

DYNAMIC IMAGING USING DEEP BILINEAR UNSUPERVISED LEARNING (DEBLUR)

Abdul Haseeb Ahmed, Prashant Nagpal, Stanley Kruger, Mathews Jacob

Department of Electrical and Computer Engineering, University of IOWA

ABSTRACT

Bilinear models such as low-rank and compressed sensing, which decompose the 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. Motivated by deep image prior, we introduce a novel bilinear model, whose factors are regularized using convolutional neural networks. To reduce the run time, we initialize the CNN parameters by pre-training them on pre-acquired data with longer acquisition time. Since fully sampled data is not available, pretraining is performed on undersampled data in an unsupervised fashion. 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— Cardiac MRI, dynamic imaging, bilinear model, unsupervised learning, image reconstruction

1. INTRODUCTION

Deep learning models are emerging as powerful approaches for image recovery in a range of static inverse problems. Direct inversion strategies, which rely on a large CNN to recover the images from undersampled data, as well as model based deep learning methods that interleave smaller CNN blocks with data-consistency enforcing optimization modules are available. 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 current deep-learning schemes to the above setting, along with end-to-end optimization, is severely limited due to the high memory demand and computational complexity of current methods. Current strategies are either restricted to fewer time frames [1] or often have to use small networks [2, 3].

Bilinear image models that factorize the dataset into spatial and temporal factors have been widely used in the dynamic/parametric imaging applications. In addition to offering good recovery, a major benefit of these schemes is the significantly reduced memory demand of these algorithms.

Specifically, the factors are significantly smaller in dimension than the dynamic dataset. While early methods have relied on calibration data to estimate one of the factors, the joint optimization of both the factors offer several advantages including improved image quality [4, 5]. These joint optimization schemes often rely on sparsity and energy regularization priors on the factors, resulting in several flavors of algorithms (e.g. low-rank, compressed sensing, manifold methods).

The main focus on this work is to use the power of convolutional neural networks (CNN) to improve the recovery of dynamic imaging data. In particular, we regularize the factors using CNN-based priors. While the factors can be recovered from the data using an unrolled optimization as in [6], the memory demand of the unrolled strategy is a concern in large scale problems (e.g. 3D + time). Hence, we propose to use the direct inversion strategy to further reduce the memory demand of the algorithm. A challenge with direct inversion schemes is the lack of consistency between the reconstructed images and the measured data. We propose to optimize the parameters of the network to match the measured data during image reconstruction as in deep image prior (DIP) [7]. While this reconstruction algorithm is associated with higher computational complexity than unrolled schemes, it keeps the memory demand of the algorithm minimal; we expect this approach to facilitate the recovery of large-scale datasets.

We initialize the generators with pre-trained networks to reduce the reconstruction time and to improve performance. Since it is impossible to acquire fully sampled ground truth datasets, we propose to pre-learn the networks from undersampled k-space data from longer acquisitions (42 seconds). Our experiments show that this initialization results in a significantly faster convergence compared to bilinear models trained using deep image priors. More importantly, the algorithm converges to a less blurred solution compared to random initialization of the network, when the reconstruction from 10 second data is considered. We attribute the improved performance to the pre-learning of the image properties (e.g sharpness of boundaries), that may be absent or difficult to estimate from short segments of data.

2. BACKGROUND

2.1. Bilinear models for dynamic MRI

Bilinear models, also termed as partially separable models, are widely used in dynamic MRI, parameter mapping, and

MR spectroscopic imaging. These schemes express the Casorati matrix of the volume \mathbf{X} as

$$\mathbf{X} = \mathbf{U}\mathbf{V}^T \quad (1)$$

where the columns of \mathbf{U} are the spatial basis functions, while that of \mathbf{V} can be interpreted as the temporal basis functions. In addition to the efficiency in representing the large dynamic dataset using few parameters, the above representation also offers computational benefits. Specifically, the measurements can be expressed as

$$\mathcal{A}_i(\mathbf{x}_i) = \mathbf{A}_i(\mathbf{U})\mathbf{v}_i \quad (2)$$

Where \mathbf{A}_i is sampling operator, while early methods relied on calibrated strategies, the joint estimation of \mathbf{U} and \mathbf{V} from the measured undersampled data offers several benefits. These schemes pose the recovery of the signals from the undersampled measurements $\mathcal{A}(\mathbf{X})$ as

$$\{\mathbf{U}^*, \mathbf{V}^*\} = \arg \min_{\mathbf{U}, \mathbf{V}} \|\mathcal{A}(\mathbf{U}\mathbf{V}^T) - \mathbf{B}\|^2 + \lambda_1 \mathcal{R}_1(\mathbf{U}) + \lambda_2 \mathcal{R}_2(\mathbf{V}) \quad (3)$$

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. Where \mathbf{B} is the measurement data.

1. Low-rank regularization: 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 (1), where $\mathcal{R}_1(\mathbf{U}) = \sum \sigma_i \|\mathbf{u}\|_i^2$ and \mathbf{V} is obtained as the eigen vectors of the graph Laplacian matrix of the graph of the data. Both calibrated and uncalibrated 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} . All of the current methods rely on carefully chosen norms to exploit specific image properties.

2.2. Deep Image Prior (DIP)

Deep image priors has been introduced to exploit the implicit property that CNN architectures favor natural images more than noise. The regularized reconstruction of an image from undersampled and noisy measurements are posed as

$$\{\boldsymbol{\theta}^*\} = \arg \min_{\boldsymbol{\theta}} \|\mathcal{A}(\mathbf{x}) - \mathbf{b}\|^2 \quad \text{such that } \mathbf{x} = \mathcal{G}_{\boldsymbol{\theta}}[\mathbf{z}] \quad (4)$$

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 above problem is often solved using stochastic gradient descent (SGD), which is often terminated early to obtain regularized recovery. Specifically, when the generator has sufficient capacity, the network will fit the measurement noise; 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.

3. DEEP BILINEAR UNSUPERVISED LEARNING (DEBLUR)

Instead of using handcrafted models, we propose to use the deep learned priors to regularize the problem as shown in the fig 1. We propose to pre-learn the priors from exemplary data, which is further optimized based on the measured undersampled data during the reconstruction. Specifically, we pose the recovery as

$$\{\mathbf{U}^*, \mathbf{V}^*\} = \min_{\boldsymbol{\theta}, \boldsymbol{\phi}} \|\mathcal{A}(\mathbf{U}\mathbf{V}^T) - \mathbf{B}\|^2, \quad \text{such that } \mathbf{U} = \mathcal{G}_{\boldsymbol{\theta}}[\mathbf{U}_0]; \mathbf{V} = \mathcal{G}_{\boldsymbol{\phi}}[\mathbf{V}_0] \quad (5)$$

Here, $\mathcal{G}_{\boldsymbol{\theta}}$ and $\mathcal{G}_{\boldsymbol{\phi}}$ are two CNN generators, parametrized by the network parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, respectively. Here, \mathbf{U}_0 and \mathbf{V}_0 are latent variables. We propose to solve the above optimization scheme using SGD. Rather than using random latent variables as in deep image prior, we choose them in an image specific way. In this work, we choose \mathbf{V}_0 as the STORM basis functions that is estimated from calibration data, while \mathbf{U}_0 is obtained as

$$\mathbf{U}_0 = (\mathcal{A}^H \mathbf{B}) \mathbf{V}_0^T = \mathcal{A}^H (\mathbf{B} \mathbf{V}_0^T) \quad (6)$$

Since we minimize (5) using SGD, we expect the specific choice of \mathbf{U}_0 and \mathbf{V}_0 to have minimal impact on the final solution. However, we expect the specific choice of initial guesses to influence the run time or the number of epochs.

3.1. Unsupervised pre-training of the generators

The generators in DIP are usually not pre-trained. To further reduce the run time of the algorithm, we propose to pre-learn the networks from exemplary data. As discussed previously, it is difficult to acquire fully sampled datasets in dynamic imaging applications. We hence rely on the unsupervised strategy:

$$\{\boldsymbol{\theta}_0, \boldsymbol{\phi}_0\} = \arg \min_{\boldsymbol{\theta}, \boldsymbol{\phi}} \sum_{i=1}^{N_{\text{train}}} \|\mathcal{A}(\mathbf{U}\mathbf{V}^T) - \mathbf{B}(i)\|^2, \quad \text{such that } \mathbf{U} = \mathcal{G}_{\boldsymbol{\theta}}[\mathbf{U}_0(i)]; \mathbf{V} = \mathcal{G}_{\boldsymbol{\phi}}[\mathbf{V}_0(i)] \quad (7)$$

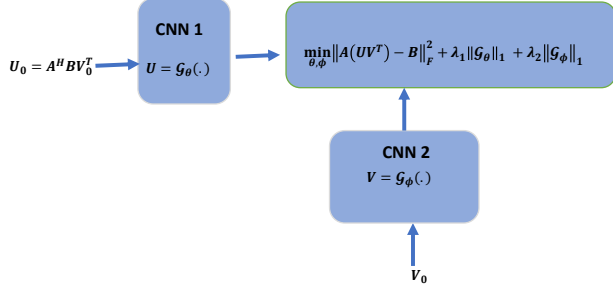


Fig. 1. Proposed method. Two CNN networks are used on the spatial and temporal prior factors. CNN 1 is initiated with the \mathbf{U}_0 as mentioned in the Eq. 6. CNN 2 is initialized with the STORM temporal basis. We have also applied l_1 norm on the network parameters to stabilize the convergence.

Here, $\mathbf{U}_0(i)$ and $\mathbf{V}_0(i)$ are initial guess factors for the i^{th} dataset, while $\mathbf{B}(i)$ is the corresponding measurements. The above training will yield initial weights ϕ_0 and θ_0 , which are used to initialize the algorithm.

3.2. Regularization of network parameters

A challenge with the optimization of the parameters in (5) is the risk of overfitting to noise, similar to the ones demonstrated in deep image priors [7]. Deep image prior uses the reduced number of iterations as the prior, which uses the property that the network structure of CNN favors images; it takes more iterations to fit noise. To minimize the risk of overfitting, we propose to add regularization priors on the network parameters:

$$\{\mathbf{U}^*, \mathbf{V}^*\} = \min_{\theta, \phi} \left\| \mathcal{A}(\mathbf{U}\mathbf{V}^T) - \mathbf{B} \right\|_F^2 + \lambda_1 \|\theta\|_{\ell_1} + \lambda_2 \|\phi\|_{\ell_1}, \quad \text{such that } \mathbf{U} = \mathcal{G}_\theta[\mathbf{U}_0]; \mathbf{V} = \mathcal{G}_\phi[\mathbf{V}_0] \quad (8)$$

The impact of the regularization parameters and their ability to minimize over-fitting issues are studied in the results section.

3.3. Data acquisition and post-processing

The experimental data was obtained using 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 study was an add-on to the routine cardiac MRI exams. Each frame was sampled by two k-space navigator spokes, oriented at 0 degrees and 90 degrees respectively. The protocol was approved by the Institutional Review Board (IRB) at the University of Iowa. The sequence parameters were: 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 lines of k-space per frame. The

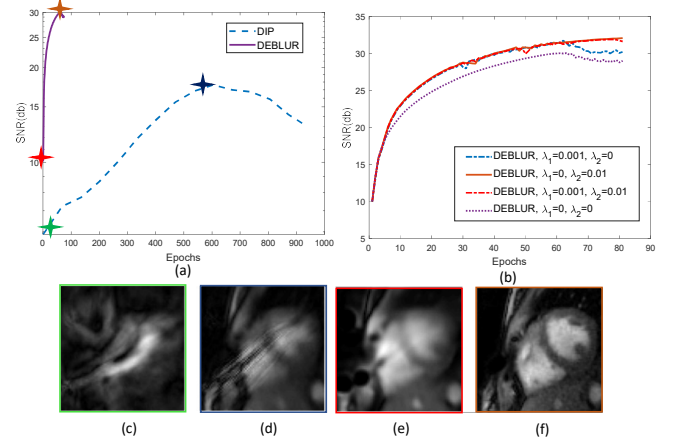


Fig. 2. (a) shows the SNR curves of DIP and DEBLUR methods. Their corresponding initial and peak values images are shown in (c)-(f). The image with the green border (c) corresponds to the DIP initialization with random weights, while the final DIP solution from 600 epochs is shown in (d). The use of the pre-trained parameter yields (e) while optimizing the parameters during reconstruction significantly improve the performance as seen from (f). We note the absence of artifacts and sharper features in (f). (b) shows the benefits of using l_1 regularization on network parameters. Dotted line shows the plot of DEBLUR method without any regularization. Other curves show l_1 regularization on U CNN network ($\lambda_1=0.001$), V CNN network ($\lambda_2=0.01$) and both ($\lambda_1=0.001$, $\lambda_2=0.01$). Use of l_1 regularization provides the benefits of improved SNR and stable convergence.

scan parameters were kept same across all patients. To reduce the computational complexity, we combined the data from 34 channels to seven using principal component analysis. For the experiments in this work, we retained the initial 10 seconds of the original acquisition.

4. RESULTS

4.1. Impact of Pre-training and benefit of training during reconstruction

To show the benefit of pre-training, we have compared our proposed method with the DIP method. Fig 2 (a) shows the plot of SNR values with respect to the number of epochs and with initial and peak SNR values are indicated by different color stars. Fig 2(c)-(f) show the corresponding images at the initial and peak values respectively. Since the DIP method is initialized with the random network weights, therefore, it starts with the low SNR as compared to the DEBLUR method, which utilizes the pre-trained network as indicated in Eq. 7. Fig 2(e) shows the output of pre-trained network which indicates the benefit of using pre-training in our method.

4.2. Impact of regularization

Fig 2(b) shows the impact of regularization on the deep learned priors. We have compared four cases as mentioned below:

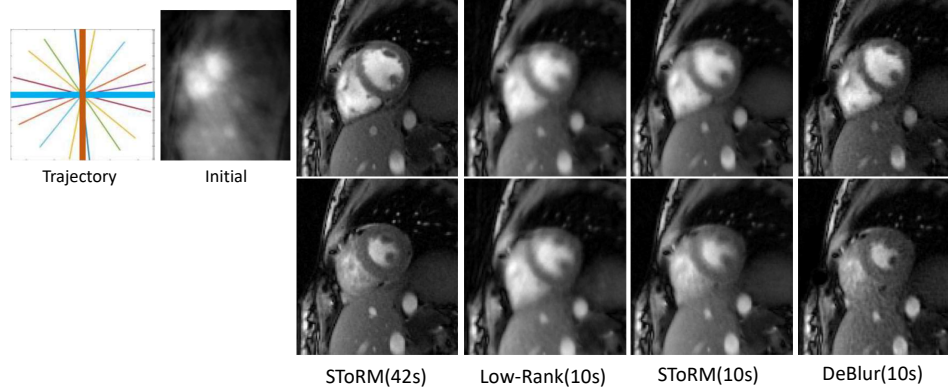


Fig. 3. Performance comparison of the low-rank method, SToRM method (10s) and the proposed method. SToRM reconstruction with 40s of acquisition data is used as a ground truth. We observe that DEBLUR gives better image quality with less blurring as compared to the other methods.

1. Without regularization and pretrained weights
2. With l_1 regularization on \mathbf{U} network parameters and pretrained weights
3. With l_1 regularization on \mathbf{V} network parameters and pretrained weights
4. With l_1 regularization on both \mathbf{U} and \mathbf{V} network parameters with pretrained weights.

Fig 2(b) clearly shows the benefit of using regularization on network parameters. It provides better SNR values and also stabilizes the convergence to avoid the early stopping requirement of the DIP method.

4.3. Comparison with other methods

To show the performance of our proposed method, we compare DEBLUR method with the SToTM and low-rank reconstructions. Since ground truth is not available, we have used SToRM 42sec acquisition as a ground truth. We have shown two frames (end of diastole and end of systole) from each method to show the comparison. DEBLUR provides better spatial quality as compared to the low-rank and SToRM (10s) methods as shown in the Fig. 3.

5. 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.

6. ACKNOWLEDGMENTS

This work is supported by grants NIH 1R01EB019961. The authors claim that there is no conflicts of interest.

7. CONCLUSION

In this paper, we have proposed a new cardiac cine MRI reconstruction method based on bilinear unsupervised learning.

Deep Regularized spatial and temporal priors used to reconstruct the undersampled MR data. Results show that our CNN priors with l_1 norm on the learning parameters give better performance as compared to the deep image prior. Reconstructed cardiac CINE images show the ability of our proposed method to give improved image quality as compared to the other methods.

8. REFERENCES

- [1] Thomas Küstner, Niccolo Fuin, Kerstin Hammernik, Aurelien Bustin, Haikun Qi, Reza Hajhosseiny, Pier Giorgio Masci, Radhouene Neji, Daniel Rueckert, René 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.
- [2] Sampurna Biswas, Hemant K Aggarwal, and Mathews 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.
- [3] Christopher M Sandino, Peng Lai, Shreyas S Vasanawala, and Joseph Y Cheng, “Accelerating cardiac cine mri using a deep learning-based espirit reconstruction,” *Magnetic Resonance in Medicine*, 2020.
- [4] Sajjan Goud Lingala, Yue Hu, Edward DiBella, and Mathews 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.
- [5] Sajjan Goud Lingala and Mathews Jacob, “Blind compressive sensing dynamic mri,” *IEEE transactions on medical imaging*, vol. 32, no. 6, pp. 1132–1145, 2013.
- [6] Hemant K Aggarwal, Merry P Mani, and Mathews Jacob, “Modl: Model-based deep learning architecture for inverse problems,” *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 394–405, 2018.
- [7] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky, “Deep image prior,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9446–9454.