

FREE-BREATHING CARDIOVASCULAR MRI USING A PLUG-AND-PLAY METHOD WITH LEARNED DENOISER

Sizhuo Liu, Edward Reehorst, Philip Schniter, and Rizwan Ahmad

The Ohio State University

ABSTRACT

Cardiac magnetic resonance imaging (CMR) is a noninvasive imaging modality that provides a comprehensive evaluation of the cardiovascular system. The clinical utility of CMR is hampered by long acquisition times, however. In this work, we propose and validate a plug-and-play (PnP) method for CMR reconstruction from undersampled multi-coil data. To fully exploit the rich image structure inherent in CMR, we pair the PnP framework with a deep learning (DL)-based denoiser that is trained using spatiotemporal patches from high-quality, breath-held cardiac cine images. The resulting “PnP-DL” method iterates over data consistency and denoising subroutines. We compare the reconstruction performance of PnP-DL to that of compressed sensing (CS) using eight breath-held and ten real-time (RT) free-breathing cardiac cine datasets. We find that, for breath-held datasets, PnP-DL offers more than one dB advantage over commonly used CS methods. For RT free-breathing datasets, where ground truth is not available, PnP-DL receives higher scores in qualitative evaluation. The results highlight the potential of PnP-DL to accelerate RT CMR.

Index Terms— Cardiac MRI, deep learning, denoising, plug-and-play algorithms

1. INTRODUCTION

Magnetic resonance imaging (MRI) is a well-established imaging modality that offers high soft-tissue contrast without the use of ionizing radiation. Cardiovascular magnetic resonance imaging (CMR) extends the application of MRI to yield static or dynamic images of the cardiovascular system. Cardiac cine, i.e., creating a movie of the beating heart, is one of the most common applications of CMR. Cardiac cine is typically performed under breath-holding conditions and requires regular heart rhythm. Under these conditions, the measured k-space data can be combined from several heartbeats, leading to a well-posed inverse problem. This approach, however, is not feasible for subjects who cannot hold their breath or are arrhythmic. For such patients, cardiac cine is performed in real-time (RT) and under free-breathing

conditions. RT cine does not combine data across heartbeats, leading to high undersampling rates.

To facilitate RT cine, several methods that combine multi-coil MRI and compressed sensing (CS) have been proposed. Selecting a sparsifying transform in CS, however, is non-trivial. Also, most commonly used transforms do not fully capture the rich structure of CMR images. More recently, deep learning (DL) methods have been shown to outperform CS methods. Some DL methods pose reconstruction as a “de-aliasing” problem, where the coil-combined aliased image is de-aliased using a convolutional neural network (CNN) [1]. These methods are computationally fast after the training phase, but ignore the multi-coil structure of the MRI data. In other methods, such as AUTOMAP [2], the training is used to learn the entire reconstruction process, from undersampled k-space to the final image, without any guidance from the forward model. The resulting CNN has fully connected layers, making it computationally prohibitive in practice. Also, such methods require extensive training to handle variations in the forward model. More recently, methods that explicitly use knowledge of the forward model [3] have gained significant attention. In these methods, the training is guided by explicit inclusion of prototypical forward models. For these methods, however, significant deviations between the forward models used for training versus testing can be problematic.

In this work, we recover CMR images from undersampled, multi-coil k-space data by iterating between data-fidelity enforcement and image denoising, i.e., via plug-and-play (PnP) recovery [4]. To fully exploit the rich spatiotemporal structure of CMR images, we use a DL-based denoiser trained specifically on CMR images. In the resulting “PnP-DL” method, training images are used to learn the denoiser, and these images are invariant to the forward model. Thus, the learned PnP-DL method is not biased by assumptions about the forward model, which can change significantly from training to testing. This unique feature of PnP-DL can lead to improved generalizability. We apply PnP-DL to reconstruct cardiac cine images and show that it outperforms traditional CS methods. More importantly, we demonstrate that PnP-DL can recover RT free-breathing cine images by using a denoiser trained on breath-held cine images. This capability is important because it is difficult, if not impossible, to obtain training data for RT free-breathing cine.

Corresponding author: Rizwan Ahmad (ahmad.46@osu.edu). This work was funded by NIH R01HL135489.

2. METHODS

2.1. PnP for Image Reconstruction

We consider the problem of recovering an image $\mathbf{x} \in \mathbb{C}^N$ from noisy k-space measurements. Most CS methods recover $\mathbf{x}$ by solving an optimization problem of the form

$$\hat{\mathbf{x}}_{\text{CS}} = \arg \min_{\mathbf{x}} \{f(\mathbf{x}) + \phi(\mathbf{x})\}, \quad (1)$$

where $f(\mathbf{x})$ is a (usually quadratic) data-fidelity term and $\phi(\mathbf{x})$ a convex, sparsity-promoting regularizer. A common algorithm to solve Eq. 1 is the “alternating directions method of multipliers” (ADMM), which reformulates Eq. 1 as

$$\hat{\mathbf{x}}_{\text{ADMM}} = \arg \min_{\mathbf{x}} \min_{\mathbf{v}} \{f(\mathbf{x}) + \phi(\mathbf{v})\} \quad \text{s.t.} \quad \mathbf{x} = \mathbf{v}, \quad (2)$$

and solves it using Algorithm 1, with $g(\mathbf{z}) = \text{prox}_{\nu\phi}(\mathbf{z})$ and

$$\text{prox}_{\nu\phi}(\mathbf{z}) \triangleq \arg \min_{\mathbf{x}} \left\{ \phi(\mathbf{x}) + \frac{1}{2\nu} \|\mathbf{x} - \mathbf{z}\|_2^2 \right\}, \quad (3)$$

which is known as the proximal operator.

Algorithm 1 The ADMM Algorithm

Require: $\nu > 0, \mathbf{x}_0 \in \mathbb{C}^N, \mathbf{u}_0 \in \mathbb{C}^N$

- 1: **for** $t = 1, 2, 3, \dots$ **do**
 - 2: $\mathbf{v}_t = \arg \min_{\mathbf{x}} \left\{ f(\mathbf{x}) + \frac{1}{2\nu} \|\mathbf{x} - (\mathbf{x}_{t-1} - \mathbf{u}_{t-1})\|_2^2 \right\}$
 - 3: $\mathbf{x}_t = g(\mathbf{v}_t - \mathbf{u}_{t-1})$
 - 4: $\mathbf{u}_t = \mathbf{u}_{t-1} + (\mathbf{v}_t - \mathbf{x}_t)$
 - 5: **end for**
 - 6: **return** $\hat{\mathbf{x}}_{\text{ADMM}} \leftarrow \mathbf{x}_t$
-

From the Bayesian perspective, the proximal operator in Eq. 3 can be interpreted as the maximum a posteriori estimator of $\mathbf{x}$ from the noisy measurement $\mathbf{z} = \mathbf{x} + \mathbf{w}$, where $\mathbf{w}$ is ν -variance additive white Gaussian noise and $\mathbf{x}$ has the probabilistic prior $p(\mathbf{x}) \propto \exp(-\phi(\mathbf{x}))$. Given this denoising interpretation, Bouman et al. [4] replaced the proximal update in ADMM with a call to a sophisticated image denoising subroutine like BM3D and observed a dramatic improvement in the quality of image recovery. The approach was called “plug-and-play” (PnP) because essentially any image denoiser $g(\cdot)$ can be “plugged into” ADMM, including those based on DL.

2.2. Learned Denoiser

Compared to PnP methods that use generic denoisers (e.g., BM3D or BM4D), we propose to use a DL denoiser specifically trained for cardiac cine. First, it has been shown that application-specific DL denoisers can outperform generic denoisers [5]. Second, CNN-based denoisers can be efficiently implemented on a GPU, unlike generic denoisers (e.g., BM4D), which can be prohibitively slow for the image sizes encountered in MRI. Third, most generic denoisers are designed only for real-valued 2D images, leaving limited options for complex-valued or higher-dimensional images. In contrast, DL denoisers can be trained and implemented for images in any domain or dimension.

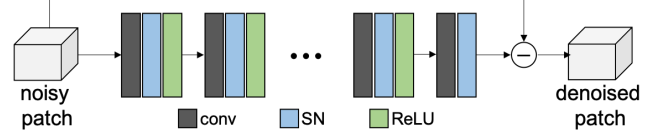


Fig. 1. The spatiotemporal DL denoiser used in this work. Here, conv, SN, and ReLU stand for 3D convolution, spectral normalization, and rectified linear unit, respectively.

The DL-based denoiser used in this work has the architecture shown in Fig. 1. The denoiser is constructed with five 3D-convolution layers, each with $64 \ 3 \times 3 \times 3$ kernels. Spectral normalization (SN) [6] was used to control the Lipschitz constant of each convolution layer and to provide some control of the overall Lipschitz constant of the network. We observed that the use of SN was critical for the stability of PnP [7]. The denoiser was trained using spatiotemporal patches extracted from high-quality, complex-valued cine images.

3. EXPERIMENT AND RESULTS

3.1. Experimental Setup

For training, we acquired 50 fully sampled, breath-held cine datasets from eight healthy volunteers. To promote diversity in the training data, 28 short-axis, seven two-chamber, seven three-chamber, and eight four-chamber views were acquired. The reference images were reconstructed by taking the inverse FFT along phase and frequency encoding directions and combining the resulting coil images using the method [8] by Walsh et al. Then, complex-valued i.i.d. Gaussian noise was added to these “noise-free” reference images to simulate noisy images with an SNR of 26 dB. For training, we cropped each image into patches of size $55 \times 55 \times 15$, with the last dimension representing time. The resulting noisy and noise-free patches were assigned as input and target output when training the DL denoiser. We fed the real and imaginary parts into separate channels, set the minibatch size to four, and used the ADAM optimizer with a learning rate of 10^{-4} over 500 epochs on an NVIDIA GPU (GeForce RTX 2080 Ti).

For testing, nine breath-held datasets were collected from four different healthy volunteers, with four in the short-axis view, two in the two-chamber view, two in the four-chamber view, and one in the axial view. Note, the axial view was not included in the training datasets. All datasets were retrospectively undersampled at three different acceleration rates: $R = 6, 8$, and 10 . We also collected ten RT free-breathing datasets with prospective undersampling at $R = 9$. Pseudo-random Cartesian masks were used to perform undersampling; an example is shown in Fig. 2. Before reconstruction, all datasets were compressed to 12 virtual coils for faster processing. For PnP-DL, the reconstruction followed Algorithm 1, with $f(\mathbf{x})$ representing the SENSE-based forward model. The coil sen-

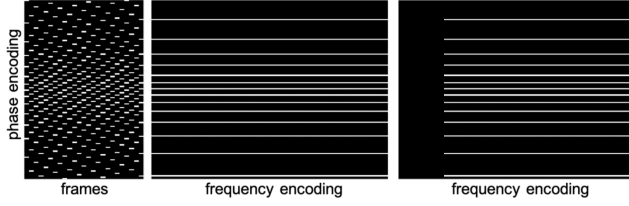


Fig. 2. A representative sampling pattern. Left: phase encoding and frame (time) dimensions shown for one frequency encoding location. Middle: phase encoding and frequency encoding dimensions shown for the last frame. Right: same as the middle column but with asymmetric echo, which was only used in prospective undersampling.

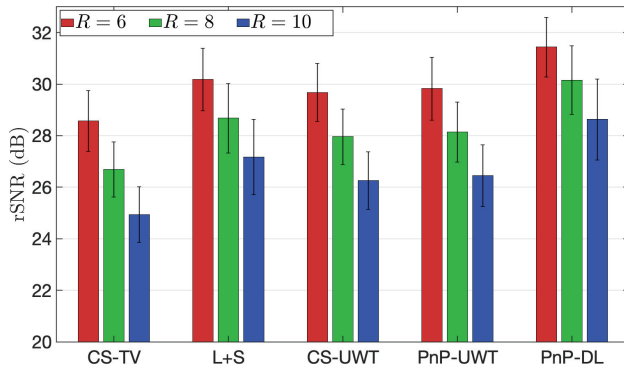


Fig. 3. Reconstruction SNR (rSNR) for breath-held datasets retrospectively undersampled at $R = 6, 8$, and 10 .

sitivity maps were estimated using the method [8] by Walsh et al. For the recruitment and consent of human subjects used in this study, the ethical approval was given by an Internal Review Board (2005H0124) at The Ohio State University.

3.2. Results from Breath-held Data

In the study with retrospective undersampling, we evaluated the reconstructed images using reconstruction SNR (rSNR), defined as $20 \log_{10}(\|x\|_2 / \|x - \hat{x}\|_2)$, where $\hat{x}$ and x represent reconstructed and reference cine images, respectively. The results from PnP-DL and PnP-UWT were compared to those from CS-TV [9], CS-UWT [10], and L+S [11], which represent CS with spatiotemporal finite difference as the sparsifying transform, CS with undecimated wavelet transform (UWT) as the sparsifying transform, and low-rank plus sparse reconstruction, respectively. For PnP-UWT, denoising was accomplished by soft-thresholding the UWT coefficients. Therefore, CS-UWT and PnP-UWT solve the same optimization problem, but CS-UWT used bFISTA [10] while PnP-UWT used ADMM from Algorithm 1. PnP-DL also used ADMM, but with the trained DL-based denoiser shown in Fig. 1.

Fig. 3 shows rSNR values aggregated over eight datasets

	CS-TV	L+S	CS-UWT	PnP-UWT	PnP-DL
$R = 6$	30.6	32.1	31.6	31.9	33.5
$R = 8$	29.0	31.1	30.2	30.5	32.5
$R = 10$	27.1	29.4	28.5	28.8	31.3

Table 1. Reconstruction SNR (rSNR) for the retrospectively undersampled dataset collected in the axial view.

and for three acceleration rates. The dataset collected in the the axial view was not included. As expected, the performances of PnP-UWT and CS-UWT were similar. The average advantage of PnP-DL over L+S (second best) was 1.3 dB, 1.5 dB, and 1.5 dB for $R = 6, 8$, and 10 , respectively.

Table 1 compares rSNR of the ninth dataset, which was collected in the axial view. Note, the axial view was not included when training the denoiser. Even for this “unseen” view, PnP-DL maintains its advantage over the CS methods. The superiority of PnP-DL is also evident from the example recoveries in Fig. 4, which show that PnP-DL did better in preserving the mitral valve and papillary muscles (yellow arrows), as well as in suppressing artifacts (green arrows).

3.3. Results from Free-breathing Data

In the study with prospective undersampling, we reconstructed ten RT cine datasets collected under free-breathing conditions. Each dataset was reconstructed using PnP-DL, PnP-UWT, and L+S. To evaluate the reconstructed images, each cine series was blindly reviewed by an expert, with seven years of experience in CMR, who assigned a qualitative score on a five-point Likert scale (1: non-diagnostic, 2: poor, 3: adequate, 4: good, 5: excellent). L+S, PnP-UWT, and PnP-DL received average scores of 3.2, 3.5, and 3.9, respectively. Also, in nine out of ten cases, the PnP-DL recoveries were deemed best among the three reconstruction methods. As seen in Fig. 5, which shows a representative frame from one of the RT free-breathing datasets, PnP-DL was able to preserve details lost with PnP-UWT (green arrows) and was more effective in removing the noise that is evident in the L+S recoveries (yellow arrows).

4. CONCLUSIONS

In this work, we proposed and validated a reconstruction method for real-time cardiac MRI. The method utilizes an application-specific deep-learning denoiser within the plug-and-play framework. The learning process, which entails training a denoiser on spatiotemporal image patches, is totally decoupled from the forward model. Our preliminary results suggest that PnP-DL outperforms CS in both quantitative and qualitative assessments. More importantly, PnP-DL was also effective in reconstructing real-time, free-breathing cine images even when the denoiser was trained using only breath-held cine images.

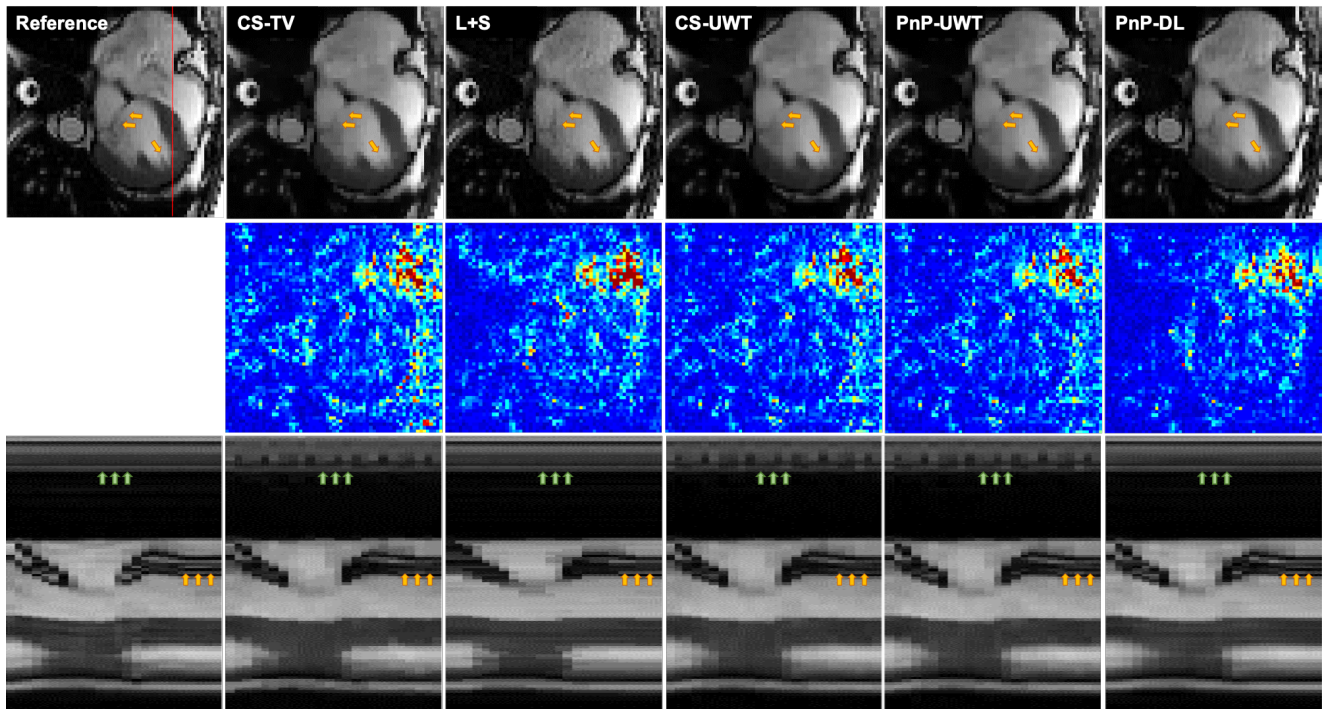


Fig. 4. Top row: A representative frame from the retrospectively undersampled dataset collected in the axial view. The arrows highlight the details that are better preserved in PnP-DL. Middle row: Error map with five-fold amplification. Bottom row: Temporal profile along the red line shown on the Reference image in the top row.

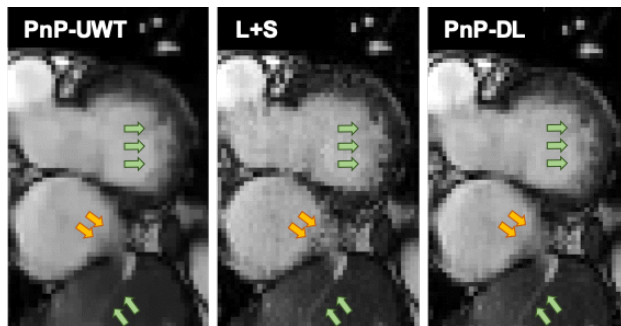


Fig. 5. A representative frame from an RT dataset.

5. REFERENCES

- [1] Hyun et al., “Deep learning for undersampled MRI reconstruction,” *PMB*, vol. 63, no. 13, pp. 135007, 2018.
- [2] Zhu et al., “Image reconstruction by domain-transform manifold learning,” *Nature*, vol. 555, no. 7697, pp. 487, 2018.
- [3] Hammernik et al., “Learning a variational network for reconstruction of accelerated MRI data,” *MRM*, vol. 79, no. 6, pp. 3055–3071, 2018.
- [4] Venkatakrishnan et al., “Plug-and-play priors for model based reconstruction,” in *2013 IEEE GlobalSIP*. IEEE, 2013, pp. 945–948.
- [5] Zhang et al., “Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising,” *IEEE TIP*, vol. 26, no. 7, pp. 3142–3155, 2017.
- [6] Miyato et al., “Spectral normalization for generative adversarial networks,” *preprint arXiv:1802.05957*, 2018.
- [7] Ryu et al., “Plug-and-play methods provably converge with properly trained denoisers,” *preprint arXiv:1905.05406*, 2019.
- [8] Walsh et al., “Adaptive reconstruction of phased array MR imagery,” *MRM*, vol. 43, no. 5, pp. 682–690, 2000.
- [9] Lustig et al., “Sparse MRI: The application of compressed sensing for rapid MR imaging,” *MRM*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [10] Ting et al., “Fast implementation for compressive recovery of highly accelerated cardiac cine MRI using the balanced sparse model,” *MRM*, vol. 77, no. 4, pp. 1505–1515, 2017.
- [11] Otazo et al., “Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components,” *MRM*, vol. 73, no. 3, pp. 1125–1136, 2015.