack of fully sampled training data makes it challenging to apply it to the free-breathing and ungated imaging application, with at least an order higher number of image frames (e.g. 300 or more in our setting). While we have recently used an unrolled architecture for the free-breathing and ungated setting, we had to severely restrict the number of frames ( $\approx 100$ ) and had to use considerably simpler networks that severely restricted the performance [27]. Another challenge associated with these schemes is the lack of fully sampled training data for training these models, especially in the free-breathing and ungated mode.

The main focus of this work is to introduce a memory-efficient model for dynamic MRI called Deep Bi-Linear Unsupervised Representation (DEBLUR). This approach exploits the power of CNNs to further improve the performance, while retaining the memory efficiency of bilinear factor representations. We use the structural bias of CNNs [19] as priors on the factors; this approach is an alternative to current schemes that penalize energy [9], sparsity [12], or kernel priors [16], [22] on the factors. In particular, we assume that the spatial and temporal factors are generated by two CNN generators, whose parameters are estimated from the measured data. The CNN parameters are learned such that the k-space measurements of the bilinear representation matches the multi-channel measurements. The proposed scheme directly learns the compressed representation (factors) from the undersampled k-t space data of each specific subject. Because the full image data need not be stored or used in the computation, the proposed approach is more memory efficient in multidimensional applications compared to current CNN approaches that require the image frames to be stored [25]–[27]. However, the improved memory efficiency comes at the expense of compute time during inference, compared to current deep learning solutions. The parameters of the CNN and the latent vectors are learned during inference; the computational complexity of the proposed scheme is comparable to classical bilinear based methods [7]–[10].

This paper generalizes our earlier conference version [28] in multiple ways. First of all, we follow a completely unsupervised strategy; we pretrain the CNN bilinear generators from SToRM reconstruction of undersampled k-space data rather than using exemplar datasets. In addition, the generator of temporal basis functions is significantly different from [28]; the inputs to the generator are learned during the optimization in the current setting, which offers improved performance compared to keeping them fixed. More importantly, the approach is now validated with several datasets compared to the limited comparisons in [28], along with ablation studies to determine the impact of the regularization terms. We note that an unsupervised approach was recently introduced in [29] for free-breathing dynamic MRI data. Unlike the proposed scheme that uses deep image priors on the factors, the above approach

learns the parameters of the generator and assumes the latent vectors to be random. We compare the performance of the proposed scheme with [29].

## II. BACKGROUND

### A. Dynamic MRI: Problem Setup

We consider the recovery of an image time series  $\mathbf{x}_i; i = 1, \dots, N$  from their noisy undersampled measurements

$$\mathbf{b}_i = \mathcal{A}_i(\mathbf{x}_i) + \mathbf{n}_i. \quad (1)$$

Here,  $\mathcal{A}_i$  is the undersampled multi-channel MRI measurement operator for the  $i^{\text{th}}$  frame and  $\mathbf{n}_i$  is the measurement noise. The undersampling pattern varies from frame to frame. It is a common practice to arrange the frames in a multidimensional Casorati matrix

$$\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N], \quad (2)$$

where  $\mathbf{x}_i$  denotes the vectorized version of the  $i^{\text{th}}$  frame in the dataset.

### B. Bilinear Factor Models for Dynamic MRI

Bilinear factor models are widely used in multidimensional applications, including dynamic MRI [7], [8], parameter mapping [30], [31], and MR spectroscopic imaging [32], [33]. These approaches represent the Casorati matrix in (2) in terms of the factors  $\mathbf{U}$  and  $\mathbf{V}$ :

$$\mathbf{X} = \underbrace{[\mathbf{u}_1, \dots, \mathbf{u}_r]}_{\mathbf{U}} \underbrace{[\mathbf{v}_1, \dots, \mathbf{v}_r]}_{\mathbf{V}^T}^T. \quad (3)$$

The columns of  $\mathbf{U}$  denoted by  $\mathbf{u}_i$  are identified as the spatial basis functions, while those of  $\mathbf{V}$  are the temporal basis functions. The main benefit in the above representation is the significant reduction in the number of free parameters that need to be estimated, which translates to reduced data demand [27]. In the context of large multidimensional applications, another key benefit is the memory and computational benefits. In particular, the measurements can be expressed completely in terms of  $\mathbf{U}$  and  $\mathbf{V}$  as:

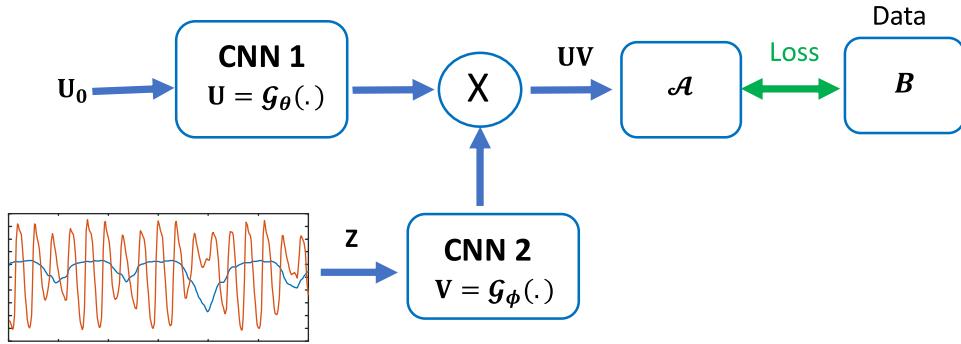
$$\mathcal{A}_i(\mathbf{x}_i) = \mathbf{A}_i(\mathbf{U})\mathbf{v}_i \quad (4)$$

This approach eliminates the need for computing and storing the image frames  $\mathbf{x}_i$  themselves during image reconstruction; post-recovery, the desired frame can be retrieved as  $\mathbf{x}_i = \mathbf{U}\mathbf{v}_i$ .

Many schemes [7], [8] estimate the temporal basis functions  $\mathbf{V}$  from k-space navigators, followed by the estimation of the spatial factors  $\mathbf{U}$  in (3). By contrast, the joint estimation of  $\mathbf{U}$  and  $\mathbf{V}$  from the measured data offers improved performance and reduces the need for specialized acquisition schemes with k-space navigators. The joint approaches pose the recovery of the signals from the undersampled measurements  $\mathcal{A}(\mathbf{X})$  as:

$$\begin{aligned} \{\mathbf{U}^*, \mathbf{V}^*\} = \arg \min_{\mathbf{U}, \mathbf{V}} & \left\| \mathcal{A}(\mathbf{U}\mathbf{V}^T) - \mathbf{B} \right\|^2 \\ & + \lambda_1 \mathcal{R}_1(\mathbf{U}) + \lambda_2 \mathcal{R}_2(\mathbf{V}) \end{aligned} \quad (5)$$

Here,  $\mathbf{B}$  denotes the measured data. We note that the representation in (3) is linear in  $\mathbf{U}$  and  $\mathbf{V}$  independently; the joint



**Fig. 1.** Outline of the DEBLUR model. Two CNN generative networks  $\mathcal{G}_\theta$  and  $\mathcal{G}_\phi$  are used to generate the spatial ( $\mathbf{U}$ ) and temporal  $\mathbf{V}$  factors, respectively. The Casorati matrix of the dataset is modeled as  $\mathbf{X} = \mathbf{UV}^T$ , where the columns of  $\mathbf{X}$  correspond to the temporal frames. Rather than using random inputs as in [19], the  $\mathcal{G}_\theta$  generator is fed with the  $\mathbf{U}$  factor matrix from SToRM, denoted by  $\mathbf{U}_0$ . The input to  $\mathcal{G}_\phi$  network are the latent vectors  $\mathbf{Z}$ , which are also learned from the data. We expect the joint learning procedure to result in interpretable latent vectors, which will capture the temporal motion components (e.g., cardiac and respiratory motion).

optimization in (5) can be viewed as a bilinear optimization. Here,  $\mathcal{R}_1$  and  $\mathcal{R}_2$  are regularization functionals. Depending on the specific form of the regularization functions, one would obtain different flavors of reconstruction algorithms.

- 1) Low-rank regularization [9]: Here, one would choose  $\mathcal{R}_1(\mathbf{U}) = \|\mathbf{U}\|^2$  and  $\mathcal{R}_2(\mathbf{V}) = \|\mathbf{V}\|^2$ .
- 2) Blind compressed sensing: Here, one would choose  $\mathcal{R}_1(\mathbf{U}) = \|\mathbf{U}\|_{\ell_1}$  and  $\mathcal{R}_2(\mathbf{V}) = \|\mathbf{V}\|^2$ .
- 3) Smoothness regularization on manifolds (SToRM): The SToRM scheme also relies on a factorization as in (3), where  $\mathcal{R}_1(\mathbf{U}) = \sum \sigma_i \|\mathbf{u}_i\|_i^2$  and  $\mathbf{V}$  is obtained as the eigenvectors of the graph Laplacian matrix of the graph of the data. Both calibrated [34] and uncalibrated [17] formulations are available.

The performance of the above methods critically depends on the specific choice of the priors  $\mathcal{R}_1$  and  $\mathcal{R}_2$  to estimate  $\mathbf{U}$  and  $\mathbf{V}$ .

### C. Deep Image Prior (DIP)

The deep image prior approach has been introduced in inverse problems to exploit the structural bias of CNNs to natural image structure rather than noise [19]. The regularized reconstruction of a static image from undersampled and noisy measurements is posed as

$$\{\boldsymbol{\theta}^*\} = \arg \min_{\boldsymbol{\theta}} \|\mathcal{A}(\mathbf{x}) - \mathbf{b}\|^2 \quad \text{such that } \mathbf{x} = \mathcal{G}_{\boldsymbol{\theta}}(\mathbf{z}) \quad (6)$$

where  $\mathbf{x} = \mathcal{G}_{\boldsymbol{\theta}^*}(\mathbf{z})$  is the recovered image, generated by the CNN generator  $\mathcal{G}_{\boldsymbol{\theta}^*}$  whose parameters are denoted by  $\boldsymbol{\theta}^*$ .

The constraint that the image is generated by a CNN provides implicit regularization, which facilitates the recovery of  $\mathbf{x}$  in challenging inverse problems. Here,  $\mathbf{z}$  is a random latent variable, which may or may not be optimized. The structural bias of untrained CNN generators towards natural images is exploited to yield good recovery. The above problem is often solved using stochastic gradient descent (SGD), which is often terminated early to obtain regularized recovery. Few iterations of the above model are observed to represent images reasonably well, while the model will learn the noise in the measurements, resulting in poor reconstructions with more

iterations, provided the generator has sufficient capacity. Early termination is often used to avoid this and thus regularize the recovery. Alternate approaches including alternatives to SGD have been introduced to avoid the early stopping strategies.

### III. DEEP BI-LINEAR REPRESENTATION (DEBLUR)

The main focus of this work is to develop a multidimensional image reconstruction algorithm that inherits the memory efficiency of bilinear models; we express the data as  $\mathbf{X} = \mathbf{UV}^T$  as in (3). Unlike current bilinear models that use energy and sparsity priors, we use CNN priors on the factors. In particular, we propose to represent the factors as

$$\mathbf{U} = \mathcal{G}_\theta(\mathbf{U}_0) \quad (7)$$

$$\mathbf{V} = \mathcal{G}_\phi(\mathbf{Z}). \quad (8)$$

Here,  $\mathcal{G}_\theta$  and  $\mathcal{G}_\phi$  are CNN generators whose weights are denoted by  $\theta$  and  $\phi$ , respectively. Here,  $\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_N]$  are time-dependent latent vectors that are also learned from the measured k-space data. Here,  $\mathbf{U}_0$  is an initial approximate reconstruction that is often obtained using a simple algorithm, which is refined by the generator. See Fig. 1 for details. We observe that feeding an initial reconstruction of the  $\mathbf{U}$  factors to  $\mathcal{G}_\theta$  offers faster reconstructions compared to the DIP strategy of feeding noise to the image generator. Similar to DIP [19], we expect to capitalize on the structural bias of the CNNs towards smooth natural images.

We propose to recover the factors by solving the optimization problem:

$$\{\mathbf{U}^*, \mathbf{V}^*\} = \min_{\theta, \phi, \mathbf{Z}} \|\mathcal{A}(\mathbf{UV}^T) - \mathbf{B}\|^2, \\ \text{such that } \mathbf{U} = \mathcal{G}_\theta[\mathbf{U}_0]; \quad \mathbf{V} = \mathcal{G}_\phi[\mathbf{Z}], \quad (9)$$

which is solved using SGD. Note that we also optimize for the latent vectors  $\mathbf{Z}$ .

Note that the recovery scheme in (9) recovers the factors  $\mathbf{U}$  and  $\mathbf{V}$  directly from the multi-channel measurements  $\mathbf{B}$ . Once  $\mathbf{U}$  and  $\mathbf{V}$  are recovered, the  $i^{\text{th}}$  frame of the time series can be generated as

$$\mathbf{x}_i = \mathbf{U} \underbrace{\mathcal{G}_\phi[\mathbf{z}_i]}_{(10)}$$

Similarly, one can obtain the measurements of the  $i^{\text{th}}$  frame as  $\mathbf{b}_i = \mathcal{A}(\mathbf{U})\mathbf{v}_i$  during training, without having to evaluate the frames. This approach significantly reduces the memory demand and computational complexity of the algorithm, especially in applications involving multidimensional time series. As an example, in our case with 300 frames, we choose the number of columns in  $\mathbf{U}$  and  $\mathbf{V}$  as 30. Thus, our setting only requires the storage of 30 images and 30 1-D vectors. By contrast, unrolled approaches [23], [24], [27] require the storage of the entire time series and their intermediate copies at each unroll. With 10 unrolls, this would translate to the storage of 3000 images, which is an order of 100 times larger. Unfortunately, these methods are not directly applicable in our setting.

#### A. Unsupervised Pretraining of the Generators

The generators in DIP are usually initialized with random weights. Unlike past convex strategies, the CNN-based algorithm in (9) is not guaranteed to converge to the global minimum of the cost function. The final solution will be dependent on the initialization. To improve performance and to reduce the computational complexity, we propose to initialize the  $\mathcal{G}_\theta$  and  $\mathcal{G}_\phi$  networks.

In this work, we chose the SToRM [15] derived basis vectors to initialize the network. In particular, SToRM performs kernel low-rank regularization of the k-space navigators to derive the temporal basis functions  $\mathbf{V}_{\text{SToRM}}$ . Once the temporal basis functions are available, one can perform a fast and approximate reconstruction of the SToRM factor  $\mathbf{U}_0 = \mathcal{A}^H(\mathbf{D}\mathbf{B}\mathbf{V}^H)$ , where  $\mathbf{D}$  is a diagonal matrix corresponding to density compensation. Note that this approximation is not as computationally expensive (around 20 seconds) compared to the SToRM algorithm [15], which involves a conjugate gradients iteration. We note that reconstructions from other approaches (e.g. low-rank and time-DIP [29]) can be used to also initialize the DEBLUR scheme. We use these fast estimate to pre-train  $\mathcal{G}_\theta$  and  $\mathcal{G}_\phi$  generators independently as

$$\theta_0 = \arg \min_{\theta} \|\mathcal{G}_\theta(\mathbf{U}_0) - \mathbf{U}_0\|^2, \quad (11)$$

$$\{\phi_0, \mathbf{Z}_0\} = \arg \min_{\phi, \mathbf{Z}} \|\mathcal{G}_\phi(\mathbf{Z}) - \mathbf{V}_{\text{SToRM}}\|^2. \quad (12)$$

These initial guesses of the network weights ( $\theta, \phi$ ) and latent vectors  $\mathbf{Z}$  are used as initialization in (9). We study the impact of the initialization on the algorithms in the results section. The input latent vectors  $\mathbf{Z}_0$  was initialized randomly during the pre-training phase.

#### B. Regularization Penalties

The DIP approach, as well as our extension to the dynamic setting in (9), is vulnerable to noise overfitting. In particular, if the generator networks have sufficient capacity, they will learn the noise in the measurements when the number of iterations is large [19]. To further improve the robustness of the algorithm to noise, we propose to additionally penalize the  $\ell_1$  norm of the network weights. The regularized recovery is

posed as

$$\begin{aligned} & \{\mathbf{U}^*, \mathbf{V}^*\} \\ &= \min_{\theta, \phi, \mathbf{Z}} \left\| \mathcal{A}(\mathbf{U}\mathbf{V}^T) - \mathbf{B} \right\|^2 \\ &+ \underbrace{\lambda_1 \|\theta\|_{\ell_1} + \lambda_2 \|\phi\|_{\ell_1}}_{\text{network regularization}} + \underbrace{\lambda_3 \|\nabla_t \mathbf{Z}\|_{\ell_1}}_{\text{temporal regularization}}, \\ & \text{such that } \mathbf{U} = \mathcal{G}_\theta[\mathbf{U}_0]; \quad \mathbf{V} = \mathcal{G}_\phi[\mathbf{Z}] \end{aligned} \quad (13)$$

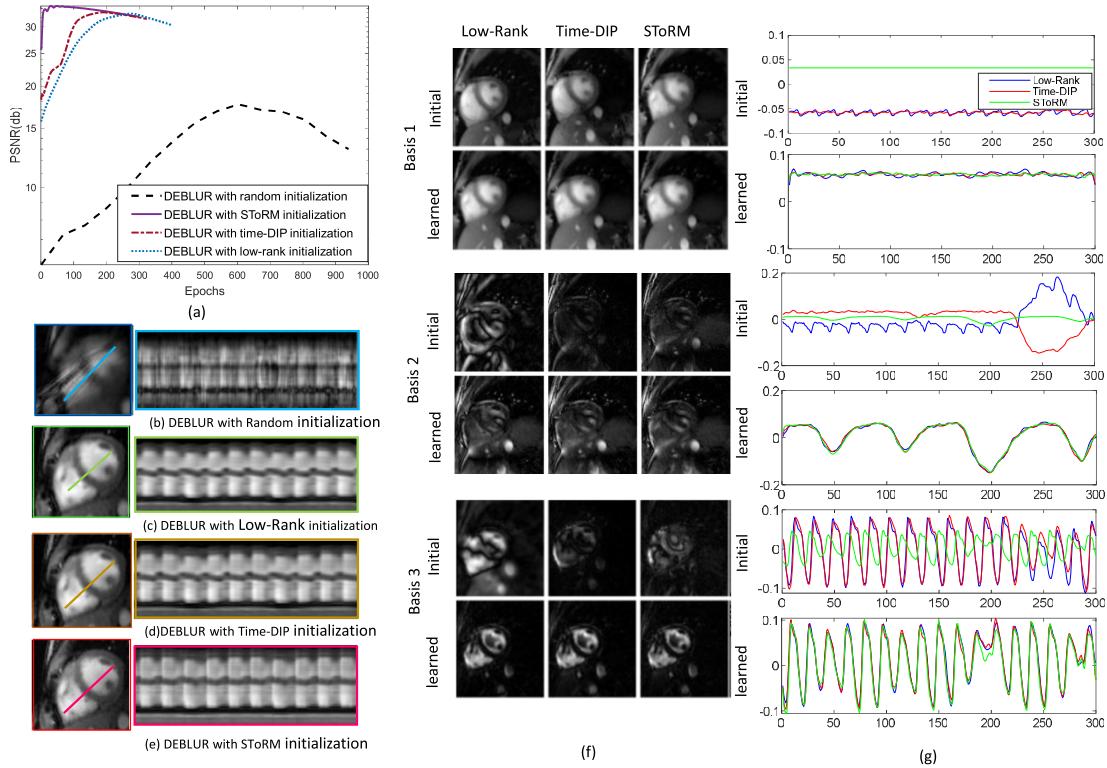
We note that the above optimization is performed for each dataset. We expect the weight-regularization strategy to learn networks with fewer non-zero weights, which translates to smaller capacity; this approach is expected to reduce the vulnerability of the network to overfitting. The sparsity of the weights of  $\mathcal{G}_\theta$  promotes the learning of local spatial factors similar to [10], which is more efficient than global low-rank minimization. We note that the use of sparsity regularization is qualitatively equivalent to drop-out approaches, which are widely used in CNN training instead of explicit regularization. Unlike the explicit definition of the blocks in [10], the definition of the local neighborhoods in the proposed scheme is more data-driven. The images in the time series often vary smoothly with time in dynamic imaging applications; the norm of the temporal derivatives of the images is a widely used prior. Our goal is to directly recover the compressed representation from the data without using the actual images. We note from (10) that the  $i^{\text{th}}$  image in the time series is dependent on the latent vector  $\mathbf{z}_i$ . The last term in (13) is the norm of the temporal gradients of  $\mathbf{Z}$ , which will encourage the temporal smoothness of the time series. The algorithm in (13) is also initialized with the pretraining strategy discussed in the previous section. The impact of the regularization parameters and their ability to minimize overfitting issues are studied in the results section.

## IV. IMPLEMENTATION DETAILS

#### A. Data Acquisition and Post-Processing

The experimental data was obtained using the FLASH sequence on a Siemens 1.5T scanner with 34 coil elements total (body and spine coil arrays) in the free-breathing and ungated mode from cardiac MRI patients(five patients with 10-14 slices) with a scan time of 42 seconds per slice. The patients with cardiac abnormalities were recruited from those who were referred for routine clinical examinations. The protocol was approved by the Institutional Review Board (IRB) at the University of Iowa.

**1) Pulse Sequence Details:** All datasets were acquired using a 34-channel cardiac array on Siemens Aera scanners at the University of Iowa. We used a radial GRE sequence with the following parameters: TR/TE 4.68/2.1 ms, FOV 300 mm, base resolution 256, slice thickness 8 mm. A temporal resolution of 46.8 ms was obtained by sampling 10 k-space spokes per frame. Each temporal frame was sampled by two k-space navigator spokes (out of 10 spokes/frame), oriented at 0 degrees and 90 degrees, respectively. The remaining spokes were chosen with a golden-angle view ordering. The scan parameters were kept the same across all patients. The subjects were asked to breathe freely, and the data was acquired in an ungated



**Fig. 2.** Impact of initialization of basis functions on image quality: (a) shows the SER curves of the DEBLUR scheme with random initialization, low-rank initialization, Time-DIP initialization and SToRM initialization. Their corresponding peak values images and time profiles are shown in (b)-(e). Three spatial basis vectors of each methods are shown in (f), both initial and learned basis. Their corresponding temporal basis vectors are shown in (g). These results show that the proposed method can offer qualitatively equivalent results when initialized using any of the existing algorithms. In particular, the spatial and temporal basis functions converge to the same ones, irrespective of the initialization. By contrast, the approach converges to a poor local minimum when initialized with random weights.

fashion. The complete data acquisition lasted 42 seconds for each slice. We extracted the data corresponding to the first 14s of the 42 second data from each slice to demonstrate the ability of the algorithm to reduce the data demand; the remaining 28s of data is discarded in our approach. By contrast, the classical SToRM algorithm requires the data from 42 second acquisitions to offer good recovery. We note that a cardiac cycle is less than 1 second in duration; 14 second acquisition often has 15-20 cardiac cycles and several respiration cycles.

**2) Coil Selection and Compression:** To improve the image reconstruction quality, we excluded the coils with low sensitivities in the region/slice of interest. We used an automatic coil selection algorithm to pick the five best coil images, which provided the best signal-to-noise ratio (SNR) in the heart region. Our experiments (not included in this paper) show that this coil combination has minimal impact on image quality. The main motivation for the combination was to reduce the memory requirement so that it fit on our GPU device, which significantly reduced the computational complexity. All the results were generated using a single node of the high-performance Argon Cluster at the University of Iowa, equipped with Titan V 32GB of memory.

**3) Performance Metrics:** We used the following four quantitative metrics to compare our method against the existing schemes:

- Signal to error ratio (SER):

$$\text{SER} = 20 \log_{10} \frac{\|\mathbf{x}_{\text{orig}}\|_2}{\|\mathbf{x}_{\text{orig}} - \mathbf{x}_{\text{rec}}\|_2}, \quad (14)$$

where  $\|\cdot\|_2$  denotes the  $\ell_2$  norm, and  $\mathbf{x}_{\text{orig}}$  and  $\mathbf{x}_{\text{rec}}$  denote the original and the reconstructed images, respectively.

- Peak Signal to Noise Ratio (PSNR):

$$\text{PSNR} = 20 \log_{10} \frac{\max\{\mathbf{x}_{\text{orig}}\}}{\|\mathbf{x}_{\text{orig}} - \mathbf{x}_{\text{rec}}\|_2}. \quad (15)$$

- Normalized High Frequency Error (HFEN) [35]: This measures the quality of fine features, edges, and spatial blurring in the images and is defined as:

$$\text{HFEN} = 20 \log_{10} \frac{\|\text{LoG}(\mathbf{x}_{\text{orig}}) - \text{LoG}(\mathbf{x}_{\text{rec}})\|_2}{\|\text{LoG}(\mathbf{x}_{\text{orig}})\|_2}, \quad (16)$$

where LoG is a Laplacian of the Gaussian filter that captures edges. We use the same filter specifications as did Ravishankar *et al.* [35]: kernel size of  $15 \times 15$  pixels, with a standard deviation of 1.5.

- The Structural Similarity index (SSIM) is a perceptual metric introduced in [36], whose implementation is publicly available. We used the default contrast values, Gaussian kernel size of  $11 \times 11$  pixels with a standard deviation of 1.5 pixels.

- BRISQUE is a referenceless measure of image quality, where a smaller score indicates better perceptual quality. BRISQUE estimates the score using a support vector regression (SVR) model with the help of an image database with distorted and undistorted images to derive the differential mean opinion score values. The distortions include compression artifacts, blurring, and additive noise [37].

Higher values of the above-mentioned performance metrics correspond to better reconstruction, except for the HFEN, where a lower value is better.

**4) State-of-the-Art Algorithms for Comparison:** We compare the proposed scheme against the following algorithms:

- SToRM [15]: The manifold Laplacian is estimated from the self-gating navigators acquired in k-space. Once the Laplacian matrix is obtained from navigators, the high-resolution images are recovered using kernel low-rank-based framework. The SToRM basis functions were estimated from navigator data with 70 iterations. Following the estimation of the navigators, the algorithm was run till convergence with a minimum of 60 iterations, as described in [15].
- Low-Rank [7], [9]: The image time series is recovered by nuclear norm minimization. The nuclear norm minimization approach models the images as points living on a subspace.

### B. Architecture of the Generators

We refer to  $\mathcal{G}_\theta$  as the spatial generator. We use a four-layer CNN with ReLU activation. The number of channels of the input and output is equal to twice the number of basis functions, to account for the real and imaginary parts of the basis. In our work, we use 30 basis functions, and hence the number of input and output channels is 60. We refer to  $\mathcal{G}_\phi$  as the temporal generator, where we use a four-layer CNN with ReLU activation. The inputs to the temporal generator are the latent vectors  $\mathbf{Z} \in \mathbb{R}^{d \times n_f}$ , where  $d$  represents the latent dimension and  $n_f$  denotes the number of frames in the dynamic dataset. In our work, we observe that two-dimensional latent vectors are sufficient to obtain good reconstructions. The outputs of the temporal generator are the temporal basis functions  $\mathbf{V}$  of dimension  $r \times n_f$ , where  $r$  is the number of temporal basis functions in (3).

## V. EXPERIMENTS AND RESULTS

We describe the experiments and our observations in this section.

### A. Impact of Pretraining

To demonstrate the benefit of pre-training, we compare the algorithms described in (9) with random initialization and with the initialization scheme denoted in (11) & (12). Fig 2.(a) shows the plot of SER values with respect to the number of epochs. The color of the borders of the images in Fig 2.(b)-(e) correspond to the color of the markers in the plots, which denotes the initial and maximum SER values. We note that

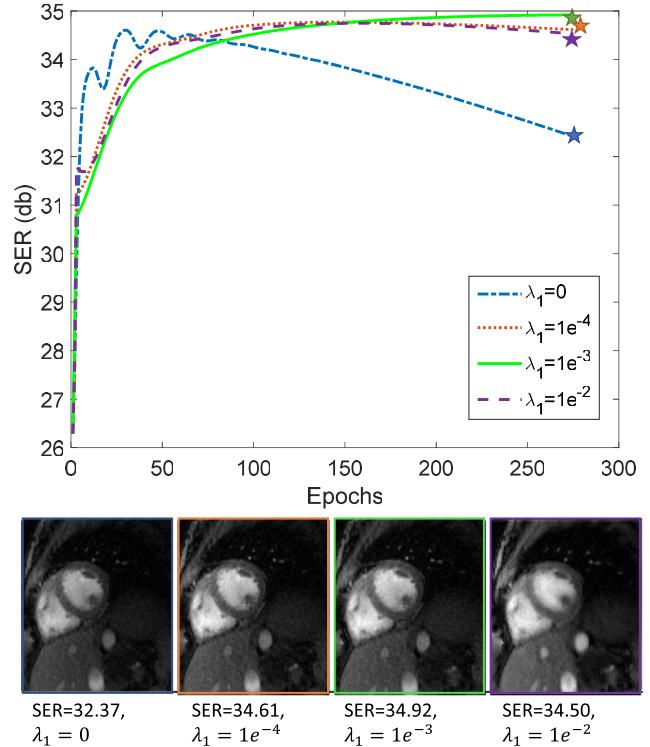


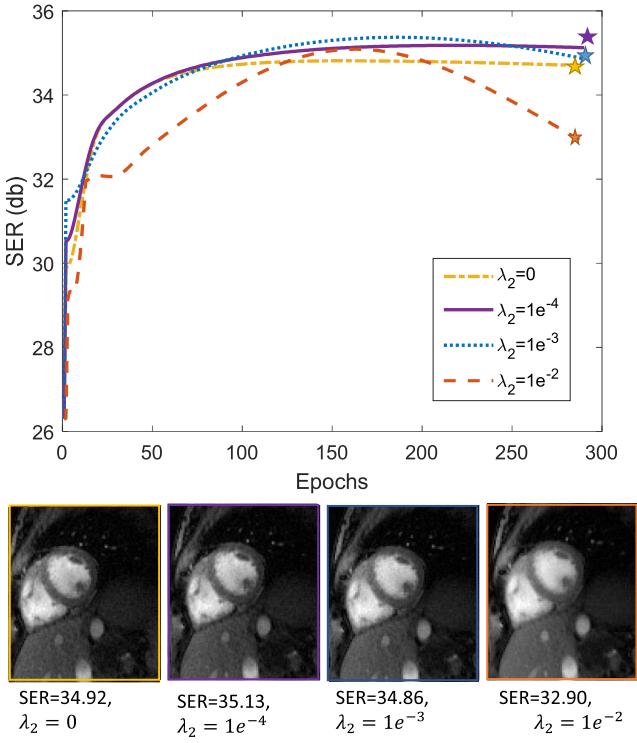
Fig. 3. shows the SER curves with different  $\lambda_1$  values in DEBLUR method in (13), which regularizes the generator  $\mathcal{G}_\theta$  used to generate  $\mathbf{U}$ . The un-regularized setting is denoted by the blue curve, which is the zoomed version of the red-curve in Fig. 2. We note that higher regularization parameters control the decay of SER with iterations, which indicates improved generalization of the network to unseen k-space samples. We note that the best performance is achieved with  $\lambda_1 = 10^{-3}$ . Larger regularization parameters (e.g.,  $\lambda = 10^{-2}$  denoted by the red curve) translate to slight oversmoothing of the spatial factors.

the initialization with random weights starts with the low SER as seen from Fig. 2.(a), compared to the other initializations. As seen from the plots in (a), the DEBLUR approach with low-rank, SToRM, and time-DIP converged rapidly to a peak value, while the one with random initialization converged to a poor solution with significantly more iterations. While SToRM initialization results in a slightly higher peak SNR, we note that the reconstructions shown in (c)-(e) are qualitatively equivalent. The plot of the  $\mathbf{U}$  and  $\mathbf{V}$  basis functions shown in (f) and (g) shows that the DEBLUR approach yielded similar final basis functions, irrespective of the specific initialization. We note that the computational complexity of the approximate low-rank and SToRM reconstructions are very low, compared to the time-DIP approach. We consider the initialization by SToRM basis vectors for the rest of the experiments considered in this work.

### B. Impact of Regularization Parameters

We study the impact of the regularization priors in (13) in Figures 3, 4 and 5.

**1) Regularization of  $\mathbf{U}$  Network Parameters:** We initialize the network with the SToRM initialization, described by (7) & (8). We set  $\lambda_2 = \lambda_3 = 0$  and consider four different  $\lambda_1$  values in this study. We note from Fig. 3 that the network with

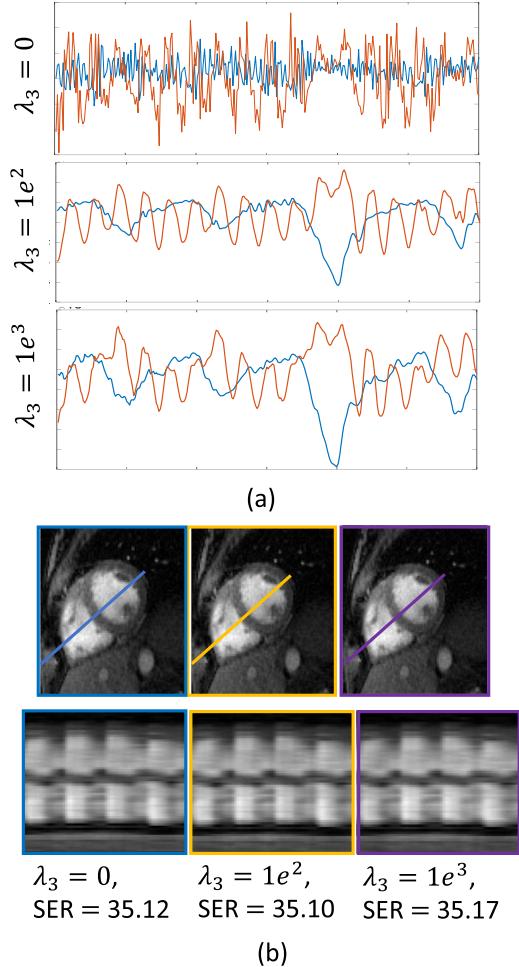


**Fig. 4.** shows the SER curves with different  $\lambda_2$  values in DEBLUR method.  $\lambda_2$  regularizes  $\mathbf{V}$  network parameters. Lower  $\lambda_2$  value allows the network to learn noisy temporal information in the data as illustrated by the yellow curve. A higher value of  $\lambda_2$  oversmooths the temporal basis as depicted by the orange color. Empirical findings show  $\lambda_2 = 1e^{-4}$  gives better image SER as compared to higher and lower values. Secondly, the performance does not deteriorate if it runs for more epochs, as shown by the purple color.

$\lambda_1 = 0$  results in the performance peaking after a few iterations, similar to the case in Fig. 2. With more iterations, the SER drops because of the overfitting to noise. By contrast, we observe that  $\lambda_1 = 10^{-3}$  results in saturating performance, which is around 2.5 dB superior to the peak performance obtained from  $\lambda_1 = 0$ . We also observe that  $\lambda_1$  values that are slightly higher and lower than the optimal values result in somewhat similar performance, which indicates that the network is not very sensitive on the optimal values.

**2) Regularization of  $\mathbf{V}$  Network Parameters:** In Fig. 4, we study the impact of  $\lambda_2$  on the results. We keep the best  $\lambda_1 = 10^{-3}$  value from Fig. 3 and set  $\lambda_3 = 0$ . We vary  $\lambda_2$  in this study and plot the change in SER with iterations. We note that  $\lambda_2 = 10^{-4}$  offers the best final performance, resulting in around 0.2 dB improvement in performance compared to  $\lambda_2 = 0$ . We note that the joint optimization of  $\mathbf{V}$  and  $\mathbf{U}$  networks results in around 8-9 dB improvement in performance over the SToRM initialization. The networks learned by DEBLUR result in a bilinear representation that is more optimal in representing the data when compared to the classical bilinear methods.

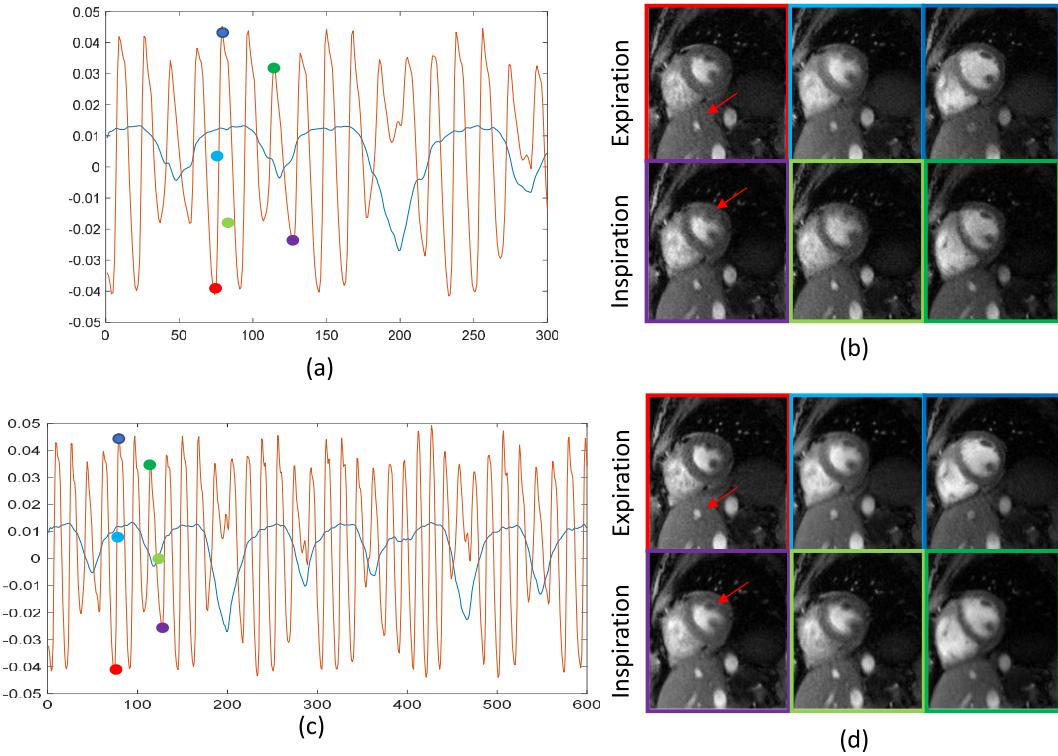
**3) Regularization of Latent Vectors  $\mathbf{Z}$ :** In Fig. 5 we study the impact of  $\lambda_3$  on the results. We keep  $\lambda_1 = 10^{-3}$  and  $\lambda_2 = 10^{-4}$ , which were the best values that we determined in the previous subsections. We consider different values of  $\lambda_3$  and plot the corresponding latent vectors. We observe that



**Fig. 5.** shows the latent vectors with different  $\lambda_3$  values in the DEBLUR method.  $\lambda_3$  applies temporal smoothness on latent vectors to achieve meaningful cardiac and respiratory motion. Lower  $\lambda_3$  gives noisy latent vectors as illustrated by  $\lambda_3 = 0$ . A higher value of  $\lambda_3$  oversmooths the latent vectors as shown by  $\lambda_3 = 1e^3$ . Empirical findings show  $\lambda_3 = 1e^2$  gives better separation of latent vectors that represent cardiac and respiratory motion. Secondly, the SER changes between the different values of  $\lambda_3$  are marginal. If no regularization is added, the latent vectors will capture alias artifacts resulting from under-sampling. Since cardiac and respiratory motion induced changes will be on a slower time-scale, we impose a temporal regularization to encourage the latent vectors from learning physiologically relevant parameters rather than noise. In our current work, the main focus is on accurate recovery of the images rather than disentangling the different latent variables.

the latent vector regularization had marginal impact on the SER. We note that the optimal value is  $\lambda_3 = 1e^2$ . We observe that  $\lambda_3 = 0$  resulted in noisy latent vectors, while the one with  $\lambda_3 = 1e^2$  had good reconstructions.

**4) Benefits of Using Latent Vectors to Generate Temporal Basis:** We note from Fig 5 and Fig. 6 that the  $\lambda_3$  parameter is not very sensitive to image quality. The  $\mathbf{V}$  basis functions are derived as a non-linear function of  $\mathbf{Z}$ . When  $\lambda_3 \rightarrow 0$ ,  $\mathcal{G}_\phi$  learns to generate the basis functions from noise-like latent vectors. We note that cardiac and respiratory motion is expected to be smooth. With smoothness regularization (increasing  $\lambda_3$ ), we note that the latent vectors capture physiologically relevant parameters. In the datasets we considered, we note that the latent vectors separate into a fast-changing latent vector that



**Fig. 6.** Comparison of the DEBLUR method using 14 s and 28 s of acquired data. **Fig 6(a)** shows the two latent vectors of 14 s data, where the blue vector represents respiratory motion and the red vector represents cardiac motion in the data. Images at different time points, indicated by color dots, are shown in **Fig 6(b)**. It also shows the ability of our latent vector-based approach to capture images at different time instants in a series. For example, the red dot in **Fig 6(a)** captures the image at the systole phase and at the expiration state. We have also compared 14 s data with 28 s data to show the improvement in the performance of the DEBLUR method. Differences in image quality with less and more data are subtle in the diastole phase but become prominent in the systole phase, as indicated by the red arrows.

captures cardiac motion and a slow one that captures respiratory motion. Post-reconstruction, the latent variables may be used to bin the images into different cardiac and respiratory states. The top row corresponds to the latent vectors and reconstructions from 14 s of data. The red box (corresponding to the red lines in the latent vector plot in (a)) corresponds to the image frame in the systole phase and expiration state, while the blue box corresponds to an image in the diastole phase and inspiration phase. **Fig 6(c)** shows the latent vectors estimated from 28 s of data. The results show that similar decomposition of latent vectors can be obtained from more data, with moderate improvement in image quality.

### C. Comparison With State-of-the-Art Methods

We compare the proposed DEBLUR method with the SToRM and low-rank reconstructions. **Fig 7** shows the comparison between the DEBLUR and SToRM methods using five different datasets. We consider the recovery from 42 s and 14 s of data.

We observe that the DEBLUR(42s) visually offers similar or improved image quality when compared to SToRM(42s). However, we note that the SToRM approach results in significant blurring and degradation in image quality when only 14 s of data is available. Both SToRM algorithms were iterated until convergence as reported in [15]. We note that the SToRM approach reported in our earlier study relied on 42s of data [15]; the reduced data is the main reason for the blurring

**TABLE I**  
QUANTITATIVE COMPARISON OF THE DEBLUR METHOD  
WITH SOTA METHODS

Metric	Low-Rank	SToRM(14s)	DEBLUR(14s)
SER	$18.93 \pm 0.48$	$22.41 \pm 0.78$	$35.99 \pm 4.98$
SSIM	$0.38 \pm 0.02$	$0.59 \pm 0.04$	$0.96 \pm 0.04$
HFEN	$1.23 \pm 0.02$	$0.82 \pm 0.05$	$0.18 \pm 0.13$
Brisque	$40.82 \pm 2.1$	$32.88 \pm 4.3$	$26.83 \pm 5.81$

seen in the SToRM reconstructions. The DEBLUR reconstructions are observed to minimize the blurring, resulting in reconstructions that are comparable to the 42 s reconstructions. The improved performance may be attributed to the spatial and temporal regularization of the factors offered by DEBLUR.

Since ground truth is not available, we have used SToRM(42s) acquisition as reference data for SER and SSIM comparisons in **Table I**. We also compare the methods using the BRISQUE score in **Table I** and **II**, where we show the BRISQUE scores of the individual datasets.

In **Fig 8**, we compare the methods that recover the images from 14 s of data against SToRM reconstructions from 42 s. We have shown two frames (end of diastole and end of systole) from each method, along with error maps in **Fig 8(a)**. Time profiles are depicted in **Fig 8(b)**. We note that DEBLUR provides the best spatial and temporal quality and improved details, which are comparable to the SToRM reconstructions from 42 s.

TABLE II

PERFORMANCE COMPARISON OF THE DEBLUR METHOD USING THE BRISQUE SCORE ON MULTIPLE DATA. WE HAVE ALSO SHOWN THE BENEFIT OF USING MORE ACQUIRED DATA. CORRESPONDING IMAGES ARE SHOWN IN FIG 7

Method	Data1	Data2	Data3	Data4	Data5
DEBLUR(42s)	28.30 $\pm$ 3.7	30.33 $\pm$ 3.3	29.90 $\pm$ 3.8	28.11 $\pm$ 3.7	17.53 $\pm$ 5.0
SToRM(42s)	34.44 $\pm$ 3.9	26.0 $\pm$ 2.9	29.96 $\pm$ 4.3	34.0 $\pm$ 5.0	17.32 $\pm$ 5.4
DEBLUR(14s)	29.5 $\pm$ 5.3	25.94 $\pm$ 3.4	28.61 $\pm$ 4.1	32.87 $\pm$ 5.5	17.28 $\pm$ 5.2
SToRM(14s)	36.26 $\pm$ 0.9	32.63 $\pm$ 1.3	35.72 $\pm$ 3.5	34.57 $\pm$ 0.8	25.42 $\pm$ 2.3

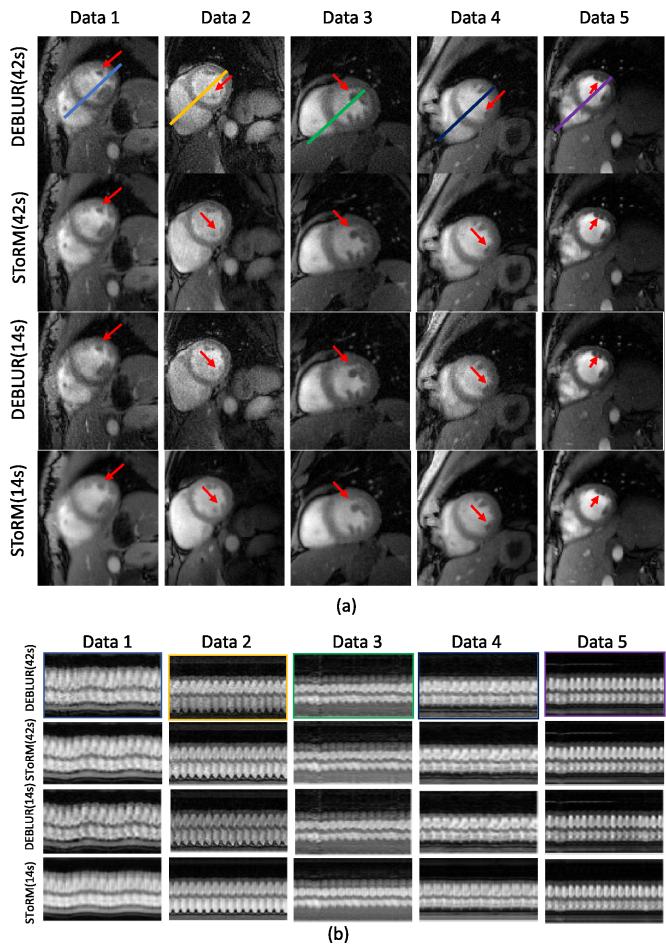


Fig. 7. Comparison of the DEBLUR and SToRM methods using 14 s and 42 s of acquired data. We have used five different datasets to compare the performance of the DEBLUR and SToRM methods. With less data (14 s), images have less sharpness than they do with 42 s of data. However, DEBLUR(14s) gives better image quality than the SToRM(14s) method. DEBLUR(42s) achieves better image contrast with enhanced features, as depicted by red arrows in the Figure.

## VI. DISCUSSION

The experiment in Section V-A clearly shows the benefit of initializing the network parameters by (11) and (12), respectively. As seen from Fig. 2, the optimization process is able to offer around 8-9 dB improvement in performance over the SToRM initialization. We note that the proposed framework of recovering the bilinear representation from

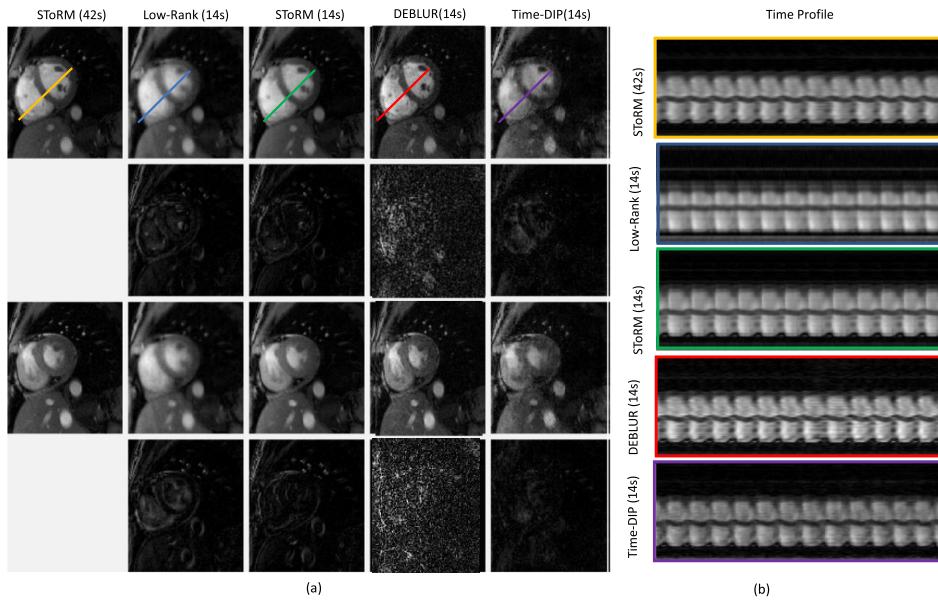
10 spokes/frame is significantly more challenging than the traditional DIP strategy in [19]. A reasonable initialization of the network can offer improved performance compared to random initialization. Fig. 2 also shows the need for early stopping of the unregularized setting in (9). In particular, the performance of the algorithm decays with increasing iterations, indicating overfitting to the noise in the k-space measurements.

The experiments in Section V-B show the benefit of regularizing the generator parameters while fitting to undersampled data. We observe from Fig. 3 & 4 that regularizing the network parameters improves the generalization performance of the network to unobserved k-space samples. Specifically, the network parameters are learned from few measured k-space data. The regularization of these networks reduces the impact of overfitting, thus improving the degradation in performance with iterations.

In this work, we have only used smoothness regularization on the latent vectors  $\mathbf{Z}$ . With smoothness regularization, the latent vectors are observed to learn the temporal variations seen in the data, including cardiac and respiratory motion. Interpretable latent vectors can aid in visualization of the results, as showcased in Fig. 6. In particular, the slow-changing latent vector captures the respiratory motion, while the fast latent vector captures the cardiac motion. The data can be sorted into the respective phases based on the latent vectors. We note that disentangling the latent variables is an active research area [38], [39]. The use KL divergence priors to encourage the latent vectors to follow specific probability distributions may offer improved disentanglement of the latent variables. For instance, this prior may reduce the slight variations in the cardiac latent vectors resulting from respiratory motion seen in Fig. 7. This will be a focus of our future work; the main focus of this work is on obtaining good image quality rather than deriving interpretable representations.

The comparison of the proposed scheme with the classical approaches shows that the proposed DEBLUR(42s) can offer comparable performance to SToRM(42s), while the proposed scheme from 14 s of data significantly outperforms the SToRM(14s) approach.

The main challenge associated with the proposed scheme is the high computational complexity of the algorithm during inference. Unlike current supervised deep learning strategies that do not perform any training during inference, the network parameters  $\theta$  and  $\phi$  are estimated during inference. The computational complexity of the algorithm is hence more comparable to traditional factor based methods that estimate the factors using iterative optimization for each dataset. Low-rank method takes 58 minutes on CPU, on GPU, Storm takes 12 minutes and DEBLUR takes roughly 10 minutes. While this approach is ideally suited for applications with limited datasets for supervised training, there is currently no systematic way for learning prior information from other datasets. In the future, we plan to address the unsupervised learning of the network from the measured k-t space data of multiple datasets as in [40]. If successful, this strategy eliminates the need for fully sampled data and is expected to considerably reduce the reconstruction time. The network



**Fig. 8.** Comparison of the DEBLUR(14s) method with state-of-the-art methods. Since ground truth is not available, we have used SToRM(42s) as ground truth. (a) To compare the image quality spatially, we have shown two frames (diastole and systole) from each method, along with their error maps. (b) Time profiles are shown.

parameters may be fine-tuned using the data of each subject, as considered in this work, to further improve performance.

## VII. CONCLUSION

We introduced a deep unsupervised bilinear algorithm to reconstruct dynamic MRI from undersampled measurements. We represented the spatial and temporal factors using two CNN-based generators, which are learned from the undersampled k-space data of each subject. The initialization of the network weights using an existing bilinear model (SToRM) is observed to both reduce the run-time as well as offer improved performance compared to initialization by random weights. The weights of the networks are regularized with  $l_1$  penalty to minimize the overfitting of the network to the noise in the measurements. The weight regularization is observed to minimize the degradation in performance with iterations. The implicit and learned regularization offered by the proposed scheme offers improved image quality compared to current methods, especially while recovering the data from shorter acquisition strategies. The proposed scheme directly learns a compressed image representation from the measured data, making it considerably more memory efficient than current approaches. The high memory efficiency makes it readily applicable to large-scale dynamic imaging (e.g., 3D+time) applications. In addition, the unsupervised strategy also eliminates the need for fully sampled training data, which is often not available in large-scale imaging problems.

## VIII. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted using human subject data. The Institutional Review Board at the local institution approved the acquisition of the data, and written consent was obtained from the subject.

## ACKNOWLEDGMENT

The authors claim that there is no conflict of interest.

## REFERENCES

- [1] S. B. Gay, C. L. Sistrom, C. A. Holder, and P. M. Suratt, "Breath-holding capability of adults: Implications for spiral computed tomography, fast-acquisition magnetic resonance imaging, and angiography," *Investigative Radiol.*, vol. 29, no. 9, pp. 848–851, Sep. 1994.
- [2] Z.-P. Liang, H. Jiang, C. P. Hess, and P. C. Lauterbur, "Dynamic imaging by model estimation," *Int. J. Imag. Syst. Technol.*, vol. 8, no. 6, pp. 551–557, 1997.
- [3] B. Sharif, J. A. Derbyshire, A. Z. Faranesh, and Y. Bresler, "Patient-adaptive reconstruction and acquisition in dynamic imaging with sensitivity encoding (PARADISE)," *Magn. Reson. Med.*, vol. 64, no. 2, pp. 501–513, 2010.
- [4] 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.*, vol. 50, no. 5, pp. 1031–1042, 2003.
- [5] M. Lustig, J. M. Santos, D. L. Donoho, and J. M. Pauly, " $k$ - $t$  SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity," in *Proc. 13th Annu. Meeting ISMRM*, Seattle, WA, USA, vol. 2420, 2006, p. 1.
- [6] 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, 2016.
- [7] B. Zhao, J. P. Haldar, A. G. Christodoulou, and Z.-P. Liang, "Image reconstruction from highly undersampled ( $k$ ,  $t$ )-space data with joint partial separability and sparsity constraints," *IEEE Trans. Med. Imag.*, vol. 31, no. 9, pp. 1809–1820, Sep. 2012.
- [8] 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.*, vol. 61, no. 1, pp. 103–116, 2009.
- [9] 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.
- [10] F. Ong *et al.*, "Extreme MRI: Large-scale volumetric dynamic imaging from continuous non-gated acquisitions," *Magn. Reson. Med.*, vol. 84, no. 4, pp. 1763–1780, Oct. 2020.
- [11] R. Otazo, E. J. 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, 2014.
- [12] S. G. Lingala and M. Jacob, "Blind compressive sensing dynamic MRI," *IEEE Trans. Med. Imag.*, vol. 32, no. 6, pp. 1132–1145, Jun. 2013.
- [13] M. S. Asif, L. Hamilton, M. Brummer, and J. Romberg, "Motion-adaptive spatio-temporal regularization for accelerated dynamic MRI," *Magn. Reson. Med.*, vol. 70, no. 3, pp. 800–812, 2013.

- [14] S. G. Lingala, E. DiBella, and M. Jacob, "Deformation corrected compressed sensing (DC-CS): A novel framework for accelerated dynamic MRI," *IEEE Trans. Med. Imag.*, vol. 34, no. 1, pp. 72–85, Jan. 2015.
- [15] 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.
- [16] 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.
- [17] A. H. Ahmed, R. Zhou, Y. Yang, P. Nagpal, M. Salerno, and M. Jacob, "Free-breathing and ungated dynamic MRI using navigator-less spiral SToRM," *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 3933–3943, Dec. 2020.
- [18] W. Jiang *et al.*, "Motion robust high resolution 3D free-breathing pulmonary MRI using dynamic 3D image self-navigator: Motion robust high-resolution 3D pulmonary MRI," *Magn. Reson. Med.*, vol. 79, no. 6, pp. 2954–2967, Jun. 2018.
- [19] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.
- [20] K. Hammerink *et al.*, "Learning a variational network for reconstruction of accelerated MRI data," *Magn. Reson. Med.*, vol. 79, no. 6, pp. 3055–3071, 2017.
- [21] 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. 2017.
- [22] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-based deep learning architecture for inverse problems," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 394–405, Feb. 2019.
- [23] C. Sandino, F. Ong, and S. Vasanawala, "Deep subspace learning: Enhancing speed and scalability of deep learning-based reconstruction of dynamic imaging data," in *Proc. Annu. Meeting ISMRM*, vol. 599, 2020.
- [24] W. Huang *et al.*, "Deep low-rank plus sparse network for dynamic MR imaging," *Med. Image Anal.*, vol. 73, Oct. 2021, Art. no. 102190.
- [25] T. Küstner *et al.*, "CINENet: Deep learning-based 3D cardiac CINE MRI reconstruction with multi-coil complex-valued 4D spatio-temporal convolutions," *Sci. Rep.*, vol. 10, no. 1, pp. 1–13, Dec. 2020.
- [26] C. M. Sandino, P. Lai, S. S. Vasanawala, and J. Y. Cheng, "Accelerating cardiac cine MRI using a deep learning-based ESPiRiT reconstruction," *Magn. Reson. Med.*, vol. 85, no. 1, pp. 152–167, Jan. 2021.
- [27] 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, 2019.
- [28] A. H. Ahmed, P. Nagpal, S. Kruger, and M. Jacob, "Dynamic imaging using deep bilinear unsupervised learning (DEBLUR)," in *Proc. IEEE 18th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2021, pp. 1099–1102.
- [29] J. Yoo, K. H. Jin, H. Gupta, J. Yerly, M. Stuber, and M. Unser, "Time-dependent deep image prior for dynamic MRI," *IEEE Trans. Med. Imag.*, vol. 40, no. 12, pp. 3337–3348, Dec. 2021.
- [30] B. Zhao, F. Lam, and Z.-P. Liang, "Model-based MR parameter mapping with sparsity constraints: Parameter estimation and performance bounds," *IEEE Trans. Med. Imag.*, vol. 33, no. 9, pp. 1832–1844, Sep. 2014.
- [31] S. Bhave, S. G. Lingala, C. P. Johnson, V. A. Magnotta, and M. Jacob, "Accelerated whole-brain multi-parameter mapping using blind compressed sensing," *Magn. Reson. Med.*, vol. 75, no. 3, pp. 1175–1186, 2016.
- [32] F. Lam and Z.-P. Liang, "A subspace approach to high-resolution spectroscopic imaging," *Magn. Reson. Med.*, vol. 71, no. 4, pp. 1349–1357, Apr. 2014.
- [33] I. Bhattacharya and M. Jacob, "Compartmentalized low-rank recovery for high-resolution lipid unsuppressed MRSI," *Magn. Reson. Med.*, vol. 78, no. 4, pp. 1267–1280, Oct. 2017.
- [34] 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.
- [35] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1028–1041, May 2010.
- [36] 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.
- [37] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
- [38] C. P. Burgess *et al.*, "Understanding disentangling in  $\beta$ -VAE," 2018, *arXiv:1804.03599*.
- [39] I. Higgins *et al.*, " $\beta$ -VAE: Learning basic visual concepts with a constrained variational framework," in *Proc. ICLR*, 2016, pp. 1–13.
- [40] H. K. Aggarwal, A. Pramanik, and M. Jacob, "ENSURE: A general approach for unsupervised training of deep image reconstruction algorithms," 2020, *arXiv:2010.10631*.