

DAILY QA

Finished before your first cup of coffee

MONTHLY QA

Never re-learn workflow again

ANNUAL QA

Confidence with fewer gray hairs

STEREOTACTIC

When millimeters matter most

ABSOLUTE DOSIMETRY

"Pinpoint" is our margin of error

PATIENT DOSIMETRY

Pre-treatment is just the beginning



QA Solutions.

Our expertise is at your service.

STANDARDIMAGING®



MORE THAN JUST QA PRODUCTS, FIND QA SOLUTIONS AT

www.standardimaging.com

HYDRA: Hybrid deep magnetic resonance fingerprinting

Pingfan Song^{a)}

Department of Electronic and Electrical Engineering, Imperial College, London, UK

Yonina C. Eldar

Faculty of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, Israel

Gal Mazor

Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa, Israel

Miguel R. D. Rodrigues

Department of Electronic and Electrical Engineering, University College London, London, UK

(Received 12 February 2019; revised 28 May 2019; accepted for publication 5 July 2019; published 10 September 2019)

Purpose: Magnetic resonance fingerprinting (MRF) methods typically rely on dictionary matching to map the temporal MRF signals to quantitative tissue parameters. Such approaches suffer from inherent discretization errors, as well as high computational complexity as the dictionary size grows. To alleviate these issues, we propose a HYbrid Deep magnetic Resonance fingerprinting (HYDRA) approach, referred to as HYDRA.

Methods: HYDRA involves two stages: a model-based signature restoration phase and a learning-based parameter restoration phase. Signal restoration is implemented using low-rank based de-aliasing techniques while parameter restoration is performed using a deep nonlocal residual convolutional neural network. The designed network is trained on synthesized MRF data simulated with the Bloch equations and fast imaging with steady-state precession (FISP) sequences. In test mode, it takes a temporal MRF signal as input and produces the corresponding tissue parameters.

Results: We validated our approach on both synthetic data and anatomical data generated from a healthy subject. The results demonstrate that, in contrast to conventional dictionary matching-based MRF techniques, our approach significantly improves inference speed by eliminating the time-consuming dictionary matching operation, and alleviates discretization errors by outputting continuous-valued parameters. We further avoid the need to store a large dictionary, thus reducing memory requirements.

Conclusions: Our approach demonstrates advantages in terms of inference speed, accuracy, and storage requirements over competing MRF methods. © 2019 American Association of Physicists in Medicine [https://doi.org/10.1002/mp.13727]

Key words: deep learning, magnetic resonance fingerprinting, nonlocal residual convolutional neural network, quantitative magnetic resonance imaging, self-attention

1. INTRODUCTION

Magnetic resonance fingerprinting (MRF)^{1–8} has emerged as a promising quantitative magnetic resonance imaging (QMRI) approach, with the capability of providing multiple tissue's intrinsic spin parameters simultaneously, such as the spin-lattice magnetic relaxation time (T1) and the spin-spin magnetic relaxation time (T2). Based on the fact that the response from each tissue with respect to a given pseudorandom pulse sequence is unique, MRF exploits pseudo-randomized acquisition parameters to create unique temporal signal signatures, analogous to a “fingerprint,” for different tissues. A dictionary matching operation is then performed to map an inquiry temporal signature to the best matching entry in a precomputed dictionary, leading to multiple tissue parameters directly.

The temporal signatures are generated by varying the acquisition parameters of a pseudorandom excitation pulse sequence, such as repetition time (TR), time of echo (TE), and radio frequency flip angle (FA) over time. The dictionary is composed of a large number of entries that are usually

simulated by the Bloch equations given pseudorandom pulse sequences. Each entry represents a unique temporal signature associated with a specific tissue and its quantitative parameters, such as the T1 and T2 relaxation times. Thus, once the best matching (i.e., most correlated) entry is found, it directly leads to multiple quantitative parameters simultaneously via a lookup table operation.

Magnetic resonance imaging physics and physiological constraints make the MR scanning procedure time consuming. To shorten acquisition time, subsampling is commonly performed in k-space (a.k.a conjugate Fourier transform domain) in order to reduce the number of samples and accelerate imaging speed. However, such k-space subsampling results in temporal signatures that are corrupted by aliasing, blurring, and noise. This hampers the accuracy associated with the estimation of the tissue parameters using a dictionary matching procedure. In order to alleviate the impact of such distortion and corruption, de-aliasing operations are often exploited to restore cleaner signatures before performing signature-to-parameter mapping. Therefore, MRF reconstruction

usually involves two operations: signature restoration and parameter restoration.

Inspired by the successful application of sparsity-driven image processing approaches in MRI reconstruction,^{9–12} several works^{3–6} suggest to incorporate prior knowledge, such as sparsity and low rank, to attenuate distortion and corruption, improving the signature restoration performance, during the initial MRF reconstruction stage. This is then followed by a dictionary matching operation, performing mapping from purified temporal signatures to tissue's quantitative parameters. However, such dictionary matching-based signature-to-parameter mapping exhibits several drawbacks.^{13,14} Since the simulated dictionary and lookup table contain a finite number of elements, they can only cover a limited number of discrete values for each type of tissue parameter. We refer to the difference between a continuous-valued tissue parameter and its closest available discrete value on a lattice as the discretization error. For example, a pair of dictionary and lookup table that contain 101 elements will lead to a discretization error of maximum 25 ms if they cover the range of 0–5000 ms with a fixed interval of 50 ms for a specific tissue parameter, for example, T1. To reduce the discretization error, a huge dictionary that is composed of a large number of entries is needed to represent tissues with fine granularity over the entire value range of target tissue parameters. However, storing a large dictionary becomes prohibitively memory consuming, as the dictionary size and density often increase exponentially with the number of tissue parameters. Specifically, the number of entries in a dictionary will be P^s for s parameters each containing P values, since every combination of these s parameters determines a specific tissue which is characterized by a specific signature. For example, given T1, T2 relaxation times, that is, $s = 2$, if each of them contains 1000 values, the dictionary will have 1000^2 entries. In addition, finding the best matching entry becomes computationally intense for a large dictionary, considerably limiting the inference speed.

In this paper, we propose an alternative approach to dictionary matching-based on deep neural networks (a.k.a. deep learning),^{15,16} which we refer to as HYDRA: HYbrid Deep magnetic Resonance fingerprinting. The motivation derives from the fact that a well designed and tuned deep neural network is capable of approximating complex functions, leading to state-of-the-art results in a number of tasks such as image classification, super-resolution, speech recognition, and more.^{17–23} Recent work^{13,14} proposed to exploit neural networks to replace the dictionary and lookup table used in conventional MRF reconstruction approaches. These proposed neural networks suffer from two limitations: First, these approaches are based on neural network models containing only three layers, thus suffer from limited capacity of capturing complex mapping functions. Second, these methods focused exclusively on parameter restoration stage (the second stage in MRF reconstruction), but not on signature restoration (the first stage in MRF reconstruction). Therefore, these techniques rely on fully sampled data instead of typically available subsampled k-space data.

Different from Cohen *et al.*'s fully connected feed-forward neural network,¹³ and Hoppe *et al.*'s vanilla convolutional

neural network (CNN),¹⁴ the proposed HYDRA involves both a signature restoration and a parameter restoration phase. Signature restoration is implemented using a low-rank based de-aliasing method adapted from Mazor *et al.*⁶ while parameter restoration is implemented using a deep nonlocal residual convolutional neural network developed for this purpose. Our key contributions with respect to prior work are:

1. HYDRA is, to the best of our knowledge, the first deep network approach to combine model-based de-aliasing and learning-based parameter mapping. HYDRA eliminates the requirement for the memory and time-consuming dictionary matching operation, thus significantly improving inference speed without compromising on reconstruction performance.
2. A one-dimensional (1D) nonlocal residual convolutional neural network is designed to capture the mappings from temporal MRF signals to tissue parameters. Owing to residual learning and a self-attention mechanism, our network is deeper and more sophisticated than competing network models. This allows to capture complex parameter mappings more effectively, and output continuous parameters to alleviate discretization issues.
3. The designed network is trained on synthesized MRF data simulated with the Bloch equations, but is still applicable to anatomical data. This contributes to eliminating the requirement for a large amount of real MRF data.
4. A low-rank based de-aliasing technique is developed in order to take advantage of temporal similarity for signature restoration.
5. The low-rank based signature restoration is organically combined with the learning-based parameter restoration to achieve fast and accurate MRF reconstruction. Such strategy enables HYDRA to handle both fully sampled k-space data and more importantly subsampled k-space data.
6. A series of numerical experiments are conducted to evaluate the proposed approach on both synthetic and anatomical data. The results demonstrate improved inference speed, accuracy, and discretization errors over competing methods.^{1–6,13,14}

The rest of the paper is organized as follows. In Section 2, we formulate the MRF reconstruction problem, introduce related methods, and present our approach, involving the use of a low-rank based signature restoration procedure together with a deep network for parameter restoration. Section 3 is devoted to experimental results, followed by a discussion in Section 4 and a conclusion in Section 5.

2. MATERIALS AND METHODS

2.A. The MRF problem formulation

Magnetic resonance fingerprinting data are composed of multiple frames sampled in k-space over time. A series of

such frames are vectorized and then stacked together along the temporal dimension to construct a measurement matrix $\mathbf{Y} \in \mathbb{C}^{Q \times L}$, where Q is the number of k-space samples in each frame, and L is the number of frames. Due to k-space subsampling, every column vector $\mathbf{Y}_{:,i}$ represents a subsampled Fourier transform of a vectorized image frame $\mathbf{X}_{:,i}$:

$$\mathbf{Y} = [\mathbf{Y}_{:,1}, \dots, \mathbf{Y}_{:,L}] = [F_u\{\mathbf{X}_{:,1}\}, \dots, F_u\{\mathbf{X}_{:,L}\}], \quad (1)$$

where $F_u\{\cdot\}$ denotes a subsampled two-dimensional (2D) Fourier transform.

Each column $\mathbf{X}_{:,i}$ represents a MR contrast acquired with RF sequence parameters:

$$\Theta_{:,i}^{TRE} = [TR^i, TE^i, FA^i]^T, \quad i \in [1, L] \quad (2)$$

where TR^i and TE^i denote the repetition time and echo time, respectively, and FA^i denotes the flip angle of the RF pulse during sampling the i -th contrast. Every row $\mathbf{X}_{j,:}$ represents a temporal signature, that is, temporal signal evolution of a specific tissue at the j -th image pixel. The signature depends on the tissue's relaxation times, such as T1 and T2, grouped as a row vector:

$$\Theta_{j,:}^{T12} = [T1^j, T2^j], \quad j \in [1, N] \quad (3)$$

where, N denotes the number of pixels in each image frame. Note that, j is the spatial index while i is the temporal index throughout. Given RF sequence parameters $\Theta_{j,:}^{TRE}$, and parameters $\Theta_{j,:}^{T12}$ of a specific tissue, its temporal signature $\mathbf{X}_{j,:}$ can be derived as:

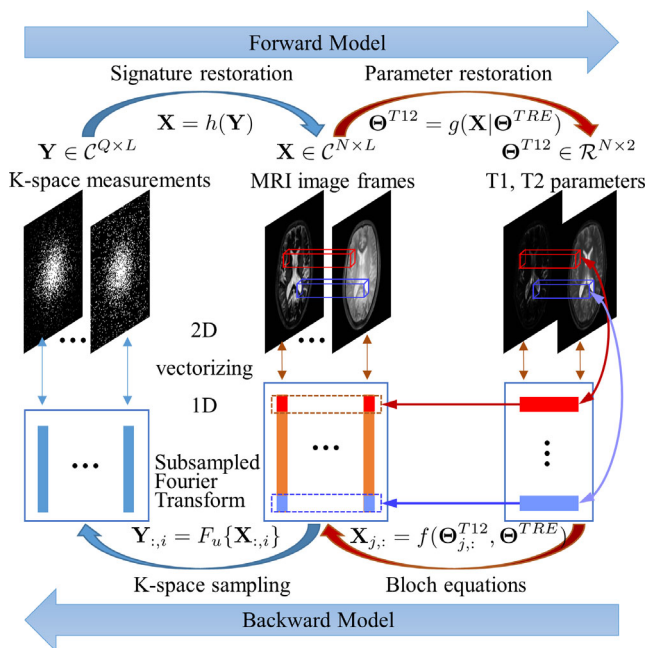


FIG. 1. Relationship between key variables. The MR contrast matrix \mathbf{X} is associated with the k-space measurements \mathbf{Y} per column by the subsampled Fourier transform. It is related to tissue parameters Θ^{T12} per row by the Bloch equations. Given Θ^{TRE} and \mathbf{Y} , the image stack \mathbf{X} is commonly first reconstructed from \mathbf{Y} , referred to as signature restoration, and then mapped to tissue parameters Θ^{T12} via dictionary matching, referred to as parameter restoration. [Color figure can be viewed at wileyonlinelibrary.com]

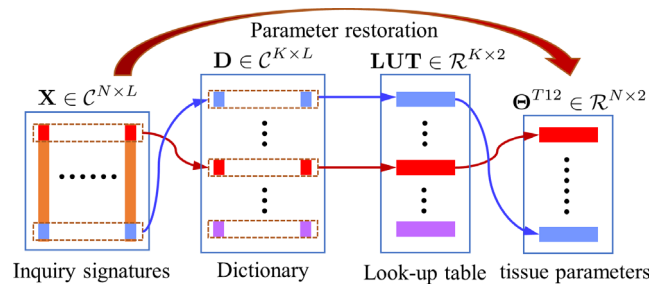


FIG. 2. Parameter restoration using dictionary matching. Given an inquiry temporal signature, dictionary matching computes its inner product with each dictionary entry and selects the most correlated one with the highest inner product as the best matching signature. Once the best matching entry is found, it directly leads to multiple tissue parameters, such as T1, T2, simultaneously, via searching a lookup table. [Color figure can be viewed at wileyonlinelibrary.com]

$$\mathbf{X}_{j,:} = f(\Theta_{j,:}^{T12}, \Theta^{TRE}) \quad (4)$$

where $f(\cdot)$ denotes the Bloch equations. This MR contrast matrix \mathbf{X} is associated with the k-space measurements \mathbf{Y} per column by the subsampled Fourier transform, and it is related to tissue parameters Θ^{T12} per row by the Bloch equations, as illustrated in Fig. 1.

Given RF sequence parameters Θ^{TRE} and k-space measurements \mathbf{Y} , the goal of MRF reconstruction is to estimate the tissue parameters Θ^{T12} . Typically, the image stack \mathbf{X} is first reconstructed from \mathbf{Y} , referred to as signature restoration, and then mapped to tissue parameters Θ^{T12} via dictionary matching, referred to as parameter restoration.¹⁻⁶ This process is illustrated in Fig. 2.

The dictionary is a collection of temporal signatures that are usually simulated by the Bloch equations for various typical tissues, given the pseudorandom RF pulse sequences and tissue parameters. Given an inquiry temporal signature, dictionary matching computes the inner product between the temporal signature with each dictionary entry, selecting the entry in the dictionary exhibiting the highest correlation with the inquiry one as the best matching signature. Once the best entry is found, it directly leads to multiple tissue parameters, such as T1, T2, simultaneously, via searching a lookup table.

Let $\text{LUT} \in \mathbb{R}^{K \times 2}$ denote a lookup table composed of K tissues, each containing two parameters, that is, T1 and T2 relaxation times*. Let $\mathbf{D} \in \mathbb{C}^{K \times L}$ denote the corresponding dictionary simulated using Bloch equations given the RF sequence parameters Θ^{TRE} , formulated as $\mathbf{D}_{k,:} = f(\text{LUT}_{k,:}, \Theta^{TRE})$. Since each temporal signature $\mathbf{D}_{k,:}$ is linked with the k -th tissue's parameters $\text{LUT}_{k,:}$, the choice of a large dictionary size K can in principle provide enough granularity to capture a range of possible tissue values.

*Note that the off resonance parameter, which appeared in the original MRF paper,¹ has been omitted here, since the sequence used in our experiments is derived from the FISP sequence, which is insensitive to off resonance effects.^{2,6}

In conclusion, existing MRF reconstruction approaches involve two stages: signature restoration and parameter restoration, that can be succinctly written as

$$\Theta_{j,:}^{T12} = g(h(\mathbf{Y})_{j,:} | \Theta^{TRE}), j \in [1, N], \quad (5)$$

where the function $\mathbf{X} = h(\mathbf{Y})$ represents the signature restoration operation such as sparsity or low-rank based de-aliasing and denoising methods, whereas $g(\mathbf{X}_{j,:} | \Theta^{TRE})$ denotes the parameter restoration operation, such as dictionary matching-based methods.¹⁻⁶

Our approach aims to perform signature restoration via low-rank based de-aliasing and parameter restoration via a neural network in order to achieve improved MRF reconstruction performance. We highlight that our method only requires a simulated dictionary during network training. Once the network is trained, the dictionary is not needed anymore. In addition, our approach also eliminates a simulated dictionary for signature restoration, which is a key difference from FLOR^{5,6} during signature restoration.

Algorithm 1 Original MRF method¹

Input:

A set of subsampled k-space images: \mathbf{Y}

A pre-simulated dictionary: \mathbf{D}

An appropriate lookup-table: \mathbf{LUT}

Output:

Magnetic parameter maps: \hat{T}_1, \hat{T}_2

Step 1. Restore signatures:

$$\hat{\mathbf{X}}_{:,i} = F_u^H \{ \mathbf{Y}_{:,i} \}, \forall i$$

Step 2. Restore parameters for every j via dictionary matching:

$$\hat{k}_j = \arg \max_k \frac{\text{Re} \langle \mathbf{D}_{k,:}, \hat{\mathbf{X}}_{j,:} \rangle}{\|\mathbf{D}_{k,:}\|_2^2}, \quad \hat{T}_1^j, \hat{T}_2^j = \mathbf{LUT}(\hat{k}_j)$$

2.B. Previous methods

2.B.1. Dictionary matching-based MRF approaches

The original MRF reconstruction algorithm¹ is based on dictionary matching, as presented in Algorithm 1. It finds the best matching dictionary entry for the acquired temporal signature according to their inner product and then searches the lookup table to obtain corresponding tissue parameters. Here, $F_u^H \{ \cdot \}$ denotes the inverse Fourier transform operating on the zero filled k-space data where zeros are filled at the unknown frequencies and symbol $\text{Re} \langle \mathbf{a}, \mathbf{b} \rangle$ represents the real part of the inner product of two vectors \mathbf{a} and \mathbf{b} .

Exploiting the nature of signals, by using appropriate prior knowledge, can often contribute to improved signal processing performance. In this spirit, later works suggested to incorporate sparsity in MRF reconstruction to further improve performance, inspired by successful applications of sparsity in MRI reconstruction.⁹⁻¹¹ Davies et al.³ proposed BLoch response recovery via Iterative Projection (BLIP) which

exploits sparsity in the dictionary domain. BLIP consists of iterating between two main steps: (a) a gradient step which enforces consistency with the measurements, based on the Projected Landweber Algorithm (PLA) generalized from the iterative hard thresholding method; (b) a projection which matches each row of \mathbf{X} to a single dictionary atom. Instead of exploiting sparsity in the dictionary domain, Wang et al.⁴ suggested to leverage sparsity in the wavelet domain of each imaging frame, $\mathbf{X}_{:,i}$. They further replaced the Euclidean norm with the Mahalanobis distance for dictionary matching. Considering that adjacent MR image frames along the temporal dimension should exhibit high resemblance, Mazor et al.^{5,6} proposed a magnetic resonance Fingerprint with LOW-Rank prior for reconstructing the image stack and quantitative parameters, referred to as FLOR, which achieved state-of-the-art performance. The algorithm, described in Algorithm 2, relies on two priors: a low rank prior on the matrix \mathbf{X} , and the fact that the rows of \mathbf{X} lie in the column space of the dictionary \mathbf{D} .

Algorithm 2 FLOR⁶

Input:

A set of subsampled k-space images: \mathbf{Y} ; A pre-simulated dictionary: \mathbf{D} ; An appropriate lookup-table: \mathbf{LUT} ; Parameters μ for gradient step and λ for regularization

Output:

Magnetic parameter maps: \hat{T}_1, \hat{T}_2

Initialization:

$\hat{\mathbf{X}}^0 = 0, \mathbf{P} = \mathbf{D}^\dagger \mathbf{D}$, where \mathbf{D}^\dagger is the pseudo-inverse of \mathbf{D} .

Step 1. Restore signatures via iterating until convergence:

- Gradient step for every i :

$$\hat{\mathbf{Z}}_{:,i}^{t+1} = \hat{\mathbf{X}}_{:,i}^t - \mu F_u^H \{ F_u \{ \hat{\mathbf{X}}_{:,i}^t \} - \mathbf{Y}_{:,i} \} \quad (6)$$

where the superscript t represents the index of iterations.

- Project onto the dictionary subspace:

$$[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \text{svd}(\hat{\mathbf{Z}}^{t+1} \mathbf{P}) \quad (7)$$

where svd denotes the singular-value decomposition operation, and $\mathbf{S} = \text{diag}(\{\sigma_j\})$ is a rectangular diagonal matrix with singular values $\{\sigma_j\}$ on its diagonal.

- Soft-threshold the non-zero singular values with $\lambda\mu$ and reconstruct signatures $\hat{\mathbf{X}}^{t+1}$:

$$\sigma'_j = \max\{\sigma_j - \lambda\mu, 0\}, \quad \hat{\mathbf{X}}^{t+1} = \mathbf{U} \mathbf{S}' \mathbf{V}^H \quad (8)$$

where $\mathbf{S}' = \text{diag}(\{\sigma'_j\})$.

Step 2. Restore parameters for every j via dictionary matching:

$$\hat{k}_j = \arg \max_k \frac{\text{Re} \langle \mathbf{D}_{k,:}, \hat{\mathbf{X}}_{j,:} \rangle}{\|\mathbf{D}_{k,:}\|_2^2}, \quad \hat{T}_1^j, \hat{T}_2^j = \mathbf{LUT}(\hat{k}_j)$$

2.B.2. Learning-based MRF approaches

The above techniques all use dictionary matching to perform mapping from temporal signatures to tissue parameters. Therefore, these methods suffer from drawbacks such as discretization error, slow inference speed, and memory-

consuming storage. In order to alleviate these issues, recent works^{13,14} propose to exploit neural networks to replace dictionaries and lookup tables used in conventional MRF reconstruction approaches. Cohen et al. suggest a fully connected feed-forward neural network (FNN).¹³ Since the input layer of the FNN is fully connected with the input temporal signature, the number of neurons in the input layer corresponds to the length of the input temporal signature. This makes the network structure less flexible, as a FNN network trained on temporal signatures with a certain length is not applicable to temporal signatures with a different length. In addition, the fully connected structure results in rapid increase in the number of parameters along with the growth of depth, making the network more susceptible to overfitting. Hoppe et al.¹⁴ propose a three-layer vanilla CNN for parameter restoration. Both^{13,14} focus exclusively on learning the signature-to-parameter mapping from a pair of dictionary and lookup table simulated using the Bloch equations. During the validation, they assume that clean temporal signatures are available as input into the trained networks. However, since temporal signatures obtained from k-space subsampled MRF data are always contaminated by aliasing and noise, their approaches, when applied directly in such k-space subsampling situations, suffer from heavy artifacts introduced during the signature restoration phase, leading to poor performance.

2.C. Proposed methods

The proposed hybrid deep magnetic resonance fingerprinting (HYDRA) approach, summarized in Algorithm 3, consists of two stages: signature restoration and parameter restoration, [see also (5)]. As illustrated in Fig. 3, a low-rank based de-aliasing method is used to restore signatures, and then a 1D nonlocal residual convolutional neural network is

used to map each restored signature to corresponding tissue parameters.

In particular, given Θ^{TRE} and k-space samples \mathbf{Y} , in our proposed approach, the function $\mathbf{X} = h(\mathbf{Y})$ in Eq. (5) represents a signature restoration operation using low-rank based de-aliasing techniques without requiring a dictionary. The function $\Theta_{j,:}^{T12} = g(\mathbf{X}_{j,:} | \Theta^{TRE})$ in Eq. (5) represents a parameter restoration operation that exploits a trained neural network to map each restored signature $\mathbf{X}_{j,:}$ to corresponding tissue parameters $\Theta_{j,:}^{T12}$ directly. In the subsequent sections, we provide a detailed description of both stages of our technique.

2.C.1. Low-rank based signature restoration

Since MRF data consist of multiple frames exhibiting temporal similarity, the imaging contrasts matrix \mathbf{X} is typically a low-rank matrix.⁶ Therefore, $h(\cdot)$ leverages a low-rank prior to denoising and de-aliasing, formulated as

$$h(\mathbf{Y}) = \arg \min_{\mathbf{X}} \frac{1}{2} \sum_i \|\mathbf{Y}_{:,i} - F_u\{\mathbf{X}_{:,i}\}\|_2^2, \quad (9)$$

s.t. $\text{rank}(\mathbf{X}) < r$

where the parameter r is the rank of the matrix, a fixed pre-chosen parameter. Since typically r is not known in advance, we consider a relaxed regularized version:

$$h(\mathbf{Y}) = \arg \min_{\mathbf{X}} \frac{1}{2} \sum_i \|\mathbf{Y}_{:,i} - F_u\{\mathbf{X}_{:,i}\}\|_2^2 + \lambda \|\mathbf{X}\|_*, \quad (10)$$

where $\|\mathbf{X}\|_*$ denotes the nuclear norm²⁴ of \mathbf{X} , defined as the sum of the singular values of \mathbf{X} , and λ is the Lagrangian multiplier manually selected for balancing data fidelity and the rank. Problem (10) can be solved using the incremental

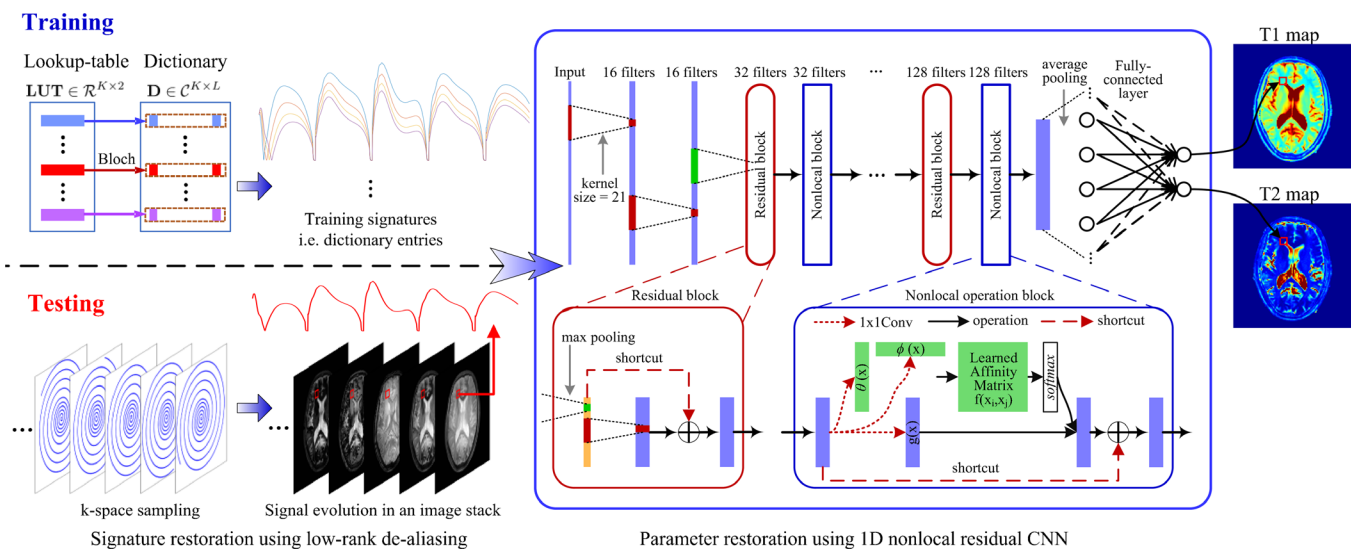


FIG. 3. Diagram of the proposed magnetic resonance fingerprinting reconstruction approach. During the training stage, synthesized dictionary entries are used as training signatures to train the designed one-dimensional nonlocal residual convolutional neural network until the outputs approximate parameter values in LUT well. In this way, the network captures the signature-to-parameter mapping. During the testing stage, a low-rank based algorithm is used to restore the image stack, a matrix containing signatures in rows, from k-space measurements. Then the restored signatures are input into the trained network to obtain corresponding tissue parameters directly. [Color figure can be viewed at wileyonlinelibrary.com]

subgradient proximal method,²⁵ similar to FLOR.⁶ The procedure for solving (5) is shown in Algorithm 3.

One of the differences from FLOR⁶ is the fact that we removed the operation of projecting the temporal signal onto a dictionary. This allows to eliminate the requirement for a simulated dictionary in the signature restoration stage, which also alleviates the memory consumption issue. In addition, the computational complexity is reduced by $N \cdot L^2$ floating-point operations in each iteration, where L is the dimension of a dictionary element, and N is the number of pixels in each image frame. On the other hand, the gained benefits are at the price of requiring more iterations to converge. Another difference from FLOR⁶ is that we exploit a network, instead of dictionary matching, for signature-to-parameter mapping.

2.C.2. Learning-based parameter restoration

Once the imaging contrasts matrix \mathbf{X} is recovered from the k-space samples \mathbf{Y} , each temporal signature $\mathbf{X}_{j,:}$ is input into the trained network for parameter restoration, formulated as:

$$\Theta_{j,:}^{T12} = g(\mathbf{X}_{j,:} | \Theta^{TRE}), j \in [1, N], \quad (11)$$

where $g(\cdot)$ denotes the trained network, Θ^{TRE} denotes the fixed RF sequence parameters. We next describe the network structure, training, and testing procedures.

Algorithm 3 Proposed MRF reconstruction approach: HYDRA

Input:

A set of subsampled k-space images: \mathbf{Y}

The trained network: g

Parameters μ for gradient step and λ for regularization

Output: Magnetic parameter maps \hat{T}_1, \hat{T}_2

Initialization: $\hat{\mathbf{X}}^0 = 0$

Step 1. Restore signatures via iterating until convergence:

- Gradient step for every i , the same as (6).
- Perform SVD:

$$[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \text{svd}(\hat{\mathbf{Z}}^{t+1})$$
- Soft-threshold the non-zero singular values $\{\sigma_j\}$ of \mathbf{S} with parameter $\lambda\mu$ and reconstruct signatures $\hat{\mathbf{X}}^{t+1}$, the same as (8).

Step 2. Restore parameters for every j via the trained network:

$$\hat{T}_1^j, \hat{T}_2^j = g(\hat{\mathbf{X}}_{j,:})$$

Network structure: The proposed network has a 1D non-local residual CNN architecture with shortcuts for residual learning and nonlocal operations for achieving a self-attention mechanism. As illustrated in Fig. 3, it starts with two 1D convolutional layers before connecting with four residual/nonlocal operation blocks, and finally ends with a global-average-pooling layer followed by a fully connected layer. Every residual block is followed by a nonlocal operation block. Four such blocks are interspersed with each other.

Each residual block contains a max-pooling layer with stride 2, two convolution layers and a shortcut that enforces

the network to learn the residual content. The filter size is set to be equal to 21 throughout convolutional layers. The number of channels, a.k.a feature maps, in the first two convolutional layers is set to 16 and then is doubled in the subsequent four residual blocks until 128 is reached in the final residual block. The size of feature maps in the next block halves in contrast with the previous one due to max-pooling. In this way, we gradually reduce temporal resolution while extracting more features to increase content information.

Inspired by the self-attention scheme and nonlocal neural networks,^{26,27} nonlocal operations are incorporated into the designed network to achieve the attention mechanism, in order to capture long-range dependencies with fewer layers. In contrast to the progressive behavior of convolutional operations that process one local neighborhood at a time, the non-local operations compute the response at a position as a weighted sum of the features at all positions in the feature maps. Formally, the nonlocal operation is formulated as²⁷:

$$\mathbf{y}_i = \frac{1}{C(\mathbf{x})} \sum_{\forall j} f(\mathbf{x}_i, \mathbf{x}_j) g(\mathbf{x}_j). \quad (12)$$

Here, i is the index of an output position and j is the index that enumerates all possible temporal positions, \mathbf{x} is the input temporal signal or its features and \mathbf{y} is the output signal of the same size as \mathbf{x} . A pairwise function f computes a scalar between i and all j to represent the affinity relationship of these two positions. The unary function g computes a representation of the input signal at position j . The response is normalized by a factor $C(\mathbf{x})$. There exists a few instantiations for function f and g . For simplicity, the unary function g is chosen as a linear embedding: $g(\mathbf{x}_j) = \mathbf{W}_g(\mathbf{x}_j)$, where \mathbf{W}_g is a weight matrix to be learned. Regarding the affinity matrix f , we adopt the embedded Gaussian to compute similarity in an embedding space, which is formulated as: $f(\mathbf{x}_i, \mathbf{x}_j) = e^{\theta(\mathbf{x}_i)^T \phi(\mathbf{x}_j)}$. Here, $\theta(\mathbf{x}_i) = \mathbf{W}_\theta \mathbf{x}_i$ and $\phi(\mathbf{x}_j) = \mathbf{W}_\phi \mathbf{x}_j$ are two learned embeddings. The normalization factor is set as $C(\mathbf{x}) = \sum_{\forall j} f(\mathbf{x}_i, \mathbf{x}_j)$. For a given i , $\frac{1}{C(\mathbf{x})} f(\mathbf{x}_i, \mathbf{x}_j)$ becomes the softmax computation along the dimension j , which leads to $\mathbf{y} = \text{softmax}(\mathbf{x}^T \mathbf{W}_\theta^T \mathbf{W}_\phi \mathbf{x}) g(\mathbf{x})$, which is the self-attention form.²⁶

The nonlocal behavior in Eq. (12) is due to the fact that all positions (j) are considered in the operation. As a comparison, a convolutional operation sums up the weighted input in a local neighborhood.²⁷ It implies that the nonlocal operation directly captures long-range dependencies in the temporal dimension via computing interactions between any two points, regardless of their positional distance. In this way, the network is able to extract global features and take advantage of the full receptive field in each layer.

The global-average-pooling layer is used to average each feature map in order to integrate information in every channel for improved robustness to corrupted input data. This global-average-pooling layer also reduces the number of parameters significantly, thus lessening the computation cost as well as preventing overfitting. The last fully connected layer outputs estimated parameters — T_1 and T_2 relaxation times. The

designed network contains around 0.27 million parameters. The weights are initialized using He normal distribution.²⁸ The max-norm kernel constraint²⁹ is exploited to regularize the weight matrix directly in order to prevent overfitting. The designed network can also be adapted for various MRF sequences, such as the original MRF sequence — inversion recovery balanced steady-state free precession (IR-bSSFP) sequence, that depends also on the intrinsic T_2 parameter. It is possible to adjust the number of outputs to adapt to more parameters, such as proton density, B_0 .

To summarize, our network is motivated and inspired by recent successful applications of convolutional neural networks and variants. Convolutional neural networks have been proved to be a powerful model to capture useful features from signals and images. By introducing convolution, local receptive field and weight sharing design, a CNN is capable of taking advantage of local spatial coherence and translation invariance characteristics in the input signal, thus become especially well suited to extract relevant information at a low computational cost.^{17–22} On the other hand, the residual network architecture^{18,19} provides an effective way to design and train a deeper model, since it alleviates the gradient vanishing or exploding problems by propagating gradients throughout the model via shortcuts, a.k.a skip connections. By leveraging a nonlocal operation-based attention mechanism, neural networks are endowed with the capability of extracting global features and capturing long-range dependencies.

Network training: The designed network is trained on a synthesized dictionary \mathbf{D} and corresponding lookup table \mathbf{LUT} to learn the signature-to-parameter mappings $\mathbf{LUT}_{k,:} = g(\mathbf{D}_{k,:} | \Theta^{TRE})$.

The training dataset is synthesized as follows. First, we determine the range of tissue parameters. For example, one may set T_1 relaxation times to cover a range of [1, 5000] ms and T_2 relaxation times to cover a range of [1, 2000] ms with an increment of 10 ms for both. Thus, the T_1 and T_2 values constitute a grid with dimension 500×200 , in which each point represents a specific combination of T_1 and T_2 values, and hence characterizes a specific tissue. Points corresponding to $T_1 < T_2$ have been excluded as such combinations have no physical meaning. All the valid points are stacked together to generate a lookup table. For instance, the above setting for T_1 and T_2 leads to a lookup table of dimension $80\,100 \times 2$. The RF pulse sequences used in our work are fast imaging with steady-state precession (FISP) pulse sequences with parameters that have been used in previous publications in the field of MRF.^{2,6,8} Given the lookup table and RF pulse sequences, dictionary entries can be synthesized by solving the Bloch equations using the extended phase graph formalism.^{30,31}

When the training dataset is ready, the dictionary entries are used as input signals and corresponding lookup table entries serve as the groundtruth. All the dictionary entries are input into the designed network batch by batch which outputs estimated parameters. The root mean square errors (RMSE) of the

outputs are calculated with respect to corresponding ground-truth. The resulting RMSE loss is then backpropagated from the output layer to the first layer to update the weights and bias by using Adam³² as the optimization algorithm. More training details are provided in the subsequent experiment section. Once the training procedure is completed, given an inquiry signal evolution $\mathbf{X}_{j,:}$, it is able to map such a time sequence directly to corresponding tissue parameters, as formulated in Eq. (11), implying that no dictionary or lookup table are required during the inference. Since we only need to store the trained network which is a compact model, it consumes less memory than storing the dictionary and lookup table.

We emphasize that even though the network is trained on a grid of tissue values, it is expected to capture the mapping function from temporal signatures to tissue parameters. Thus, the trained network is capable of outputting tissue values not existing in the grid of training values. Detailed results can be found in Fig. 7 and Table II. This feature is favorable, as it implies that well designed and trained networks have an ability to overcome discretization issues. The overall procedures for solving (10) and (11) are shown in Algorithm 10.

3. EXPERIMENTAL RESULTS

In this section, we conduct a series of experiments to evaluate our approach, comparing it with other state-of-the-art MRF methods.^{1,3,6,13,14}

The experiments are categorized into a few types: training, testing on synthetic data, testing on anatomical data using variable density Gaussian sampling patterns and spiral sampling patterns at different sampling ratios, and the number of time frames, as described in Table I. For the network training, synthesized temporal signatures, that is, simulated dictionary entries of \mathbf{D} shown as Fig. 4, are used as input signals and corresponding parameter values in the lookup table \mathbf{LUT} serve as the groundtruth. The proposed network

TABLE I. Brief description of experiment types and settings.

Experiment	Settings
Training	Input: \mathbf{D} , size $K \times L = 80\,100 \times 200$. Groundtruth: \mathbf{LUT} , size $80\,100 \times 2$. k-space subsampling factor β : not available.
Testing on synthetic data	Input: \mathbf{X} , size $N \times L = 80\,000 \times 200$. Groundtruth: Θ^{T12} , size $80\,000 \times 2$. k-space subsampling factor β : not available.
Testing on anatomical data	Input: \mathbf{Y} , size $Q \times L = 16\,384\beta \times 200$ or $16\,384\beta \times 1000$. Reference: Θ^{T12} , size $N \times 2 = 16\,384 \times 2$. k-space subsampling factor β : 70%, 15% using Gaussian patterns.
Testing on anatomical data	Input: \mathbf{Y} , size $Q \times L = 16\,384\beta \times 1000$. Reference: Θ^{T12} , size $N \times 2 = 16\,384 \times 2$. k-space subsampling factor β : 9% using spiral trajectories.

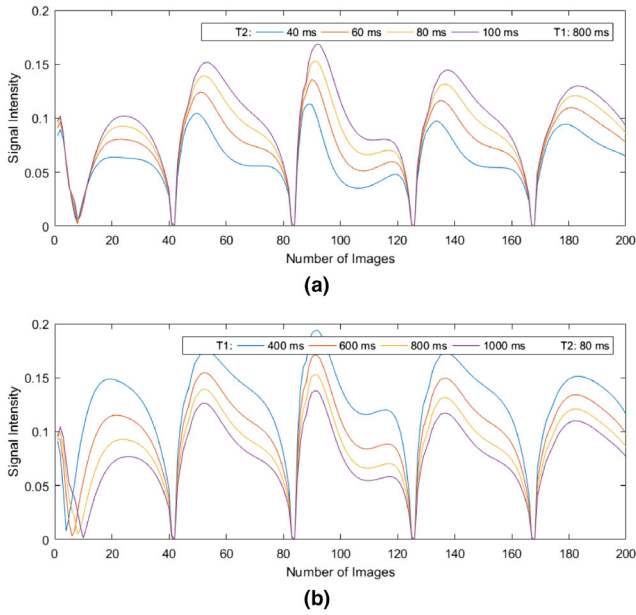


FIG. 4. Synthetic magnetic resonance fingerprinting temporal signatures with 200 time frames. (a) Temporal signatures corresponding to parameter values $\{(T1, T2)\} \text{ ms} = \{(800, 40), (800, 60), (800, 80), (800, 100)\}$ ms. (b) Temporal signatures corresponding to parameter values $\{(T1, T2)\} \text{ ms} = \{(400, 80), (600, 80), (800, 80), (1000, 80)\}$ ms. [Color figure can be viewed at [wileyonlinelibrary.com](#)]

is trained to capture the signature-to-parameter mappings. For testing on synthetic data, synthesized temporal signatures in \mathbf{X} are used as input signals and corresponding parameter values in Θ^{T12} serve as the groundtruth. The aim is to test the parameter restoration performance only. For testing on anatomical data, the k-space measurements \mathbf{Y} which are derived from the Fourier transform (for Gaussian patterns) or nonuniform FFT (for spiral trajectories)³³ of \mathbf{X} , are used as input and corresponding parameter values in Θ^{T12} serve as reference. When there is no k-space subsampling, the aim is to test the parameter restoration performance only. When there exists k-space subsampling, the aim is to test the overall performance, including both signature restoration and parameter restoration. More detailed descriptions are provided in each subsection.

3.A. Training

As mentioned in Section 2.C.2, the designed network is trained on a pair of synthesized dictionary \mathbf{D} and lookup table \mathbf{LUT} , simulated using Bloch equations and FISP pulse sequences.^{2,6}

The FISP pulse sequence used in our experiments was designed with parameters $\Theta_{:,i}^{TRE} = [TR^i, TE^i, FA^i]^T, i \in [1, L]$ that have been used in previous publications in the field of MRF.^{2,6,8} The echo time TE^i was a constant of 2 ms. The repetition time TR^i was randomly varied in the range of 11.5–14.5 ms with a Perlin noise pattern. All the flip angles $FA^i, i \in [1, L]$ constituted a sinusoidal variation in the range of 0° – 70° to ensure smoothly varying transient state of the magnetization, as shown in Fig. 5.

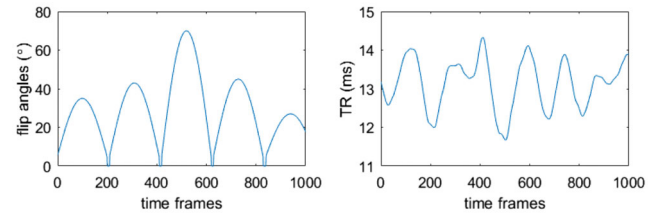


FIG. 5. FISP pulse sequence parameters. All the flip angles (FA) constituted a sinusoidal variation in the range of 0° – 70° to ensure smoothly varying transient state of the magnetization. The repetition time (TR) was randomly varied in the range of 11.5–14.5 ms with a Perlin noise pattern. [Color figure can be viewed at [wileyonlinelibrary.com](#)]

For the range of tissue parameters, T1 relaxation times are set to cover a range of $[1, 5000]$ ms and T2 relaxation times to cover a range of $[1, 2000]$ ms with an increment of 10 ms for both. Such parameter ranges cover the relaxation time values that can be commonly found in a brain scan.³⁴ All the valid combinations of T1 and T2 values are stacked together, generating a lookup table \mathbf{LUT} of dimension $K \times 2$, where $K = 80 \times 100$. Given the lookup table and RF pulse sequences, dictionary entries are synthesized by solving the Bloch equations using the extended phase graph formalism, leading to a dictionary of dimension $K \times L$, where $L = 200$ or 1000 is the number of time frames.

When the training dataset is ready, the dictionary entries are used as input signals and corresponding lookup table entries serve as the groundtruth to train the designed network, as mentioned in Section 2.C.2. The model was trained for 50 epochs. It takes around 30 sec for running one epoch on average, thus around 25 min for completing 50 epochs, on a NVIDIA GeForce GTX 1080 Ti GPU. In each training epoch, 20% of the training samples are separated aside for validation dataset. The learning rate decays from $1e-2$ to $1e-6$ every 10 epochs. Each batch was experimentally set to contain 256 time sequences in order to balance the convergence rate and weights updating rate well. For comparison purposes, we also implemented Hoppe et al.'s CNN referring to Ref. ¹⁴ and Cohen et al.'s FNN referring to Ref. ¹³ with the same structure and parameters as specified in their papers. Then, we use the same GPU and training dataset to train their networks with specified learning rate and number of epochs until convergence.

We adopt a few widely used metrics, such as root mean square error (RMSE), signal-to-noise ratio (SNR), and peak signal-to-noise ratio (PSNR) to evaluate the image quality quantitatively. The definitions of RMSE, SNR, and PSNR are given as follows:

$$\begin{aligned} RMSE &= \sqrt{\frac{\|\mathbf{X} - \hat{\mathbf{X}}\|_F^2}{N}}, \\ SNR &= 20 \log_{10} \frac{\|\mathbf{X}\|_F^2}{RMSE}, \\ PSNR &= 20 \log_{10} \frac{PeakVal}{RMSE}, \end{aligned} \quad (13)$$

where matrices \mathbf{X} and $\hat{\mathbf{X}}$ denote the ground truth signal and its reconstructed version, respectively, N denotes the total

number of elements in the signal and $\|\cdot\|_F$ denotes the Frobenius norm. PeakVal stands for the pixel peak value in an image, for example, 1 for a normalized signal.

3.B. Testing on synthetic dataset

In this subsection, we evaluate the performance of HYDRA on a synthetic testing dataset. The procedures of constructing a synthetic testing dataset is similar to the construction of the training dataset: 500 different T1 values are randomly selected from 1 to 5000 ms, while 200 different T2 values are randomly selected from 1 to 2000 ms, using random permutation based on uniformly distributed pseudorandom numbers. All the valid combinations from the selected T1 and T2 values are stacked together, generating a parameter matrix Θ^{T12} of dimension $80\,000 \times 2$ with $N = 80\,000$.

TABLE II. Testing on synthetic dataset. Comparing parameter restoration performance, in terms of peak signal-to-noise ratio (PSNR), signal-to-noise ratio (SNR), root mean square error (RMSE), and correlation coefficient.

	Dict. match. T1/T2	CNN ¹⁴ T1/T2	FNN ¹³ T1/T2	Proposed T1/T2
PSNR (dB)	59.15/52.31	62.96/49.64	58.97/54.96	79.30/72.99
SNR (dB)	55.23/47.15	59.05/44.49	55.06/49.81	75.38/67.83
RMSE (ms)	5.515/4.847	3.554/6.591	5.63/3.57	0.542/0.448
CorrCoef	1.00/1.00	1.00/1.00	1.00/1.00	1.00/1.00
Time cost (s)	464.10	2.87	1.58	8.2

Bold represents the best value in terms of the corresponding metric. For example, bold represents the largest value for PSNR, SNR, CorrCoef while bold represents the least value for RMSE and Time cost.

TABLE III. Testing on synthetic dataset involving detailed T1/T2 examples that are not on the training grid and their intervals are much smaller than the training grid intervals. D.M. denotes dictionary matching. T1 and T2 errors are defined as the difference between estimated values and groundtruth values.

Truth	T1 estimation				T1 errors			
	D.M.	¹⁴	¹³	Ours	D.M.	¹⁴	¹³	Ours
1005.0	1001.0	1002.9	1009.3	1004.8	-4.0	-2.1	4.3	-0.2
1005.5	1001.0	1003.3	1010.0	1005.3	-4.5	-2.3	4.5	-0.2
1006.0	1011.0	1003.6	1010.6	1005.8	5.0	-2.4	4.6	-0.2
1006.5	1011.0	1004.1	1011.2	1006.3	4.5	-2.5	4.7	-0.2
1007.0	1011.0	1004.5	1011.8	1006.8	4.0	-2.5	4.8	-0.3
RMSE	-	-	-	-	4.4	2.3	4.6	0.2

Truth	T2 estimation				T2 errors			
	D.M.	¹⁴	¹³	Ours	D.M.	¹⁴	¹³	Ours
505.0	501.0	513.6	504.3	505.2	-4.0	8.6	-0.7	0.2
505.5	511.0	514.1	504.8	505.7	5.5	8.6	-0.7	0.2
506.0	501.0	514.7	505.3	506.2	-5.0	8.6	-0.7	0.2
506.5	511.0	515.2	505.9	506.8	4.5	8.7	-0.6	0.3
507.0	511.0	515.7	506.4	507.3	4.0	8.7	-0.6	0.3
RMSE	-	-	-	-	4.6	8.6	0.7	0.2

Bold represents the best value in terms of the corresponding metric.

The RF pulse sequences are the same as in the training stage. Given the parameter matrix and RF pulse sequences, input signal signatures are synthesized by solving the Bloch equations using the extended phase graph formalism, leading to a signature matrix \mathbf{X} of dimension $N \times L = 80\,000 \times 200$, with each row representing a temporal signature corresponding to a specific combination of T1 and T2 values. The signature matrix \mathbf{X} and parameter matrix Θ^{T12} constitute the synthetic testing dataset, with \mathbf{X} as input and Θ^{T12} as the groundtruth.

We input the synthetic testing signatures \mathbf{X} into Hoppe et al.'s CNN,¹⁴ Cohen et al.'s FNN,¹³ and the network of HYDRA to compare the outputs with groundtruth T1 and T2 values in Θ^{T12} . We also compare with dictionary matching methods¹⁻⁶ which exploit the same dictionary \mathbf{D} and lookup table \mathbf{LUT} to find the best matching entry for each signature in \mathbf{X} and then estimate parameter values by searching the lookup table. As shown in Tables II, and III, Figs. 6 and 7, the estimated parameter values using the proposed network obtained outstanding agreement with the groundtruth, yielding higher PSNR, SNR, and smaller RMSE than the dictionary matching method,¹⁻⁶ as well as competing networks.^{13,14}

In particular, to illustrate in detail how well neural networks tackle the discretization issue inherent to dictionary matching, we show the testing performance on continuous-valued T1, T2 parameters which have small intervals, for example, 0.5 ms, that is 20 times smaller than the training grid intervals of 10 ms, between neighboring values in Table III. Since these values and their corresponding MRF signatures do not exist in the training dictionary and lookup table[†], the dictionary matching methods report a T1 and T2 value — the closest discretized value present in the dictionary — that can be quite distinct from the groundtruth. In contrast, the various neural network approaches can potentially learn an underlying mapping from the temporal signatures to the respective T1 and T2 values, leading to estimates that are much closer to the groundtruth. Interestingly, our approach outperforms previous networks^{13,14} as shown in Table III and Fig. 6. Evidently, neural networks demonstrate much better robustness to discretization issues, leading to improved parameter restoration in comparison to dictionary-based methods.

Another impressive advantage of HYDRA is the fast inference speed. HYDRA takes only 8.2 s to complete the mapping operation for 80,000 temporal signatures, that is, $53 \times$ faster than dictionary matching. Furthermore, the inference speed of HYDRA is subject to the network topology. That is, once the network structure is fixed, the complexity is fixed. In contrast, the complexity of dictionary matching is limited by the dictionary density. This implies that our advantage will

[†]As mentioned in the experiment setting in section 3.A, in the training dataset, T1 relaxation times are set to cover a range of [1, 5000] ms and T2 relaxation times to cover a range of [1, 2000] ms with an increment of 10 ms for both, that is, T1 values = {1, 11, 21, ..., 4991}, and T2 values = {1, 11, 21, ..., 1991}.

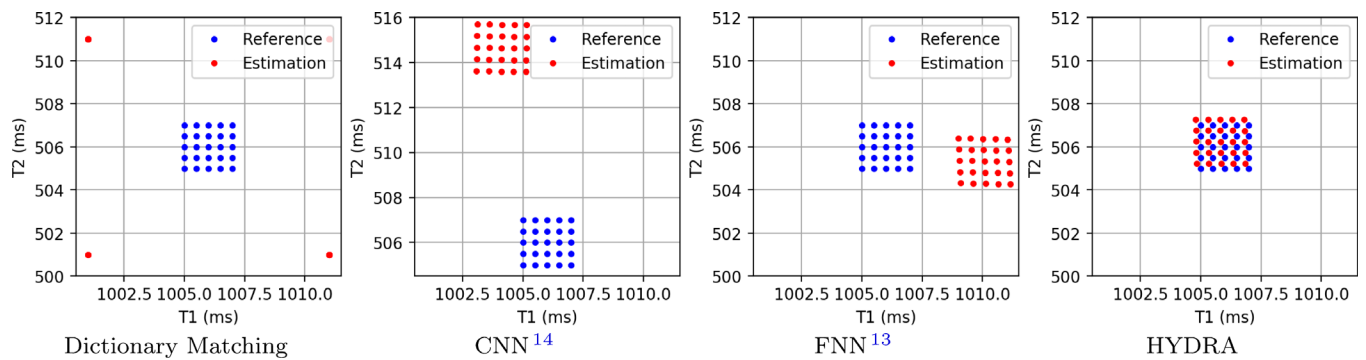


FIG. 6. Testing on synthetic dataset involving detailed T1 / T2 examples that are not on the training grid and their intervals are much smaller than the training grid intervals. Dictionary matching finds best adjacent values from the dictionary, that is, 1001, 1011 for T1, and 501, 511 for T2. In contrast, owing to the captured mapping functions, neural networks output continuous values. Proposed HYDRA leads to the smallest deviations and bias. [Color figure can be viewed at wileyonlinelibrary.com]

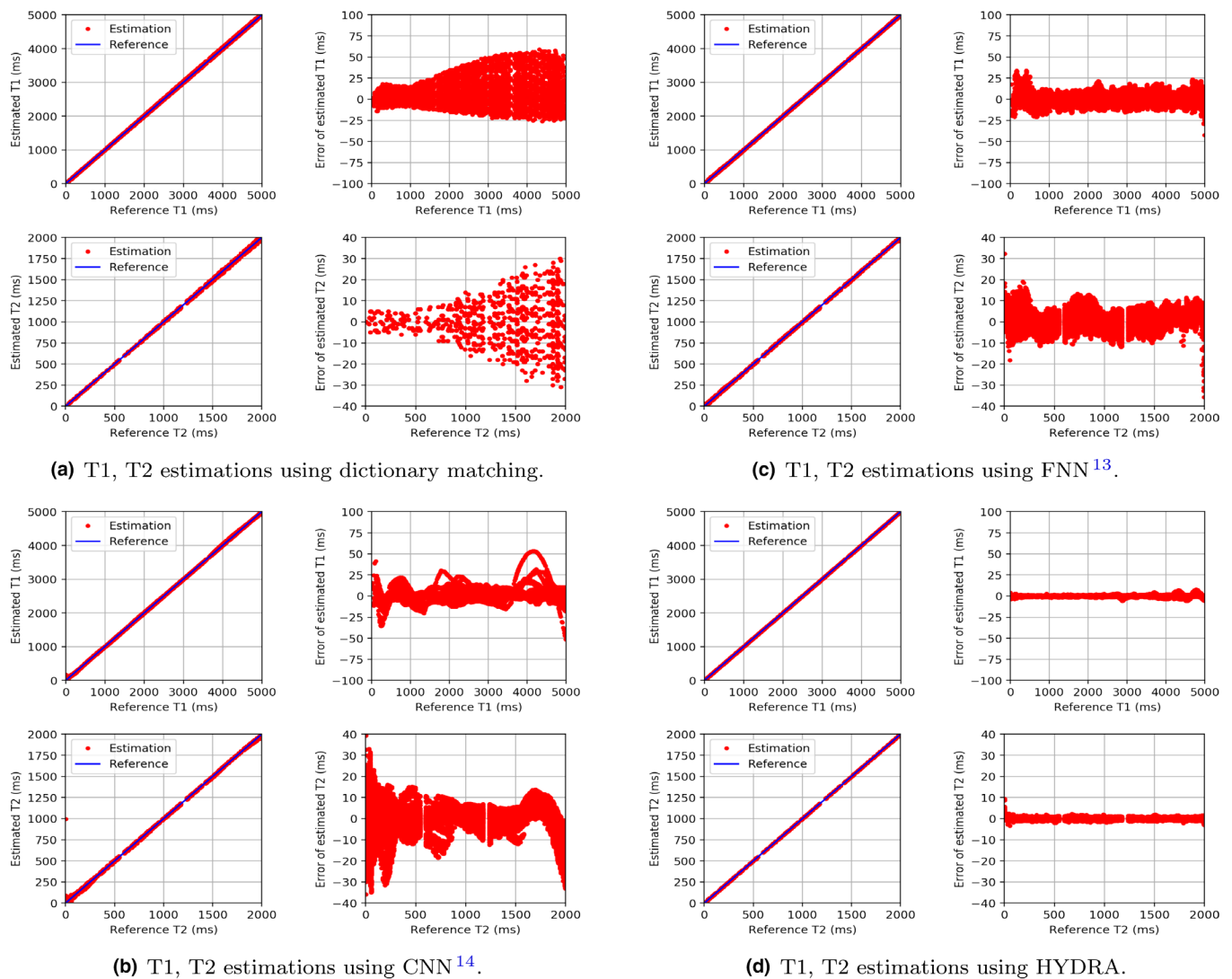


FIG. 7. Testing on the synthetic dataset for comparing parameter restoration performance. Subfig. (a)–(d) show the results using dictionary matching,^{1–6} FNN,¹³ convolutional neural network¹⁴ and HYDRA. In each subfigure, the left figure compares the estimated T1 or T2 values (marked with red dot) with groundtruth values (marked with blue line), and the right figure shows the deviations of the estimation from the groundtruth. Parameter mapping performance of HYDRA is much better than competing methods, in the entire value range of T1 and T2 parameters, resulting in smaller deviations. [Color figure can be viewed at wileyonlinelibrary.com]

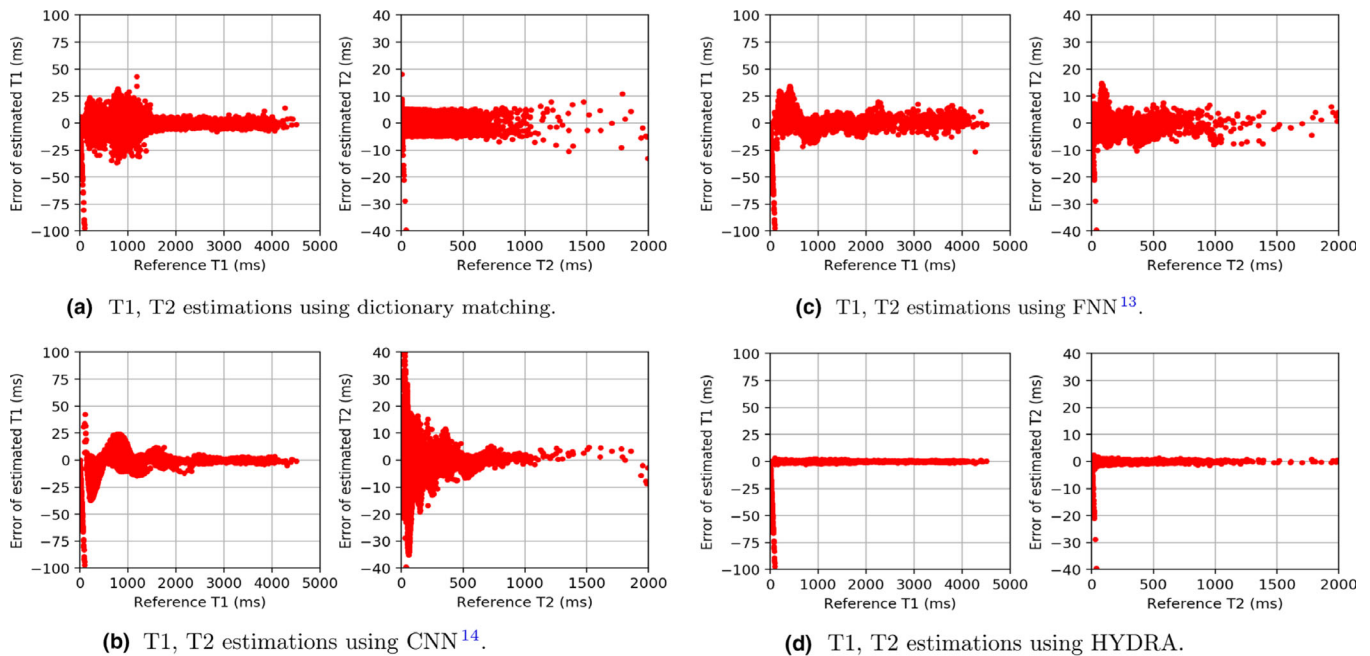


FIG. 8. Testing on the anatomical dataset with full k-space sampling for comparing parameter restoration performance. Subfig. (a)–(d) show the results using dictionary matching,^{1–6} FNN,¹³ convolutional neural network¹⁴ and HYDRA. Each subfigure shows the deviations of the estimation from the reference. Parameter mapping performance using HYDRA outperforms competing methods significantly, resulting in smaller deviations. The performance is also verified by quantitative metrics, as shown in Table 4. [Color figure can be viewed at wileyonlinelibrary.com]

be more prominent in comparison with competing techniques using a dictionary with higher density.

3.C. Testing on anatomical dataset

In this subsection, we evaluate our approach on an anatomical testing dataset. We construct the dataset from brain scans that were acquired with GE Signa 3T HDXT scanner from a healthy subject.[‡] Since there are no ground-truth parameter values for the T1 and T2 parameter maps, we obtain gold standard data by acquiring Fast Imaging Employing Steady-state Acquisition (FIESTA) and Spoiled Gradient Recalled Acquisition in Steady State (SPGR) images, at four different flip angles ($3^\circ, 5^\circ, 12^\circ$, and 20°), and implementing corrections³⁵ followed by DESPOT1 and DESPOT2³⁶ algorithms. The constructed gold standard T1, T2 parameter maps have a dimension of 128×128 for each map, accordingly leading to a parameter matrix Θ^{T12} of size $16\,384 \times 2$ by stacking vectorized T1, T2 maps together. Based on the parameter matrix Θ^{T12} and predefined RF pulse sequences, we generate temporal signatures using Bloch equations, the same mechanism as generating the synthetic testing dataset, leading to a signature matrix \mathbf{X} of dimension $N \times L = 16\,384 \times 200$. The signature matrix \mathbf{X} and parameter matrix Θ^{T12} constitute the anatomical testing dataset, with \mathbf{X} as input and Θ^{T12} as the gold standard reference.

[‡]The experiment procedures involving human subjects described in this paper were approved by the Institutional Review Board of Tel-Aviv Sourasky Medical Center, Israel.

Note that, since the gold standard T1, T2 maps exhibit spatial structures in the image domain, the resulting signature matrix \mathbf{X} can be regarded as a stack of $L = 200$ vectorized image frames, where each frame exhibits specific spatial structures. Therefore, it makes sense to perform Fourier transform and k-space subsampling for each column of \mathbf{X} to get k-space measurements \mathbf{Y} . This is the key difference between the anatomical dataset and the synthetic dataset.

We first explore the case with full k-space sampling in order to evaluate the parameter restoration performance of HYDRA. Then, we consider situations with k-space subsampling in order to evaluate both the signature restoration and the parameter restoration performance of HYDRA.

3.C.1. Full k-space sampling

In the first case, the fully sampled k-space measurements \mathbf{Y} , derived from the Fourier transform of \mathbf{X} , are used as input to obtain the estimated Θ^{T12} . This is equivalent to inputting \mathbf{X} into the network of HYDRA, or performing dictionary matching based on \mathbf{X} directly, since the inverse Fourier transform of the fully sampled measurements \mathbf{Y} is exactly the same as \mathbf{X} . The aim is to test the parameter restoration performance only. In the experiment, corresponding parameter values in Θ^{T12} serve as the gold standard reference. For comparison, dictionary matching methods^{1–6} exploit the same dictionary \mathbf{D} and lookup table \mathbf{LUT} as in our training stage to find the best matching entry and estimate parameter values for each signature in \mathbf{X} .

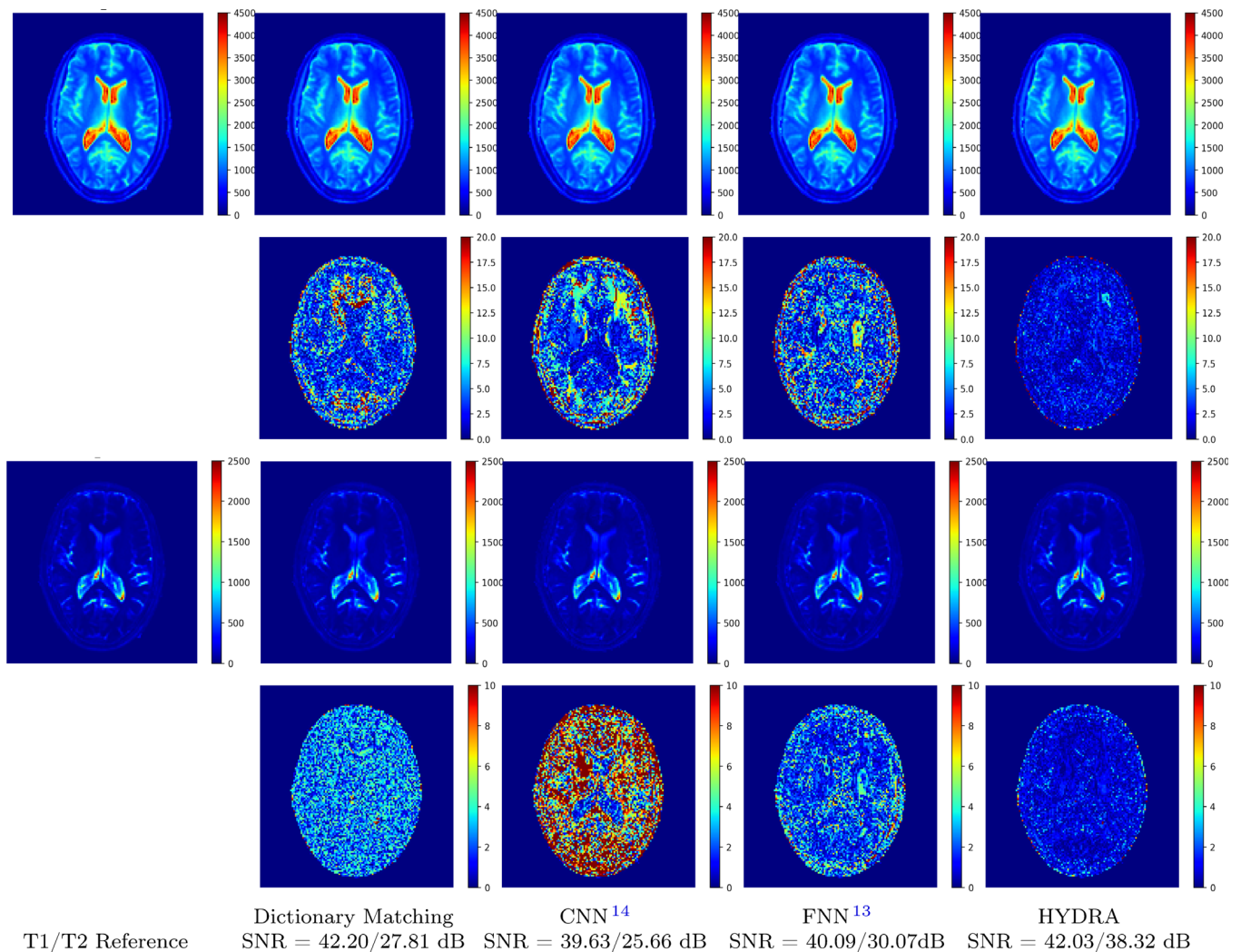


FIG. 9. Visual results of testing on anatomical dataset with full k-space sampling for comparing parameter restoration performance. Top two rows correspond to T1 maps and residual errors while bottom two rows correspond to T2 maps and residual errors. Proposed HYDRA results in comparable performance for T1 mapping and yields much better performance for T2 mapping, obtaining 10dB higher SNR gains than competing dictionary matching-based methods.¹⁻⁶ HYDRA also outperforms previous networks, such as convolutional neural network by Hoppe et al.¹⁴ and FNN by Cohen et al.¹³ [Color figure can be viewed at wileyonlinelibrary.com]

Visual and quantitative results are shown in Figs. 8, 9 and Table IV. It can be seen that our basic version of HYDRA outperforms dictionary matching,¹⁻⁶ yielding better visual and quantitative performance, for example, 7.9 dB SNR gains for T2 mapping. The RMSE of T2 mapping is also reduced to 2.498 from 6.252, accordingly. Our nonlocal version of HYDRA achieves even better performance, leading to 10 dB SNR gains with RMSE as small as 1.86. This is owing to the advantage that the trained network is a powerful function approximator, which is able to output well-estimated parameter values based on learnt mapping functions, even though these values do not exist in the training dictionary and lookup table. In contrast, dictionary matching only matches signatures to discrete parameters existing in the training dataset. In other words, if there are no exact matching dictionary element and parameter values for an inquiry MRF signature, it will find adjacent values as approximations, thus introducing discretization error. On the other hand, the advantage of

HYDRA over dictionary matching on T1 mapping is not as significant as on T2 mapping quantitatively. But the visual improvements are evident. A similar trend is observed when comparing our network with competing networks such as Hoppe et al.'s CNN¹⁴ and Cohen et al.'s FNN.¹³ In addition, HYDRA takes around 2 s to accomplish the mapping for 16 384 signatures, 40× faster than dictionary matching.¹⁻⁶

3.C.2. k-space subsampling using Gaussian patterns

In k-space subsampling situations, the developed low-rank based de-aliasing method is applied to restore the signature matrix \mathbf{X} from the measurements matrix \mathbf{Y} . Then, the reconstructed \mathbf{X} is used as input into the network for parameter mapping to obtain the corresponding tissue parameter values. In the experiments, the subsampling factor β is set to be 70% and 15%. For $\beta = 15\%$, 15% k-space data are acquired by a series of 2D random Gaussian sampling patterns, shown in

TABLE IV. Testing on anatomical dataset with full k-space sampling. Comparing parameter restoration performance, in terms of peak signal-to-noise ratio (PSNR), signal-to-noise ratio (SNR), root mean square error (RMSE), and correlation coefficient.

	Dict. Match. T1/T2	CNN ¹⁴ T1/T2	FNN ¹³ T1/T2	Proposed basic T1/T2	Proposed nonlocal T1/T2
PSNR (dB)	56.64/52.04	54.06/49.88	54.53/54.36	56.59/60.01	56.47/62.56
SNR (dB)	42.20/27.81	39.63/25.66	40.09/30.07	42.15/35.76	42.03/38.32
RMSE (ms)	6.623/6.252	8.912/8.015	8.45/4.78	6.661/2.498	6.76/1.86
CorrCoef	1.00/1.00	1.00/1.00	1.00/1.00	1.00/1.00	1.00/1.00
Time cost (s)	84.56	0.69	0.41	1.6	2.1

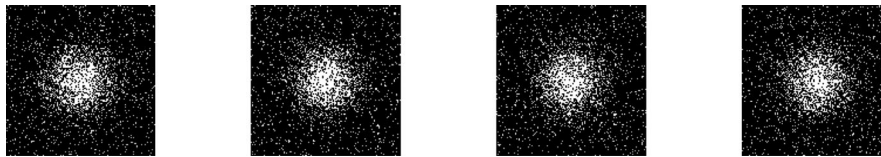


FIG. 10. A series of Gaussian patterns used for k-space subsampling.

TABLE V. Testing on anatomical dataset with k-space subsampling ratio 70% and 15% using Gaussian patterns and 200 time frames.

		Ma et al. ¹	BLIP ³	FLOR ⁶	CNN ¹⁴	FNN ¹³	Proposed	
							Basic	Nonlocal
k-space subsampling factor $\beta = 70\%$								
PSNR (dB)	23.69/38.17	45.67/47.84	50.11/50.85	49.71/45.48	50.15/51.08	50.79/51.59	49.87/57.57	
SNR (dB)	8.73/13.84	31.28/23.49	35.67/26.48	35.26/21.19	35.70/26.67	36.34/27.19	35.42/33.30	
RMSE (ms)	294.32/30.87	23.42/10.14	14.01/7.17	14.71/13.31	13.99/6.98	12.99/6.57	14.44/3.31	
Time cost (s)	72.88	75.70	85.35	23.72	23.53	24.85	26.3	
k-space subsampling factor $\beta = 15\%$								
PSNR (dB)	27.94/32.84	35.45/39.25	44.95/46.11	43.74/35.98	45.03/45.90	45.23/44.44	45.39/51.32	
SNR (dB)	13.50/8.61	20.99/14.58	30.51/21.89	29.23/12.26	30.58/21.32	30.76/19.78	30.91/26.99	
RMSE (ms)	180.3/57.03	76.01/27.25	25.46/12.37	29.27/39.73	25.21/12.68	24.65/15.00	24.20/6.79	
Time cost (s)	106	112.8	121.7	24.54	24.36	25.67	27.31	

Fig. 10, leading to a k-space measurement matrix \mathbf{Y} of size $Q \times L = 2458 \times 200$. Similarly, $\beta = 70\%$ gives rise to a k-space measurement matrix \mathbf{Y} of size $Q \times L = 11\,469 \times 200$. A larger λ enforces lower rank for the restored signature matrix \mathbf{X} to strengthen the de-aliasing effect, while a smaller λ encourages \mathbf{X} to have a subsampled Fourier transform that approximates the k-space measurements matrix \mathbf{Y} better. Therefore, we tried a range of values from 1 to 20 for λ and experimentally select the best one $\lambda = 5$. Since the low-rank based signature restoration involves gradient descent steps, a larger step size μ accelerates gradient descent speed, but tends to result in oscillation or even divergence, while a smaller μ leads to a slower convergence. We experimentally find that $\mu = 1$ gives a good balance. The same k-space measurements \mathbf{Y} are also used by dictionary matching-based methods^{1,3,6} for comparison, and the same signature restoration approach is used to convert \mathbf{Y} onto \mathbf{X} for learning-based methods.^{13,14} The aim is to evaluate the overall performance on both signature restoration and parameter restoration.

Quantitative performance is shown in Table V. Note that the advantage of learning-based methods over dictionary matching degrades when the subsampling factor increases. This is due to the fact that the restored signatures from highly subsampled k-space data exhibit deviations and distortions, thus leading to poorer input for the trained networks. In spite of this, the proposed approach outperforms the dictionary matching-based methods^{1,3} with significant gains, and also yields better or comparable performance as the state-of-the-art methods FLOR,⁶ CNN¹⁴, and FNN.¹³ In addition, it takes around 23 s for low-rank based signature restoration and <3 s for network-based parameter restoration. Thus, the total time cost is around 26 s, almost $4.8\times$ faster than FLOR.⁶ Furthermore, the speed of our method is $60\times$ faster than FLOR⁶ for parameter restoration.

We compared the performance with/without nonlocal operations in our developed network. The results in Tables IV and V show that the proposed network with nonlocal operation-based self-attention scheme outperforms the basic

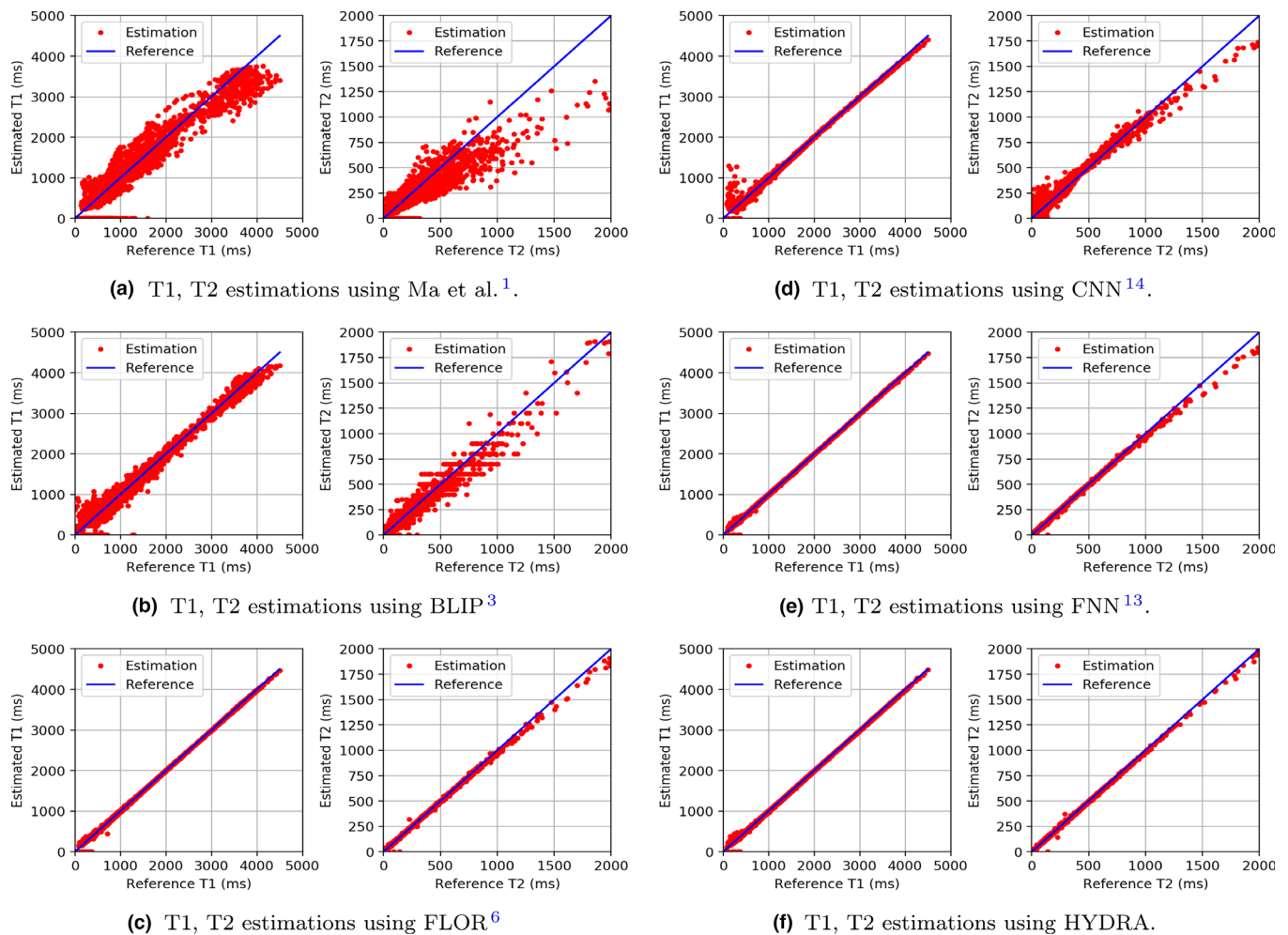


FIG. 11. Testing on the anatomical dataset with k-space subsampling factor 15% using Gaussian patterns and 1000 time frames. Subfig. (a)–(f) show the results using Ma et al.,¹ BLIP,³ FLOR,⁶ convolutional neural network by Hoppe et al.,¹⁴ FNN by Cohen et al.,¹³ and HYDRA. [Color figure can be viewed at wileyonlinelibrary.com]

counterpart. In particular, the nonlocal version achieves 6 dB gains in terms of SNR over the basic version for T2 mapping. Such a improvement demonstrates the benefits of capturing long-range dependencies and global features using the nonlocal operation-based attention scheme.

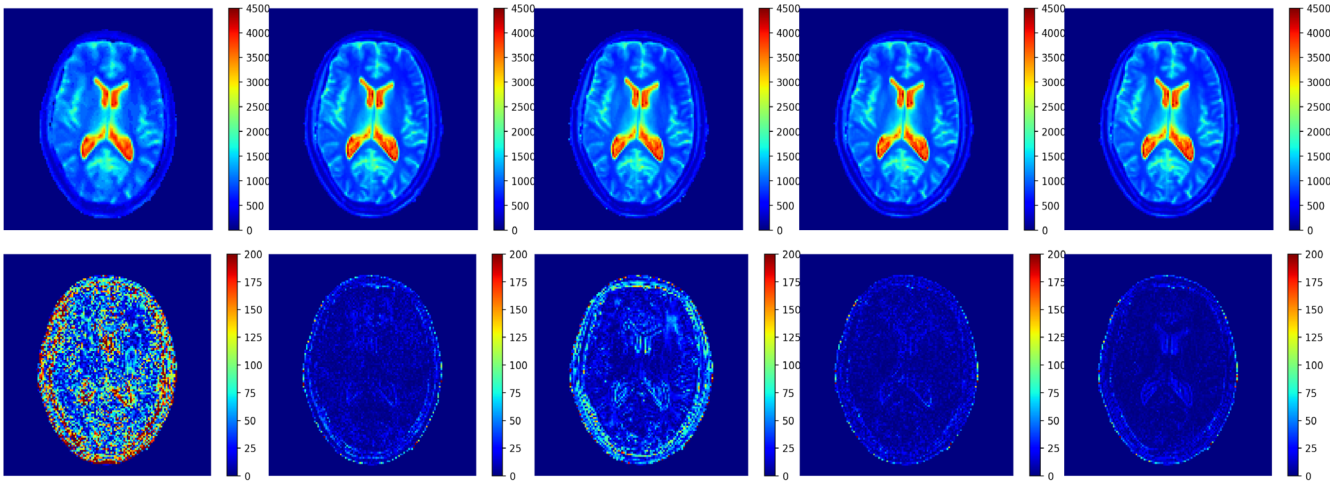
We also investigated the performance with respect to the number of time frames. In particular, we increased L from 200 to 1000 and kept other experiment settings the same as before. The quantitative results are shown in Table VI. It is noticed that given more time frames, all the methods show better performance. Moreover, the performance of learning-based methods, including CNN,¹⁴ FNN¹³, and HYDRA, improve more than model-based techniques.^{1,3,6} In particular, our approach outperforms competing algorithms quantitatively in terms of PSNR, SNR, and RMSE, as well as demonstrates visual advantage, as shown in Figs. 11 and 12.

3.C.3. k-space subsampling using spiral trajectories

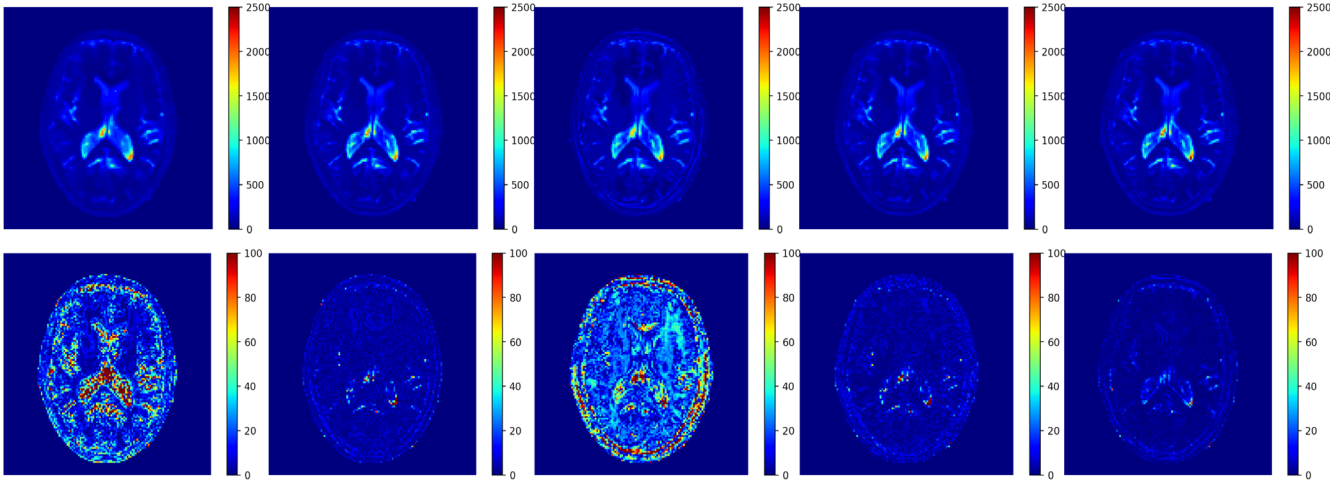
We carried out additional experiments with widely used non-Cartesian sampling patterns — variable density spiral

trajectories.^{6,37} A set of spiral trajectories used in the experiments is shown in Fig. 13. They have an FOV of 24 and a rotation angle difference of 7.5° between any two adjacent spirals to spread out the alias artifacts. Given such spiral trajectories, data were subsampled to acquire 1488 k-space samples in each time frame, leading to a subsampling ratio of 9% which is defined by the number of acquired samples in the k-space domain divided by the number of pixels in a frame. This setting closely matches the original MRF paper.¹ where each single spiral trajectory samples 1450 k-space points (leading to a subsampling ratio around 9%) and any two adjacent spiral trajectories have a rotation angle of 7.5° .

In the case of spiral subsampling, during the signature restoration, SParse Uniform ReSampling (SPURS) algorithm³⁸ was exploited to implement nonuniform Fourier transform between k-space domain and image domain, as SPURS has proved to achieve smaller approximation errors while maintaining low computational cost comparing with other resampling methods, such as nonuniform-FFT algorithm³³ and regularized Block Uniform ReSampling (rBURS).³⁹ In the experiments, 1000 density variable spiral trajectories were used for k-space subsampling, leading to



(a) T1 estimation (top row) and residual errors (bottom row).



(b) T2 estimation (top row) and residual errors (bottom row).

BLIP³ SNR = 21.06/14.87 dB FLOR⁶ SNR = 36.44/25.65 dB CNN¹⁴ SNR = 27.44/15.05 dB FNN¹³ SNR = 38.17/25.43 dB HYDRA SNR = 37.86/28.35 dB

FIG. 12. Visual results of testing on anatomical dataset with k-space subsampling factor 15% using Gaussian pattern with $L = 1000$. Comparison between BLIP,³ FLOR,⁶ convolutional neural network by Hoppe et al.,¹⁴ FNN by Cohen et al.,¹³ and HYDRA. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE VI. Testing on anatomical dataset with k-space subsampling ratio 15% using Gaussian patterns and 1000 time frames

	Ma et al. ¹	BLIP ³	FLOR ⁶	CNN ¹⁴	FNN ¹³	Proposed
PSNR (dB)	27.53/33.28	35.50/39.10	50.90/50.04	41.96/39.21	52.62/49.86	52.32/52.79
SNR (dB)	13.09/9.05	21.06/14.87	36.44/25.65	27.44/15.05	38.17/25.43	37.86/28.35
RMSE (ms)	189.09/54.21	75.53/27.74	12.83/7.87	35.91/27.37	10.52/8.04	10.89/5.74

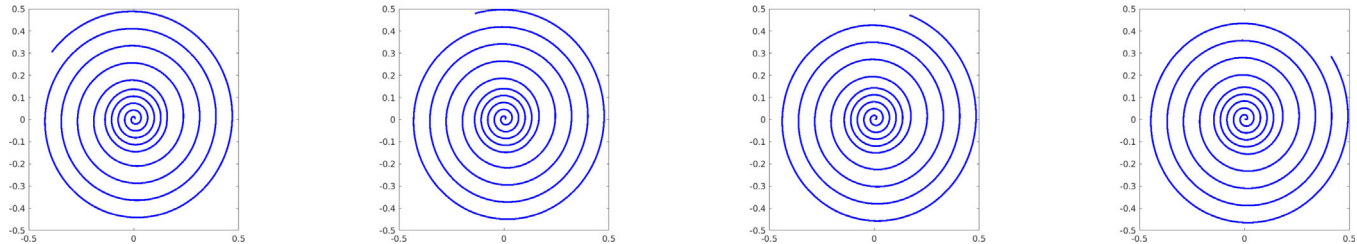
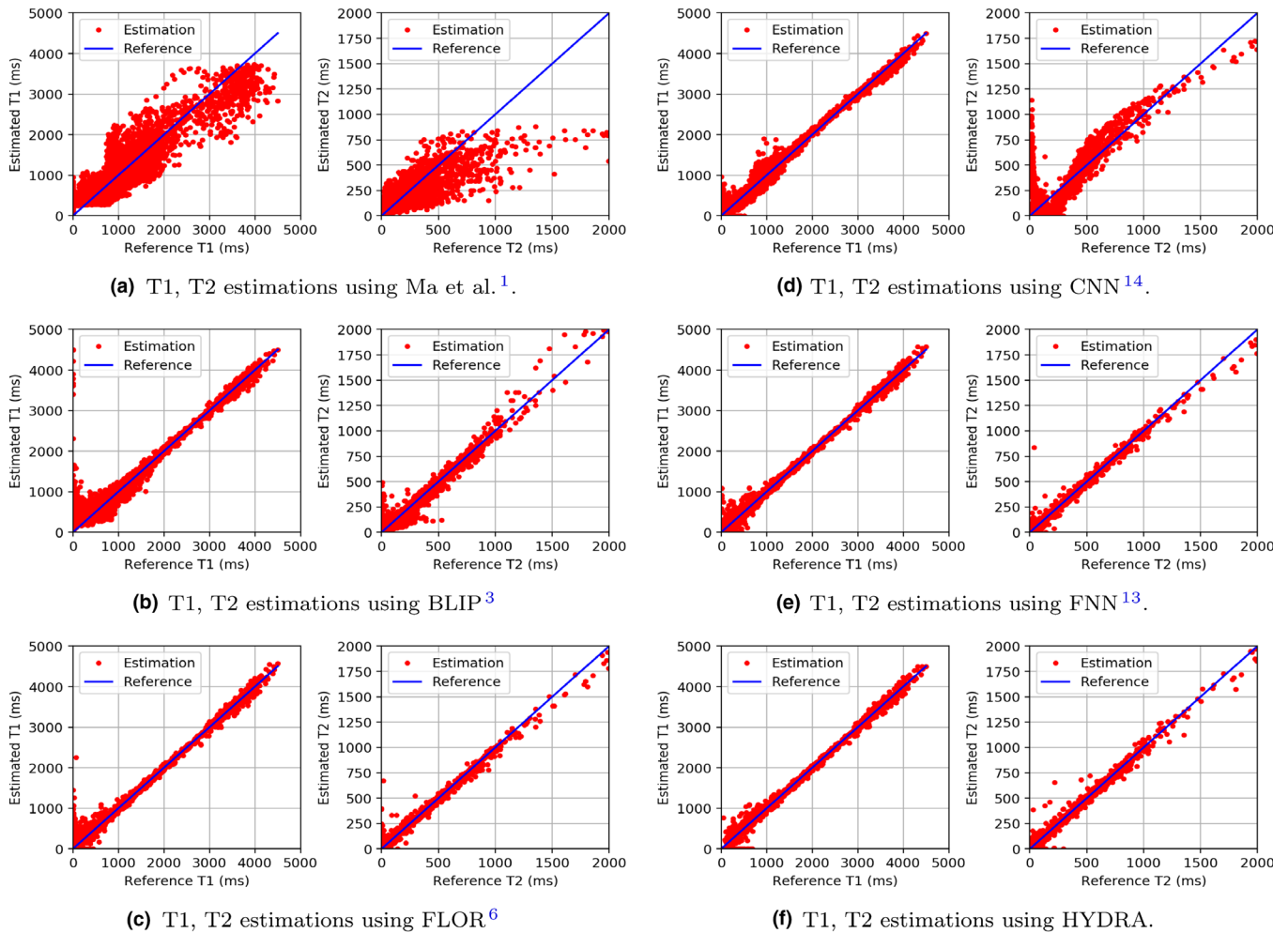


FIG. 13. A series of spiral trajectories for k-space subsampling. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE VII. Testing on anatomical dataset with k-space subsampling ratio 9% using spiral trajectories and 1000 time frames

	Ma et al. ¹	BLIP ³	FLOR ⁶	CNN ¹⁴	FNN ¹³	Proposed
PSNR (dB)	26.66/30.44	29.35/39.47	39.32/44.60	35.68/27.74	40.26/44.70	41.45/45.41
SNR (dB)	12.22/6.21	15.03/15.22	24.88/20.38	21.45/4.48	25.84/20.30	27.02/21.04
RMSE (ms)	209.01/75.18	153.37/26.57	48.67/14.72	73.96/102.57	43.65/14.55	38.08/13.41

FIG. 14. Testing on the anatomical dataset with k-space subsampling factor 9% using spiral trajectories and 1000 time frames. Subfig. (a)–(f) show the results using Ma et al.¹, BLIP³, FLOR⁶, convolutional neural network by Hoppe et al.,¹⁴ FNN by Cohen et al.¹³, and HYDRA. [Color figure can be viewed at wileyonlinelibrary.com]

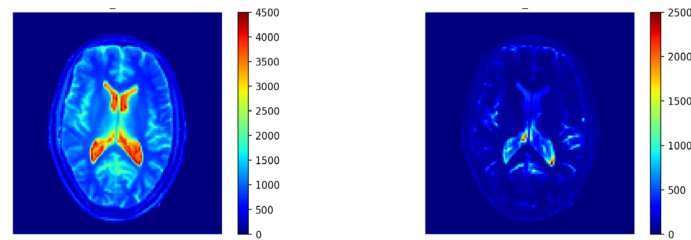
1000 time frames. The quantitative and qualitative reconstruction results demonstrate that HYDRA outperforms competing methods with smaller estimation errors, as shown in Table VII, Figs. 14 and 15.

4. DISCUSSION

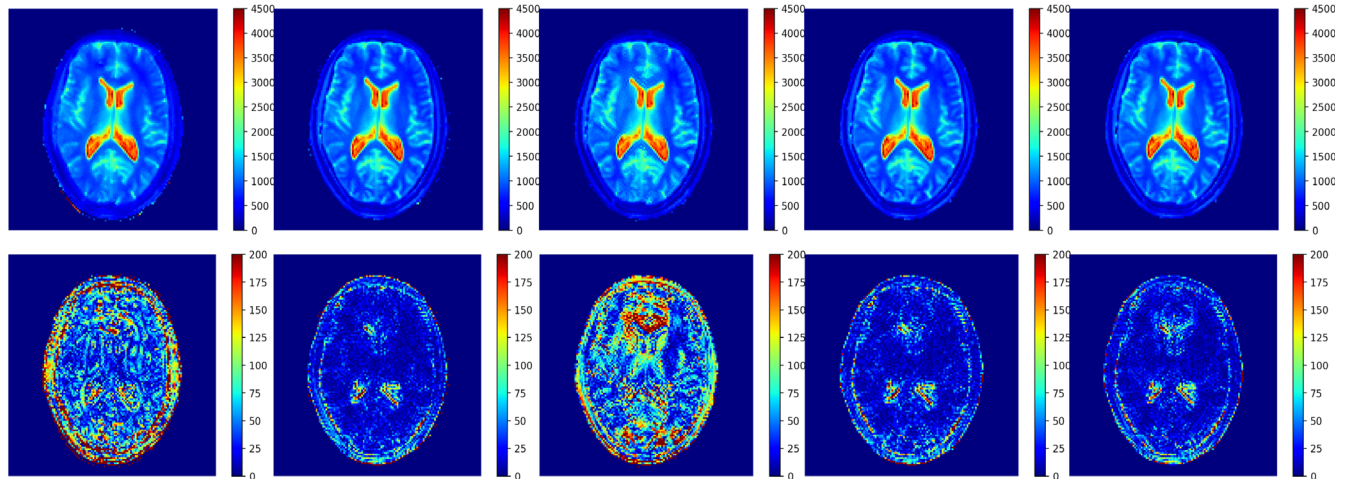
4.A. Relation to previous works

Our low-rank based signature restoration method is adapted from FLOR⁶ by removing the operation of projecting the temporal signal onto a dictionary. Thus, the signature restoration does not require a simulated dictionary, and saves

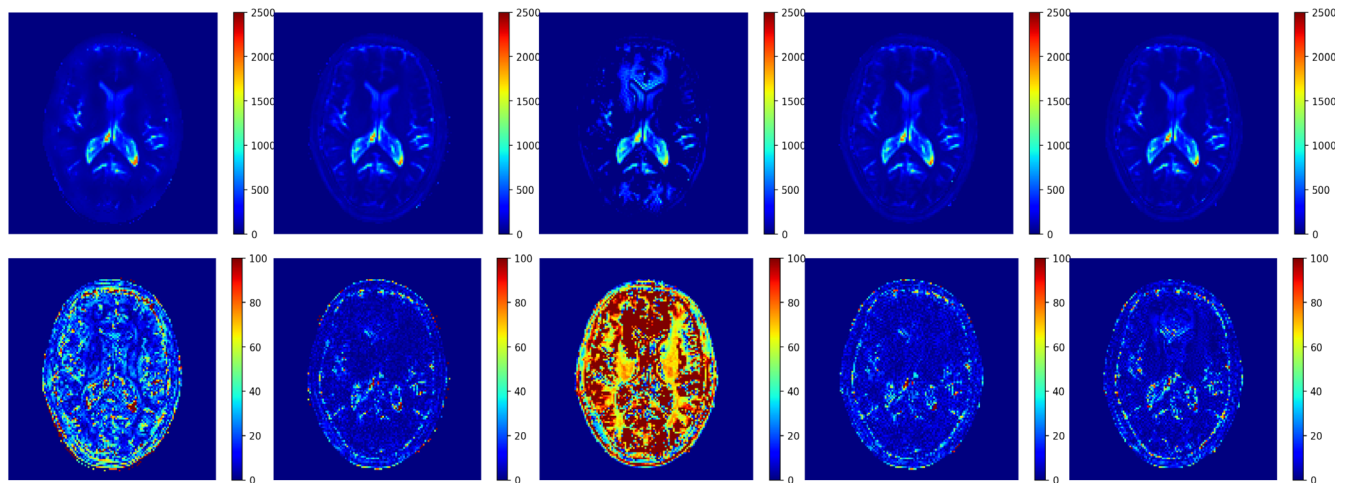
computational cost. Although recent works^{13,14} exploit neural networks to perform parameter mapping, replacing dictionaries and lookup tables used in conventional MRF reconstruction approaches, our technique is different from these methods.^{13,14} We design a deep nonlocal residual CNN for capturing signature-to-parameter mapping which is organically combined with low-rank based de-aliasing techniques for signature restoration. In this way, our algorithm can bypass some of the issues associated with other techniques: (a) The input dimension issue. The proposed approach can ingest temporal signatures with different lengths without the need to change the structure of the network. This is due to the fact that we rely on convolutional neural networks (CNNs)



(a) T1 Reference (left) and T2 Reference (right).



(b) T1 estimation (top row) and residual errors (bottom row).



(c) T2 estimation (top row) and residual errors (bottom row).

BLIP³ SNR = 15.03/15.22 dB FLOR⁶ SNR = 24.88/20.38 dB CNN¹⁴ SNR = 21.45/4.48 dB FNN¹³ SNR = 25.84/20.30 dB HYDRA SNR = 27.02/21.04 dB

FIG. 15. Visual results of testing on anatomical dataset with k-space subsampling factor 9% using spiral trajectories with $L = 1000$. Comparison between BLIP,³ FLOR,⁶ convolutional neural network by Hoppe et al.,¹⁴ FNN by Cohen et al.,¹³ and HYDRA. (a)–(c) show the reference T1/T2 maps, reconstructed T1 maps and T2 maps, respectively. [Color figure can be viewed at wileyonlinelibrary.com]

rather than fully connected neural networks (FNNs) such as the model used in Ref. ¹³ (b) The k-space subsampling issue. The proposed approach involves a hybrid of a neural network with a low-rank based de-aliasing approach. Thus, it is able to deal with correlations both over time and space via exploiting low-rank regularization and convolution operation. This

enables our work to handle k-space subsampling situations. (c) The complex mappings issue. By exploiting a residual network structure, our method can be successfully extended to deeper levels and thus obtain a better capacity to learn complex signature-to-parameter mapping functions. (d) Distortion and corruption issue. Due to the subsampling in k-

space, the restored temporal signatures suffer from local distortion and corruption. Such deviation may lead to performance degeneration in the second stage. By incorporating nonlocal operations in the network design, our method is able to capture global features and find most relevant components for inference, thereby reducing interference of local distortion and corruption.

4.B. Computational complexity

HYDRA involves two main stages: the low-rank based signature restoration stage and the network-based parameter restoration stage. Even though the time cost for parameter restoration is longer than previous methods,^{13,14} the time cost in the this stage is only a small fraction of the total time consumption, as the computational complexity is dominated by the signature restoration stage. In other words, the computational burden of HYDRA lies in the SVD calculation in the first stage. Hence, fast SVD methods can be employed to dramatically improve the efficiency of signature restoration.

4.C. Model storage requirements

Regarding the storage requirement (in double precision), HYDRA needs only 2.1 mb to store the network with 0.5 million parameters, while it requires 108 mb to store a simulated dictionary of size $80\ 100 \times 200$ and 551 mb for size $80\ 100 \times 1000$. Note that the dictionary volume will grow exponentially with the number of parameters, but the space required for storing a network is not strictly limited by the dictionary density once the topology of the network is fixed, thus significantly alleviating the storage burden inherent to the exponential growth of multidimensional dictionaries.

4.D. Impact of providing continuous T1/T2 values

Providing continuous T1/T2 values is an advantage of neural network-based parameter mapping over dictionary matching. This property may find promising applications in some practical scenarios, for example, monitoring sensitive changing of pathology condition over time, such as multiple sclerosis,^{40,41} stroke,⁴² and treatment responses,^{43,44} where the differences in T1 and T2 values between healthy and diseased tissues or between disease stages could be very small.⁴⁴ On the other hand, to fulfil this potential of network-based MRF techniques, prerequisites on the accuracy and precision of MRI measurements are needed. Taking T1/T2 quantification as an example, even for the inversion recovery spin echo (IR-SE)/multiple single-echo spin echo MRI sequences which are considered as the gold standard for T1/T2 quantification, there exist variations of 2%–9% on the measured relaxation times.⁴⁵ Such anatomical measurement uncertainties and model imperfections may weaken the advantage and clinical impact of providing continuous T1/T2 values using network-based MRF techniques to some extent. Therefore,

improving the accuracy of gold standard approaches in the future would contribute to making the most of the potential of neural networks in the MRF domain.

5. CONCLUSION

We proposed a hybrid deep MRF approach which combines low-rank based signature restoration with learning-based parameter restoration. In our approach, a low-rank based sub-sampling method is used to restore clean signatures from down-sampled k-space measurements. Then, a 1D deep nonlocal residual CNN is developed for efficient signature-to-parameter mapping, replacing the time-consuming dictionary matching operation in conventional MRF techniques. Our approach demonstrates advantages in terms of inference speed, accuracy, and storage requirements over competing MRF methods as no dictionary is needed for recovery.

ACKNOWLEDGMENTS

This work was supported by the Royal Society International Exchange Scheme IE160348, by the European Union's Horizon 2020 grant ERC-BNYQ, by the Israel Science Foundation grant no. 335/14, by ICore: the Israeli Excellence Center 'Circle of Light', by the Ministry of Science and Technology, Israel, by UCL Overseas Research Scholarship (UCL-ORS) and by China Scholarship Council (CSC) and by EPSRC grant EP/K033166/1.

CONFLICT OF INTEREST

The authors have no conflict of interest to disclose.

^{a)}Author to whom correspondence should be addressed. Electronic mail: p.song@imperial.ac.uk.

REFERENCES

1. Ma D, Gulani V, Seiberlich N, et al. Magnetic resonance fingerprinting. *Nature*. 2013;495:187.
2. Jiang Y, Ma D, Seiberlich N, Gulani V, Griswold MA. MR fingerprinting using fast imaging with steady state precession (FISP) with spiral readout. *Magn Reson Med*. 2015;74:1621–1631.
3. Davies M, Puy G, Vanderghenst P, Wiaux Y. A compressed sensing framework for magnetic resonance fingerprinting. *SIAM J Imaging Sci*. 2014;7:2623–2656.
4. Wang Z, Li H, Zhang Q, Yuan J, Wang X. Magnetic Resonance Fingerprinting with compressed sensing and distance metric learning. *Neurocomputing*. 2016;174:560–570.
5. Mazor G, Weizman L, Tal A, Eldar YC. Low rank magnetic resonance fingerprinting. in Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the. IEEE; 2016:439–442.
6. Mazor G, Weizman L, Tal A, Eldar YC. Low-rank magnetic resonance fingerprinting. *Med Phys*. 2018;45:4066–4084.
7. Liao C, Bilgic B, Manhard MK, et al. 3D MR fingerprinting with accelerated stack-of-spirals and hybrid slidingwindow and GRAPPA reconstruction. *NeuroImage*. 2017;162:13–22.
8. Cao X, Liao C, Wang Z, et al. Robust sliding-window reconstruction for Accelerating the acquisition of MR fingerprinting. *Magn Reson Med*. 2017;78:1579–1588.

9. Lustig M, Donoho DL, Santos JM, Pauly JM. Compressed sensing MRI. *IEEE Signal Process Mag.* 2008;25:72–82.
10. Weizman L, Eldar YC, Ben Bashat D. Compressed sensing for longitudinal MRI: an adaptive-weighted approach. *Med Phys.* 2015;42:5195–5208.
11. Weizman L, Eldar YC, Ben Bashat D. Reference-based MRI. *Med Phys.* 2016;43:5357–5369.
12. Eldar YC. *Sampling Theory: Beyond Bandlimited Systems*. Cambridge: Cambridge University Press; 2015.
13. Cohen O, Zhu B, Rosen MS. MR fingerprinting deep reconstruction network (DRONE). *Magn Reson Med.* 2018;80:885–894.
14. Hoppe E, Kördörfer G, Würfl T, et al. Deep learning for magnetic resonance fingerprinting: a new approach for predicting quantitative parameter values from time series. *Stud Health Technol Inform.* 2017;243:202–206.
15. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521:436.
16. Goodfellow I, Bengio Y, Courville A, Bengio Y. *Deep Learning*, Vol. 1. Cambridge: MIT Press; 2016.
17. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*; 2012:1097–1105.
18. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proc. IEEE Conf. Comput. Vision Pattern Recog.* 2016:770–778.
19. He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: *Proc. Eur. Conf. Comput. Vision*. Springer; 2016:630–645.
20. Dong C, Loy CC, He K, Tang X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 2016;38:295–307.
21. Kim J, Kwon Lee J, Mu Lee K. Deeply-recursive convolutional network for image super-resolution. In: *Proc. IEEE Conf. Comput. Vision Pattern Recog.* 2016:1637–1645.
22. Gehring J, Auli M, Grangier D, Yarats D, Dauphin YN. Convolutional sequence to sequence learning. *arXiv preprint arXiv:1705.03122*; 2017.
23. Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Process Mag.* 2012;29:82–97.
24. Cai J-F, Candes EJ, Shen Z. A singular value thresholding algorithm for matrix completion. *SIAM J Optim.* 2010;20:1956–1982.
25. Sra S, Nowozin S, Wright SJ. *Optimization for Machine Learning*. Cambridge: MIT Press; 2012.
26. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: *Advances in neural information processing systems*; 2017:5998–6008.
27. Wang X, Girshick R, Gupta A, He K. Non-local neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2018:7794–7803.
28. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: *Proc. IEEE Int. Conf. Comput. Vision*; 2015:1026–1034.
29. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res.* 2014;15:1929–1958.
30. Hennig J. Echoes – how to generate, recognize, use or avoid them in MR-imaging sequences. Part I: Fundamental and not so fundamental properties of spin echoes. *Concepts Magn Reson.* 1991;3:125–143.
31. Weigel M. Extended phase graphs: dephasing, RF pulses, and echoes—pure and simple. *J Magn Reson Imaging.* 2015;41:266–295.
32. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*; 2014.
33. Fessler JA, Sutton BP. Nonuniform fast Fourier transforms using min-max interpolation. *IEEE Trans Sig Proc.* 2003;51:560–574.
34. Vymazal J, Righini A, Brooks RA, et al. T1 and T2 in the brain of healthy subjects, patients with Parkinson disease, and patients with multiple system atrophy: relation to iron content. *Radiology.* 1999;211:489–495.
35. Liberman G, Louzoun Y, Ben Bashat D. T1 mapping using variable flip angle SPGR data with flip angle correction. *J Magn Reson Imaging.* 2014;40:171–180.
36. Deoni SC, Peters TM, Rutt BK. High-resolution T1 and T2 mapping of the brain in a clinically acceptable time with DESPOT1 and DESPOT2. *Magn Reson Med.* 2005;53:237–241.
37. Lee JH, Hargreaves BA, Hu BS, Nishimura DG. Fast 3D imaging using variable-density spiral trajectories with applications to limb perfusion. *Magn Reson Med.* 2003;50:1276–1285.
38. Kiperwas A, Rosenfeld D, Eldar YC. The SPURS algorithm for resampling an irregularly sampled signal onto a cartesian grid. *IEEE Trans Med Imaging.* 2017;36:628–640.
39. Rosenfeld D. New approach to gridding using regularization and estimation theory. *Magn Reson Med.* 2002;48:193–202.
40. Manfredonia F, Ciccarelli O, Khaleeli Z, et al. Normal-appearing brain t1 relaxation time predicts disability in early primary progressive multiple sclerosis. *Arch Neurol.* 2007;64:411–415.
41. Papadopoulos K, Tozer DJ, Fisniku L, et al. T1-relaxation time changes over five years in relapsing-remitting multiple sclerosis. *Mult Scler J.* 2010;16:427–433.
42. Bernarding J, Braun J, Hohmann J, et al. Histogram-based characterization of healthy and ischemic brain tissues using multiparametric MR imaging including apparent diffusion coefficient maps and relaxometry. *Magn Reson Med.* 2000;43:52–61.
43. McSheehy PM, Weidensteiner C, Cannet C, et al. Quantified tumor T1 is a generic early-response imaging biomarker for chemotherapy reflecting cell viability. *Clin Cancer Res.* 2010;16:212–225.
44. Weidensteiner C, Allegrini PR, Sticker-Jantschkeff M, Romanet V, Ferretti S, McSheehy PM. Tumour T1 changes in vivo are highly predictive of response to chemotherapy and reflect the number of viable tumour cells—a preclinical MR study in mice. *BMC Cancer.* 2014;14:88.
45. Jiang Y, Ma D, Keenan KE, Stupic KF, Gulani V, Griswold MA. Repeatability of magnetic resonance fingerprinting T1 and T2 estimates assessed using the ISMRM/NIST MRI system phantom. *Magn Reson Med.* 2017;78:1452–1457.