Dynamic Imaging using Deep Bi-linear Unsupervised Regularization (DEBLUR)

Abdul Haseeb Ahmed, Member IEEE, Prashant Nagpal, and Mathews Jacob, Member IEEE

Abstract—Bilinear models that decompose dynamic data to spatial and temporal factors are powerful and memory-efficient tools for the recovery of dynamic MRI data. These methods rely on sparsity and energy compaction priors on the factors to regularize the recovery. The quality of the recovered images depend on the specific priors. Motivated by deep image prior, we introduce a novel bilinear model whose factors are represented using convolutional neural networks (CNNs). The CNN parameters are learned from the undersampled data off the same subject. To reduce the run time and to improve performance, we initialize the CNN parameters. 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 low-rank and SToRM reconstructions.

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 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-hold 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. These adaptive approaches, where the signal model is learned from the data itself, are observed to be far more efficient than earlier strategies that relied on hand-crafted models. 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]. Many of these schemes rely on k-space navigator measurements to estimate the temporal factors, while the spatial factors are estimated from the entire data. The kernel approach for estimating the temporal factors is observed to be more efficient in representing the dynamics, especially in free-breathing applications. A major benefit of the bilinear methods is the significantly reduced memory demand of these algorithms, in addition to the good reconstructions they offer. Specifically, the factors are significantly smaller in dimension than the dynamic dataset, which facilitates the recovery of large 3D volumes [10], [18]. While early methods relied on calibration data to estimate one of the factors, the joint optimization of both the factors offers several advantages, including improved image quality [9], [12]. The distinction between the methods can be viewed 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].

Deep learning models are now emerging as powerful approaches for image recovery in a range of static inverse problems [20]. Direct inversion strategies, which rely on a large convolutional neural network (CNN) to recover the images from undersampled data [21], [22], as well as model-based deep learning methods [23]–[25] that interleave smaller CNN blocks with data-consistency enforcing optimization modules, have been introduced. By enforcing the data consistency, model-based methods can offer improved image quality over direct inversion strategies. Unfortunately, dynamic MRI and parametric MRI schemes often require the recovery of a large number of image frames; the direct application of the unrolled model-based deep learning schemes to the above setting is severely limited due to the high memory demand and computational complexity of current methods. Current strategies are either restricted to fewer time frames [26] or often have to use small networks [27], [28]. 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 CNN to further improve the performance while retaining the memory efficiency of bilinear representations. We use the structural bias of CNN networks \([19]\) as priors on the factors; this approach is an alternative to current schemes that penalize energy \([9]\), sparsity \([12]\), or kernel priors \([10], [25]\) 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 multichannel measurements. The ability of the proposed scheme to directly learn the compressed representation makes it more memory efficient in multidimensional applications compared to current CNN approaches \([26]–[28]\).

This paper generalizes our earlier conference version \([29]\) in multiple ways. First of all, we follow a completely unsupervised strategy: we pretrain the CNN factor 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 \([29]\); 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 \([29]\), along with ablation studies to determine the impact of the regularization terms.

II. BACKGROUND

A. Bilinear models for dynamic MRI

Bilinear models are widely used in multidimensional applications, including dynamic MRI \([7], [8]\), parameter mapping \([30], [31]\), and MR spectroscopic imaging \([32], [33]\). We have shown the significance of this approach in reconstructing dynamic images in our limited study \([29]\). These schemes express the Casorati matrix of the multidimensional dataset denoted by \(\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]\), where \(\mathbf{x}_i\) denotes the vectorized version of the \(i^{th}\) frame in the dataset as

\[
\mathbf{X} = \left[\mathbf{u}_1, \ldots, \mathbf{u}_r \mid \mathbf{v}_1, \ldots, \mathbf{v}_r\right]^T = \mathbf{Uv}^T,
\]

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:

\[
\mathcal{A}_i(\mathbf{x}_i) = \mathcal{A}_i(\mathbf{U})\mathbf{v}_i
\]

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 \(\mathbf{x}_i = \mathbf{Uv}_i\).

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 \([1]\). 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:

\[
\{\mathbf{U}^*, \mathbf{V}^*\} = \arg \min_{\mathbf{U}, \mathbf{V}} \|\mathcal{A}(\mathbf{U}\mathbf{V}^T) - \mathbf{B}\|^2 + \lambda_1\mathcal{R}_1(\mathbf{U}) + \lambda_2\mathcal{R}_2(\mathbf{V})
\]

Here, \(\mathbf{B}\) denotes the measured data. We note that the representation in \([1]\) is linear in \(\mathbf{U}\) and \(\mathbf{V}\) independently; the joint optimization in \([3]\) 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 \(\mathcal{R}_1(\mathbf{U}) = \|\mathbf{U}\|_1\) 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 (1), where $R_1(U) = \sum \sigma_i \|u_i\|^2$ and $V$ 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 $R_1$ and $R_2$ to estimate $U$ and $V$. 

B. 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 are posed as

$$\{\theta^*\} = \arg \min_{\theta} \|A(x) - b\|^2 \text{ such that } x = G_{\theta}[z]$$  (4)

where $x = G_{\theta^*}[z]$ is the recovered image, generated by the CNN generator $G_{\theta^*}$ whose parameters are denoted by $\theta$.

The constraint that the image is generated by a CNN provides implicit regularization, which facilitates the recovery of $x$ in challenging inverse problems. Here, $z^*$ 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 $X = UV^T$ as in (1). 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

$$U = G_\theta(U_0)$$  (5)
$$V = G_\varphi(Z).$$  (6)

Here, $G_\theta$ and $G_\varphi$ are CNN generators whose weights are denoted by $\theta$ and $\varphi$, respectively. Here, $Z = [z_1, \ldots, z_N]$ are time-dependent latent vectors that are also learned from the measured $k$-space data. Here, $U_0$ 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 $U$ factors to $G_\theta$ 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 optimiza-
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 (7) 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 $G_\theta$ and $G_\phi$ networks. In this work, we chose the SToRM [15] data to initialize the network. We note from our past results that the SToRM approach yields improved results compared to the other state-of-the art algorithms [15], including low-rank methods. In particular, we pretrain $G_\theta$ and $G_\phi$ generators independently using the SToRM factors, as

$$\begin{align*}
\{\theta_0, \phi_0, Z_0\} &= \arg\min_{\theta, \phi, Z} \| G_\theta(U_0) - U_{\text{SToRM}} \|^2, \\
\{\phi_0, Z_0\} &= \arg\min_{\phi, Z} \| G_\phi(Z) - V_{\text{SToRM}} \|^2.
\end{align*}$$

These initial guesses of the network weights $(\theta, \phi)$ and latent vectors $Z$ are used as initialization in (7). We study the impact of the initialization on the algorithms in the results section.

B. Regularization penalties

The DIP approach, as well as our extension to the dynamic setting in (7), is vulnerable to noise overfitting. In particular, if the generator networks have sufficient capacity, it will learn the noise in the measurements when the number of iterations is large [19]. To further improve the robustness of the algorithm...
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 $\lambda_3 = 0$. Higher value of $\lambda_3$ oversmoothes the latent vectors as shown by $\lambda_3 = 1e^3$. Empirical findings show $\lambda_3 = 1e^3$ gives better separation of latent vectors that represent cardiac and respiratory motion. Secondly, the $\text{SER}$ changes between the different values of $\lambda_3$ are marginal.

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 $G_0$ promotes the learning of local spatial factors similar to [10], which is more efficient than global low-rank minimization. Unlike the explicit definition of the blocks in [10], the definition of the local neighborhoods in the proposed scheme are 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 (8) that the $i^{th}$ image in the time series is dependent on the latent vector $z_i$. The last term in (11) is the norm of the temporal gradients of $Z$, which will encourage the temporal smoothness of the time series. The algorithm in (11) 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 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: The datasets were acquired using a 34-channel cardiac array. 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 fashion. The complete data acquisition lasted 42 seconds for each slice. To determine the ability of the algorithms to reduce the acquisition time, we retained the initial 14 seconds of the original acquisition.

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):

$$\text{SER} = 20 \log_{10} \frac{\|x_{\text{orig}}\|^2}{\|x_{\text{orig}} - x_{\text{rec}}\|^2},$$

(12)
where $\| \cdot \|_2$ denotes the $\ell_2$ norm, and $x_{\text{orig}}$ and $x_{\text{rec}}$ denote the original and the reconstructed images, respectively.

- **Peak Signal to Noise Ratio (PSNR):**

$$\text{PSNR} = 20 \log_{10} \frac{\max \{ x_{\text{orig}} \}}{\| x_{\text{orig}} - x_{\text{rec}} \|_2}.$$  

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

$$\text{HFEN} = 20 \log_{10} \frac{\| \text{LoG}(x_{\text{orig}}) - \text{LoG}(x_{\text{rec}}) \|_2}{\| \text{LoG}(x_{\text{orig}}) \|_2},$$  

where $\text{LoG}$ 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 and gets the score using a support vector regression (SVR) model with the help of an image database and corresponding differential mean opinion score values. The distorted image database such as compression artifacts, blurring, and noise images, and with pristine versions of the distorted images [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.

- **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 $G_\theta$ as the spatial generator. We use a four-layer CNN network 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 $G_\phi$ as the temporal generator, where we use a four-layer CNN network with ReLU activation. The inputs to the temporal generator are the latent vectors $Z \in \mathbb{R}^{d \times n_f}$, 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 $V$ of dimension $r \times n_f$, where $r$ is the number of temporal basis functions in $V$.

**V. Experiments and Results**

We describe the experiments and our observations in this section.

A. Impact of pretraining

To demonstrate the benefit of pretraining, we compare the algorithms described in [7] with random initialization and with the initialization scheme denoted by (5) & (10). 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)-(c) correspond to the color of the markers in the plots, which denotes the initial and maximum SER values. We note that the initialization with random weights starts with the low SER as seen from Fig 2(b), compared to the SToRM initialization in Fig 2(d). As seen from the plots in (a), the DEBLUR approach with STORM initialization converged rapidly to a peak value, while the one with random initialization converged to a poor solution with significantly more iterations. We also note that the DEBLUR approach with random initialization yielded poor results as seen from Fig 2(c), while the one with STORM initialization in 2(f) shows significantly improved results.

B. Impact of regularization parameters

We study the impact of the regularization priors in (11) in Figures 3 and 5.

1) Regularization of $U$ network parameters: We initialize the network with the SToRM initialization, described by (5) & (6). We set $\lambda_2 = \lambda_3 = 0$ and consider four different $\lambda_1$ values in this study. We note that the network with $\lambda_1 = 0$ 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 to the optimal values.

| Method          | Data1   | Data2   | Data3   | Data4   | Data5   |
|-----------------|---------|---------|---------|---------|---------|
| DEBLUR(42s)     | 28.30 ± 3.7 | 30.33 ± 3.3 | 29.90 ± 3.8 | 28.11 ± 3.7 | 17.53 ± 5.0 |
| SToRM(42s)      | 34.44 ± 3.9 | 26.0 ± 2.9  | 29.96 ± 4.3  | 34.0 ± 5.0  | 17.32 ± 5.4  |
| DEBLUR(14s)     | 29.5 ± 3.4  | 34.1 ± 4.1  | 38.6 ± 5.5   | 32.87 ± 5.2 | 17.28 ± 5.4  |
| SToRM(14s)      | 36.26 ± 0.9 | 32.63 ± 3.5 | 35.72 ± 0.8  | 34.57 ± 2.3 | 25.42 ± 2.3  |

**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 [7].
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.

2) Regularization of \( V \) 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 \( V \) and \( U \) 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 \( Z \): 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 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 \( \lambda_3 = 1e^2 \) offered the disentangling of cardiac/respiratory motion.

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. However, by properly selecting \( \lambda_3 \), we observe that the latent vectors learn physiologically relevant parameters. In the datasets we considered, we note that the latent vectors separate into a fast-changing latent vector that captures cardiac motion and a slow one that captures respiratory motion. Post-reconstruction, the latent variables can 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 expiration 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), manifested by reduced blurring and the depiction of the papillary muscles.

The SToRM approach results in significant blurring and degradation in image quality when only 14 s of data is
available. We note that the DEBLUR reconstructions are able to minimize the blurring, offering 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.

VI. Discussion

The experiment in Section V-A clearly shows the benefit of initializing the network parameters by (9) and (10), 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 (7). In particular, the performance of the algorithm decays with increasing iterations, indicating overfitting to the noise in the k-space measurements.

These 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.

The best choice of λ3 enables the learning of latent vectors Z that are interpretable. In particular, 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.

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.

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 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 enables the learning of latent vectors Z that are interpretable. In particular, 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.
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.

**IX. ACKNOWLEDGMENTS**

This work is supported by grant NIH 1R01EB019961. 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 radiology, vol. 29, no. 9, pp. 848–851, 1994.

[2] Z.-P. Liang, H. Jiang, C. P. Hess, and P. C. Lauterbur, “Dynamic imaging by model estimation,” International journal of imaging systems and technology, 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).” Magnetic Resonance in Medicine, 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.” Magnetic resonance in medicine, vol. 50, no. 5, pp. 1031–1042, 2003.

[5] M. Lustig, J. M. Santos, D. L. Donoho, and J. M. Pauly, “kt sparse: High frame rate dynamic mri exploiting spatio-temporal sparsity.” in Proceedings of the 13th Annual Meeting of ISMRM, Seattle, vol. 2420, 2006.

[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.” Magnetic resonance in medicine, vol. 73, 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 transactions on medical imaging, vol. 31, no. 9, pp. 1809–1820, 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.” Magnetic resonance in medicine, 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: kt slr.” IEEE transactions on medical imaging, vol. 30, no. 5, pp. 1042–1054, 2011.

[10] F. Ong, X. Zhu, J. Y. Cheng, K. M. Johnson, P. E. Larson, S. S. Vasanawala, and M. Lustig, “Extreme mri: Large-scale volumetric dynamic imaging from continuous non-gated acquisitions.” Magnetic resonance in medicine, vol. 84, no. 4, pp. 1763–1780, 2020.

[11] 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.” Magnetic Resonance in Medicine, vol. 73, no. 3, pp. 1125–1136, 2015.

[12] S. G. Lingala and M. Jacob, “Blind compressive sensing dynamic mri.” IEEE transactions on medical imaging, vol. 32, no. 6, pp. 1132–1145, 2013.

[13] M. S. Asif, L. Hamilton, M. Brummer, and J. Romberg, “Motion-adaptive spatio-temporal regularization for accelerated dynamic mri.” Magnetic Resonance in Medicine, 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 transactions on medical imaging, vol. 34, no. 1, pp. 72–85, 2014.

[15] S. Poddar, Y. Mohsin, D. Ansah, B. Thattaliyath, R. Ashwath, and M. Jacob, “Manifold recovery using kernel low-rank regularization: application to dynamic imaging.” IEEE Transactions on Computational Imaging, 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 transactions on medical imaging, vol. 36, no. 11, pp. 2297–2307, 2017.

[17] A. H. Ahmned, R. Zhou, Y. Yang, P. Nagpal, M. Salerno, and M. Jacob, “Free-breathing and ungated dynamic mri using navigator-less spiral storm.” IEEE Transactions on Medical Imaging, vol. 39, no. 12, pp. 3933–3943, 2020.

[18] W. Jiang, F. Ong, K. M. Johnson, S. K. Nagle, T. A. Hope, M. Lustig, and P. E. Larson, “Motion robust high resolution 3d free-breathing pulmonary mri using dynamic 3d image self-navigator.” Magnetic resonance in medicine, vol. 79, no. 6, pp. 2954–2967, 2018.

[19] D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Deep image prior.” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 9446–9454.

[20] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, and D. Liang, “Accelerating magnetic resonance imaging via deep learning.” in 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). IEEE, 2016, pp. 514–517.
[22] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, “Deep convolutional neural network for inverse problems in imaging,” IEEE Transactions on Image Processing, vol. 26, no. 9, pp. 4509–4522, 2017.

[23] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, “Learning a variational network for reconstruction of accelerated mri data,” Magnetic resonance in medicine, vol. 79, no. 6, pp. 3055–3071, 2018.

[24] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, “A deep cascade of convolutional neural networks for dynamic mri image reconstruction,” IEEE transactions on Medical Imaging, vol. 37, no. 2, pp. 491–503, 2017.

[25] H. K. Aggarwal, M. P. Mani, and M. Jacob, “Modl: Model-based deep learning architecture for inverse problems,” IEEE transactions on medical imaging, vol. 38, no. 2, pp. 394–405, 2018.

[26] T. Küstner, N. Fuin, K. Hammernik, A. Bustin, H. Qi, R. Hajhosseiny, P. G. Masci, R. Neji, D. Rueckert, R. M. Botnar et al., “Cinenet: deep learning-based 3d cardiac cine mri reconstruction with multi-coil complex-valued 4d spatio-temporal convolutions,” Scientific reports, vol. 10, no. 1, pp. 1–13, 2020.

[27] S. Biswas, H. K. Aggarwal, and M. Jacob, “Dynamic mri using model-based deep learning and storm priors: Modl-storm,” Magnetic resonance in medicine, vol. 82, no. 1, pp. 485–494, 2019.

[28] C. M. Sandino, P. Lai, S. S. Vasanawala, and J. Y. Cheng, “Accelerating cardiac cine mri using a deep learning-based espirit reconstruction,” Magnetic Resonance in Medicine, 2020.

[29] A. H. Ahmed, P. Nagpal, S. Kruger, and M. Jacob, “Dynamic imaging using deep bilinear unsupervised learning (deblur),” in 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). IEEE, 2021, pp. 1099–1102.

[30] B. Zhao, F. Lam, and Z.-P. Liang, “Model-based mr parameter mapping with sparsity constraints: parameter estimation and performance bounds,” IEEE transactions on medical imaging, vol. 33, no. 9, pp. 1832–1844, 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,” Magnetic resonance in medicine, vol. 75, no. 3, pp. 1175–1186, 2016.

[32] F. Lam and Z.-P. Liang, “A subspace approach to high-resolution spectroscopic imaging,” Magnetic resonance in medicine, vol. 71, no. 4, pp. 1349–1357, 2014.

[33] I. Bhattacharya and M. Jacob, “Compartmentalized low-rank recovery for high-resolution lipid unsuppressed mrsi,” Magnetic resonance in medicine, vol. 78, no. 4, pp. 1259–1306, 2017.

[34] S. Poddar and M. Jacob, “Dynamic mri using smoothness regularization on manifolds (storm),” IEEE transactions on medical imaging, vol. 35, no. 4, pp. 1106–1115, 2015.

[35] S. Ravishankar and Y. Bresler, “Mr image reconstruction from highly undersampled k-space data by dictionary learning,” IEEE transactions on medical imaging, vol. 30, no. 5, pp. 1028–1041, 2011.

[36] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” IEEE transactions on image processing, vol. 13, no. 4, pp. 600–612, 2004.

[37] A. Mittal, A. K. Moothy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” IEEE Transactions on Image Processing, vol. 21, no. 12, pp. 4695–4708, 2012.