

# MAGNETIC RESONANCE FINGERPRINTING USING A RESIDUAL CONVOLUTIONAL NEURAL NETWORK

Pingfan Song\* Yonina C. Eldar<sup>#</sup> Gal Mazor<sup>#</sup> Miguel R. D. Rodrigues\*

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

<sup>#</sup> Department of Electrical Engineering, Technion – Israel Institute of Technology, Israel

## ABSTRACT

Conventional dictionary matching based MR Fingerprinting (MRF) reconstruction approaches suffer from time-consuming operations that map temporal MRF signals to quantitative tissue parameters. In this paper, we design a 1-D residual convolutional neural network to perform the signature-to-parameter mapping in order to improve inference speed and accuracy. In particular, a 1-D convolutional neural network with shortcuts, a.k.a skip connections, for residual learning is developed using a TensorFlow platform. To avoid the requirement for a large amount of MRF data, the designed network is trained on synthesized MRF data simulated with the Bloch equations and fast imaging with steady state precession (FISP) sequences. The proposed approach was validated on both synthetic data and phantom data generated from a healthy subject. The reconstruction performance demonstrates a significantly improved speed – only 1.6s for reconstructing a pair of T1/T2 maps of size  $128 \times 128 - 50 \times$  faster than the original dictionary matching based method. The better performance was also confirmed by improved signal to noise ratio (SNR) and reduced root mean square error (RMSE). Furthermore, it is more compact to store a network instead of a large dictionary.

**Index Terms**— Magnetic Resonance Fingerprinting, Quantitative Magnetic Resonance Imaging, deep learning, residual Convolutional Neural Network

## 1. INTRODUCTION

Magnetic Resonance Fingerprinting (MRF) [1–7] 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 pseudo-random pulse sequence is unique, MRF exploits pseudo-randomized acquisition parameters to create unique temporal signal signatures, analogue to the "fingerprint", for different tissues. Then, a dictionary matching operation is performed to map an inquiry temporal signature to the best matching entry in a pre-computed dictionary which leads to multiple tissue parameters directly. In particular, the temporal signatures is generated by varying the acquisition parameters of a pseudo-random excitation pulse sequence, such as repetition time (TR), time of echo (TE), and radio

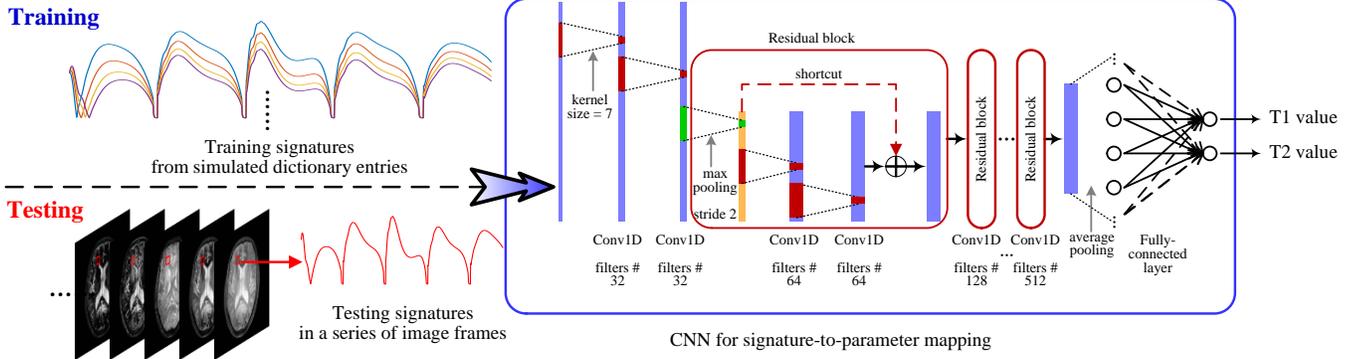
frequency flip angle (FA) over time. The dictionary is composed of a large amount of entries that are usually simulated refer to the Bloch equations and given pseudo-random pulse sequence. Each entry represents a unique temporal signature associated with a specific tissue and is also linked to its quantitative 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 simple look-up table operation.

Due to the k-space subsampling, temporal signatures are often corrupted by blurring, noising or aliasing, thus hamper the accuracy of estimating the tissue parameters using the dictionary matching. In order to alleviate the impact of corruption, denoising or 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. In particular, inspired by the successful application of sparsity in MRI reconstruction [8–10], some work [3–6] suggest to incorporate prior knowledge such as sparsity and low-rank to attenuate corruption, improving the signature restoration performance.

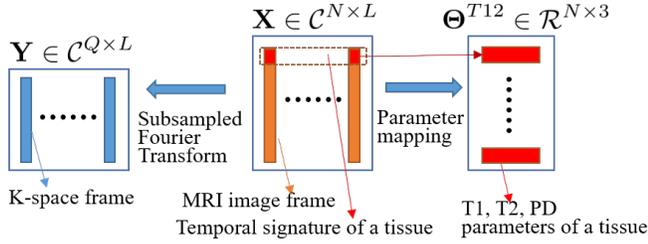
However, all of above-mentioned approaches use dictionary matching to perform mapping from purified temporal signatures to tissue’s quantitative parameters. Such dictionary matching based signature-to-parameter mapping exhibits some drawbacks [11–13]. First, it usually requires a huge dictionary composed of a large number of entries to represent tissues with fine granularity over the entire range of possible tissue values. Further, the dictionary size often grows exponentially with the number of tissue parameters, so that storing the large dictionary becomes prohibitively memory-consuming. Second, finding the best matching entry needs to compare the inquiry temporal signature with the whole dictionary. Thus, dictionary matching operation is prohibitively time-consuming for a large dictionary, considerably limit the inference speed.

In this paper, we consider a different approach to alleviate some of these issues based on deep neural networks (a.k.a. deep learning) [14, 15]. The rationale has to do with the fact that a well designed and tuned deep neural networks can approximate very well complex functions, leading to state-of-the-art results in a number of tasks such as image classification, image super-resolution, and many more [16–22]. Recent work [11–13] propose to exploit neural networks to replace the dictionary and the look-up table used in conventional MRF reconstruction approaches. Different from the fully-connected feed-forward neural network [11], and the convolutional neural network (CNN) model [12, 13], we propose a 1-D deep residual convolutional neural network to perform signature-to-parameter mapping, replacing the memory-consuming dictionary and time-consuming dictionary matching. A series of experiments demonstrate that our designed neural network is capable of severing as a compact model to represent complicated mappings.

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).



**Fig. 1.** Diagram of designed 1-D residual CNN for signature-to-parameter mapping in MRF reconstruction. During the training stage, each dictionary entry, as a 1-D time sequence, is input into the network for a training signature. During the testing stage, the signature for each pixel is extracted from a purified image stack to serve as a testing signature which is mapped to corresponding T1/T2 relaxation times by the network.



**Fig. 2.** Relationship between  $\mathbf{X}$ ,  $\mathbf{Y}$ ,  $\Theta^{T12}$ .

## 2. PROBLEM FORMULATION

Similar to a video structure, MRF data is composed of multiple frames sampled in the k-space, where each frame is an MR contrast acquired with different acquisition parameters, such as TR, TE and FA, at a different time point. A series of frames are stacked together along the temporal dimension to construct a 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, equivalent to the number of time points. Every column vector  $\mathbf{Y}_{:,i}$  represents an subsampled Fourier transform of an image frame  $\mathbf{X}_{:,i}$  from the contrasts matrix  $\mathbf{X} \in \mathbb{C}^{N \times L}$ , i.e.  $\mathbf{Y}_{:,i} = F_u\{\mathbf{X}_{:,i}\}$ , where  $F_u\{\cdot\}$  denotes an subsampled 2-D Fourier transform at the time point  $i \in [1, L]$ ,  $N$  denotes the number of pixels in each image frame. The relationship of these variables is shown in Figure 2 for better illustration.

Each image frame  $\mathbf{X}_{:,i}$  represents a response MR contrast acquired at a single time point with different acquisition parameters, stacked as a column vector  $\Theta_{:,i}^{TRE} = [TR^i, TE^i, FA^i]^T$ ,  $i \in [1, L]$ , where TR and TE denote the repetition time and echo time, respectively, and FA denotes the flip angle of the RF pulse. Each row vector  $\mathbf{X}_{j,:}$  represents a temporal signature, i.e. temporal signal evolution of a specific tissue at the  $j$ -pixel along the sampling time points. The signature depends on the tissue's relaxation times, such as T1, T2 and proton density (PD), grouped as a row vector  $\Theta_{j,:}^{T12} = [T1^j, T2^j, PD^j]$ ,  $j \in [1, N]$ . Here,  $i$  is the temporal index, while  $j$  is the spatial index. Combining temporal dimension with spatial dimension together, the temporal signature can be formulated as  $\mathbf{X}_{j,:} = f(\Theta_{j,:}^{T12}, \Theta^{TRE})$ , where  $f(\cdot)$  denotes the Bloch equations.

Let  $\mathbf{D} \in \mathbb{C}^{K \times N}$  denote a precomputed dictionary composed of  $K$  entries. Each entry  $\mathbf{D}_{k,:}$  represents a temporal signature simulated using Bloch equations for a specific tissue characterized by

certain quantitative parameters, such as T1, T2, PD, given the RF sequence parameters  $\Theta^{TRE}$ . Such a dictionary  $\mathbf{D}$  with a large  $K$  gives fine granularity over the entire range of possible tissue values. Let  $\mathbf{LUT} \in \mathbb{R}^{K \times P}$  denote the corresponding look-up table where each row  $\mathbf{LUT}_{k,:}$  contains  $P$  quantitative parameters, e.g. T1, T2 and PD relaxation times ( $P = 3$ ), for a specific tissue. In this way, the temporal signature represented by the  $k$ -th entry  $\mathbf{D}_{k,:}$  is linked with the quantitative parameters  $\mathbf{LUT}_{k,:}$  of the  $k$ -th tissue.

Given  $\Theta^{TRE}$  and k-space samples  $\mathbf{Y}$ , the goal of MRF reconstruction is to estimate the tissue quantitative parameters  $\Theta^{T12}$ . Commonly, the image stack  $\mathbf{X}$  is first reconstructed from the k-space samples  $\mathbf{Y}$ , and then the recovered  $\mathbf{X}$  is mapped to tissue parameters via pattern recognition, such as dictionary matching [1–6]. However, our approach, as described in the next section, exploits low-rank prior for signature restoration and a trained neural network to replace the dictionary matching operation in order to improve the inference speed, as well as alleviating the storage burden.

## 3. PROPOSED MRF RECONSTRUCTION

The proposed MRF reconstruction approach involves two stages: signature restoration and parameter restoration. Given  $\Theta^{TRE}$  and k-space samples  $\mathbf{Y}$ , we proposed a low-rank based de-aliasing and denoising method for restoring signatures  $\mathbf{X}$ , and a residual convolutional neural network to map each signature  $\mathbf{X}_{j,:}$  to corresponding tissue parameters  $\Theta_{j,:}^{T12}$  directly. Such signature and parameter restoration operations are formulated as

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

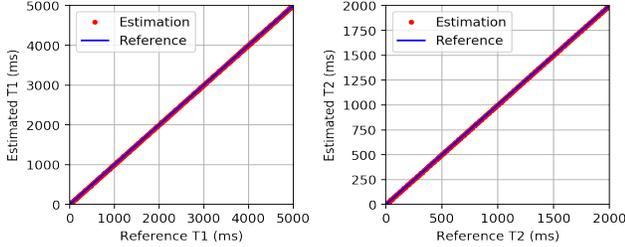
where function  $\mathbf{X} = h(\mathbf{Y})$  represents the signature restoration operator and function  $\Theta_{j,:}^{T12} = g(\mathbf{X}_{j,:} | \Theta^{TRE})$  represents the parameter restoration operator.

### 3.1. Low-rank for signature restoration

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

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

where  $\|\mathbf{X}\|_*$  denotes the nuclear norm, the relaxation form of low-rank, defined as the sum of the singular values of  $\mathbf{X}$  [23]. (2) can be solved using the incremental subgradient proximal method [24].



**Fig. 3.** Parameter restoration performance using designed network on synthetic data. Blue and red lines represent the groundtruth and estimation of corresponding parameter. It can be noticed that the trained network fits the parameters well in the whole range, yielding high correlation coefficients and low RMSE. Correlation Coefficients for T1 / T2:  $R^2 = 0.99999986/0.99999963$ ; RMSE = 0.659 / 0.491.

### 3.2. Residual CNN for parameter restoration

Once the imaging contrasts matrix  $\mathbf{X}$  is recovered, each temporal signature in  $\mathbf{X}$  is input into the designed residual convolutional neural network for parameter restoration, formulated as:

$$\Theta_{j,:}^{T12} = g(\mathbf{X}_{j,:} | \Theta^{TRE}), \forall j \in [1, N] \quad (3)$$

where  $g(\cdot)$  denotes the trained network.

By introducing convolution, local receptive field and weight sharing design, CNN can take 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 [16–21]. On the other hand, residual network architecture [17, 18] enables deeper model with more representational power to be trained effectively, as it alleviates the gradient vanishing or exploding problems by propagating gradients throughout the model via short-cuts, a.k.a skip connections.

Inspired by these network designs, the proposed network has a 1-D residual CNN architecture with short-cuts for residual learning. As illustrated in Figure 1, it starts with two 1-D convolutional layers before connecting with 4 residual blocks, and finally ends with a global-average-pooling layer followed by a fully-connected layers. 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 in each block. The filter size in each convolutional layer is set to be equal to 7. The number of channels, a.k.a feature maps, in the first two convolutional layers is set to be 32 and then is doubled in each residual block until 512 in the final residual block. The size of each filter map halves due to max-pooling in each block. In this way, we gradually reduce temporal resolution while we extract more features to increase content information. The global-average-pooling layer is used to average each feature map in order to integrate information in each 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 over-fitting. The last fully-connected layer outputs estimated parameters, such as T1 and T2 relaxation times. It is trivial to adjust the number of output to adapt to more parameters.

Regarding the training, the network is trained on synthesized dictionary and look-up table to learn the mapping  $\mathbf{LUT}_{k,:} = g(\mathbf{D}_{k,:} | \Theta^{TRE})$ . Once the network is trained, given an inquiry temporal signature  $\mathbf{X}_{j,:}$ , the network is able to map the inquiry time sequence directly to the tissue parameters as (3).

## 4. EXPERIMENTS

In this section, we conducted a series of experiments to evaluate our approach.

### 4.1. Training

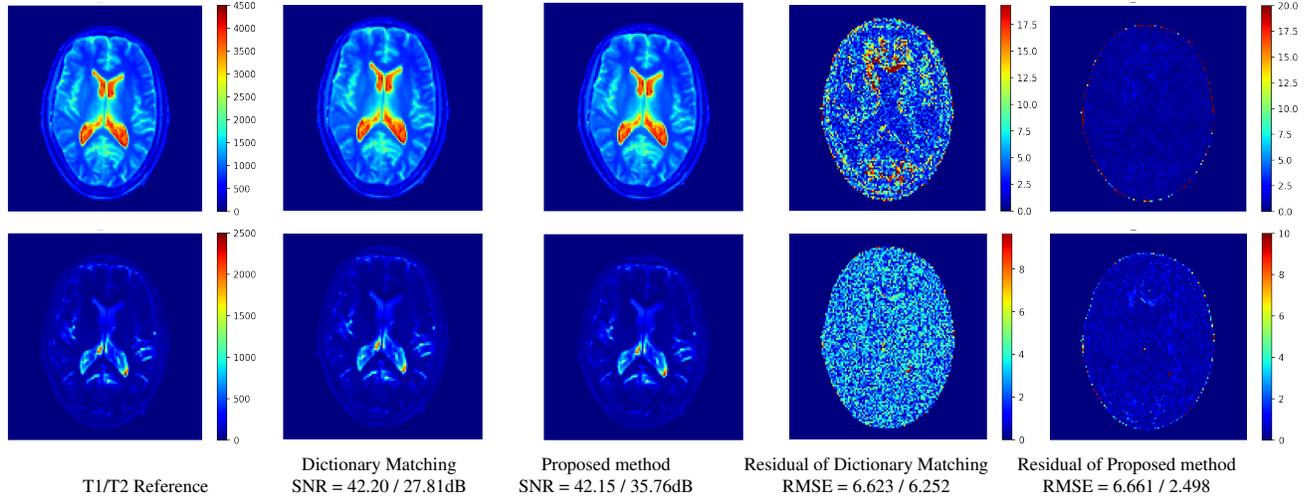
The training dataset in our numerical experiments is composed of a dictionary  $\mathbf{D}$  as input data and a Look-up-Table  $\mathbf{LUT}$  as labels.  $\mathbf{D}$  and  $\mathbf{LUT}$  were synthesized by solving the Bloch equations using the extended phase graph formalism and fast imaging with steady state precession (FISP) sequences [2, 6]. The FISP pulse sequence was designed with constant TE of 2ms, random TR values in the range of 11.5-14.5 ms, and a sinusoidal variation of FA (RF pulses) in the range of 0 - 70 degrees. We generated 80100 sequences with 200 time points as entries, which constitute the dictionary of dimension  $80100 \times 200$ . Accordingly, we obtained a Lookup-Table  $\mathbf{LUT}$  of dimension  $80100 \times 2$  for T1 and T2 parameters. The 80100 samples correspond to T1 and T2 values starting from 1 ms and increasing with an increment of 10 ms until covering the T1 parameter range of [1, 5000] ms and T2 parameter range of [1, 2000] ms. Note, those samples corresponding to  $T1 < T2$  have been excluded during the construction, as they have no physical meaning. The parameter ranges cover the relaxation time values that can be commonly found in a brain scan [25]. The model was trained for 50 training epochs. In each epoch, 20% of the training samples are separated aside for validation dataset. We use RMSE as the loss and Adam [26] as the optimization algorithm with the learning rate decaying from 1e-2 to 1e-6 every 10 epochs. The batch size was set to 256 time series. The training procedure takes around 30 seconds for one epoch in average, thus around 25 minutes for completing 50 epochs, on a NVIDIA GeForce GTX 1080 Ti GPU.

### 4.2. Testing

We evaluate the performance of a trained network on both synthetic and phantom testing dataset. The construction of the synthetic testing dataset is similar to the construction of the training dataset, but with a different starting index of the T1 and T2 parameters. Thus, the synthetic testing dataset has distinct T1 / T2 values without overlapping with the training dataset. We also construct a phantom testing dataset from the brain scans that were acquired with GE Signa 3T HDXT scanner from a healthy subject.<sup>1</sup> To generate gold standard data for the T1 and T2 parameter maps, we acquired Fast Imaging Employing Steady-state Acquisition (FIESTA) and Spoiled Gradient Recalled Acquisition in Steady State (SPGR) images, at 4 different flip angles ( $3^\circ, 5^\circ, 12^\circ$  and  $20^\circ$ ), and then implemented well known DESPOT1 and DESPOT2 [27] algorithms, after improvements as described in [28]. Based on the gold standard T1, T2 maps, we generated phantom MRF data using the same mechanism as generating the synthetic testing dataset. The sub-sampling factor is set to be 0.15, that is, only 15% k-space data is used for reconstructing image stack  $\mathbf{X}$ . The tuning parameters were experimentally set as  $\mu = 1$  and  $\lambda = 5$ .

To evaluate the parameter restoration performance of the proposed network, we input testing synthetic signatures into the network and compare its output with groundtruth T1 and T2 values. As shown in Figure 3, the estimation from the proposed network obtained outstanding agreement with the reference, yielding high correlation coefficients. In addition, it also gives higher PSNR, SNR and lower RMSE than the dictionary matching operation used in [1–

<sup>1</sup>The experiment procedures involving human subjects described in this paper were approved by the Institutional Review Board of Tel-Aviv Sourasky Medical Center, Israel.



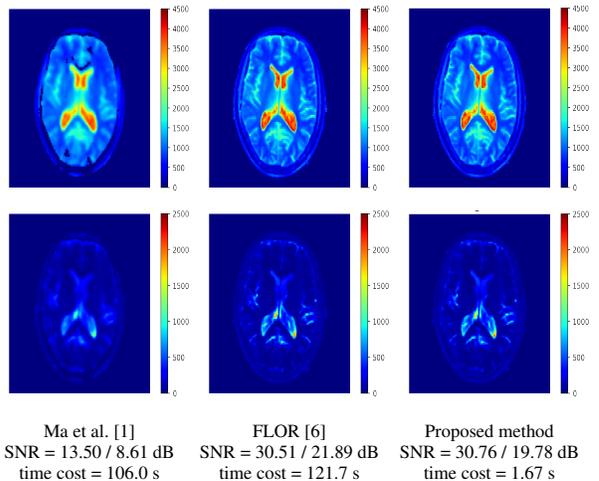
**Fig. 4.** Visual comparison between the dictionary matching method [1] and proposed method without subsampling.

6]. The most impressive advantage of the network is the fast inference speed. It is observed that the network takes only 8.2 s to complete the mapping operation for eighty thousand temporal signatures, that is,  $53\times$  faster than 464.1 s using the dictionary matching method that is based on computing inner-products of the target signature with each dictionary entry. Furthermore, since the complexity has been fixed once the network topology is fixed, the inference speed of the network is not affected by the dictionary density, i.e. the number of entries in a dictionary.

The parameter restoration performance of the proposed network is also tested on the phantom dataset without subsampling. The visual results are shown in Figure 4. It is noticed that our approach gives competitive performance for T1 mapping and yields much better performance for T2 mapping, obtaining 7.9dB SNR gains than the competing method [1]. The RMSE of T2 mapping is also reduced to 2.498 from 6.252, accordingly. This is owing to the advantage that the trained neural network is a powerful function representation that outputs continuous-valued parameters. In contrast, conventional dictionary-matching method [1] only matches signatures to discrete dictionary entries. If there is no exact parameters for an inquiry signature, dictionary-matching methods will find adjacent parameters for approximation, thus introducing mismatch. However, a well designed and trained network can still estimate the missing parameters based on captured mapping functions. In addition, the network takes only 1.6 s to accomplish the mapping for a pair of T1 / T2 parameter maps of size  $128 \times 128$ , that is,  $56\times$  faster than the dictionary matching method [1].

To evaluate both the signature restoration and the parameter restoration performance of the proposed method, we input subsampled phantom data into our model for low-rank based signature de-aliasing followed by network based parameter mapping. The same phantom data is also input into competing methods [1, 6] for comparison. The visual performance is shown in Figure 5. It is noticed that our method outperforms dictionary matching method [1] with significant gains, and also yields competitive performance as the state-of-the-art method FLOR [6]. In addition, our method is  $73\times$  faster than FLOR [6] for parameter mapping.

We also note that storing the network requires 20.3 megabytes while storing the training dictionary of size  $80100 \times 200$  requires more than 100 megabytes. Even though the dictionary volume will grow exponentially with the number of parameters, but the space



**Fig. 5.** Visual comparison between the Ma et al.'s dictionary matching [1], FLOR [6], and proposed method, with subsampling ratio 0.15.

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 multi-dimensional dictionaries.

## 5. CONCLUSION

We proposed a deep residual convolutional neural network to learn the mapping from MRF signatures to tissue parameters, combined with a low-rank based signature restoration. The proposed network plays the role of dictionary and Lookup-Table in the dictionary matching based methods. However, neural networks demonstrate various additional advantages. Owing to the feedforward characteristics, the signature-to-parameter mapping operation using the proposed network is much faster than the conventional dictionary-matching. As a neural network is a compact function representation, storing a trained network needs less memory than storing a large dictionary. Serving as a powerful function representation, neural networks are able to output continuous-valued parameters, thus perform well on estimating parameters which may not exist in a simulated dictionary.

## 6. REFERENCES

- [1] Dan Ma, Vikas Gulani, Nicole Seiberlich, Kecheng Liu, Jeffrey L Sunshine, Jeffrey L Duerk, and Mark A Griswold, “Magnetic resonance fingerprinting,” *Nature*, vol. 495, no. 7440, pp. 187, 2013.
- [2] Yun Jiang, Dan Ma, Nicole Seiberlich, Vikas Gulani, and Mark A Griswold, “MR fingerprinting using fast imaging with steady state precession (fisp) with spiral readout,” *Magnetic resonance in medicine*, vol. 74, no. 6, pp. 1621–1631, 2015.
- [3] Mike Davies, Gilles Puy, Pierre Vandergheynst, and Yves Wiaux, “A compressed sensing framework for magnetic resonance fingerprinting,” *SIAM Journal on Imaging Sciences*, vol. 7, no. 4, pp. 2623–2656, 2014.
- [4] Zhe Wang, Hongsheng Li, Qinwei Zhang, Jing Yuan, and Xiaogang Wang, “Magnetic resonance fingerprinting with compressed sensing and distance metric learning,” *Neurocomputing*, vol. 174, pp. 560–570, 2016.
- [5] Gal Mazor, Lior Weizman, Assaf Tal, and Yonina C Eldar, “Low rank magnetic resonance fingerprinting,” in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the. IEEE*, 2016, pp. 439–442.
- [6] Gal Mazor, Lior Weizman, Assaf Tal, and Yonina C Eldar, “Low-rank magnetic resonance fingerprinting,” *Medical physics*, vol. 45, no. 9, pp. 4066–4084, 2018.
- [7] Pedro A Gómez, Miguel Molina-Romero, Cagdas Ulas, Guido Bounincontri, Jonathan I Sperl, Derek K Jones, Marion I Menzel, and Bjoern H Menze, “Simultaneous parameter mapping, modality synthesis, and anatomical labeling of the brain with mr fingerprinting,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 579–586.
- [8] Michael Lustig, David L Donoho, Juan M Santos, and John M Pauly, “Compressed sensing MRI,” *IEEE signal processing magazine*, vol. 25, no. 2, pp. 72–82, 2008.
- [9] Lior Weizman, Yonina C Eldar, and Dafna Ben Bashat, “Compressed sensing for longitudinal MRI: An adaptive-weighted approach,” *Medical physics*, vol. 42, no. 9, pp. 5195–5208, 2015.
- [10] Lior Weizman, Yonina C Eldar, and Dafna Ben Bashat, “Reference-based MRI,” *Medical physics*, vol. 43, no. 10, pp. 5357–5369, 2016.
- [11] Ouri Cohen, Bo Zhu, and Matthew S Rosen, “MR fingerprinting deep reconstruction network (drone),” *Magnetic resonance in medicine*, vol. 80, no. 3, pp. 885–894, 2018.
- [12] Elisabeth Hoppe, Gregor Kördörfer, Tobias Würfl, Jens Wetzl, Felix Lugauer, Josef Pfeuffer, and Andreas Maier, “Deep learning for magnetic resonance fingerprinting: A new approach for predicting quantitative parameter values from time series,” *Stud Health Technol Inform*, vol. 243, pp. 202–206, 2017.
- [13] Fabian Balsiger, Amaresha Shridhar Konar, Shivaprasad Chikop, Vimal Chandran, Olivier Scheidegger, Sairam Geethanath, and Mauricio Reyes, “Magnetic resonance fingerprinting reconstruction via spatiotemporal convolutional neural networks,” in *International Workshop on Machine Learning for Medical Image Reconstruction*. Springer, 2018, pp. 39–46.
- [14] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436, 2015.
- [15] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio, *Deep learning*, vol. 1, MIT press Cambridge, 2016.
- [16] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vision Pattern Recog*, 2016, pp. 770–778.
- [18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Identity mappings in deep residual networks,” in *Proc. Eur. Conf. Comput. Vision*. Springer, 2016, pp. 630–645.
- [19] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, “Image super-resolution using deep convolutional networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, 2016.
- [20] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, “Deeply-recursive convolutional network for image super-resolution,” in *Proc. IEEE Conf. Comput. Vision Pattern Recog*, 2016, pp. 1637–1645.
- [21] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N Dauphin, “Convolutional sequence to sequence learning,” *arXiv preprint arXiv:1705.03122*, 2017.
- [22] Geoffrey Hinton, Li Deng, Dong Yu, George E Dahl, Abdelrahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N Sainath, et al., “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups,” *IEEE Signal processing magazine*, vol. 29, no. 6, pp. 82–97, 2012.
- [23] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [24] Suvrit Sra, Sebastian Nowozin, and Stephen J Wright, *Optimization for machine learning*, Mit Press, 2012.
- [25] Josef Vymazal, Andrea Righini, Rodney A Brooks, Margherita Canesi, Claudio Mariani, Marco Leonardi, and Gianni Pezzoli, “T1 and t2 in the brain of healthy subjects, patients with parkinson disease, and patients with multiple system atrophy: relation to iron content,” *Radiology*, vol. 211, no. 2, pp. 489–495, 1999.
- [26] Diederik P Kingma and Jimmy Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [27] Sean CL Deoni, Terry M Peters, and Brian K Rutt, “High-resolution t1 and t2 mapping of the brain in a clinically acceptable time with despot1 and despot2,” *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 53, no. 1, pp. 237–241, 2005.
- [28] Gilad Liberman, Yoram Louzoun, and Dafna Ben Bashat, “T1 mapping using variable flip angle spgr data with flip angle correction,” *Journal of Magnetic Resonance Imaging*, vol. 40, no. 1, pp. 171–180, 2014.