RARE: Image Reconstruction using Deep Priors Learned without Ground Truth

Jiaming Liu, Student Member, IEEE, Yu Sun, Student Member, IEEE, Cihat Eldeniz, Weijie Gan, Hongyu An, and Ulugbek S. Kamilov, Member, IEEE

Abstract—We propose a new method for image reconstruction that systematically enforces data consistency while also exploiting deep-learning-based image priors. The proposed regularization by artifact removal (RARE) specifies the prior through a convolutional neural network (CNN) trained to remove image-domain artifacts. Additionally, to address the lack of artifact-free training examples, the CNN in RARE is trained to restore images by looking only at corrupted examples without any ground-truth data. This makes RARE broadly applicable to challenging medical imaging problems, where it is nearly impossible to have clean training data. We validate RARE on both synthetic and experimentally collected datasets by reconstructing a free-breathing whole-body MRI scans into ten respiratory phases. The results show that RARE can form high-quality 4D images from severely undersampled measurements corresponding to acquisitions of about 1 minute in duration. Our results also highlight the potential of integrating both the forward-model and the data-driven CNN prior to yield better performance compared to purely model-based and purely CNN-based approaches.

Index Terms—Imaging inverse problems, regularization by denoising, plug-and-play priors, deep learning, MRI.

I. INTRODUCTION

IMAGE reconstruction from a set of undersampled observations is central in various imaging applications. The reconstruction is often formulated as an optimization problem

\[
\hat{x} = \arg \min_x f(x) \quad \text{with} \quad f(x) = g(x) + h(x),
\]

where \(x\) denotes the unknown image, \(g\) is the data-fidelity term that relies on the forward model to impose data-consistency, and \(h\) is the prior image used to regularize the solution for ill-posed problems. Over the years, many regularizers have been proposed as priors, including those based on transform-domain sparsity, low-rank penalty, and dictionary learning [1]-[7].

There has been considerable recent interest in applying deep learning for solving imaging inverse problems [8]-[16]. A widely used strategy is based on training a deep convolutional neural network (CNN) in a supervised fashion to learn a regularized inverse of the forward model. However, the simplicity and convenience of the conventional strategy comes with two important limitations: (a) it does not explicitly leverage the physics of the imaging system characterized by the forward model; (b) it requires a large training dataset containing “clean” ground-truth examples. There have been meaningful recent attempts to address both issues. One common approach to address the first issue introduces the forward model into the CNN architecture as a pre-defined block [17]-[21], while another uses pre-trained CNN image denoisers as priors within iterative image reconstruction [22]-[33]. On the other hand, several recent techniques, such as Noise2Noise [34] and its extensions [35]-[37], have attempted to address the second issue in the context of learning CNNs for image restoration.

In this paper, we propose a new regularization by artifact removal (RARE) method that jointly addresses both limitations above. RARE systematically enforces data-consistency with the measured data while also leveraging the power of deep learning through a CNN prior, trained without any ground-truth. We experimentally validate the method on a highly ill-posed inverse problem in free-breathing motion-compensated 4D MRI, where it is nearly impossible to obtain the ground-truth data. To the best of our knowledge, RARE is the first method proposing an integration of a CNN prior trained without ground-truth into the regularized inversion. While our experimental validations are focused on MRI, we note that RARE is broadly applicable to other biomedical imaging modalities. Our key contributions are summarized as follows:

• We extend the regularization by denoising (RED) framework [38] beyond simple additive white Gaussian noise (AWGN) removal. Specifically, we propose to use any artifact-removal CNN as an image prior within regularized inversion, which leads to a framework that is fully compatible with the conventional deep-learning reconstruction.

• We show that the ability to leverage any artifact-removal CNN naturally leads to a new learning paradigm that requires no ground-truth data. Inspired by Noise2Noise [34], we propose to train the CNN by mapping pairs of low-quality images with distinct artifact patterns.

• We validate RARE both quantitatively on synthetic and qualitatively on experimentally acquired data by reconstructing 4D MR images from highly undersampled k-space measurements. Our results highlight the ability of
RARE to significantly improve imaging quality compared to purely model-based and purely CNN-based approaches.

This paper is organized as follows. In Section II, we present the background material on imaging inverse problems. In Section III, we present the algorithmic details of RARE, including our strategy for training CNNs without ground-truth data. In Section IV, we provide simulations and experiments that illustrate the excellent performance of RARE on real MRI data. Section V concludes the paper.

II. Background

A. Inverse problem formulation

Consider a linear inverse problem of recovering an unknown image \( x \in \mathbb{C}^n \) from its noisy measurements \( y \in \mathbb{C}^m \) specified by the linear system

\[
y = Hx + e
\]

where the forward operator \( H \in \mathbb{C}^{m \times n} \) characterizes the response of the system, and \( e \in \mathbb{R}^m \) is AWGN. For example, in compressed sensing (CS) MRI [39], [40] the goal is to recover an image \( x \) from sparsely-sampled Fourier measurements. In the special case of parallel MRI [41] with a dynamic object, the forward model can be represented as

\[
H_t^{(i)} = P^{(i)}FS_i,
\]

where \( F \) denotes the Fourier transform operator, \( S_i \) is the matrix containing the per-pixel sensitivity map for the \( i \)th coil, and \( P^{(i)} \) is the k-space sampling matrix at time \( t \). If the k-space samples lie on the Cartesian grid, the fast Fourier transform (FFT) provides an efficient implementation of \( F \) and its adjoint. If non-Cartesian k-space sampling is used, then a nonuniform FFT (NUFFT) is needed [42]. When the k-space is undersampled, the inverse problem is ill-posed even in the absence of noise; therefore, it is common to formulate the solution as the regularized inversion.

Conventional methods for reconstruction from underdetermined measurements rely on the sparsity of typical images in some transform domain, such as wavelet [3], total variation (TV) [1], or exploit the spatio-temporal correlations, such as in the k-t Sparse and Low-Rank (kt-SLR) model [43]. Such regularizers generally result in non-smooth optimization problems. A variety of methods such as PGM [28], [44] and ADMM [45] have been developed for efficient minimization of nonsmooth functions, without differentiating them, by using the proximal operator, defined as

\[
\text{prox}_{\mu h}(z) := \arg \min_x \left\{ \frac{1}{2} ||x - z||_2^2 + \mu h(x) \right\},
\]

where \( \mu > 0 \) is a parameter. Note that the proximal operator can be interpreted as the regularized image denoiser for AWGN with noise of variance of \( \mu \). This mathematical equivalence led to the development of more general denoiser-based algorithms such as PnP [46] and RED [38], which we will discuss next.

Algorithm 1 Regularization by Artifact Removal

1: input: \( x^0 = s^0, \gamma > 0, \tau > 0, \rho = 1e^{-6}, \beta \in (0, 1), \) and \( \{q_k\}_{k \in \mathbb{N}} \)
2: for \( k = 1, 2, \ldots \) do
3: \( x^k \leftarrow s^{k-1} - \gamma G(s^{k-1}) \)
4: \( y \leftarrow x^k + (q_{k-1} - 1)/q_k(x^k - x^{k-1}) \)
5: \( s^k \leftarrow x^k + (\|G(x^k)\|_2 > \|G(s^{k-1})\|_2 \text{ do} \)
6: \( \gamma = \beta \gamma \)
7: \( x^k \leftarrow s^{k-1} - \gamma G(s^{k-1}) \)
8: if \( \gamma < \rho \) then
9: break
10: end if
11: end while
12: end for

B. Using denoisers as image priors

Inspired by the equivalence of the proximal operator to a denoiser, Venkatakrishnan et al. [46] have introduced the plug-and-play priors (PnP) framework by replacing the proximal map with an arbitrary image denoiser \( D_\sigma(\cdot) \), where \( \sigma > 0 \) controls the strength of denoising. This simple replacement enables PnP to regularize the problem by using advanced denoisers, such as BM3D [47] and DnCNN [48], that do not correspond to any explicit \( h \). Numerical experiments have shown that PnP achieves the state-of-the-art performance in many applications. While the original formulation is based on ADMM, other PnP algorithms have been developed using PGM and primal-dual splitting [27], [28], [49].

The RED framework is an alternative to PnP that uses an image denoiser to formulate an explicit regularization function

\[
h(x) = \frac{\tau}{2} x^T (x - D_\sigma(x)),
\]

where \( \tau > 0 \) is the regularization parameter. When the denoiser is locally homogeneous and has a symmetric Jacobian [31], [38], the gradient of the RED regularizer has a simple form

\[
\nabla h(x) = \tau (x - D_\sigma(x)),
\]

which enables the efficient implementation of RED. The excellent performance of RED under a CNN denoiser have been demonstrated in super-resolution [32], phase retrieval [33], [39], compressed sensing [30].

C. Image reconstruction using deep learning

In the recent years, deep learning has gained popularity for solving imaging inverse problems. Instead of explicitly defining an image prior, the majority of deep-learning algorithms formulate image reconstruction as an artifact-removal problem in the image domain [9], where a CNN is trained to map a set of corrupted images \( \{\tilde{x}_j\} \) to their clean target versions \( \{x_j\} \) via the following optimization

\[
\arg \min_{\theta} \sum_j \mathcal{L}(f_\theta(\tilde{x}_j), x_j),
\]

where \( \mathcal{L} \) is a loss function that measures the difference between the denoised and the clean images.
where \( f_\theta(\cdot) \) represents the CNN parametrized by \( \theta \), under the loss function \( L \). The popular loss functions include the \( \ell_2 \)-norm and \( \ell_1 \)-norm. In practice, (7) can be optimized using the family of stochastic gradient descend (SGD) methods, such as adaptive moment estimation (ADAM) [51]. For example, in the context of CS-MRI, previous methods have trained \( f_\theta(\cdot) \) for mapping a zero-filled image \( \tilde{x} \) to a reconstruction \( x \) from a fully-sampled ground-truth data [52], [53].

Noise2Noise [34] is an elegant solution to the problem when the ground-truth data is not available, where \( f_\theta(\cdot) \) is trained to learn a mapping between pairs of images that have random and independent degradations. For example, given a family of degraded images \( \{x_j + n_j, \tilde{x}_j + n'_j\} \), with random independent noise vectors \( n_j \) and \( n'_j \), one can replace \( \{x_j, \tilde{x}_j\} \) in (7) by the degraded pair and still learn the model \( f_\theta(\cdot) \) that predicts a clean image. The key assumption for this strategy is to have the expected value of the noisy input to be equal to the clean signal [34]. This central idea has been recently expanded to other restoration settings [35]–[37].

III. REGULARIZATION BY ARTIFACT REMOVAL

A. Algorithmic framework

The CNN priors in the traditional formulation of RED are typically designed for AWGN removal. This is inconvenient for applications where it is difficult to directly synthesize the AWNG corrupted images for training the CNN. RARE (summarized in Algorithm 1) circumvents this problem by replacing the denoiser \( D_\sigma(\cdot) \) in RED by an artifact-removal network \( R_\theta(\cdot) \), where \( \theta \) denotes the weights of the CNN. RARE can thus leverage any existing image restoration CNN designed for removing artifacts specific to some imaging modality.

Let \( H^\dagger \) denote the pseudoinverse of the forward operator \( H \). Then, given an initial solution provided by the traditional artifact removal network \( x^0 = R_\theta(H^\dagger y) \), RARE iteratively refines it by infusing information from both the gradient of the data-fidelity term \( \nabla g \) and the residual of the restoration CNN \((x - R_\theta(x))\). Specifically, RARE aims to find the solution \( x^* \) lying at the zero set of the operator

\[
\text{zer}(G) := \{x \in \mathbb{C}^n : G(x) = 0\} \quad \text{with} \quad G(x) := \nabla g(x) + \tau (x - R_\theta(x)),
\]

where \( \tau > 0 \) is the regularization parameter. Consider the following two sets

\[
\text{zer}(\nabla g) := \{x \in \mathbb{C}^n : \nabla g(x) = 0\} \quad \text{and} \quad \text{fix}(R_\theta) := \{x \in \mathbb{C}^n : x = R_\theta(x)\},
\]

where \( \text{zer}(\nabla g) \) is the set of all critical points of the data-fidelity term and \( \text{fix}(R_\theta) \) is the set of all fixed points of the artifact-removal CNN. Intuitively, \( \text{fix}(R_\theta) \) correspond to all the vectors that are artifact-free according to the CNN, while \( \text{zer}(\nabla g) \) correspond to the vectors that are consistent with the measured data. If \( x^* \in \text{zer}(\nabla g) \cap \text{fix}(R_\theta) \), then \( G(x^*) = 0 \) and \( x^* \) is one of the solutions of RARE. Hence, any vector that is consistent with the data for a convex \( g \) and artifact free according to \( R_\theta \) is in the solution set of RARE. On the other hand, when \( \text{zer}(\nabla g) \cap \text{fix}(R_\theta) = \emptyset \), then \( x^* \in \text{zer}(G) \) corresponds to a trade-off between the two sets, explicitly controlled via \( \tau > 0 \).

The simplest algorithm for computing (8) can be obtained by running the following fixed-point algorithm

\[
x^k \leftarrow x^{k-1} - \gamma G(x^{k-1}),
\]

where \( \gamma > 0 \) is the step-size. RARE adds two notable extensions. First, we use the backtracking line-search strategy [54] to automatically adjust \( \gamma > 0 \). The line-search starts with some large step size, which is subsequently decreased by a factor \( \beta \in (0, 1) \) to ensure monotonic decrease \( ||G(x^k)||_2 \leq ||G(x^{k-1})||_2 \). Second, we equip RARE with Nesterov’s acceleration [55] via \( \{q_k\} \) for better convergence. The acceleration terms are updated in the usual way as

\[
q_k \leftarrow \frac{1}{2} (1 + \sqrt{1 + q_{k-1}^2}).
\]

Note that when \( q_k = 1 \) for all \( k \), the algorithms reverts to the usual gradient method without acceleration.

---

(a) Artifact removal network
(b) Regularization by artifact removal (RARE)

Fig. 1: Illustration of the proposed RARE framework on the problem of reconstructing 10 motion phases from a single free-breathing MRI acquisition. (a) The artifact-removal CNN is trained on complex-valued 3D images \((x, y \text{ for space and } p \text{ for phase})\) without any ground-truth. (b) The method combines the information from the CNN with that from the forward model.

---
B. Ground-truth free training

In many biomedical imaging applications it is straightforward to acquire several random and independent views of the same object. Consider a set of unknown images \( x_j \), each acquired multiple times

\[
y_j = H_{ij} x_j + e_{ij},
\]

where \( i \) is the index of the specific acquisition and \( H_{ij} \) denotes the specific forward operator. One can then generate a dataset of degraded images by simply applying the pseudoinverse of each forward operator

\[
\hat{x}_{ij} = H^†_{ij}y_{ij}.
\]

Inspired from Noise2Noise, one can then train the artifact removal network with the following empirical risk minimization

\[
\arg\min_\theta \sum_{i,j,i′,j′} \mathcal{L}(R_\theta(\tilde{x}_{ij}), \tilde{x}_{i′j′}),
\]

where the goal is to simply map pairs of artifact-contaminated images in (11). Underlying assumption of this training strategy is that the expected value of the images \( \{x_j\} \) matches the ground-truth vector \( x_j \). While this assumption might seem idealized for some practical applications, our experiments corroborate its excellent performance in the context of a heavily ill-posed 4D MR image reconstruction under object motion. In particular, we show that the best performance is achieved by combining a pre-trained \( R_\theta(\cdot) \) with several iterations of RARE updates that also include \( \nabla g(\cdot) \).

C. Implementation for 4D MRI

The RARE framework for accelerated free-breathing MRI and the architecture for the artifact-removal CNN used in this paper are shown in Figure 1. Since the multi-coil nonuniform inverse fast Fourier transform (MCNUFFT) data are 4D complex-valued images \( x, y, z \) for space and \( p \) for respiratory phases), we used a 3D residual CNN, for dimensions \( x, y, p \), which is a simplified version of the popular DnCNN denoiser [48] as the artifact removal prior. The network consists of ten layers with three different blocks. The first block is a composite convolutional layer, consisting of a normal 3D convolutional layer and a rectified linear units (ReLU) layer, and the second block is a sequence of 8 composite convolutional layers, each having 64 filters. Those composite layers further process the feature maps generated by the first part. The third block of the network, a single convolutional layer, generates the final output image with two channels including the real and imagery part of the image. Every convolution is performed with a stride \( = 1 \), so that the intermediate feature maps have the same spatial size as the input to the CNN. It is worth noting that we also considered the U-Net [53], [56] architecture for the artifact removal CNN. However, our experiments showed no significant quality improvements from U-Net, which has a higher computational complexity than DnCNN.

Our implementation of free-breathing 4D MRI is based on Consistently Acquired Projections for Tuned and Robust
Fig. 4: Visual illustration of results from experimentally collected measurements on three distinct patients using 400, 800, and 1200 spokes (about 1, 2, and 3 minute scans) from top to bottom, respectively. The leftmost column shows the compressed sensing (CS) reconstruction using 2000 spokes, which is provided as the gold-standard. The following images correspond to the results of k-t SLR, CS, U-Net3D, RED-DnCNN*, and RARE, respectively. This figure highlights the state-of-the-art performance of RARE on 4D MRI reconstruction.

Estimation (CAPTURE) method [57], which retrospectively bins the k-space data into several motion phases. The key observation is that the binning leads to different k-space coverage patterns for different acquisition times, leading to distinct artifact patterns. Based on this observation, we learned $R_\theta(\cdot)$ by mapping pairs of complex 3D MR volumes $(x, y \text{ space } + p \text{ phase})$ acquired over different acquisition times to one another, without using artifact-free ground-truth images. Let $(\hat{x}_{ij}, \hat{x}_{i'j'})$ in (12) denotes a set of pairs of corrupted images from MCNUFFT reconstructions. The volumetric image of the whole body is obtained by applying the model $R_\theta(\cdot)$ slice by slice along $z$. We used a mixed $\ell_1/\ell_2$ loss for training (see eq. (13) below). This type of mixture loss functions have been shown to be effective in natural image restoration [58]. Specifically, the training is carried out by minimizing the loss over a training set as follows

$$\arg \min_\theta \sum_{i,j,i',j'} L(R_\theta(\hat{x}_{ij}), \hat{x}_{i'j'})$$

with

$$L = \alpha L_{\ell_1} + (1 - \alpha) L_{\ell_2}, \quad (13)$$

where $\alpha \in (0, 1)$ is used to adjust the strength of the $\ell_1/\ell_2$ loss. In addition, the diversity of artifact patterns in the training data is further increased by using for each input MCNUFFT images, an output MCNUFFT image of the same or higher number of radial spokes. For example, given a dataset with a 5 minute acquisition, we split it into 1, 2, 3, 4, and 5 minute MCNUFFT images that are all used for training the network.

IV. EXPERIMENTAL RESULTS

RARE was qualitatively and quantitatively evaluated by experiments using real in vivo liver data sets and synthetic liver MRI data corresponding to different sampling trajectories. Retrospective downsampling from acquired datasets was used to compare the reconstruction results for each method. All training and testing experiments in this paper were performed on a machine equipped with an Intel Xeon Gold 6130 Processor and an NVIDIA GeForce RTX 2080 Ti GPU.

A. In-vivo liver dataset

All acquisition processes were performed on a 3T PET/MRI scanner (Biograph mMR; Siemens Healthcare, Erlangen, Germany). The data was collected using the CAPTURE method, a recently proposed T1-weighted stack-of-stars 3D spoiled gradient-echo sequence with fat suppression that has consistently acquired projections for respiratory motion detection [57].
Fig. 5: The effect of additional data on the performance of RARE and U-Net3D for patient 5. Note that U-Net3D is trained to map the 400 line MCNUFFT images to the gold-standard provided by the 2000 line CS reconstruction (see leftmost column). On the other hand, the CNN in RARE was trained by using only MCNUFFT data. This figure highlights the ability of RARE to achieve better results with less data.

Upon the approval of our Institutional Review Board, multi-channel liver data from 10 healthy volunteers and 6 cancer patients were used in this paper. The acquisition parameters were as follows: TE/TR = 1.69 ms/3.54 ms, FOV = 360 mm × 360 mm, in-plane resolution = 1.125 × 1.125 mm, partial Fourier factor = 6/8, number of radial lines = 2000, slice resolution = 50%, slices per slab = 96 with a slice thickness of 3 mm, total acquisition time = about 5 minutes (slightly longer for larger subjects). We drop out the first 10 spokes during reconstruction in order to make sure the acquired signal reaches a steady state. Our free-breathing MRI data was then subsequently binned into 10 respiratory phases, and thus each phase view was reconstructed with 199 spokes accordingly. The image dimensions are 320 × 320 × 10 × 96 for each subject. The coil sensitivity maps were estimated from the central radial k-space spokes of each slice and were assumed to be known during experiments. Apodization was applied by using a Hamming window that covers this range in order to avoid Gibbs ringing. We then implemented MCNUFFT on those individual coil element data. We used the first 8 healthy subjects for all convolutional neural networks training and 1 for validation. The complex-valued 4 dimensional images were cropped along x-y axes in order to acquire images of parts of the anatomy. The real and imaginary parts of these images were separated into two channels as the inputs of the artifact removal neural networks. Thus, the training or validation data had dimensions in slices × phases × rows × columns × channels as 96 × 10 × 320 × 320 × 2 for each subject. For testing the proposed framework, we picked 5 slices of the remaining 1 healthy subject and of each patient. We conducted the experiments for various acquisition durations (i.e. 400, 800, 1200, and 1600 radial spokes of the original k-space data), corresponding to scan times of 1, 2, 3, 4 minutes, respectively. We note that the multi-coil sensitivity maps for each downsampling rate would be recalculated accordingly. We selected the compressed sensing (CS) variant of CAPTURE using 2000 spokes data as our “gold standard” reference [57]. CAPTURE-CS relies on two types of regularization: 1) Total Variation (TV) that imposes piece-wise smoothness across different respiratory phases and 2) Total Generalized Variation (TGV) regularization that imposes higher-order spatial piece-wise smoothness. Both regularizers are controlled via the regularization parameters that are optimized for the best performance.

B. Synthetic liver dataset

Due to the lack of fully sampled k-space data, we selected 2000 spokes CS as the ground truth images in order to quantitatively validate RARE against several baseline methods. Similarly to in-vivo data, we used the CS reconstruction of the first 8 healthy subjects for training, 1 for validation, and the rest were used for testing.

C. Baseline algorithms for comparison

We compared RARE with several image reconstruction methods such as CS, k-t SLR, and U-Net3D (x-y-phase), as well as the RED-DnCNN\(^1\). The k-t SLR algorithm\(^1\) exploits the global low-rank structure of the spatiotemporal signal [43]. Specifically, the problem is posed as a matrix recovery problem using a low-rank penalty. A sparsity prior is also used to improve the reconstruction quality. U-Net and its variants have been successfully used for many image-to-image prediction tasks. We implemented our own version of U-Net3D based on the

\(^1\)The source code for k-t SLR are available in http://user.engineering.uiowa.edu/~jcb/software.html.
Fig. 6: Quantitative evaluation of RARE on the synthetic 4D MRI data: (a) visual comparison of the fifth respiratory phase at 10× acceleration with 30 dB noise. The yellow box provides an enlarged view of the image. The green box provides the error residual, which was amplified 10× for better visualization. (b) The radial sampling masks used in this simulation, corresponding to 10×, 6.6×, and 5× accelerations of acquisition. (c) and (d) SNR and SSIM values along respiratory phases for several reference methods. This figure highlights the competitive quantitative performance of RARE compared to several algorithms. Note that the CNN in RARE was trained without the ground truth images used for synthesizing the k-space data.

code that was originally developed in [12]. Furthermore, it has an analysis and a synthesis path each with four resolution steps. In the analysis path, each layer contains three $3 \times 3 \times 3$ convolutions each followed by a ReLu, and then a $1 \times 2 \times 2$ max pooling with strides of two in $x \cdot y$ axes while keeping the phase dimension not changed. In the synthesis path, each layer consists of an up-convolution of $2 \times 2 \times 2$ by strides of two in each dimension, followed by three $3 \times 3 \times 3$ convolutions each followed by a ReLU. We treated the 5-min CS reconstruction as the “ground truth” image and trained the model separately for each different acquisition lengths. For the Gaussian denoiser used in RED-DnCNN*, we adopted and modified the network architecture proposed in [59] to fit our problems. The CNN denoiser had seven layers, including 5 hidden layers with input and output layer. We used a supervised $\ell_2$ loss for both training the U-Net3D and DnCNN*. The algorithms are compared visually, as well as quantitatively. Two image evaluation metrics were used: signal-to-noise ratio (SNR) and structural similarity (SSIM). SSIM is computed as

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{\mu_x^2 + \mu_y^2 + c_1} \frac{(2\sigma_{xy} + c_2)}{(\sigma_x^2 + \sigma_y^2 + c_2)}$$  \hspace{1cm} \text{(14)}$$

where $\mu$ and $\sigma$ are the average and variance of the image, $c_1 = (k_1L)^2$, $c_2 = (k_2L)^2$ two variables to stabilize the division with weak denominator and $L = \text{dynamic range of the pixel-values}$.

D. Results on in-vivo data

Figure 2 illustrates reconstructions for 400 radial spokes (which corresponds to roughly 1 minute acquisition) for two patients. We compare RARE with the results of direct application of artifact removal CNNs. The comparison area was highlighted with a yellow box for each patient. The MCNUFFT reconstruction of the highlighted area is also shown, along with the gold-standard CS result using 2000 spokes. In the figure, we refer to the artifact removal CNN as N2N*. The U-Net3D was trained by mapping 400 spoke MCNUFFT reconstructions to the gold-standard CS images. RARE was initialized with the result of the N2N*. It is clear from Figure 2, that all methods yield significant improvements over the MCNUFFT reconstruction. However, a closer inspection shows that U-Net3D leads to image blurring, which reduces the contrast of the tumor pinpointed by the red arrow. On the other hand, the
N2N\textsuperscript{*} image still has some streaking artifacts that we attribute to training without clean data. RARE significantly boosts the performance of N2N\textsuperscript{*}, combining it with the forward model, which leads to sharper results. We also tested CS, kt-SLR, and RED-DnCNN\textsuperscript{*} methods on the same data, but omitted them from the figure, since CS and kt-SLR have even more streaking artifact compared to N2N\textsuperscript{*}, while RED-DnCNN\textsuperscript{*} provides very similar results with U-Net3D. We thus chose U-Net3D and N2N\textsuperscript{*} as the representative methods for this figure.

Figure 3 demonstrates the evolution of the solution generated by RARE over 9 additional iterations after initialization by N2N\textsuperscript{*}. The final result of RARE shown here is identical to the one in Figure 2. The optimal parameters of RARE in this experiment correspond to $\tau = 0.009$ and $\lambda = 0.98$. The enlarged regions of each testing representative patient are highlighted in the yellow box along with the differences and residual with respect to the 9th iteration. The difference images were amplified $20 \times$ for better visualization of artifact patterns. In general, we observed no significant improvements after 10 iterations of RARE. It is evident that RARE is able to reduce the streaking artifacts apparent in the N2N\textsuperscript{*} solution, while maintaining the sharpness.

Figures 4 and 5 present several visual results comparing the performance of RARE on different patients and acquisition lengths against several baseline methods, including CS, kt-SLR, U-Net3D, and RED-DnCNN\textsuperscript{*}. We conducted experiments for 400, 800, 1200, and 1600 radial spokes, corresponding to roughly 1, 2, 3, and 4 minute acquisitions, respectively. The solution of N2N\textsuperscript{*} was used as the initialization for RARE using the step size $\gamma = 0.98$, while the regularization parameter $\tau$ was respectively set to 0.009, 0.005, 0.002 and 0.001 for each undersampling rate. For RED-DnCNN\textsuperscript{*}, we trained the CNN denoiser for the removal of complex-valued AWGN with $\sigma \in \{0.5, 1, 1.5, 2\}$ and the same other parameters as RED. The parameters for CS were set as suggested in [57]. For kt-SLR, we fine tuned all the parameters for best visual results relative to the gold standard images, which were $\mu_1 = 10^{-6}$, $\mu_2 = 10^{-9}$, $\rho = 0.11$. Note that all methods achieve similar visual results for 1600 spokes, and thus we choose 400, 800, and 1200 spokes reconstruction as the representative examples in the figure. For 400 and 800 spokes reconstruction, kt-SLR and CS still show streaking artifact. Although U-Net3D and RED-DnCNN\textsuperscript{*} are able to remove these artifact, they also result in additional smoothing. RARE not only successfully removes the artifact, but also produce significantly clearer results. For 1200 spokes, all the algorithms show significant improvements, but RARE shows slightly better reconstruction results compared with other learning based methods. U-Net3D and RED-DnCNN\textsuperscript{*} still lead to blurring in some parts of the reconstructions while RARE retains high-frequency details and reconstructs more realistic texture with sharper edges. Figure 5 highlights a noticeable improvement in quality for 800 spokes for both RARE and U-Net3D, when compared to 400 spokes, but the quality remains relatively stable after 800 spokes. This figure highlights that RARE can consistently outperform U-

### Table I: Average SNR and SSIM obtained for several reference methods on synthetic liver data.

| Schemes | SNR | SSIM |
|---------|-----|------|
|         | 10× | 6.6× | 5×  | 10× | 6.6× | 5×  |
| Methods | 30 dB | 40 dB | 30 dB | 40 dB | 30 dB | 40 dB | 30 dB | 40 dB | 30 dB | 40 dB | 30 dB | 40 dB | 30 dB | 40 dB | 30 dB | 40 dB | 30 dB | 40 dB |
| ZF      | 14.84 | 14.97 | 16.19 | 16.28 | 18.43 | 18.59 | 0.763 | 0.768 | 0.801 | 0.805 | 0.855 | 0.861 |
| CS      | 23.18 | 23.52 | 24.83 | 25.43 | 26.33 | 27.37 | 0.944 | 0.948 | 0.955 | 0.961 | 0.968 | 0.972 |
| kt-SLR  | 23.24 | 23.61 | 24.79 | 25.32 | 26.51 | 27.29 | 0.945 | 0.947 | 0.956 | 0.960 | 0.967 | 0.972 |
| U-Net3D | 23.76 | 24.10 | 24.95 | 25.62 | 26.70 | 27.50 | 0.946 | 0.950 | 0.957 | 0.964 | 0.969 | 0.974 |
| RED-DnCNN\textsuperscript{*} | 24.10 | 24.36 | 25.74 | 25.99 | 27.47 | 28.09 | 0.952 | 0.956 | 0.964 | 0.971 | 0.972 | 0.976 |
| RARE    | 24.07 | 24.38 | 25.77 | 26.02 | 27.38 | 27.95 | 0.951 | 0.956 | 0.966 | 0.970 | 0.971 | 0.975 |
E. Results on synthetic data

For the experiments on synthetic data, we perform retrospective radial undersampling in order to simulate a practical single-coil acquisition scenario. The forward model \( y = Hx + e \), where \( e \) is an AWGN vector and \( H \) is a matrix corresponding to the radially-subsampled Fourier transform. We set the sampling rate to 10\%, 15\%, and 20\% of the original k-space data (corresponding to 10×, 6.6×, 5× acceleration) with AWGN corresponding to input SNR of 30 dB and 40 dB.

In this experiment, we trained the artifact removal network using pairs of two independent undersampled images corresponding to the sampling rate of 40\% with AWGN corresponding to 30 dB and 40 dB using the \( \ell_2 \) loss. For the CNN denoiser, it was trained for the removal of AWGN at four noise levels corresponding to \( \sigma \in \{1, 3, 5, 10\} \). For each experiment, we selected the denoiser achieving the highest SNR value. Several instances of U-Net3D were trained for each sampling rates and input SNR levels. For k-t SLR and CS, we performed grid search to identify the optimal parameters for the data. Since the synthetic ground truth images were real magnitude images, we only kept the real part of the gradient step during each iteration for both RARE and RED-DnCNN∗.

Figure 7 and Table I summarize the quantitative results in terms of SNR and SSIM. Overall, all methods achieved good performance, learning-based methods achieving the best performances. Moreover, RARE achieves similar reconstruction results with RED-DnCNN∗, which is impressive considering the fact that RARE doesn’t use the ground-truth in training. Figure 6 shows some representative visual results for the synthetic liver test data, using various reconstruction methods at 10× acceleration with input SNR = 30 dB. The error plots were magnified by 10× and highlighted with gray color map. In general, the reconstructions of learning based methods outperform the k-t SLR and CS. We observe that k-t SLR produces cartoon-like features in the image. Although CS shows high denoising quality with sharp edges due to sparsity enforcement, many details are not preserved compared with learning based methods.

V. CONCLUSION

We proposed RARE as an iterative method that combines a learned CNN prior with the data consistency in the under-sampled k-space domain. RARE offers best of model-based and learning-based worlds, while avoiding the need for ground-truth training datasets. Extensive experimental results show that the proposed algorithm provides competitive image quality compared to several baseline methods. While our experiments focused on MRI, the method is broadly applicable to many other imaging modalities, where it is easy to collect several distinct views of an object for training.

REFERENCES

[1] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” Physica D, vol. 60, no. 1-4, pp. 259–268, November 1992.
[2] M. A. T. Figueiredo and R. D. Nowak, “Wavelet-based image estimation: An empirical Bayes approach using Jeffreys’ noninformative prior,” IEEE Trans. Image Process., vol. 10, no. 9, pp. 1322–1331, September 2001.
[3] ——, “An EM algorithm for wavelet-based image restoration,” IEEE Trans. Image Process., vol. 12, no. 8, pp. 906–916, August 2003.
[4] Y. Hu, S. G. Lingala, and M. Jacob, “A fast majorize-minimize algorithm for the recovery of sparse and low-rank matrices,” IEEE Trans. Image Process., vol. 21, no. 2, pp. 742–753, February 2012.
[5] M. Elad and M. Aharon, “Image denoising via sparse and redundant representations over learned dictionaries,” IEEE Trans. Image Process., vol. 15, no. 12, pp. 3736–3745, December 2006.
[6] K. Degraux, U. S. Kamilov, P. T. Boufounos, and D. Liu, “Online convolutional dictionary learning for multimodal imaging,” in Proc. IEEE Int. Conf. Image Proc. (ICIP), Beijing, China, September 17-20, 2017, pp. 1617–1621.
[7] U. S. Kamilov, “A parallel proximal algorithm for anisotropic total variation minimization,” IEEE Trans. Image Process., vol. 26, no. 2, pp. 539–548, February 2017.
[8] A. Mousavi, A. B. Patel, and R. G. Baraniuk, “A deep learning approach to structured signal recovery,” in Proc. Allerton Conf. Communication, Control, and Computing, Allerton Park, IL, USA, September 30-October 2, 2015, pp. 1336–1343.
[9] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, “Deep convolutional neural network for inverse problems in imaging,” IEEE Trans. Image Process., vol. 26, no. 9, pp. 4509–4522, Sep. 2017.
[10] E. Kang, J. Min, and J. C. Ye, “A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction,” Medical Physics, vol. 44, no. 10, pp. e360–e375, 2017.
[11] J. C. Ye, Y. Han, and E. Cha, “Deep convolutional framelets: A general deep learning framework for inverse problems,” SIAM J. Imag. Sci., vol. 11, no. 2, pp. 991–1048, 2018.
[12] Y. Sun, Z. Xia, and U. S. Kamilov, “Efficient and accurate inversion of multiple scattering with deep learning,” Opt. Express, vol. 26, no. 11, pp. 14678–14688, May 2018.
[13] Y. Li, Y. Xue, and L. Tian, “Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media,” Optica, vol. 5, no. 10, pp. 1181–1190, Oct 2018.
[14] J. Yoo, S. Sabir, D. Heo, K. H. Kim, A. Wahab, Y. Choi, S. Lee, E. Y. Chae, H. H. Kim, Y. M. Bae, Y. Choi, S. Cho, and J. C. Ye, “Deep learning diffuse optical tomography,” IEEE Trans. Med. Imaging, pp. 1–1, 2019.
[15] A. Goy, G. Rughooobur, S. Li, K. Arthur, A. I. Akinwande, and G. Barbarasthis, “High-resolution limited-angle phase tomography of dense layered objects using deep neural networks,” Proceedings of the National Academy of Sciences, vol. 116, no. 40, pp. 19 848–19 856, 2019.
[16] E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, “Deep-storm: super-resolution single-molecule microscopy by deep learning,” Optica, vol. 5, no. 4, pp. 458–464, Apr 2018.
[17] U. S. Kamilov and H. Mansour, “Learning optimal nonlinearities for iterative thresholding algorithms,” IEEE Signal Process. Lett., vol. 23, no. 5, pp. 747–751, May 2016.
[18] E. Bostan, U. S. Kamilov, and L. Waller, “Learning-based image reconstruction via parallel proximal algorithm,” IEEE Signal Process. Lett., vol. 25, no. 7, pp. 989–993, Jul. 2018.
[19] W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, and X. Lu, “Denoising prior driven deep neural network for image restoration,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 10, pp. 2305–2318, Oct 2019.
[59] R. Ahmad, C. A. Bouman, G. T. Buzzard, S. H. Chan, E. Reehorst, and P. Schniter, “Plug and play methods for magnetic resonance imaging.” 
CoRR, vol. abs/1903.08616, 2019.