MR fingerprinting Deep RecOnstruction NEtwork (DRONE)

Ouri Cohen[1,2,3] | **Bo Zhu[1,2,3]** | **Matthew S. Rosen[1,2,3]**

[1] Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, Massachusetts

[2] Department of Radiology, Harvard Medical School, Boston, Massachusetts

[3] Department of Physics, Harvard University, Cambridge, Massachusetts

Correspondence
Ouri Cohen, Athinoula A. Martinos Center for Biomedical Imaging, Room 2320, Department of Radiology, Massachusetts General Hospital, 149 13th Street, Charlestown, MA 02129.
Email: ouri@nmr.mgh.harvard.edu

Purpose: Demonstrate a novel fast method for reconstruction of multi-dimensional MR fingerprinting (MRF) data using deep learning methods.

Methods: A neural network (NN) is defined using the TensorFlow framework and trained on simulated MRF data computed with the extended phase graph formalism. The NN reconstruction accuracy for noiseless and noisy data is compared to conventional MRF template matching as a function of training data size and is quantified in simulated numerical brain phantom data and International Society for Magnetic Resonance in Medicine/National Institute of Standards and Technology phantom data measured on 1.5T and 3T scanners with an optimized MRF EPI and MRF fast imaging with steady state precession (FISP) sequences with spiral readout. The utility of the method is demonstrated in a healthy subject in vivo at 1.5T.

Results: Network training required 10 to 74 minutes; once trained, data reconstruction required approximately 10 ms for the MRF EPI and 76 ms for the MRF FISP sequence. Reconstruction of simulated, noiseless brain data using the NN resulted in a RMS error (RMSE) of 2.6 ms for T_1 and 1.9 ms for T_2. The reconstruction error in the presence of noise was less than 10% for both T_1 and T_2 for SNR greater than 25 dB. Phantom measurements yielded good agreement ($R^2 = 0.99/0.99$ for MRF EPI T_1/T_2 and 0.94/0.98 for MRF FISP T_1/T_2) between the T_1 and T_2 estimated by the NN and reference values from the International Society for Magnetic Resonance in Medicine/National Institute of Standards and Technology phantom.

Conclusion: Reconstruction of MRF data with a NN is accurate, 300- to 5000-fold faster, and more robust to noise and dictionary undersampling than conventional MRF dictionary-matching.

KEYWORDS
deep learning, DRONE, EPI, MR fingerprinting, neural network, optimization

1 | INTRODUCTION

Magnetic resonance fingerprinting (MRF) [1] is an acquisition strategy that uses a variable schedule of RF excitations and delays to induce differential signal evolution in tissue of differing types. Quantitative tissue parameter maps are then obtained by matching the acquired signal to a precomputed dictionary consisting of the time evolution of the magnetization for different values of the set of tissue parameters. Multiple quantitative tissue parameter maps can be simultaneously obtained from a single experiment, significantly reducing the total scan time.

Correction added after online publication 16 April 2018. Due to a publisher's error, not all of the authors' corrections were made prior to publication and have been updated in this version.
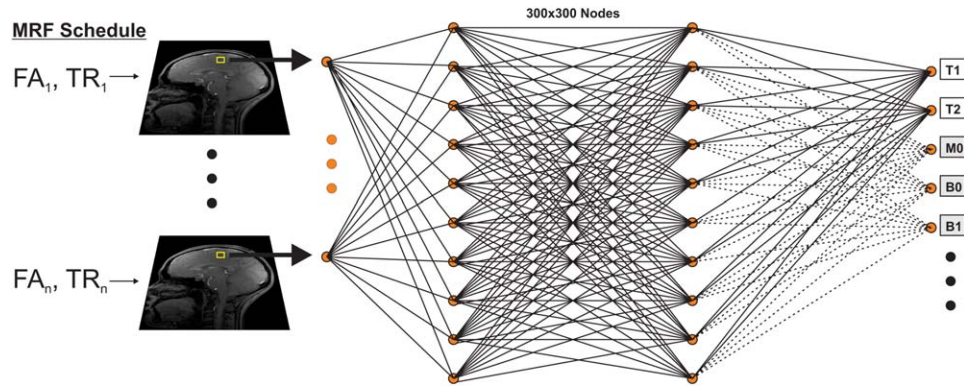
FIGURE 1 Schematic of the reconstruction approach used in this study. MRI data acquired with the MRF sequence is fed voxelwise to a 4-layer neural network containing two 300×300 hidden layers. The network is trained by a dictionary generated with the extended phase graph algorithm, with the tanh function used as activation function of the hidden layers. The network then outputs the underlying tissue parameters T_1 and T_2. Additional tissue parameters, such as M_0, B_0, and B_1 (gray boxes), can similarly be obtained by training the network with a suitable dictionary

FA, flip angle; MRF, MR fingerprinting; M_0, proton density.

To avoid errors in the reconstructed tissue maps, the reconstruction dictionary is typically computed with fine granularity over the entire range of possible tissue values. Dictionary size, however, grows exponentially as the number of tissue parameters (i.e., the dictionary dimension) is increased, which can quickly result in prohibitively large dictionaries that require extensive computational resources to process.[2] This increased memory, storage, and computational burden is a limiting factor for clinical adoption of MRF methods and is particularly pernicious in innovative high-dimensional applications of MRF.[3,4] Reducing the dictionary density is a poor solution for this problem because it limits the a priori accuracy of the reconstruction even before experimental factors are accounted for.

Existing methods address aspects of this problem, but important challenges remain. Dictionary compression[5,6] uses the compressibility of the fingerprints to reduce the dimensionality of the dictionary, leading to faster postprocessing. Unfortunately, to create the compressed dictionary, the full fine-grained dictionary must first be generated prior to decomposition with the singular value decomposition, itself a computationally expensive operation. Recent work by Yang et al.[7] used a randomized singular value decomposition with an iterative polynomial fit to reduce the memory requirements, albeit at the cost of increased processing time. Additionally, when optimized acquisition schedules are used,[8,9] the compressibility of the dictionary may be significantly reduced, rendering these methods less effective. Optimizing the acquisition schedule can indeed reduce the number of image frames needed for accurate reconstruction, but the reduction is smaller than the exponential growth of the dictionary with increasing dimensions.

In recent years, the availability of inexpensive graphical processing units has led to significant advances in neural networks (NN) and deep learning algorithms used to train these networks.[10] Mathematical work in NN theory has shown that any Borel measurable function can be represented by a NN with a finite number of neurons,[11] which can thus offer a compact representation of complicated functions. In this paper, we exploit this property and describe a novel method that reframes the MRF reconstruction problem as learning an optimal function that maps the acquired signal magnitudes to the corresponding tissue parameter values, trained on a sparse set of dictionary entries.[12] The trained neural network reconstruction function is remarkably compact ($\sim$20 times smaller than typical MRF dictionaries), and reconstruction is nearly instantaneous ($\sim$300- to 5000-fold faster than conventional dictionary-matching techniques) due to its rapid feed-forward processing. We call our method the MRF Deep RecOnstruction NEtwork (DRONE) and we validate it by numerical simulations and phantom experiments at 2 different magnetic field strengths and demonstrate its utility in the brain of a healthy subject scanned at 1.5T.

2 | METHODS

2.1 | Neural network

A 4-layer, fully connected NN composed of input and output layers and 2 hidden layers was defined using the TensorFlow framework,[13] as shown in Figure 1. The input layer consisted of 25 or 50 nodes to correspond to the magnitude images acquired with our schedule optimized (EPI) MRF sequence[8] or 571 nodes to correspond to images acquired with a conventional MRF sequence with sliding-window filtering.[14,15] Complex-valued images can also be processed by this network topology by splitting the real and imaginary channels, or through the use of a complex-valued network.[16] In this proof-of-concept study, only T_1 and T_2 were considered; therefore, the output layer consisted of 2 nodes. Reconstruction of additional parameters would require a larger output

layer. Each of the 2 hidden layers had 300 nodes. The network size was empirically selected to ensure accurate functional mapping while avoiding the risk of overfitting. The chosen size represented a good compromise between the required training time, storage space, and accuracy of the resultant reconstructions. The network was trained by the adaptive moment estimation stochastic gradient descent algorithm,[17] with the learning rate set to 0.001 and the loss function (cost) defined as the mean squared error:

$$LF = \frac{1}{n} \sum_{k=1}^{n} \left(P^k_{training} - P^k_{recon} \right)^2, \qquad (1)$$

where k ranges over the n training samples and P is the training or reconstructed tissue parameter of interest (T_1 or T_2 in this study). Two different activation functions were defined. A hyperbolic tangent (tanh) function was used for the hidden layers with a sigmoid function used for the output layer. In total, the NN required storage of $300 \times 300 = 90\,000$ coefficients.

2.2 | Pulse sequence

Numerical simulations and phantom and in vivo experiments in this study used a modified gradient-echo EPI MRF pulse sequence with Cartesian sampling for which the flip angles and TR were set according to an optimized measurement schedule, shown in Supporting Information Figure S1, as previously described.[8,18,19] To illustrate the flexibility of the proposed reconstruction, a conventional MRF fast imaging with steady state precession (FISP) pulse sequence,[20–23] with a variable density spiral readout and 600 time-points (571 after sliding-window filtering), was also used[14,15] for the phantom experiments.

2.3 | Network training

Numerical simulations and phantom experiments were carried out using a training dictionary of ~69 000 entries, consisting of T_1 and T_2 in the range 1 to 2500 ms in increments of 2 ms between 1 to 300 ms and increments of 10 ms between 300 to 2500 ms, excluding entries in which $T_1 < T_2$. An additional set was also defined, consisting of T_1 and T_2 values absent from the training dictionary, to test the network's ability to reconstruct tissue parameters values outside its training set. The same range and increments of T_1 and T_2 values were used for the test set but with a different starting point to ensure that there was no overlap with the training dictionary. Reconstruction of the test set data was used to verify the efficacy of training by comparing the resultant values to the true values. A larger range dictionary was used for the in vivo experiments, which consisted of T_1 in the range 1 to 5000 ms, in increments of 5 ms in the range 1 to 300 ms, increments of 10 in the range 300 to 1500 ms, and increments of 50 ms in the range 1500 to 5000 ms. The same set

of increments was used for T_2 but only up to 3000 ms. The magnetization due to each (T_1, T_2) pair was calculated using the extended phase graph formalism.[20,24] Gaussian noise with 1% SD and 0 mean was added to the training dictionary to promote robust learning, as previously exhibited with denoising autoencoders.[25] This dictionary was used to train the network to convergence, requiring approximately 10 to 74 mins on an Nvidia K80 GPU (Nvidia Corp., Santa Clara, CA) with 2 GB of memory. The size of the dictionary required approximately 13 MB of storage for the MRF EPI sequence and 313 MB for the MRF FISP sequence.

2.4 | Numerical simulations

The performance of the network in reconstruction of realistic brain T_1 and T_2 values was assessed using the BrainWeb digital brain phantom (http://brainweb.bic.mni.mcgill.ca/brainweb/).[26] An MRF EPI acquisition was simulated, as described above, and the resulting signal used as an input to the network.

2.4.1 | SNR versus reconstruction error

Monte Carlo simulations were used to test the network reconstruction of noisy data. Complex Gaussian random noise was added to the data for variable levels of SNR. The SNR was defined as $20 \cdot \log_{10}(S/N)$, where S is the average white matter signal intensity in the acquisition and N is the noise SD. The SNR was varied from 10 to 40 dB in increments of 5 dB, and the noisy data reconstructed with the proposed network. This was repeated 100 times for each SNR level. The resulting T_1 and T_2 maps were compared to their ground truth values according to the formula: Error = $100 \times |$True − Measured$|$/True and the mean and SD of the percent error calculated.

2.4.2 | Training dictionary density versus reconstruction error

The effect of the training dictionary density on the resulting reconstruction error was measured by subsampling the initial dictionary variously from 2- to 60-fold. The network was then trained with each subsampled dictionary and used to reconstruct the initial, fully sampled dictionary for which the entries were corrupted by 0 mean Gaussian noise with SD of either 0 (i.e., noiseless) or 1%. No noise was injected into the training dictionary for reconstruction of the noiseless data. This was repeated 5 times for each undersampling factor and the RMSE calculated for each repetition. The DRONE reconstruction was also compared to a conventional MRF dictionary-matching reconstruction by matching the initial, fully sampled dictionary to each subsampled dictionary and calculating the resultant error. The mean and SD of the RMSEs,

across the 5 repetitions, of the reconstructed T_1 and T_2 maps of each reconstruction method were then calculated as a function of the dictionary subsampling factor.

2.5 | MRI

All experiments with the optimized MRF EPI sequence were conducted on a 1.5T whole-body scanner (Avanto, Siemens Healthineers, Erlangen, Germany). The manufacturer's body coil was used for transmit, and a 32-channel head coil array used for receive. The TI/TE/BW was set to 19 ms/13 ms/2009 Hz/pixel. The slice thickness was set to 5 mm, and the in-plane resolution set to $2 \times 2 \, \text{mm}^2$, with a matrix of 128×128 and an acceleration factor $R = 2$, for a total scan time of ~3 seconds for the 25 frames acquired with the optimized schedule. A higher in-plane resolution could be obtained by increasing the number of phase-encoding steps, hence the echo time, or alternatively by increasing the acceleration factor at the cost of decreased SNR. To ensure sufficient SNR in the in vivo experiments, the same acquisition parameters were used but with a slightly longer (~5 s) acquisition. The images were reconstructed online using the GRAPPA[27] method.

The MRF FISP experiments were conducted on a 3T whole-body scanner (Prisma, Siemens Healthineers) with 2-channel parallel transmit and 20-channel parallel receive array. The TI/TE/BW was 20 ms/2.5 ms/261 Hz/pixel, with a 5-mm slice thickness, an in-plane resolution of $1.2 \times 1.2 \, \text{mm}^2$, and a matrix size of 200×200. A 30-point sliding-window filter was applied to the 600 time points to yield the 571 images used in the DRONE reconstruction, as described in Ref. 14.[14] The total acquisition time was ~7.5 s for the 600 frames acquired.

2.6 | Phantom

The accuracy and precision of the DRONE reconstruction was assessed using the International Society for Magnetic Resonance in Medicine/National Institute of Standards and Technology multi-compartment phantom, with calibrated T_1 and T_2 values similar to those of the human brain.[28] The phantom was scanned at the 2 magnetic field strengths with the pulse sequences described above and the images reconstructed with both the proposed NN as well as conventional dictionary-matching. The resulting T_1/T_2 maps were compared to the true phantom values, which were characterized by NIST for each magnetic field strength and calculated using gold-standard NMR spectroscopy inversion-recovery and Carr-Purcell-Meiboom-Gill sequences.[28,29]

2.7 | In vivo human

A healthy 35-year-old male subject was recruited for this study and provided informed consent prior to the experiment in accordance with our institution human research committee. Following DRONE reconstruction of the measured data, regions of interest of 10 pixels were defined corresponding to gray matter, white matter and CSF. The mean ± SD of the T_1/T_2 values within those regions of interest were calculated and compared to values from the literature.

3 | RESULTS

3.1 | Network training

The reconstructed test data is shown in Figure 2 in comparison to the true values. Excellent agreement was obtained between the true and reconstructed T_1 and T_2 values, yielding a correlation coefficient of $R^2 = 0.99$ for both T_1 and T_2, with a negligible bias of 1.6 ms for T_1 and 3.2 ms for T_2. The RMSE was 3.8 ms for T_1 and 16 ms for T_2. The T_1 and T_2 values at the edge of the training dictionary range showed increased deviation from the true values, likely due to the vanishing gradient of the activation function in these regions,[30] as described in the Discussion section below.

3.2 | Numerical simulation

The true and reconstructed T_1 and T_2 maps of the noiseless numerical brain phantom are shown in Figure 3, along with the associated error map calculated as the absolute value, that is, Error = |True-Reconstructed|. The RMSE for each map was 2.6 and 1.9 ms for T_1 and T_2, respectively. The largest error in the brain was less than 0.5% in T_1 and less than 3% in T_2.

3.2.1 | SNR versus reconstruction error

The mean T_1 and T_2 error as a function of SNR is shown in Figure 4. The error was relatively large (~15% and ~48% for T_1 and T_2) at the lowest SNRs but dropped rapidly with increasing SNR, reaching less than 2% for the highest SNR levels tested.

3.2.2 | Training dictionary density versus reconstruction error

The mean T_1 and T_2 RMSEs across the 5 repetitions are shown as a function of the dictionary undersampling factor in Supporting Information Figure S6 for the different noise levels tested. For a noiseless acquisition with small dictionary undersampling, the error was similar for both the NN and MRF dictionary-matching reconstructions. The error increased with increasing undersampling factors for both methods but at a significantly higher rate for the dictionary-matching. Indeed, at the largest undersampling factor tested (×60), the error in the NN reconstruction was 2-fold smaller

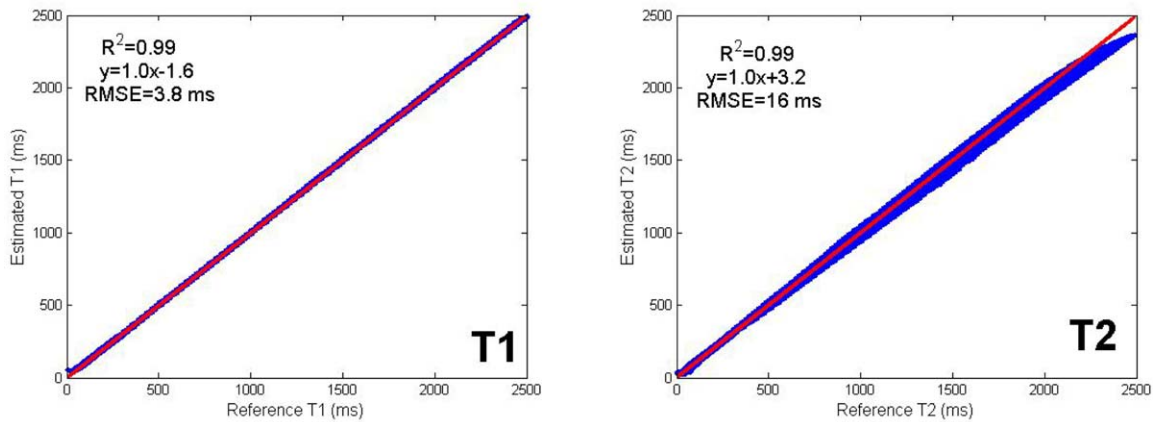**FIGURE 2** Shown is a comparison between the true T_1 and T_2 and those reconstructed by a network trained on a distinct dictionary. The red line indicates the least-squares fit curve. The reconstructed T_1 and T_2 values showed excellent agreement ($R^2 = 0.99$), with the true values with a negligible bias in T_1 and T_2 of 1.6 and 3.2 ms, respectively, validating the feasibility of the proposed approach. Very short and very long T_1 and T_2 values showed increased deviation from the true values due to vanishing gradient of the activation function used for those ranges

RMSEs, RMS error.

than the dictionary-matching for T_1 and 4-fold smaller for T_2 (Supporting Information Figures S2, S3, S6). Moreover, the dictionary-matching performance degraded rapidly in the presence of noise. In particular, dictionary-matching of short (< 11 ms) T_2 tissues resulted in large errors that dominated the total RMSE (Supporting Information Figures S4, S5). Such T_2 values were not estimable by conventional dictionary-matching (but were estimable by DRONE reconstruction) and were therefore not included in the RMSE calculations shown in Supporting Information Figure S6, unlike the DRONE reconstruction, which included all tissues. Despite the smaller set of tissues used in calculation of the dictionary-matching error, the DRONE reconstruction error was still 7% to 44% smaller for T_1 and 2% to 8% smaller for T_2 in the presence of noise.

3.3 | Phantom

The DRONE reconstruction accuracy was evaluated by estimating T_1 and T_2 values in the well-characterized calibrated International Society for Magnetic Resonance in Medicine/National Institute of Standards and Technology phantom. The measured T_1 and T_2 values were derived from the mean T_1 and T_2 values estimated within each compartment.
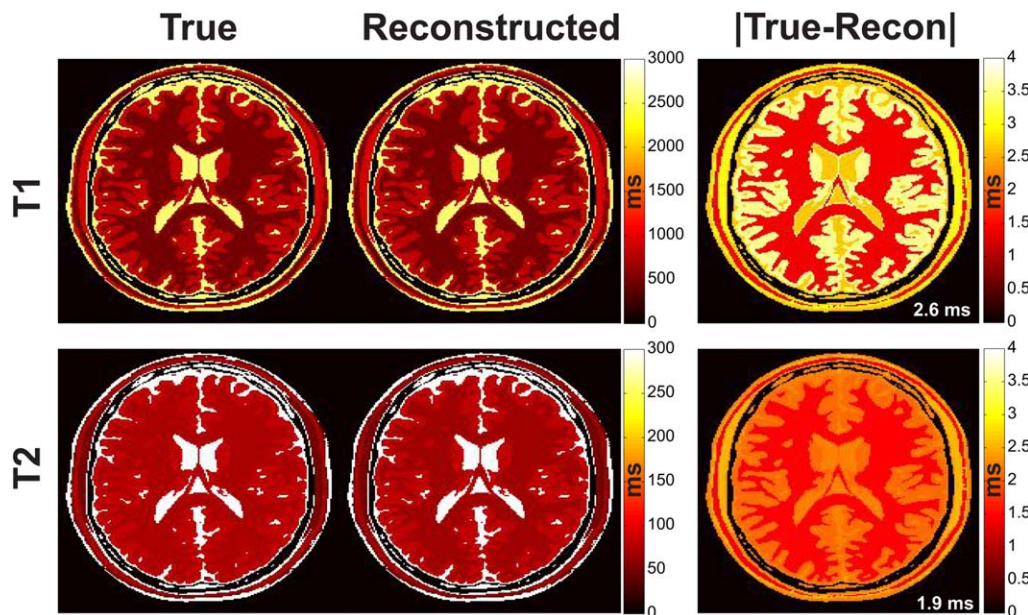

FIGURE 3 True and reconstructed T_1 and T_2 images from the numerical brain phantom shown on a common ms scale and the associated absolute error map. The true phantom values approximate those of the in vivo brain, for example, white matter $T_2 > 70$ ms. Note the close agreement between the reconstructed and true maps. The T_1 and T_2 RMSEs of 2.6 and 1.9 ms, respectively, are shown inset in white in the error map for this noiseless acquisition
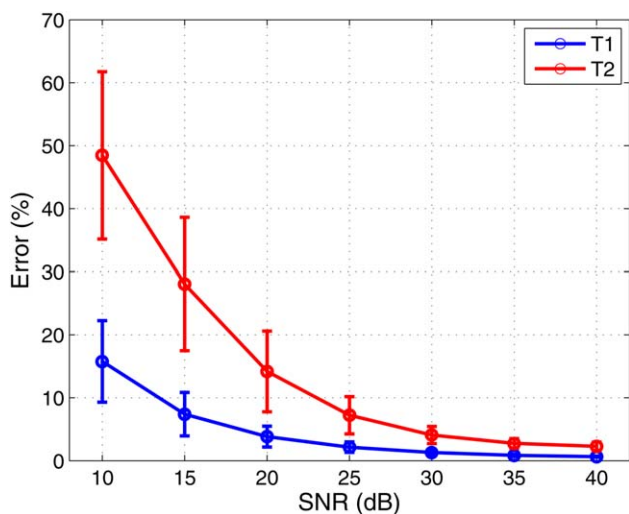
FIGURE 4 The T_1 and T_2 reconstruction error in percent as a function of the SNR. The mean (circles) and SD (whiskers) were calculated across the 100 Monte Carlo iterations. Reconstructions at the lowest SNR level showed significant error that nevertheless dropped rapidly with increasing SNR and was less than 2% for T_1 and T_2 at the highest SNR level tested

3.3.1 | MRF EPI

The estimated T_1 and T_2 values from all compartments (Figure 5) showed good agreement to the true phantom values ($R^2 = 0.99$) and a small estimation bias of 6.3 ms for T_1 and 15 ms for T_2. The calculated T_1 and T_2 RMSEs were 49 and 42 ms, respectively. In contrast, conventional dictionary-matching yielded higher RMSEs (89 and 56 ms for T_1 and T_2). The reconstruction of the $128 \times 128\, T_1$ and T_2 maps required $\sim$10 ms with the NN, which was $\sim$300-fold faster than the $\sim$3 s required with conventional dictionary-matching using a $\sim$69 000 entries dictionary.

3.3.2 | MRF FISP

The T_1 and T_2 values estimated with the MRF FISP sequence also showed good agreement to the true phantom values ($R^2 = 0.94$ for T_1 and 0.98 for T_2) with an estimation bias of 49 ms for T_1 and 3.8 ms for T_2. The T_2 RMSE was similar to that of the MRF EPI acquisition at 59 ms, but the T_1 RMSE (150 ms) was larger. Although B_1 correction was not applied to either sequence, the B_1 inhomogeneity is greater at 3T, which likely contributed to the larger error in the MRF FISP sequence. The larger number of time points in the MRF FISP sequence naturally required a longer time for reconstruction (76 ms) compared to the MRF EPI acquisition. Remarkably, despite the nearly 23-fold increase in the number of time points, the increase in reconstruction time was only $\sim$7 fold. Conventional dictionary-matching with the same size dictionary yielded RMSEs of 141 and 80 ms for T_1 and T_2 and required $\sim$380 seconds or 5000 times longer.

3.4 | In vivo human brain

The T_1 and T_2 maps reconstructed by DRONE are shown in Figure 6, along with the regions of interest chosen. The mean $\pm$ SD T_1 and T_2 for each tissue compartment are listed in Table 1 and are similar to values obtained from the literature.[1,31]

4 | DISCUSSION

MRF enables quantitative tissue mapping in a short acquisition time at the cost of increased complexity in the reconstruction. Although computational resources are typically cheaper and more accessible than scanner time, the large dictionaries required for MRF applications can overburden even the most advanced hardware. To avoid this problem, multi-parametric MRF dictionaries are by necessity undersampled in some dimensions, entailing an a priori reduction in accuracy. Instead, in this work, a 4-layer NN capable of modeling the history-dependent Bloch equations used in MRF sequences was demonstrated. Unlike conventional dictionary-matching in which the acquired signals can only be matched to the discrete entries computed in the dictionary, the proposed method relies on the functional representation of the NN, which yields continuous-valued parameter outputs. This is a notable advantage of the NN reconstruction because the reconstruction accuracy is no longer strictly limited by the dictionary density. Moreover, the network training process results in a signal-to-parameter mapping that is more robust to noise than a conventional dictionary-matching approach (Figure 4 and Supporting Information Figure S6) because the mapping is forced to be expressed in low-dimensional space and is thus insensitive to small corruptive input perturbations. This mapping need not result in a loss of accuracy provided that the salient features of the signal are adequately preserved by the NN, in analogy with truncation of small singular values in the singular value decomposition.[32] Because conventional MRF dictionaries are generated from combinations of tissue parameters, each T_1 and T_2 value can be repeated multiple times in the dictionary. Despite this natural redundancy, the error in the DRONE reconstruction was still smaller for increasing dictionary undersampling than with conventional dictionary-matching.

Conventional dictionary-matching does not learn a functional mapping, relying instead on the similarity between the normalized measured data and the corresponding normalized dictionary entry. An unfortunate side effect of the normalization is that noisy signals, arising from tissues with short T_2, for example, are amplified and then matched to some dictionary entry, leading to increased reconstruction errors (Supporting Information Figure S4). The
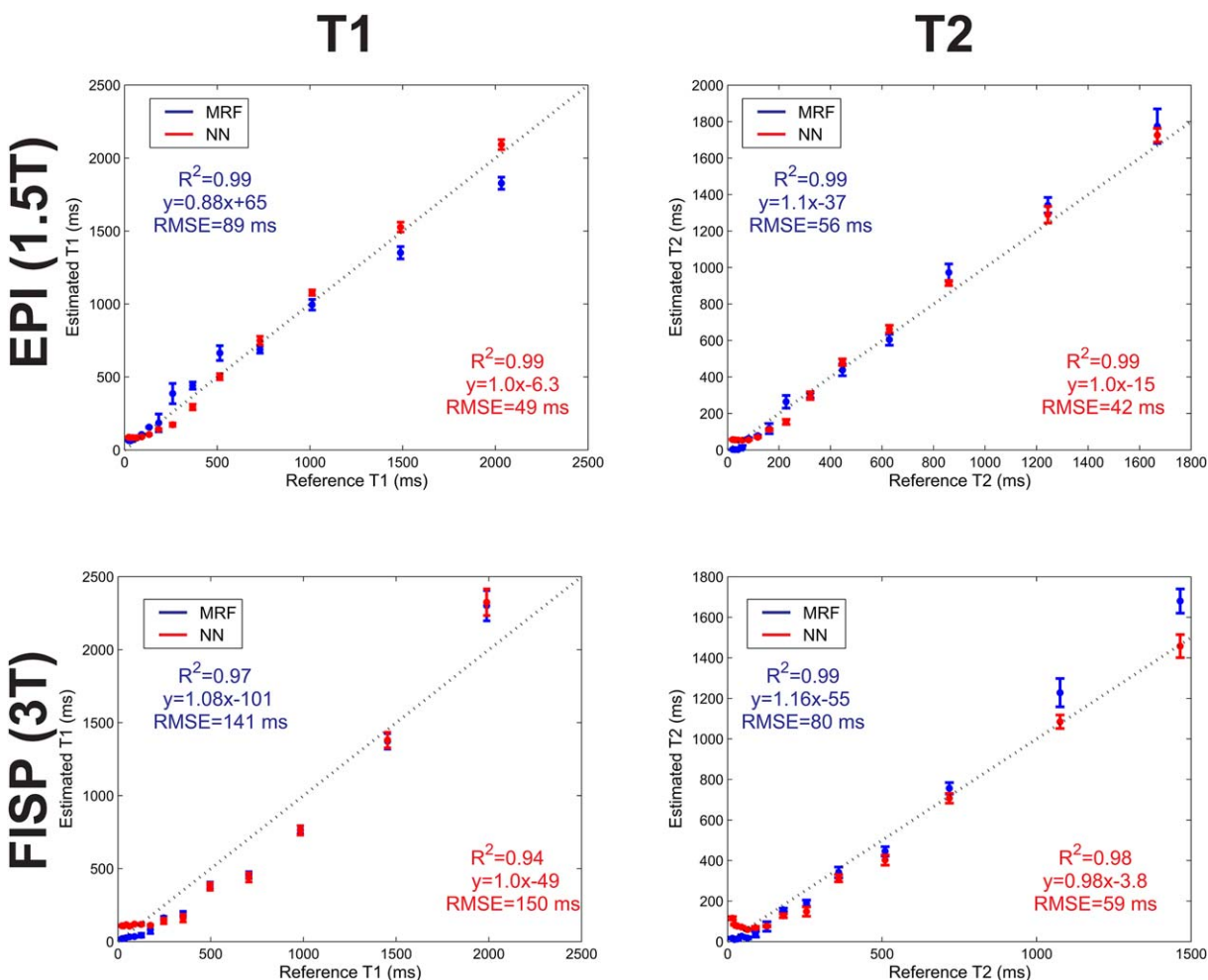
FIGURE 5 NN and MRF dictionary-matching reconstructions of the ISMRM/NIST phantom for the sequences tested. Shown is a comparison between the reference and measured compartment T_1 and T_2 values for data acquired with the optimized MRF EPI sequence at 1.5T (top) and with the MRF FISP sequence at 3T (bottom). The reference values were calculated by NIST using spectroscopic inversion-recovery and Carr-Purcell-Meiboom-Gill sequences. The dashed line is the identity line and the error bars represent the SD of the measured T_1 and T_2 values within each compartment. The uncorrected larger B_1 inhomogeneity at 3T may have contributed to the larger T_1 RMSE with the MRF FISP sequence

FISP, fast imaging with steady-state precession; ISMRM, International Society for Magnetic Resonance in Medicine; NIST, National Institute of Standards and Technology; NN, neural network.
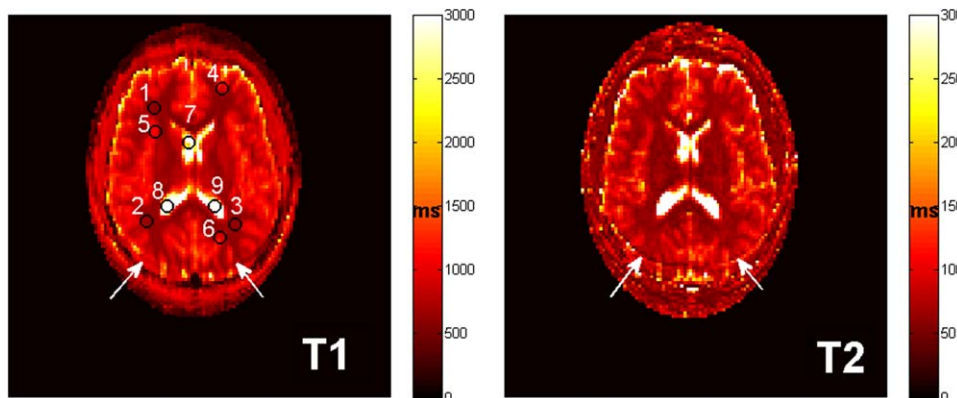

FIGURE 6 In vivo quantitative T_1 and T_2 maps from the brain of a healthy subject reconstructed with DRONE. Because no fat suppression was applied, a mild chemical shift artifact is visible in the images (white arrows). The numbered black circles indicate the locations of the gray matter, white matter, and CSF ROIs used to calculate the mean T_1 and T_2 values shown in Table 1

ROI, region of interest.

TABLE 1 Mean ± SD T_1 s and T_2 s values of the regions of interest selected

ROI	T_1 (ms)	T_2 (ms)
White matter		
1	608 ± 9	62 ± 3
2	642 ± 12	67 ± 4
3	633 ± 9	66 ± 3
Average	627 ± 10	65 ± 3
Gray matter		
4	895 ± 153	92 ± 14
5	1182 ± 65	102 ± 10
6	1241 ± 187	120 ± 17
Average	1106 ± 135	105 ± 14
CSF		
7	3292 ± 515	1475 ± 118
8	3665 ± 413	1710 ± 332
9	4026 ± 123	1544 ± 147
Average	3815 ± 424	1576 ± 199

dictionary-matching was greatly influenced by the low SNR of the short T_2 values, rendering them nonestimable. Those values, corresponding to T_2 s smaller than 11 ms, were excluded from the calculations of the dictionary-matching error but were included in those of the DRONE reconstruction. It should be noted that as the noise is increased, the smallest detectable T_2 value will become correspondingly larger. Because the NN was trained on noisy signals, it yielded smaller error for both T_1 and T_2 (Figure 4 and Supporting Information Figures S4, S5, S6).

The in vivo MRF EPI T_1 and T_2 maps reconstructed by DRONE (Figure 6) correspond well to the known anatomy, and the average values of the associated regions of interest are similar for gray and white matter to those reported by Ma et al. and references therein.[1] The average CSF value obtained in our study (~1600 ms for T_2) was similar to other studies[33,34] but significantly larger than the 550 ms reported by Ma et al. Susceptibility of the MRF bSSFP sequence to out-of-plane flow, as proposed in that paper, may explain the discrepancy, although recent work by Daoust et al.[34] suggests that flow only has a small effect on the T_2 of CSF at 1.5T.

A successful application of deep learning networks requires large and high-quality training data. In a clinical context, large and high-quality datasets may be difficult to obtain and expensive to generate. Training the network on simulated dictionary data eliminates this concern and permits generating arbitrarily large training sets. Our results (Figures 5–6) show that networks trained on simulated data can accurately reconstruct measured data despite the presence of inevitable noise and other sources of errors in the measurements.

The compact size of the NN solves many of the problems inherent to conventional dictionary-matching. Specifically, DRONE required merely ~5% of the storage and memory needed for storing even the small training dictionary used; larger dictionaries would reduce this fraction further. Because of its feedforward structure, reconstruction with the network was 300-fold faster than conventional dot product dictionary-matching of the optimized MRF EPI data and up to 5000-fold faster for the larger MRF FISP data. Accelerated matching techniques, such as that reported by Cauley et al.,[5] still necessitated 2 s per slice for reconstruction, which was ~16-fold longer than what would be required with a DRONE reconstruction of an equivalent number of time points. Although the dictionary used in that study was larger than the one used in this work, the number of entries in the training dictionary has no effect on the final reconstruction time once the network is fully trained. Because the network topology is fixed, additional training entries simply modify the weights/biases of the network but do not increase the reconstruction time. Similarly, because the architecture of our network (300×300 fully connected hidden layers) theoretically allows up to 600^2 degrees of freedom (1 weight and 1 bias per node), the inclusion of additional parameters in the reconstruction would only require increasing the size of the training dictionary, which would not affect the post-training reconstruction time. In theory, a larger network can better represent complicated functional mappings, albeit at the cost of increased reconstruction time and increased risk of overfitting the data. The relatively simple network architecture used in this study provided accurate reconstructions at near instantaneous processing time.

This work represents an initial proof of concept for MRF reconstruction by a NN and can be optimized to further improve the results. For instance, although the training time for the network used in this study was relatively small (~10–74 min), alternative methods may yield faster training. The size of the network and the small number of images used with the optimized MRF EPI schedule contributed to the short training time, but conventional MRF sequences that require a greater number of acquisitions (10- to 100-fold higher) will require a longer training time, as will simultaneous reconstruction of additional tissue parameters given the larger training dictionary needed. The sigmoid and tanh activation functions used in this study are a common choice for NN training[35] but skew the accuracy of the network toward the middle of the training dictionary range (Figure 2), where the gradient is largest and the back-propagation algorithm is thus most effective. This problem is well known in the machine-learning literature, with a number of techniques

available for addressing it.[36] Alternative training strategies and activation functions such as softmax and rectifier linear unit are beyond the scope of this work and will be examined in future studies.

5 | CONCLUSION

We have demonstrated the feasibility of using deep learning networks for reconstruction of MRF data. The proposed approach yields fast and accurate reconstruction with a limited storage requirement, despite training on sparse dictionaries, and can therefore resolve the technical issues inherent to the exponential growth of multi-dimensional dictionaries.

REFERENCES

[1] Ma D, Gulani V, Seiberlich N, et al. Magnetic resonance fingerprinting. *Nature*. 2013;495:187-192.

[2] Ma D, Coppo S, Chen Y, et al. Slice profile and B1 corrections in 2D magnetic resonance fingerprinting. *Magn Reson Med*. 2017;78:1781-1789.

[3] Su P, Mao D, Liu P, et al. Multiparametric estimation of brain hemodynamics with MR fingerprinting ASL. *Magn Reson Med*. 2017;78:1812-1823.

[4] Cohen O, Huang S, McMahon MT, Rosen MS, Farrar CT. Rapid and quantitative chemical exchange saturation transfer (CEST) imaging with magnetic resonance fingerprinting (MRF). 2017. arXiv preprint. arXiv:1710.06054.

[5] Cauley SF, Setsompop K, Ma D, et al. Fast group matching for MR fingerprinting reconstruction. *Magn Reson Med*. 2015;74:523-528.

[6] McGivney D, Pierre E, Ma D, et al. SVD compression for magnetic resonance fingerprinting in the time domain. *IEEE Trans Med Imaging*. 2014;33:2311-2322.

[7] Yang M, Ma D, Jiang Y, et al. Low rank approximation methods for MR fingerprinting with large scale dictionaries. *Magn Reson Med*. 2018;79:2392-2400.

[8] Cohen O, Rosen MS. Algorithm comparison for schedule optimization in MR fingerprinting. *Magn Reson Imaging*. 2017;41:15-21.

[9] Cohen O, Polimeni JR. Optimized inversion-time schedules for quantitative T1 measurements based on high-resolution multi-inversion EPI. *Magn Reson Med*. 2018;79:2101-2112.

[10] Schmidhuber J. Deep learning in neural networks: An overview. *Neural Netw*. 2015;61:85-117.

[11] Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. *Neural Netw*. 1989;2:359-366.

[12] Cohen O, Zhu B, Matthew R. Deep Learning for Fast MR Fingerprinting Reconstruction. In Proceedings of the 25th Annual Meeting of ISMRM, Honolulu, HI, 2017. p. 7077.

[13] Abadi M, Agarwal A, Barham P, et al. Tensorflow: large-scale machine learning on heterogeneous distributed systems. 2016. arXiv preprint. arXiv:1603.04467.

[14] 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.

[15] Liao C, Bilgic B, Manhard MK, et al. 3D MR fingerprinting with accelerated stack-of-spirals and hybrid sliding-window and GRAPPA reconstruction. *Neuroimage*. 2017;162:13-22.

[16] Virtue P, Yu SX, Lustig M. Better than real: complex-valued neural nets for MRI fingerprinting. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Beijing, Peoples Republic of China, 2017.

[17] Kingma D, Ba J. Adam: A method for stochastic optimization. 2014. arXiv preprint. arXiv1412.6980.

[18] Cohen O, Sarracanie M, Armstrong BD, Ackerman JL, Rosen MS. Magnetic resonance fingerprinting trajectory optimization. In Proceedings of the 22nd Annual Meeting of ISMRM, Milan, Italy, 2014. p. 0027.

[19] Cohen O, Sarracanie M, Rosen MS, Ackerman JL. In Vivo Optimized MR Fingerprinting in the Human Brain. In Proceedings of the 24th Annual Meeting of ISMRM, Singapore, 2016. p. 0430.

[20] 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.

[21] Pierre EY, Ma D, Chen Y, Badve C, Griswold MA. Multiscale reconstruction for MR fingerprinting. *Magn Reson Med*. 2016;75:2481-2492.

[22] Rieger B, Zimmer F, Zapp J, Weingärtner S, Schad LR. Magnetic resonance fingerprinting using echo-planar imaging: joint quantification of T1 and T2∗ relaxation times. *Magn Reson Med*. 2017;78:1724-1733.

[23] Ye H, Ma D, Jiang Y, Cauley SF, Du Y, Wald LL, Griswold MA, Setsompop K. Accelerating magnetic resonance fingerprinting (MRF) using t-blipped simultaneous multislice (SMS) acquisition. *Magn Reson Med*. 2016;75:2078-2085.

[24] Weigel M. Extended phase graphs: Dephasing, RF pulses, and echoes-pure and simple. *J. Magn Reson Imaging*. 2015;41:266-295.

[25] Vincent P, Larochelle H, Bengio Y, Manzagol P-A. Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland, 2008. p. 1096-1103.

[26] Collins DL, Zijdenbos AP, Kollokian V, et al. Design and construction of a realistic digital brain phantom. *IEEE Trans Med Imaging*. 1998;17:463-468.

[27] Griswold MA, Jakob PM, Heidemann RM, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med*. 2002;47:1202-1210.

[28] Keenan KE, Stupic KF, Boss MA, et al. Multi-site multi-vendor comparison of T1 measurement using ISMRM/NIST system phantom. In Proceedings of the 24th Annual Meeting of ISMRM, Singapore, 2016. p. 3290.

[29] 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.

[30] Hochreiter S, Bengio Y, Frasconi P, et al. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. In: Kremer SC, Kolen JF, eds. *A Field Guide to Dynamical Recurrent Neural Networks*. Piscataway, NJ: IEEE Press; 2001;14:237-245.

[31] Deoni SC, Peters TM, Rutt BK. High-resolution T1 and T2 mapping of the brain in a clinically acceptable time with DES-POT1 and DESPOT2. *Magn Reson Med*. 2005;53:237-241.

[32] Lawless JF, Wang P. A simulation study of ridge and other regression estimators. *Commun Stat Theory Methods*. 1976;14:307-323.

[33] Schmitt P, Griswold MA, Jakob PM, et al. Inversion recovery TrueFISP: quantification of T1, T2, and spin density. *Magn Reson Med*. 2004;51:661-667.

[34] Daoust A, Dodd S, Nair G, et al. Transverse relaxation of cerebrospinal fluid depends on glucose concentration. *Magn. Reson Imaging*. 2017;44:72-81.

[35] Karlik B, Olgac AV. Performance analysis of various activation functions in generalized MLP architectures of neural networks. *Int J Artif Intell Expert Syst*. 2011;1:111-122.

[36] Hochreiter S. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *Int J Uncertain Fuzziness Knowl Based Syst*. 1998;6:107-116.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article.

FIGURE S1 Schedule of FAs and TRs used for the optimized MRF EPI and the MRF FISP sequences.

FIGURE S2 MRF dictionary matching reconstruction of the noiseless T_1 and T_2 values tested in comparison to the true values for the lightly undersampled (X2) and 60-fold undersampled (X60) training dictionaries.

FIGURE S3 NN reconstruction of the noiseless T_1 and T_2 values tested in comparison to the true values for the lightly undersampled (X2) and 60-fold undersampled (X60) training dictionaries.

FIGURE S4 MRF dictionary matching reconstruction of the T_1 and T_2 values corrupted by 1% Gaussian noise in comparison to the true values for the lightly undersampled (X2) and 60-fold undersampled (X60) training dictionaries. The dictionary matching of noisy data was particularly susceptible to short (<11 ms) T_2 values for the noise level tested resulting in large errors.

FIGURE S5 NN reconstruction of the T_1 and T_2 values corrupted by 1% Gaussian noise in comparison to the true values for the lightly undersampled (X2) and 60-fold undersampled (X60) training dictionaries. Because the NN reconstruction was trained on noisy data it was more robust to the effects of noise.

FIGURE S6 RMSE of the MRF dictionary matching (open circles) and NN reconstruction (closed circles) for the different undersampling factors and noise levels tested. For the noiseless acquisition (blue curves) the error in the NN reconstruction was 2 fold smaller for T_1 and 4 fold smaller for T_2 at the largest undersampling factor tested. For the noisy acquisition (red curves) only tissues with T_2 s >11 ms were included in the error calculation for the MRF dictionary matching whereas all tissues were included in the NN reconstruction error. Nevertheless, the NN reconstruction error was still smaller or equal to the MRF error for all undersampling factors tested.