A deep learning approach for magnetic resonance fingerprinting: Scaling capabilities and good training practices investigated by simulations

Physica Medica xxx (xxxx) 1–12

Contents lists available at ScienceDirect

Physica Medica

journal homepage: www.elsevier.com/locate/ejmp

Original paper

A deep learning approach for magnetic resonance fingerprinting: Scaling capabilities and good training practices investigated by simulations.

Marco Barbieri [a, b], Leonardo Brizi [a, c], Enrico Giampieri [d], Francesco Solera [e], David Neil Manners [f], Gastone Castellani [d], Claudia Testa [a, c, *], Daniel Remondini [a, c]

[a] Department of Physics and Astronomy "Augusto Righi", University of Bologna, Bologna, Italy
[b] Department of Radiology, Stanford University, CA, USA
[c] INFN, Sezione di Bologna, Bologna, Italy
[d] Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, Bologna, Italy
[e] Deep Vision consulting, Modena, Italy
[f] IRCCS Istituto delle Scienze Neurologiche Bologna, Functional and Molecular Neuroimaging Unit, Bologna, Italy

ARTICLE INFO

Keywords:
MR fingerprinting
Deep learning
qMRI
Parameter mapping

ABSTRACT

MR fingerprinting (MRF) is an innovative approach to quantitative MRI. A typical disadvantage of dictionary-based MRF is the explosive growth of the dictionary as a function of the number of reconstructed parameters, an instance of the curse of dimensionality, which determines an explosion of resource requirements. In this work, we describe a deep learning approach for MRF parameter map reconstruction using a fully connected architecture. Employing simulations, we have investigated how the performance of the Neural Networks (NN) approach scales with the number of parameters to be retrieved, compared to the standard dictionary approach. We have also studied optimal training procedures by comparing different strategies for noise addition and parameter space sampling, to achieve better accuracy and robustness to noise. Four MRF sequences were considered: IR-FISP, bSSFP, IR-FISP-B_1, and IR-bSSFP-B_1. A comparison between NN and the dictionary approaches in reconstructing parameter maps as a function of the number of parameters to be retrieved was performed using a numerical brain phantom. Results demonstrated that training with random sampling and different levels of noise variance yielded the best performance. NN performance was at least as good as the dictionary-based approach in reconstructing parameter maps using Gaussian noise as a source of artifacts: the difference in performance increased with the number of estimated parameters because the dictionary method suffers from the coarse resolution of the parameter space sampling. The NN proved to be more efficient in memory usage and computational burden, and has great potential for solving large-scale MRF problems.

1. Introduction

Magnetic Resonance Fingerprinting (MRF) is a fast quantitative MRI method able to obtain multi-parametric maps with a one-shot measurement [1]; many applications of the technique have been investigated since its birth, ranging from brain imaging [2] to body MRI [3,4].

The key idea behind MRF is to generate signals that vary as much as possible for different tissues so that, for each voxel (a pixel in the image), a so-called fingerprint of the tissue is acquired. With this aim, MRF sequences apply a train of RF pulses whose flip angle (FA) and repetition time (TR) vary according to a pattern designed to render the magnetization evolution sensitive to several MR parameters simultaneously. MRF usually employs a non-Cartesian k-space sampling scheme using variable density spirals [1,5] or radial waveforms to allow data to be acquired rapidly [6,7].

The original data processing method is based on a pixel-wise template matching to deal with the lack of an analytical model to describe the magnetization response. The experimental fingerprint is matched against a precomputed dictionary of simulated signals, and the scalar product is usually adopted as a similarity measure. This process is commonly referred to as a dictionary matching procedure. The main challenges of the MRF framework, that need to be overcome before it is accepted in a clinical setting, include large dictionary size, accuracy in parameter evaluation, robustness to noise.

For example, confounding factors such as B_0 and B_1^+ field inhomogeneities are known to be a source of artifacts in MRI [8], and they have

* Corresponding author at: Physics and Astronomy Department, University of Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy.
E-mail address: claudia.testa@unibo.it (C. Testa).

https://doi.org/10.1016/j.ejmp.2021.07.013
Received 26 February 2021; Received in revised form 12 July 2021; Accepted 13 July 2021
1120-1797/

to be considered in MRF to improve parameter estimation [9,10]. However, the more parameters one aims to encode using the MRF framework (T_1, T_2, B_0 off resonances ($\Delta\nu_0$), B_1^+, etc…), the bigger is the size of the precomputed dictionary. Large dictionaries are hard to handle because they are costly both in memory usage efficiency, occupying up to 150 GB, and in the computation time of the matching procedure [10]. Moreover, generating an entire high-resolution dictionary can result in high computational costs when, for example, simulating the full 3D voxel profiles to perform single-voxel proton spectroscopy with MRF [11]. At the same time, the accuracy of parameter estimation depends directly on the resolution of the simulated dictionary used to perform the matching procedure. Adding new parameters to the dictionary leads to an exponential growth of the number of entries that have to be inserted without affecting resolution, the so-called curse of dimensionality [12]. Since both computational and memory usage limitations have to be taken into account, the number of entries needs to be kept under control. This increases sparsity in the dictionary, which can produce high biases in nearest neighbor algorithms [13].

The problems arising from managing large dictionaries have driven the MRF community toward methods that compress the size of the dictionaries by applying Single Value Decomposition (SVD) to generate a low-rank approximation of the fingerprinting signals, and reducing the entries of the dictionaries by using polynomial fitting [14–16].

Another strategy to overcome the limitations of dictionary-based template matching is to exploit Machine Learning. Using a supervised training procedure, a pre-trained model can predict the MR parameters given the experimental fingerprint as input. Among the family of machine learning algorithms, the Neural Network (NN) algorithm is particularly suited to such a task, as NNs are universal function approximators given enough training data and model complexity [17]. A sufficiently large NN can learn an approximation of the Inverse Transfer Function (ITF) that maps the experimental MRF fingerprint into the MR parameter space. Once the NN is trained, there is no need to exhaustively search a dictionary, making the prediction operation computationally more efficient with respect to template matching. More importantly, under optimal training conditions, a NN is able to generalize and predict MR parameters of signals that were not seen during training. A NN approach in principle can limit quantization artifacts that can arise from a dictionary approach [18] in which template matching approximates MR parameters to those present in the dictionary.

Depending on the task and the selected architecture, NN can require thousands of training data items, which constitutes a challenge in the case of MRI, where data acquisition is normally both expensive and time-consuming. However, in the MRF framework, the NN model can be trained using simulated data. Simulating input data is the standard way in which dictionaries are produced in MRF, and the reliability of such dictionaries has been widely demonstrated [1]. Few studies have investigated the feasibility of applying NNs to MRF in conjunction with numerical simulations [19–23], with phantoms [24,25] and in vivo acquisitions [24,26–30,25]. Among possible deep learning architectures, fully-connected neural networks (FCNN) have features that make them particularly suited to circumventing the curse of dimensionality in MRF while maintaining most of the desirable characteristics of the original approach. Two-dimensional convolutional neural networks (2D-CNN), for example, which require spatial structure to be present in the training data set, take advantage of local connections and correlations [31]. Thus, real acquired data rather than simulated data are usually adopted in 2D-CNN approaches for MRF [26,27]. However, NNs are at risk of overfitting if the training data are insufficiently numerous or variable. 2D-CNNs add the spatial structure as a source of variability, and the generalization capabilities of the 2D-CNNs proposed for MRF applications have not been well characterized. In contrast, FCNNs used to perform pixel-wise prediction tasks, without exploring neighboring pixels, can easily be trained with simulated data. Because the network does not take advantage of spatial information, a training data set created with simulated MRF signals using a simple MR simulator is sufficient. Combining pixel-wise processing with a training procedure based on simulated data has two desirable features: the approach is inherently structure independent, since no spatial connections are taken into account; the lack of training data will never be a problem. For these reasons, this work focuses on the application of FCNNs to MRF. FCNNs will be referred to below simply as NNs.

Although the feasibility of NN approaches has been investigated in previous works, [24,20,32,29] such applications are still in their infancy, and the capacity of NNs to circumvent the curse of dimensionality in MRF has not been deeply studied. In fact, the number of parameters predicted by the NNs has always been limited to the minimum possible based on the MRF pulse sequence being considered. For example, in [24,32], where an IR Fast Imaging with Steady state Precession (IR-FISP)[5] sequence was used, only T_1 and T_2 were retrieved, and in [20], where a IR balanced Steady State Precession (IR-bSSFP) [1] sequence was used, only T_1, T_2 and B_0 off-resonances ($\Delta\nu_0$) were retrieved. It is important to study the scalability performance of NN approaches as this allows one to assess the actual benefits of NN approaches compared to the dictionary approach. Previously, little consideration has been given to aspects of the training pipeline such as training set design or noise addition strategies during training that promote accurate and robust learning for a given MRF pulse sequence. Virtue et al. [19] constructed the training data set by sampling the MR parameter space using a random uniform distribution. In contrast, Hoppe et al. [20], Cohen et al. [24] and Golbabee et al. [32] used a grid sampling of the MR parameter space. Different strategies have been used to consider noise: in [19,20] no noise was added during the training procedure, while in [24] white Gaussian noise with zero mean and 1% standard deviation was added during the training phase.

The current work makes use of simulations and has two main goals. The primary aim is to investigate how the performance of the NN approach scales with the number of parameters to be retrieved, compared to the standard dictionary approach. Secondarily, different training strategies are investigated to identify the best practices for training a NN to allow it to learn an accurate and robust ITF for a given MRF pulse sequence.

To achieve these aims, we first investigated how the method used to sample parameter space, either uniform random sampling or grid sampling, affected the generalization capabilities of the NN, a factor of known importance [33,34]. We then studied noise robustness by applying different data augmentation strategies and adding white Gaussian noise to MRF training signals. To show the generalization ability of the NN approach and how it scales with the number of predicted MR parameters, in comparison with the dictionary approach, the NN has been applied to different MRF pulse sequences. They were the IR-FISP, its variant for B_1^+ estimation [9], the IR-bSSFP and a modification of it to simultaneously estimate T_1, T_2, B_0 off-resonances ($\Delta\nu_0$) and B_1^+ field inhomogeneities. A numerical brain phantom has been used to assess the performance of the NN and the standard dictionary approaches in reconstructing MR parameter maps. Moreover, robustness to undersampling has been studied by using different degrees of acceleration factor with radial sampling.

The reported results may help to highlight the advantages of NN approaches, which can be used to estimate other relevant MR parameters such as T_2^* [35], water-fat ratio [36], and magnetization transfer [37], without suffering the curse of dimensionality associated with dictionary-based approaches.

2. Methods

2.1. MRF pulse sequences and simulations

Simulations were performed using four MRF pulse sequences: the MRF IR-FISP pulse sequence as described in [5] that encodes T_1 and T_2,

and its variant, IR-FISP B_1, which accounts for B_1 inhomogeneity by adding abrupt flip angle changes as described in [9]; the MRF IR-bSSFP as described in the original MRF article [1], which encodes T_1, T_2 and $\Delta\nu_0$, and a modification of it, IR-bSSFP B_1, to take B_1 field inhomogeneities into account. The latter sequence adds abrupt changes to the end of the IR-bSSFP sequence, based on the same criterion used for the IR-FISP B_1 sequence. Flip angle and time repetition patterns are depicted in Fig. 1.

All sequence simulations were carried out using MATLAB (MathWorks). The Bloch equations were used for IR-bSSFP-type sequences, while the Extended Phase Graph (EPG) algorithm [38] was used for IR-FISP-type sequence simulations.

2.2. Noise

The noise affecting the MRF signal was modeled as white Gaussian noise, whose variance is expressed in terms of Signal to Noise Ratio (SNR). SNR has been defined in two ways as reported in Eqs. 1 and 2. In Eq. 1, P_{signal} and P_{noise} represent the average power of the MRF signal and the noise respectively, and $E(signal)$ indicates the expectation value of the MRF signal. Eq. 2 is used when noise is added to a brain numerical phantom. In this case, following [39] A_{signal} indicates the average signal intensity within a white matter region in the first image of the MRF time series, while σ_{noise} represents the standard deviation of the noise.

Eq. 2 expresses SNR in decibels (dB), but in this work, all SNRs are expressed on a linear scale by applying the proper conversion. The definition of SNR used during the different experiments of this work has been indicated explicitly in the corresponding sections.

$$SNR = \frac{P_{signal}}{P_{noise}} = \frac{E(signal)}{\sigma^2_{noise}} \tag{1}$$

$$SNR = 20 \times \log_{10}\left(\frac{A_{signal}}{\sigma_{noise}}\right) \tag{2}$$

Only white Gaussian noise was used as a noise model in this work, since it is the most common noise model for NMR simulations, and it is widely used in MRF literature [40,39,41,24]. As the simulated MRF signal is complex, identically distributed Gaussian noise was added to the real and imaginary parts. This indeed simulates two-channel acquisition and preserves the Rician distribution when the magnitude of the signal is considered [42].

2.3. Fully connected neural network models

The Deep Neural Network application was developed using the Python package Keras with the TensorFlow [43] backend.

Two NN models were defined: one architecture to process IR-FISP data (M1), and another architecture to process IR-bSSFP data (M2). Both of them are feed forward nets with 9 fully connected layers, and the Rectified Linear Unit (ReLU) was used as the activation function for the neurons in the first 8 layers, while a linear activation function was chosen for the output layer. The NN architectures are reported in Fig. 2. M1 takes the magnitude of the complex MR signal produced by an IR-FISP or IR-FISP B_1 sequence as input, while M2, which is used to estimate $\Delta\nu_0$, takes as input the concatenated real and imaginary parts of the complex MR signal produced by the IR-bSSFP or IR-bSSFP B_1 sequences. The output layer size matches the number of MR parameters to be estimated. These are: T_1 and T_2 for IR-FISP; T_1, T_2 and B_1^+ for IR-FISP B_1;T_1, T_2 and $\Delta\nu_0$ for IR-bSSFP and $T_1, T_2, \Delta\nu_0$ and B_1^+ for IR-bSSFP

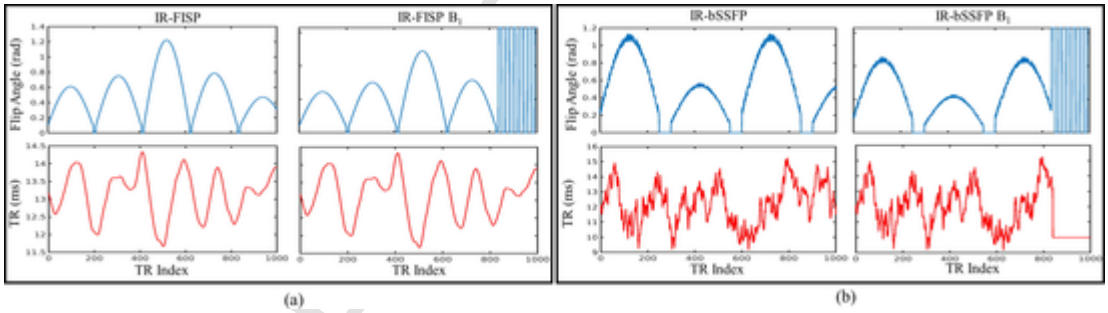

Fig. 1. MRF pulse sequences used to generate the synthetic signals. (a) MRF FISP type sequences: the IR-FISP sequence (first column); the IR-FISP B_1 sequence (second column). (b) MRF FISP type sequences: the IR-bSSFP sequence (first column); the IR-bSSFP B_1 sequence (second column)..

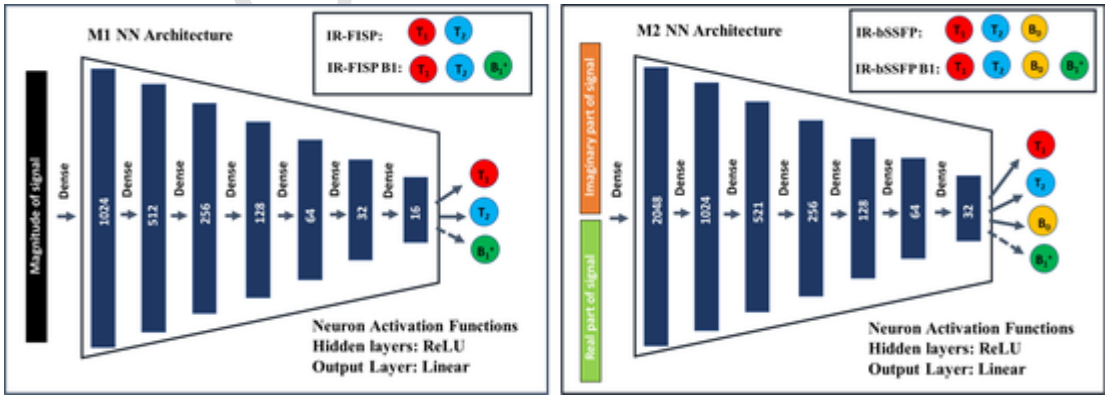

Fig. 2. NN architectures for model M1 (left box) and model M2 (right box). The two models are fully connected NNs with the Rectified Linear Unit (ReLU) used as a neuron activation function for hidden layers, while a linear activation function has been used for the output layer. Blue boxes represent layers and the numbers inside indicate the number of neurons. Model M1 was used for IR-FISP and IR-FISP B_1 pulse sequences, whereas model M2 was used for IR-bSSFP and IR-bSSFP B_1 sequences.

B_1, where $\Delta\nu_0$ indicates stationary field off-resonances, and B_1^+ indicates B_1 excitation field inhomogeneities expressed in terms of a correction ratio, i.e. the ratio between actual and nominal flip angles.

Preliminary experiments, not presented in the current work, in which different network sizes, batch sizes, and learning rates were tested, guided the selected architectures. However, it is useful to highlight some of the motivations that have driven the design of the NN model architectures described above. A bottleneck shape was selected for two main reasons. The first is to avoid an exploding number of model parameters: many layers lead to overfitting since fully connected layers with a large number of neurons rapidly increase the number of model parameters. The second is that a bottleneck shape forces deeper layers of the network to encode more meaningful representations [31]. Hence, fixing the number of hidden layers to seven, the size of the first hidden layer was selected to have around the same number of entries as the expected input, which is a 1D vector of length 1000 for model M1. In each successive layer, the number of neurons was halved. Further trials were performed by doubling all the neurons in every layer and halving the number of neurons per layer. We observed that increasing the NN size did not significantly increase the performance of the NN model, whereas halving the size of the NN decreased its performance. The same criterion was used for model M2. Hence, the architecture used for IR-bSSFPs sequences works well even for IR-FISPs sequences, whereas the contrary does not hold. However, because computational time increases as the size of the network, the smallest architecture was selected for IR-FISPs sequences.

The training procedure was supervised, using the Mean Squared Error (MSE) between NN-estimated parameters and ground truth parameters as a loss function. The Adam algorithm [44] was used for model weight optimization. In particular, 500 epochs were used, with 1000 gradient steps for each epoch and a fixed batch size of 500. Initial learning rates were set to 3×10^{-5}, for model M1, and 1×10^{-4}, for model M2. The application was run on a cluster with 16 dual-core CPUs.

2.4. Training strategies: training sets, test sets, and data augmentation

A critical step in building an NN application is the training procedure. While the selection of the NN architecture determines model complexity, the goal of the training procedure is to perform model regularization[45] to promote learning of the target function while preventing overfitting on unseen data. Training data set creation, [33,34] and data augmentation through added noise [46] are important forms of model regularization, especially when data can be simulated, as in the case of MRF. The sampling and data augmentation strategies are described below.

2.4.1. Parameter space sampling, random uniform and grid sampling

To assess the network models performance in learning the ITF depending on the training set distribution, two training sets of the same size were generated for each pulse sequence: a random uniform set, in which parameter space is sampled using a random uniform distribution; a gridded set, in which parameter space is sampled with a variable mesh grid.

Table 1 summarizes the training set characteristics, where label *R* refers to random sampling, whereas label *G* refers to grid sampling.

2.4.2. Data augmentation and preprocessing

Three data augmentation strategies were tested, by training the NN models with different noise adding procedures:

- *W/O Noise*: using no data augmentation during training, which means that only noise-free examples are fed to the NN models, as in reference [19];
- *Fixed variance*: feeding the NN with noisy inputs affected by noise with 1% standard deviation, equal to a variance of 10^{-4}, as in reference [24]. This means that the network during the training sees just one fixed SNR given a set of MR parameters;
- *Variable variance*: feeding the NN with inputs affected by noise with different variances, expressed in terms of SNRs. For each training batch, a vector of SNRs was generated by randomly sampling the SNR values in the range 2 to 100, so that the batch contains data with different SNRs, and the variances of the noise added to the batch examples are estimated using Eq. 1.

Table 1
Summary of the training sets used to train the NN models. For G labeled sets, the parameter spaces were sampled using the following grids, imposing the constraint $T_2<T_1$. *) T_1 and T_2 were incremented with steps of 10 ms; **) T_1 = [10:10:800, 850:50:1000, 1100:100:2000, 2500:500:4000] ms, T_2 = [1, 10:10:300, 350:50:1000, 600:100:1000, 1500:500:3000] ms and B_1^+ = [0.5:0.02:1.5]; ***) T_1 = [10, 20:20:500, 600:50:1000, 1100:100:2000, 2500:500:4000] ms, T_2 = [1, 10:10:500, 550:50:1000, 600:100:1000, 1500:500:3000] ms and $\Delta\nu_0$ = [-400:50:150, −100:10:-70, −60:2:60, 70:10:100, 150:50:400] Hz; ****) T_1 = [10, 20:20:300, 350:50:500, 600:100:1000, 1250:250:2000, 2500:500:4000], T_2 = [1, 10:10:300, 350:50:1000, 600:100:1000, 1500:500:3000] ms. $\Delta\nu_0$ = [-400:50:150, −100:10:-70, −60:2:60, 70:10:100, 150:50:400] Hz and B_1^+ = [0.5:0.1:1.5].

Model	Sequence	Set Label	T_1 (ms)	T_2 (ms)	$\Delta\nu_0$ (Hz)	B_1^+	Count
M1	IR-FISP	R1	[10 ÷ 4000]	[1 ÷ 3000]	0	1	75 555
		G1	[10 ÷ 4000]*	[1 ÷ 3000]*	0	1	
M1	IR-FISP B_1	R2	[10 ÷ 4000]	[1 ÷ 3000]	0	[0.5 ÷ 1.5]	164 475
		G2	[10 ÷ 4000]**	[1 ÷ 3000]**	0	[0.5 ÷ 1.5]**	
M2	IR-bSSFP	R3	[10 ÷ 4000]	[1 ÷ 3000]	[−400 ÷ 400]	1	175 041
		G3	[10 ÷ 4000] * * *	[1 ÷ 3000]] * * *	[−400 ÷ 400]* * *	1	
M2	IR-bSSFP B_1	R4	[10 ÷ 4000]	[1 ÷ 3000]	[−400 ÷ 400]	[0.5 ÷ 1.5]	396 550
		G4	[10 ÷ 4000] * * **	[1 ÷ 3000] * * **	[−400 ÷ 400] * * **	[0.5 ÷ 1.5] * * **	

The data augmentation step was done on-line. Hence, given a training set, it was unnecessary to store any new data either in hard or in flash memories to carry out the three data augmentation strategies. After the data augmentation step, each input was then normalized to the unit norm.

For each model, the performance in predicting an MR parameter was estimated in terms of Mean Absolute Percentage Error (MAPE) of the predicted parameter, evaluated on a test set composed of 30000 fingerprints generated using the same pulse sequence used for the training procedure, and sampling parameter space with a random uniform distribution. To test noise robustness, the prediction procedure described was repeated giving each model data with different SNRs = [3,5,10,20, 30,40,50]. Moreover, to assess the variance of the MAPE as a function of the SNR, for each noise level the operation was repeated 10 times. The MAPE standard deviation was considered as a measure of variability. The MAPE was computed following Eq. (3), where $\hat{p_k}$ is the estimated parameter value, while p_k is the ground truth parameter value and N is the number of examples considered.

Other error measurements used in this work are the Mean Percentage Error (MPE), the Absolute Error (AE), and the Root Mean Squared Error (RMSE) defined in Eqs. (4)–(6) respectively. Lin's concordance coefficient (ρ_c) [47] has been also used as a quantitative measure of agreement.

$$MAPE\,(\%) = \sum_{k=1}^{N} \left| \frac{\hat{p_k} - p_k}{p_k} \right| \times 100 \tag{3}$$

$$MPE\,(\%) = \sum_{k=1}^{N} \frac{\hat{p_k} - p_k}{p_k} \times 100 \tag{4}$$

$$AE = \hat{p_k} - p_k \tag{5}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (\hat{p_k} - p_k)^2} \tag{6}$$

2.4.3. Interpolation and extrapolation capabilities of trained NN models

To address the performance of NN models M1 and M2 in interpolation and extrapolation tasks, the models were required to predict MR parameters where the parameters ranges exceed those used during training. This experiment was conducted both for IR-FISP and IR-bSSFP sequences, using models M1 and M2, respectively. White Gaussian noise, with variance such that each test signal had SNR equal to 100, was added before the prediction task.

2.5. Numerical Brain Phantom Simulations

Realistic $T_1, T_2, \Delta\nu_0$, and B_1^+ maps were obtained by processing real acquisitions downloaded from the Multi-Modal MRI Reproducibility Resource repository. Landman et al. [48] performed MRI acquisitions with standard quantitative protocols at 3 T: variable flip angle (VFA) imaging for T_1 mapping; double echo time imaging for T_2 mapping; two sequential 2D gradient-echo with different echo times for B_0 ($\Delta\nu_0$) mapping and Actual Flip-Angle Imaging (AFI) for B_1^+ mapping. The in vivo brain images acquired with these protocols were then processed with MATLAB to obtain the quantitative maps. In particular, the qMRLab software [49] was used to process the VFA applying B_1^+ correction, since B_1^+ is known to be a confounding factor for the estimation of T_1 in VFA imaging [50]. Once obtained, these quantitative maps, reported in Fig. 3, were used as ground truth to simulate the MRF acquisition at the single-pixel level, using the four pulse sequences described in Section 2.1. B_1^+ is expressed as the ratio between the actual and nominal flip angles, thus it is dimensionless. Complex white Gaussian noise was added to simulated data using the SNR defined in Eq. 2.

Of note, all MR parameters were encoded in the MRF simulations. Hence, the B_1^+ map was taken into account for all four sequences; in the IR-FISP and IR-FISP $B_1, \Delta\nu_0$ was not be considered because it was not encoded by the sequence, while in the IR-bSSFP sequences $\Delta\nu_0$ is encoded and thus it has been considered in the simulations.

For each pulse sequence, the parameter maps were reconstructed by processing MRF data with both the trained NN models and the usual dictionary method based on the dot product. For the latter algorithm, the data sets G1, G2, G3, and G4 were used as dictionaries for the corresponding pulse sequences. Moreover, four dictionaries with around 400000 entries each were created to test the performance of the dictionary approach using dictionaries with higher resolution than those of data sets G1 - G4. MAPE and RMSE were used as global measures of reconstruction quality, while the AE was used as a local measure. To have a complete overview of the performance, all three errors must be assessed. MAPE gives information about the mean relative error of parameter estimation, which has the advantage of giving an immediate sense of global performance, but it can be misleading when very small values, such as $\Delta\nu_0$ off-resonances, are taken into account because small absolute errors give high relative errors. In such cases, a more reliable global measure is the RMSE, which expresses a global error in the same unit of measurement as the parameter considered. It is also interesting to check a measure of local error, such as the AE, which can be used to assess whether the error has a spatially uncorrelated distribution in the image or whether it shows spatial structure.

3. Results

3.1. Training data distribution: Grid Vs Random sampling

Fig. 4 reports the ground truth MR parameters of the test set against those predicted by the NN model M2, when trained either with a grid (top raw) or randomly (bottom raw) sampled training data, for the IR-bSSFP sequence. Data points are colored according to the value of their $\Delta\nu_0$ off-resonances. When the NN model is trained with a grid sampled training set, very weak agreement was found between the ground truth values and those predicted by the NN model. Lin's concordance coefficients were equal to 0.5, 0.1 and 0.4 for T_1, T_2 and $\Delta\nu_0$, respectively. From the scatter plots it is possible to observe that: i) the $\Delta\nu_0$ parameter was correctly estimated when $B_0 = [-60 \div 60]$ Hz (with sampling step 2 Hz in the training set); ii) T_1 generally showed good agreement with GT values when $\Delta\nu_0$ was restricted to this range (green color labeled data points in Fig. 4a); iii) T_2 had the worst level of agreement among the MR parameters (Fig. 4b). When the training data set was created by uniformly random sampling the MR parameter space, the M2 NN model predictions showed strong agreement with the ground truth values. Lin's concordance coefficients were equal to 0.99, 0.95 and 1 for T_1, T_2 and $\Delta\nu_0$, respectively. Although not reported here, it is important to note that for the IR-bSSFP-B1 sequence, the NN model M2 showed very low agreement between predicted and ground truth MR parameters when trained using a data set built by sampling the MR parameter space using a grid, whereas good agreement was found when the MR parameter space was sampled using a uniform random distribution.

Fig. 5 reports, for the IR-FISP-B_1 sequence, the ground truth MR parameters of the test set against those predicted by the NN model M1, for the cases in which the model was trained by sampling the parameter space either with a grid (top raw) or randomly (bottom raw). Data points are colored according to the value of their B_1^+ inhomogeneity. When the NN model was trained with a grid sampled training set a weak/moderate-strong agreement was found between the ground truth values and those predicted by the NN model. Lin's concordance coefficients were equal to 0.99, 0.5 and 0.8, for T_1, T_2 and $\Delta\nu_0$, respectively. From the scatter plots it is possible to observe that T_2 had the worst level of agreement among the MR parameters (Fig. 5b) although the majority of the dispersion is due to data points where the GT B_1^+ inho-

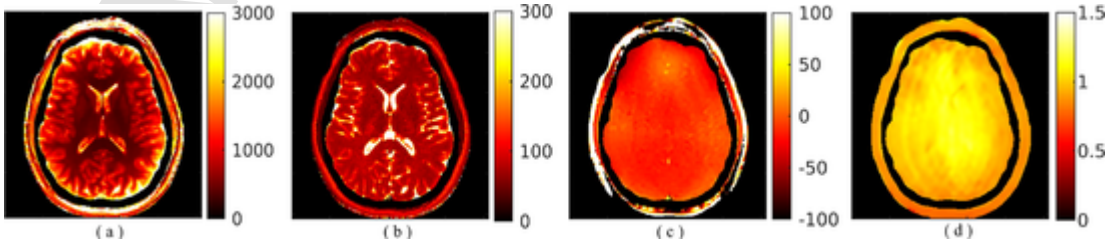

Fig. 3. Quantitative brain maps computed from the Multi-Modal MRI Reproducibility Resource: (a) T_1 (ms), where T_1 values from 3000 ms to 4000 ms are shown as equal to 3000 ms; (b) T_2 (ms), where T_2 values from 300 ms to 2500 ms are shown as equal to 300 ms; (c) off-resonances $\Delta\nu_0$ (Hz), where $\Delta\nu_0$ values less than |500| Hz are shown as equal to |100| Hz. $\Delta\nu_0$; (d) B_1^+, which is expressed as the ratio between the actual and nominal flip angles, thus it is dimensionless.

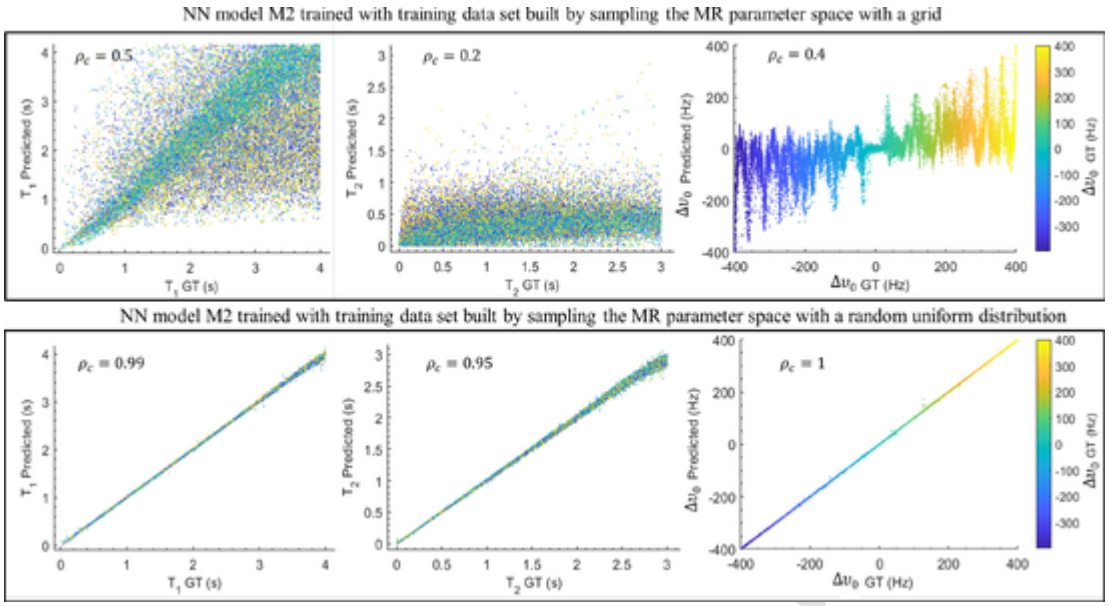

Fig. 4. Predicted MR parameters against Ground Truth (GT) parameters using the NN model trained using examples generated using the IR-bSSFP pulse sequence. Top row, NN model trained using examples sampled following a grid in the MR parameter space. Bottom row, NN model trained using examples sampled following a uniform random distribution in the MR parameter space. The number of the examples are the same in both the training sets. Lin's concordance coefficients are also reported as quantitative measures of agreement. Note that the reported predictions refer to test data without added noise.

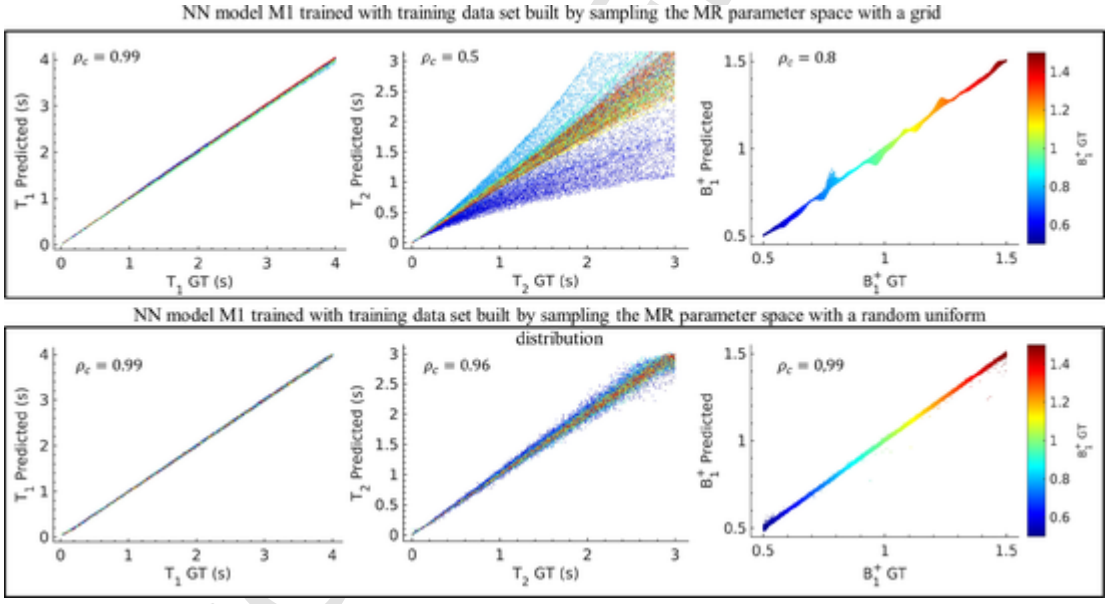

Fig. 5. Predicted MR parameters against Ground Truth (GT) parameters using the NN model trained using examples generated using the IR-FISP-B1 pulse sequence. Top row, NN model trained using examples sampled following a grid in the MR parameter space. Bottom row, NN model trained using examples sampled following a uniform random distribution in the MR parameter space. The number of the examples are the same in both the training sets. Lin's concordance coefficients are also reported as quantitative measures of agreement. Note that the reported predictions refer to test data without added noise. B_1^+ is expressed as the ratio between the actual and nominal flip angles, thus it is dimensionless.

mogeneity is below 1. When the training data set was created by uniformly random sampling the MR parameter space, the M1 NN model predictions showed a strong agreement with the ground truth values. Lin's concordance coefficients were equal to 0.99, 0.96 and 0.99 for T_1, T_2 and B_1^+, respectively. Although not reported here, it is important to note that for the IR-FISP sequence, which encodes only T_1 and T_2 the NN model M1 showed strong agreement between predicted and ground truth MR parameters regardless of the way in which the training set was sampled.

In Fig. S1 of Supplementary material, the training and test losses for models M1 and M2 are plotted as a function of the training epoch.

3.2. NN noise robustness using different noise adding strategies

MAPE of MR parameters, evaluated on test sets for different data augmentation strategies as a function of SNR, are reported in Fig. 6 for the four MRF sequences. The error bar indicates two standard deviations of the MAPE, which comes from 10 repetitions of the prediction procedure on noisy data. For all the MRF sequences and for all the predicted MR parameters, the NN models trained with the *variable variance* strategy had the lowest increase in MAPE as SNR decreases. In particular, T_2 MAPE never went over 12% and decreases rapidly under 10%, while, for the other parameters, MAPE was always lower than 5%. Not

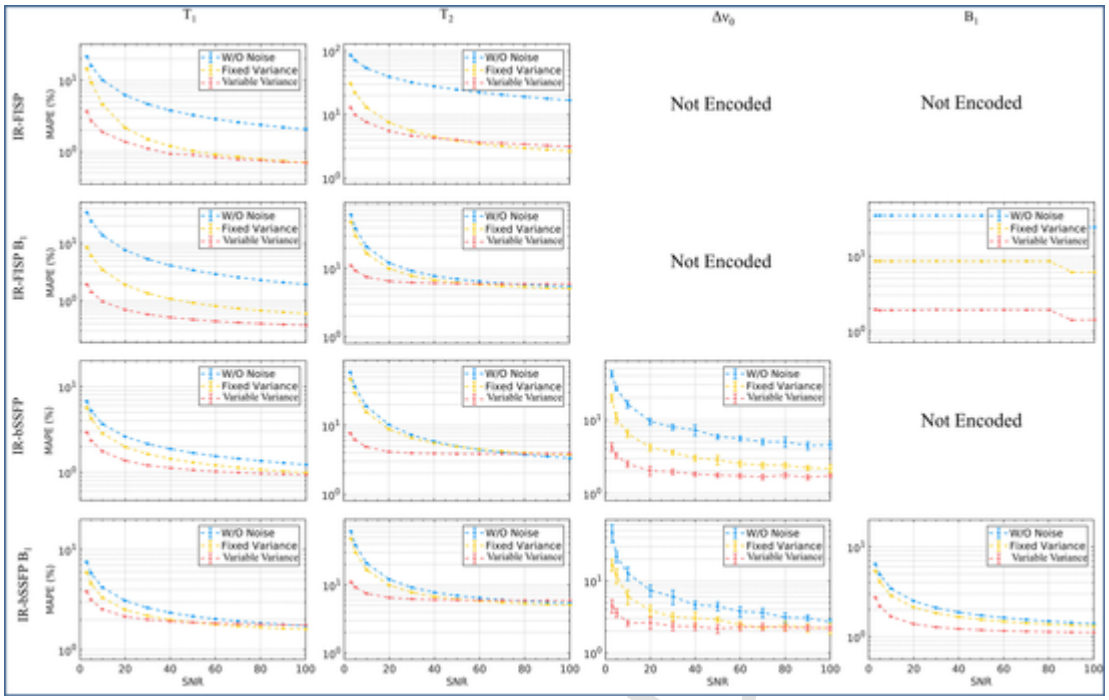

Fig. 6. Comparison of parameter MAPEs evaluated on test sets for NN models trained with different data augmentation strategies. Each column presents the comparison plots for the four pulse sequences for a specific MR parameter, i.e. T_1 (first column), T_2 (second column), $\Delta\nu_0$ (third column) and B_1^+ (fourth column).

using any noise adding strategy during training produced the worst results in terms of MAPE, while NN models trained using the *fixed variance* strategy, produced MAPEs that, for each SNR level, were higher than those reported for the *variable variance* noise addition strategy, but lower than the case in which the *W/O noise* strategy was employed. All models showed low variance over experiment repetitions. The standard deviation values were in the order of $_{10}{}^{-2}\%$, and are barely noticeable in the plots.

MPE as a function of SNR for the three different noise strategies and for the four MRF sequences is shown in Fig. 7. When the *variable variance* strategy was used MPEs were around 0% regardless of the SNR level. The NN models trained using the *fixed variance* and the *W/O noise* strategies showed increased MPE with a reduction of SNR, with the NN

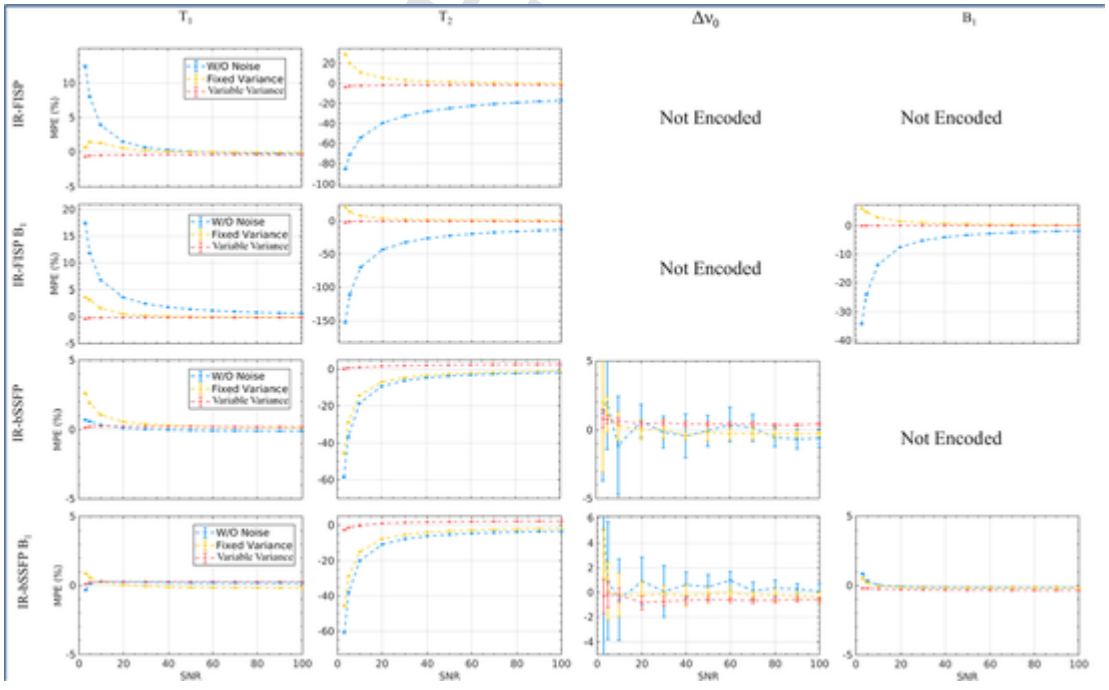

Fig. 7. Comparison of the parameter mean percentage errors (MPEs) evaluated on test sets for the NN models trained with different data augmentation strategies. Each column presents the comparison plots for the four pulse sequences for a specific MR parameter, i.e. T_1 (first column), T_2 (second column), $\Delta\nu_0$ (third column) and B_1^+ (fourth column). The *variable variance* strategy produces unbiased NNs models regardless of the SNR (MPE around 0%), whereas the other strategies show increased MPE as SNR decreases.

models trained using the *W/O noise* strategy showing the highest deviation from a 0% MPE.

In Fig. S2 of Supplementary material, the training and test losses for models M1 and M2 using the three noise adding strategies are plotted as a function of the training epoch.

3.3. Interpolation and extrapolation capabilities of trained NN models

In Figs. 8 and 9 the interpolation and extrapolation capabilities of the NN models, M1 and M2, are assessed. Considering Fig. 8 for model M1, the results showed a very good agreement when the test examples, never encountered during training, had MR parameters within the training range. Lin's coefficients were 0.99 for T_1 and T_2, whereas agreement decreased when the test signals had MR parameters partially outside the training ranges ($\rho_c = 0.55$ for T_1, and $\rho_c = 0.94$ for T_2). When all the MR parameters characterizing the test signals were outside the training ranges, the agreement was even worse ($\rho_c = 0.3$ for T_1, and $\rho_c = 0.03$ for T_2). It is worth pointing out that Lin's coefficient for T_2 was significantly higher than that for T_1 when at least one parameter was inside the training ranges because there were many more test examples for which T_1 was outside the training range, while T_2 was inside the

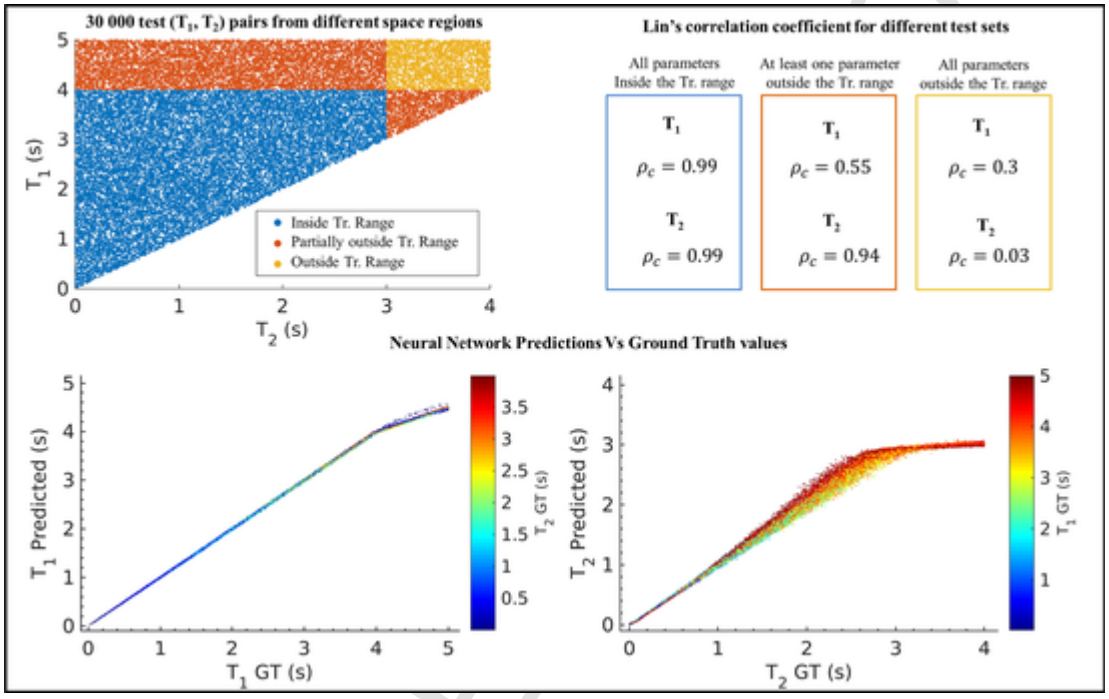

Fig. 8. IR-FISP interpolation and extrapolation capabilities. The top left plot represents the (T_1, T_2) pairs used to generate the test sets, which were colored depending on whether the (T_1, T_2) pair was completely inside (blue), partially outside (orange) and completely outside (yellow) the range of T_1 and T_2 used to train the NN. On the bottom, scatter plots between NN predicted MR parameters: T_1 (left) and T_2 (right). (top right) the Lin's correlation coefficients regarding the scatter plots, which were computed for the different test sets.

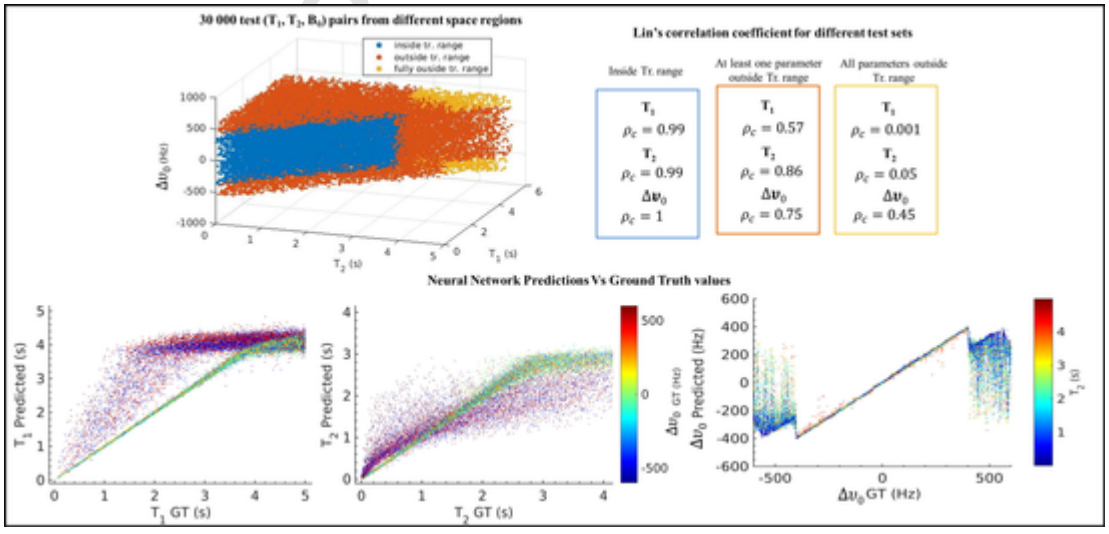

Fig. 9. IR-bSSFP interpolation and extrapolation capabilities. The top left plot represents the $(T_1, T_2, \Delta\nu_0)$ pairs used to generate the test sets, which were colored depending on whether the $(T_1, T_2, \Delta\nu_0)$ pair was completely inside (blue), partially outside (orange) and completely outside (yellow) the range of T_1, T_2 and $\Delta\nu_0$ used to train the NN. On the bottom, scatter plots between NN predicted MR parameters: T_1 (left), T_2 (central) and $\Delta\nu_0$ (right). (top right) the Lin's correlation coefficients regarding the scatter plots, which were computed for the different test sets.

training range, than vice versa. Looking at Fig. 9, the considerations made for model M1 apply equally to model M2.

3.4. Brain map reconstruction: Neural Networks vs Dictionaries

In Fig. 10(a) the absolute error maps for each MR parameter, relative to NN reconstruction and dictionary matching are reported. They were evaluated by computing, pixel-wise and for each parameter map, the absolute error between the ground-truth and the predicted parameter value. For each map, two global error estimators, and the Lin's concordance coefficient are reported at the corners of the images: the Lin's concordance coefficient between the ground-truth pixel values and those reconstructed (upper left corner); the MAPE between the ground-truth pixel values and those reconstructed (upper right corner); the RMSE between the ground-truth pixel values and those reconstructed (lower right corner). All the estimators refer to the brain, excluding the scalp region. Although only the reconstructions using noisy data with SNR = 5 are reported here, in Fig. S3 of Supplementary Material, the RMSE between reconstructed and ground-truth values as a function of SNR is reported.

Considering the NN reconstructions, Fig. 10(a), when B_1^+ was taken into account during training, the performance increased, especially in reconstructing the T_2 parameter map, for both the IR-FISP and the IR-bSSFP type sequences. The Lin's concordance coefficients did not change substantially (ρ_c greater than or equal to 0.98 for all the T_2 maps), but T_2 MAPE decreased from 13% to 3% for the IR-FISP type sequences (Fig. 10(a) (e) and (f)), and from 15% to 6% for the IR-bSSFP type sequences (Fig. 10(a) (g) and (h)). Generally, all the NN reconstructions had a strong level of agreement, with a ρ_c always greater than or equal to 0.97. There was no difference in terms of T_1 MAPEs when B_1^+ was considered during training, compared to the case in which it was not considered, both for NN model M1 (See Fig. 10(a) and (b)) and for M2 (See Fig. 10(a) (c) and (d)). Lastly, considering the $\Delta\nu_0$ reconstructions (Fig. 10(a) (i) and (l)), the MAPEs were 36% and 39% for the IR-bSSFP and IR-FISP-B1 sequences, respectively, but RMSE was in the order of 2 Hz. It is worth noting that $\Delta\nu_0$ values were close to 0 Hz and varied in the range [-60, 60] Hz within the numerical brain phantom.

Considering the MR parameter maps reconstructions using the dictionary-based approach, Fig. 10(b), for the IR-FISP type sequence when B_1^+ was encoded into the dictionary estimation accuracy in reconstructing the T_2 parameter map increased. With reference to Fig. 10(b) (e) and (f), the Lin's concordance coefficients did not change (ρ_c = 0.98), but T_2 MAPE decreased from 11% to 3% and RMSE from 31 ms to 22 ms, when the dictionary encoded B_1^+ inhomogeneity. On the contrary, passing from the IR-bSSFP sequence (third column in the figure) to the IR-bSSFP-B1 sequence (fourth column), the performance in terms of reconstruction accuracy decreased both for T_1 (MAPE increased from 3% to 6%), T_2 (MAPE increased from 13% to 16%) and $\Delta\nu_0$ (MAPE increased from 12% to 25%). Overall, there was good agreement between ground truth and dictionary matched parameters for IR-FISP and IR-FISP-B1 sequences (ρ_c grater than or equal to 0.98 for all the parameters), but there was weak agreement between ground truth and reconstructed T_2 values for either the IR-bSSFP or the IR-bSSFP-B1 sequences (ρ_c less than or equal to 0.66).

Comparing the NN and the dictionary-based map reconstructions, Fig. 10(a) and (b), respectively, it is possible to note that when only two parameters were retrieved, as in the case for the IR-FISP, the accuracy of the two reconstructions was similar. On the other hand, when four parameters were retrieved, as in the case for the IR-bSSFP B1 sequence, the NN produced more accurate results with less error dispersion. Moreover, in the case of IR-bSSFP type sequences the dictionary-based reconstruction estimated CSF T_2 values that were less than or equal to 600 ms, while the ground truth values went up to 2000 ms. The AEs in the figures were clipped to 100 ms for the T_2 to enhance the contrast, so the above-reported behavior is not detectable the AE maps. However, the global RMSE values, which were in the order of 100 ms, are evidence of this phenomenon (See Fig. 10(b) (g) and (h). The NN reconstruction did not show this bias in the CSF region. Global RMSE was in the order of 30 ms (See Fig. 10(a) (g) and (h).

Fig. 11 reports the errors in brain parameter maps reconstructed using four dictionaries where the number of entries was kept fixed at around 400000 entries, i.e. at a higher resolution than those used to reconstruct the parameter maps in 10(b). Comparing Fig. 10(a) and Fig. 11, it is apparent that the performance of the dictionary approach reached that of the NN approach for the IR-bSSFP sequence in terms of

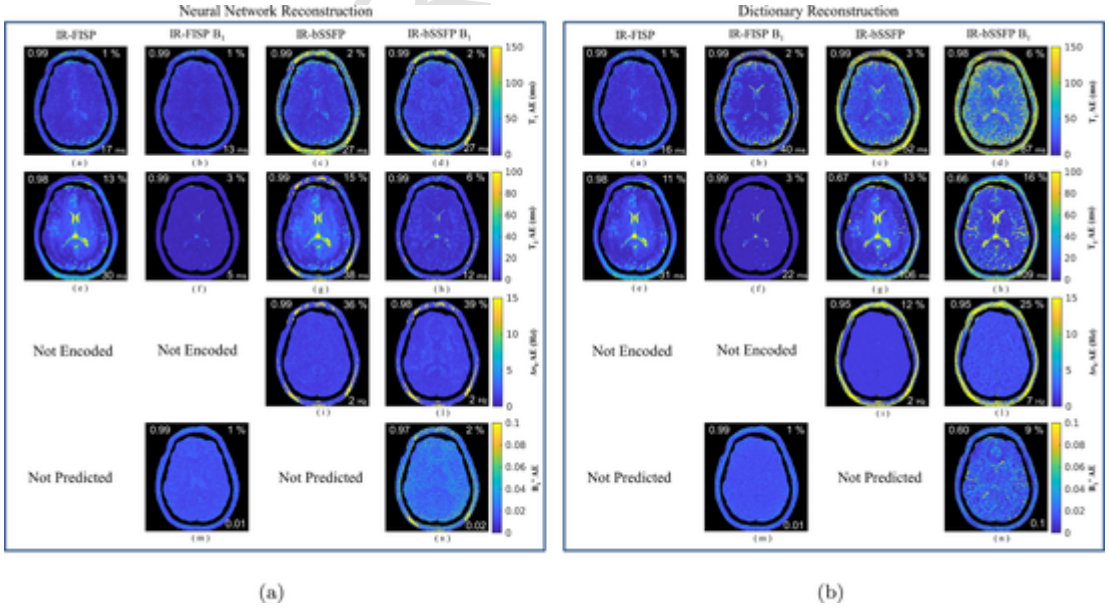

Fig. 10. Absolute error maps between reconstructed and ground truth parameter maps using NN (a) and dictionary (b) approaches. Global error estimators are reported on the four corners of each image: upper left corner: MAPE; upper right corner: MAPE evaluated without considering the scalp region; lower left corner: RMSE; lower right corner: RMSE evaluated without considering the scalp region. B_1^+ is expressed as the ratio between the actual and nominal flip angles, thus it is dimensionless.

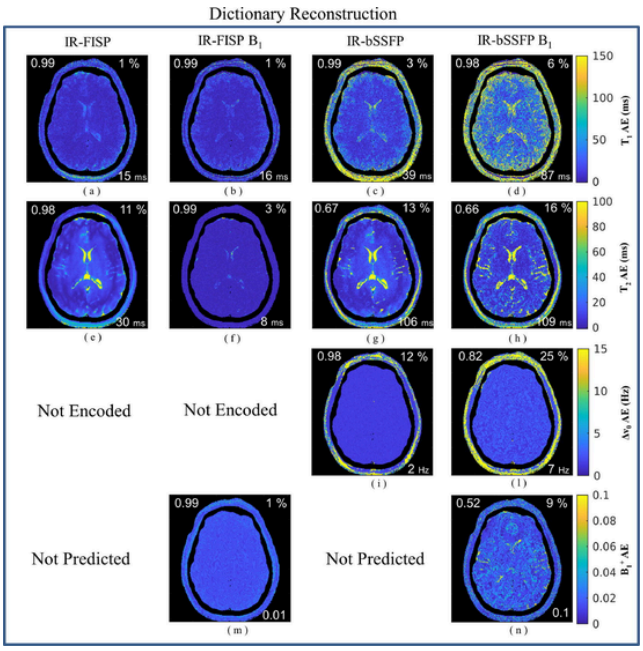

Fig. 11. Absolute error maps between NN reconstructed and ground truth parameter maps with original B_1^+ field inhomogeneities map at SNR = 5. All dictionaries are composed of 400 000 entries. Global error estimators are reported at the four corners of each image: the MAPE (upper left corner); MAPE evaluated after excluding the scalp region (upper right corner); the RMSE (lower left corner), and RMSE evaluated after excluding the scalp region (lower right corner). B_1^+ is expressed as the ratio between the actual and nominal flip angles, thus it is dimensionless.

accuracy. However, the T_2 bias in the CSF is still present. For the IR-bSSFP B_1 sequence, the NN maintained its superior accuracy.

To process 35,000 pixels, the NN reconstruction took about 1.5 s for IR-FISP-type sequences and 4 s for IR-bSSFP-type sequences. No significant changes in processing time were observed when estimating 2 or 3 parameters with the IR-FISP-type sequences, or 3 or 4 parameters with the IR-bSSFP-type sequences. The computational time required to generate the maps reported in Fig. 10(b) with the dictonary approach were 30 s, 66 s, 71 s and 163 s for the IR-FISP, IR-FISP B_1, IR-bSSFP and IR-bSSFP B_1 respectively. The computational time required to generate the maps in Fig. 11 using the dictionary approach was around 160 s.

4. Discussion

In this work, we propose a deep fully-connected neural network approach applied to voxel-wise MRF parameter map reconstruction to circumvent the curse of dimensionality that affects the gold-standard dictionary approach. We tested our approach on simulated data considering four MRF pulse sequences: the IR-FISP sequence and its variant designed to account for B_1^+ field inhomogeneities, the IR-bSSFP, and a newly proposed variant designed to account for B_1^+ field inhomogeneities.

We first studied how the design of the training set would affect the generalization capabilities of the NN models. We found that generating the training data set by sampling the MR parameter space using a grid leads to low generalizability. The method introduces a strong bias due to its regularity, and this leads the NN to overfit the data, as shown for the IR-FISP-B_1 and IR-bSSFP type sequences. Moreover, the finding that for the IR-bSSFP sequence test examples characterized by $\Delta\nu_0$ in the range [-60, 60] Hz (where the sampling step in the training set was 2 Hz) is a hint that the overfitting problem with the grid sampling is present when the grid does not have high enough sampling resolution. On the contrary, sampling the MR parameter space using random uniform distributions yielded excellent generalization capabilities regardless of the pulse sequence used. Although grid sampling is a reasonable way to build dictionaries, it is not well suited to training NNs. This result is in accordance with the NN literature, where it has been shown that grid sampling normally requires many more training examples than random sampling to reach comparable performance [34]41. In the context of applying NN to MRF, this is an important finding since several published works have built training data sets using grid sampling [20,24,32]. However, our findings may suggest that the authors were able to efficiently train the NN models because they either limited parameter estimation to T_1 and T_2 using the IR-FISP sequence, [24] or limited the MR parameter range and used a densely resolved grid to construct the training set [32].

We have characterized the interpolation and extrapolation capabilities of the NN models under test. Those experiments showed that NN models are reliable only for interpolation tasks, i.e. where the values of MR parameters of unknown signals are within the ranges used during training.

Three different data augmentation strategies were then tested with regard to robustness to white Gaussian noise. The experiments showed that feeding the network models with noisy data using different noise variances yielded the best results in terms of noise robustness. The explanation for this behavior is straightforward: with the *Variable variance* strategy, the NNs see more data characterized by heterogeneous levels of noise, improving generalization, while with the *Fixed variance* strategy, the models eventually overfit for data with high SNRs, and, as expected, with the *W/O Noise* strategy the models perform least well in terms of noise robustness. This is an important finding since the *Fixed variance* and *W/O Noise* strategies have been used as data augmentation in previously published work regarding fully connected NN applied to MRF [19,24]. Moreover, since the noise addition is performed during training, and no new data are stored either in hard or flash memory, the data augmentation strategy does not affect memory usage efficiency. It is worth pointing out that by looking at Fig. 6, the MR parameter that shows the highest MAPE is the transverse relaxation time T_2, even when SNR is high. However, those MAPEs are acceptable considering the wide range of tested SNRs, and especially the wide range of estimated T_2 values (up to 3000 ms). MRF sequences considered in this work have TRs in the order of 10 ms, and, considering the NMR transverse relaxation process, it is expected that sequences with such TRs will have higher sensitivity when the T_2 values are in the order of the TRs employed. When T_2 relaxation times are orders of magnitude longer, we expect the pulse sequence to be less sensitive to T_2 variations, since there will be less transverse magnetization decay in each TR. A loss of sensitivity is also expected when T_2 relaxation times are much shorter than pulse TRs. In Fig. S4 of the Supplementary Material, the results of simulations run using the IR-FISP sequence are summarized for three sets of examples with the following T_2 relaxation times: shorter, comparable, and much longer than the TRs used in the pulse sequence (T_1 was kept fixed at 1.5 s and no noise was added in the simulations), which confirm a decreased sensitivity of the MRF pulse sequence when the T_2 values are some order of magnitudes shorter or longer than the TRs used in the sequence.

The results of the comparison between NN and dictionary-based reconstructions of the numerical brain phantom show that NN performance is equal to or greater than that of the dictionary-based approach, depending on the resolution of the dictionary used to perform the reconstruction. However, an increase in the resolution of the dictionary comes at the cost of longer computational time, whereas the NN approach has a constant computational time independent of the number of parameters to be retrieved. A paradigmatic example of the limits of dictionary-based approaches in the context of large MRF problems can be seen in the fourth column of Fig. 10(b). From that figure, one can conclude that taking B_1^+ field inhomogeneities into account does not improve parameter estimation for IR-bSSFP sequences. In fact, errors were

higher when B_1^+ field inhomogeneities were predicted in the dictionary, which was the opposite of the expected result, and in contrast to what was found using the NN approach. However, this finding is only due to the number of entries of the dictionary, which was insufficient to guarantee an accurate sampling of the 4D MR parameter space. This is an effect of the curse of dimensionality of nearest-neighbor algorithms. With the dictionary approach, there is a trade-off between keeping the number of dictionary entries below a reasonable number and guaranteeing a good resolution. When larger dictionaries were used the performance of the dictionary approach improved in accordance with the degree of improvement in the dictionary resolution, but at the cost of longer computational time. With the NN approach, the model learns an approximation of the ITF from a set of training examples. Once trained, the processing is computationally fast and efficient, not requiring a large dictionary stored in memory.

In MRF, the improvement of parameters estimation accuracy can be obtained considering as many scanner imperfections as possible, such as B_0 and B_1^+ inhomogeneities. With the classical dictionary-based approach to MRF, adding a parameter requires an increase in the dictonary size following the power of the number of parameters to be retrieved. This challenges MRF applications that consider other relevant MR parameters, such as T_2^* [35], water-fat ratio [36], or magnetization transfer [37]. These results show the advantages of NN approaches to MR Fingerprinting parameter estimation. Thousands of training examples might be required to learn an accurate and robust ITF with NN, as shown for the IR-bSSFP and IR-bSSFP B_1 sequences. However, because the NN model is trained in batches, the data can be stored in the hard disk, and just a fraction are loaded into the RAM in each step of the training. Moreover, because the training examples are synthetic data, one can even build the simulation step into the training pipeline, without the need to store any data even on the hard disk. The code for the simulation step should be fast and efficient, or it becomes a bottleneck in terms of the time needed to train the network model. However, parallel coding is well suited to this kind of problem and can be used to speed up computation.

Another interesting finding that emerges from the work is that for the IR-bSSFP sequence, the dictionary-based reconstruction produced a systematic reduction in the estimated T_2 of the CSF region (always less than 600 ms) compared to ground truth values (around 2000 ms). Such T_2 shortening was not alleviated when the resolution of the dictionary was increased, and was not observed using NN reconstruction. In the original article [1], which used the same IR-bSSFP sequence as considered here, the authors indeed found a shortened T_2 value in the CSF region (550 ms) in their in vivo experiments. This was explained as being probably due to out-of-plane flow in the two-dimensional imaging experiment. However, in our work, no imaging scheme was used and this suggests that such shortening could be introduced by the dictionary matching procedure, that used the dot product as a similarity measure (the gold-standard similarity measure used in dictionary-based MRF). The fact that the NN model, which learns the ITF, does not present this T_2 shortening suggests that NN may be beneficial not only from a computational perspective but also in terms of accuracy. Future works will investigate in greater depth the origin of such a difference between the two approaches in estimating the T_2 value of CSF with the IR-bSSFP MRF sequence.

A limitation of the current study is that it did not consider the artifacts that arise from k-space undersampling, which could require a fine-tuning of the training pipeline here reported. In future works, different noise models could be investigated to assess the generalization performance of the NN models, as well as data preprocessing methods for noise reduction, such as SVD or the method recently proposed by Bo et al. [51], in which high-quality time-series images are reconstructed from highly-aliased images. Preliminary results of this on-going study have recently been presented [29], in which an SVD pre-processing filter in the time domain was proposed to mitigate the effect of undersampling for in vivo undersampled MRF FISP data. Other recent contributions have investigated the application of FCNNs to in vivo MRF data [28].

5. Conclusion

The set of experiments reported here allowed us to systematically identify the best practices for training fully connected NN models to be applied to MR parameter reconstruction with MRF. We have demonstrated how the NN approach can circumvent the curse of dimensionality, which inherently limits the scalability of dictionary-based approaches. We have also demonstrated that the dictionary-based approach could introduce a shortening in the T_2 estimation of the CSF region when using the IR-bSSFP sequence while the NN approach does not.

Overall, the above results evince the promise of NN approaches to MRF applications. This work may help to increase the trust of the community working in MRF in deep learning approaches, pushing fingerprinting pulse sequence design beyond the limitations affecting the dictionary approach due to the curse of dimensionality, and allowing the simultaneous consideration of more meaningful MR parameters.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Italian National Institute for Nuclear Physics (Progetto AIM – Artificial Intelligence in Medicine).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, athttps://doi.org/10.1016/j.ejmp.2021.07.013.

References

[1] Ma D, Gulani V, Seiberlich N, Liu K, Sunshine JL, Duerk JL, et al. Magnetic resonance fingerprinting. Nature 2013;495:187–92. https://doi.org/10.1038/nature11971.
[2] Lemasson B, Pannetier N, Coquery N, Boisserand LSB, Collomb N, Schuff N, et al. MR vascular fingerprinting in stroke and brain tumors models. Sci Rep 2016;6:37071. https://doi.org/10.1038/srep37071.
[3] Chen Y, Jiang Y, Pahwa S, Ma D, Lu L, Twieg MD, et al. MR fingerprinting for rapid quantitative abdominal imaging. Radiology 2016;279(1):278–86. https://doi.org/10.1148/radiol.2016152037.
[4] Chen Y, Panda A, Pahwa S, Hamilton JI, Dastmalchian S, McGivney DF, et al. Three-dimensional MR fingerprinting for quantitative breast imaging. Radiology 2018. https://doi.org/10.1148/radiol.2018180836.
[5] 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 2014;74(6):1621–31. https://doi.org/10.1002/mrm.25559.
[6] Cloos MA, Knoll F, Zhao T, Block KT, Bruno M, Wiggins GC, et al. Multiparametric imaging with heterogeneous radiofrequency fields. Nat Commun 2016;7:12445. https://doi.org/10.1038/ncomms12445; article.
[7] Assländer J, Cloos MA, Knoll F, Sodickson DK, Hennig J, Lattanzi R. Low rank alternating direction method of multipliers reconstruction for MR fingerprinting. Magn Reson Med 2018;79(1):83–96. https://doi.org/10.1002/mrm.26639.
[8] Wang J, Mao W, Qiu M, Smith MB, Constable RT. Factors influencing flip angle mapping in MRI: RF pulse shape, slice-select gradients, off-resonance excitation, and B0 inhomogeneities. Magn Reson Med 2006;56(2):463–8. https://doi.org/10.1002/mrm.20947.
[9] Buonincontri G, Sawiak SJ. MR fingerprinting with simultaneous B1 estimation. Magn Reson Med 2015;76(4):1127–35.
[10] Ma D, Coppo S, Chen Y, McGivney DF, Jiang Y, Pahwa S, et al. Slice profile and B1 corrections in 2D magnetic resonance fingerprinting. Magn Reson Med 2017;78(5):1781–9. https://doi.org/10.1002/mrm.26580.
[11] Kulpanovich A, Tal A. The application of magnetic resonance fingerprinting to single voxel proton spectroscopy. NMR Biomed 2018;31(11):e4001. https://doi.org/10.1002/nbm.4001.
[12] Bellman R. Adaptive control processes: a guided tour. Princeton University Press;

1961.
[13] Friedman JH. On Bias, Variance, 0/1–Loss, and the Curse-of-Dimensionality. Data Min Knowl Discov 1997;1(1):55–77. https://doi.org/10.1023/A:1009778005914.
[14] McGivney DF, Pierre E, Ma D, Jiang Y, Saybasili H, Gulani V, et al. SVD compression for magnetic resonance fingerprinting in the time domain. IEEE Trans Med Imaging 2014;33(12):2311–22.
[15] Mazor G, Weizman L, Tal A, Eldar YC. Low Rank Magnetic Resonance Fingerprinting. ArXiv e-prints 2017 arXiv:1701.07668.
[16] Yang M, Ma D, Jiang Y, Hamilton J, Seiberlich N, Griswold A, et al. Low rank approximation methods for MR fingerprinting with large scale dictionaries. Magn Reson Med 2017;79(4):2392–400. https://doi.org/10.1002/mrm.26867.
[17] Hornik K. Approximation capabilities of multilayer feedforward networks. Neural Netw 1991;4(2):251–7.
[18] Hamilton JI, Seiberlich N. Machine learning for rapid magnetic resonance fingerprinting tissue property quantification. Proc IEEE 2020;108(1):69–85. https://doi.org/10.1109/JPROC.2019.2936998.
[19] Virtue P, Yu SX, Lustig M. Better than real: complex-valued neural nets for MRI fingerprinting. 2017 IEEE International Conference on Image Processing (ICIP); 2017. p. 3953–7. https://doi.org/10.1109/ICIP.2017.8297024.
[20] Hoppe E, Körzdörfer G, Würfl T, Wetzl J, Lugauer F, Pfeuffer J, et al. Deep learning for magnetic resonance fingerprinting: a new approach for predicting quantitative parameter values from time series. Stud Health Technol Inform 2017;243:202–6.
[21] Chen D, Golbabaee M, Gomez PA, Menzel MI, Davies ME. A fully convolutional network for MR fingerprinting. ArXiv 2019;abs/1911.09846.
[22] Song P, Eldar YC, Mazor G, Rodrigues MRD. HYDRA: hybrid deep magnetic resonance fingerprinting. Med Phys 2019;46(11):4951–69. https://doi.org/10.1002/mp.13727.
[23] Oksuz I, Cruz G, Clough J, Bustin A, Fuin N, Botnar RM, et al. Magnetic resonance fingerprinting using recurrent neural networks. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019); 2019. p. 1537–40. https://doi.org/10.1109/ISBI.2019.8759502.
[24] Cohen O, Zhu B, Rosen MS. MR fingerprinting Deep RecOnstruction NEtwork (DRONE). Magn Reson Med 2018;80(3):885–94. https://doi.org/10.1002/mrm.27198.
[25] Cao P, Cui D, Vardhanabhuti V, Hui ES. Development of fast deep learning quantification for magnetic resonance fingerprinting in vivo. Magn Reson Imaging 2020;70:81–90. https://doi.org/10.1016/j.mri.2020.03.009.
[26] Balsiger F, Shridhar Konar A, Chikop S, Chandran V, Scheidegger O, Geethanath S, et al. Magnetic resonance fingerprinting reconstruction via spatiotemporal convolutional neural networks. In: Knoll F., Maier A., Rueckert D., editors. Machine Learning for Medical Image Reconstruction. Springer International Publishing: Cham; 2018. pp. 39–46. ISBN 978-3-030-00129-2.
[27] Fang Z, Chen Y, Liu M, Xiang L, Zhang Q, Wang Q, et al. Deep learning for fast and spatially-constrained tissue quantification from highly-accelerated data in magnetic resonance fingerprinting. IEEE Trans Med Imaging 2019;1. https://doi.org/10.1109/TMI.2019.2899328.
[28] Golbabaee M, Pirkl CM, Menzel MI, Buonincontri G, Gomez PA. Deep MR Fingerprinting with total-variation and low-rank subspace priors; 2019. Proc Intl Soc Magn Reson Med 2019;27.
[29] Barbieri M, Lee PK, Brizi L, Giampieri E, Toews AR, Castellani G, et al. Deep Learning for Magnetic Resonance Fingerprinting: Data Augmentation with Phase Encoding and SVD Preprocessing for Accurate Parameter Reconstruction of FISP Data; 2019. Proc Intl Soc Mag Reson Med 27.
[30] Hoppe E, Thamm F, Körzdörfer G. Syben Christopher Schirrmacher F, Nittka M, Pfeuffer J, et al. Magnetic resonance fingerprinting reconstruction using recurrent neural networks. Stud Health Technol Inform 2019. 10.3233/SHTI190816; pMID: 31483264
[31] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.
[32] Golbabaee M, Chen D, Gómez PA, Menzel MI, Davies ME. Geometry of deep learning for magnetic resonance fingerprinting. ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2019. p. 7825–9.
[33] Tong F, Liu X. Samples selection for artificial neural network training in preliminary structural design. Tsinghua Sci. Technol. 2005;10(2):233–9. https://doi.org/10.1016/S1007-0214(05)70060-2.
[34] R DGL, Pedergnana M, Garcia SG. Smart sampling and incremental function learning for very large high dimensional data. Neural Netw 2016;78:75–87. doi: 10.1016/j.neunet.2015.09.001; special Issue on Neural Network Learning in Big Data.
[35] Wyatt CR, Smith TB, Sammi MK, Rooney WD, Guimaraes AR. Multi-parametric T2* magnetic resonance fingerprinting using variable echo times. NMR Biomed 2018;31(9):e3951. https://doi.org/10.1002/nbm.3951.
[36] Ostenson J, Damon BM, Welch EB. MR fingerprinting with simultaneous T1, T2, and fat signal fraction estimation with integrated B0 correction reduces bias in water T1 and T2 estimates. Magn Reson Imaging 2019;60:7–19. https://doi.org/10.1016/j.mri.2019.03.017.
[37] Hilbert T, Xia D, Block KT, Yu Z, Lattanzi R, Sodickson DK, et al. Magnetization transfer in magnetic resonance fingerprinting. Magn Reson Med 2020;84(1):128–141. https://doi.org/10.1002/mrm.28096.
[38] Jürgen H. 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(3):125–43. https://doi.org/10.1002/cmr.1820030302.
[39] Zhao B, Setsompop K, Ye H, Cauley SF, Wald LL. Maximum likelihood reconstruction for magnetic resonance fingerprinting. IEEE Trans Med Imaging 2016;35(8):1812–23.
[40] Cline CC, Chen X, Mailhe B, Wang Q, Pfeuffer J, Nittka M, et al. AIR-MRF: accelerated iterative reconstruction for magnetic resonance fingerprinting. Magn Reson Imaging 2017;41:29–40.
[41] Sommer K, Amthor T, Doneva M, Koken P, Meineke J, Börnert P. Towards predicting the encoding capability of MR fingerprinting sequences. Magn Reson Imaging 2017;41:7–14. https://doi.org/10.1016/j.mri.2017.06.015.
[42] Gudbjartsson H, Patz S. The rician distribution of noisy mri data. Magn Reson Med 1995;34(6):910–4. https://doi.org/10.1002/mrm.1910340618.
[43] Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow. Large-Scale Machine Learning on Heterogeneous Systems 2015 URL: https://www.tensorflow.org/; software available from tensorflow.org.
[44] Kingma D, Ba J. Adam: a method for stochastic optimization. ArXiv e-prints 2014; arXiv:1412.6980.
[45] Goodfellow I, Bengio Y, Courville A. Deep Learning, Chapter 7. MIT Press; 2016 URL: http://www.deeplearningbook.org.
[46] Vincent P, Larochelle H, Bengio Y, Manzagol PA. Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th International Conference on Machine Learning. ICML '08; New York, NY, USA: ACM. pp. 1096–1103, ISBN 978-1-60558-205-4; 2008. DOI: 10.1145/1390156.1390294.
[47] Lin LIK. A concordance correlation coefficient to evaluate reproducibility. Biometrics 1989;45(1):255–68 URL: http://www.jstor.org/stable/2532051.
[48] Landman BA, Huang AJ, Gifford A, Vikram DS, Lim IAL, Farrell JA, et al. Multi-parametric neuroimaging reproducibility: a 3-T resource study. Neuroimage 2011;54(4):2854–66 URL: http://www.sciencedirect.com/science/article/pii/S1053811910015259 https://doi.org/10.1016/j.neuroimage.2010.11.047.
[49] Cabana JF, Gu Y, Boudreau M, Levesque IR, Atchia Y, Sled JG, et al. Quantitative magnetization transfer imaging made easy with qMTLab: Software for data simulation, analysis, and visualization. Concepts Magn Reson Part A Bridg Educ Res 2015;44A(5):263–77. https://doi.org/10.1002/cmr.a.21357.
[50] Boudreau M, Tardif CL, Stikov N, Sled JG, Lee W, Pike GB. B1 mapping for bias-correction in quantitative T1 imaging of the brain at 3T using standard pulse sequences. J Magn Reson Imaging 2017;46(6):1673–82. https://doi.org/10.1002/jmri.25692.
[51] Bo Z, Kawin S, Elfar A, Borjan G, Huihui Y, Dan M, et al. Improved magnetic resonance fingerprinting reconstruction with low-rank and subspace modeling. Magn Reson Med 2017;79(2):933–42. https://doi.org/10.1002/mrm.26701.