TU Delft

Delft University of Technology

Recurrent inference machines as inverse problem solvers for MR relaxometry

Sabidussi, E. R.; Klein, S.; Caan, M. W.A.; Bazrafkan, S.; den Dekker, A. J.; Sijbers, J.; Niessen, W. J.; Poot, D. H.J.

DOI
10.1016/j.media.2021.102220

Publication date
2021

Document Version
Final published version

Published in
Medical Image Analysis

Citation (APA)
Sabidussi, E. R., Klein, S., Caan, M. W. A., Bazrafkan, S., den Dekker, A. J., Sijbers, J., Niessen, W. J., & Poot, D. H. J. (2021). Recurrent inference machines as inverse problem solvers for MR relaxometry. *Medical Image Analysis*, *74*, Article 102220. https://doi.org/10.1016/j.media.2021.102220

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Medical Image Analysis 74 (2021) 102220

Contents lists available at ScienceDirect

Medical Image Analysis

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

Recurrent inference machines as inverse problem solvers for MR relaxometry

E.R. Sabidussi[a,*], S. Klein[a], M.W.A. Caan[b], S. Bazrafkan[c], A.J. den Dekker[c], J. Sijbers[c,d], W.J. Niessen[a,e], D.H.J. Poot[a]

[a] Erasmus MC University Medical Center, Department of Radiology and Nuclear Medicine, Rotterdam, the Netherlands
[b] Amsterdam UMC, Biomedical Engineering and Physics, University of Amsterdam, Amsterdam, the Netherlands
[c] imec-Vision Lab, Department of Physics, University of Antwerp, Antwerp, Belgium
[d] μ NEURO Research Centre of Excellence, University of Antwerp, Antwerp, Belgium
[e] Delft University of Technology, Delft, the Netherlands

ARTICLE INFO

Article history:
Received 10 December 2020
Revised 10 June 2021
Accepted 26 August 2021

Keywords:
Quantitative MRI
Relaxometry
Deep learning
Mapping
Recurrent inference machines

ABSTRACT

In this paper, we propose the use of Recurrent Inference Machines (RIMs) to perform T_1 and T_2 mapping. The RIM is a neural network framework that learns an iterative inference process based on the signal model, similar to conventional statistical methods for quantitative MRI (QMRI), such as the Maximum Likelihood Estimator (MLE). This framework combines the advantages of both data-driven and model-based methods, and, we hypothesize, is a promising tool for QMRI. Previously, RIMs were used to solve linear inverse reconstruction problems. Here, we show that they can also be used to optimize non-linear problems and estimate relaxometry maps with high precision and accuracy. The developed RIM framework is evaluated in terms of accuracy and precision and compared to an MLE method and an implementation of the Residual Neural Network (ResNet). The results show that the RIM improves the quality of estimates compared to the other techniques in Monte Carlo experiments with simulated data, test-retest analysis of a system phantom, and in-vivo scans. Additionally, inference with the RIM is 150 times faster than the MLE, and robustness to (slight) variations of scanning parameters is demonstrated. Hence, the RIM is a promising and flexible method for QMRI. Coupled with an open-source training data generation tool, it presents a compelling alternative to previous methods.

© 2021 The Author(s). Published by Elsevier B.V.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

1. Introduction

MR relaxometry is a technique used to measure intrinsic tissue properties, such as T_1 and T_2 relaxation times. Compared to qualitative weighted images, quantitative T_1 and T_2 maps are much less dependent on variations of hardware, acquisition settings, and operator (Cercignani et al., 2018). Additionally, because measured T_1 and T_2 maps are more tissue-specific than weighted images, they are promising biomarkers for a range of diseases (Cheng et al., 2012; Conlon et al., 1988; Erkinjuntti et al., 1987; Larsson et al., 1989; Lu, 2019).

Thanks to their low dependence on hardware and scanning parameters, quantitative maps are highly reproducible across scanners and patients (Weiskopf et al., 2013), presenting variability comparable to test-retest experiments within a single center (Deoni et al., 2008). The low variability allows for direct comparison of tissue properties between patients and across time (Cercignani et al., 2018). However, to ensure that quantitative maps are reproducible, mapping methods must produce estimates with low variance and bias.

Conventionally, quantitative maps are estimated by fitting a known signal model to every voxel of a series of weighted images with varying contrast settings. The Maximum Likelihood Estimator (MLE) is a popular statistical method used to estimate parameters of a probability density by maximizing the likelihood that a signal model explains the observed data and is extensively used in quantitative mapping (Ramos-Llorden et al., 2017; Smit et al., 2013; Sijbers and Den Dekker, 2004). Usually, MLE methods estimate parameters independently for each voxel. This may lead to high variability for scans with low signal-to-noise ratio (SNR). Spatial regularization can be added to the MLE (referred to as the Maximum

* Corresponding author.
E-mail address: e.ribeirosabidussi@erasmusmc.nl (E.R. Sabidussi).

https://doi.org/10.1016/j.media.2021.102220
1361-8415/© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

a Posteriori (MAP) estimation) to enforce spatial smoothness, but demands high domain expertise. Additionally, for most signal models, MLE/MAP methods require an iterative non-linear optimization, which is relatively slow for clinical applications and might demand complex algorithm development.

Despite the current success of deep learning methods in the medical field, their application to Quantitative MRI (QMRI) is still affected by the lack of large in-vivo training datasets. Specifically in MR relaxometry, the use of neural networks is still limited. Previous works successfully applied deep learning in cardiac MRI (Jeelani et al., 2020) and knee (Liu et al., 2019), but they required the scans of many subjects to train the networks and were dependent on alternative mapping methods to generate training labels. This limitation was addressed in Cai et al. (2018) and Shao et al. (2020) by using the Bloch equations to generate simulated data to train convolutional neural networks in T_1 and T_2 mapping. However, estimation precision, a central metric in QMRI, was not reported. It is unclear, therefore, how well these methods would perform with noisy in-vivo data.

In this paper, we propose a new framework for MR relaxometry based on the Recurrent Inference Machines (RIMs) (Putzky and Welling, 2017). RIMs employ a recurrent convolutional neural network (CNN) architecture and, unlike most CNNs, learn a parameter inference method that uses a signal model, rather than a direct mapping between input signal and estimates. This hybrid framework combines the advantages of both data-driven and model-based methods, and, we hypothesize, is a promising tool for QMRI.

Previously, RIMs were used to solve linear inverse problems to reconstruct undersampled MR images (Lønning et al., 2019) and radio astronomy images (Morningstar et al., 2019). In both works, synthetic, corrupted training signals (i.e. images) were generated from high-quality image labels using the forward model.

A significant limitation on the use of deep learning in MR relaxometry is the lack of large publicly available datasets. The acquisition of in-vivo data is a costly and time consuming process, limiting the size of training datasets and reducing flexibility in terms of the pulse sequence and scanning parameters. Using a model-based strategy for data generation (in contrast to costly acquisitions) allows the creation of arbitrarily large training sets, where observational effects (e.g., acquisition noise, undersampling masks) and fixed model parameters are drawn from random distributions. This represents an essential advantage over other methods that rely entirely on acquired data. Yet, the lack of high-quality training labels (i.e. ground-truth T_1 and T_2 maps) limits the variability of training signals. Here, we also generate synthetic training labels to achieve sufficient variation in the training set.

We compared the proposed framework with an MLE method and an implementation of the Residual Neural Network (He et al., 2016) as a baseline for conventional deep learning QMRI methods. In contrast to MLE methods with user-defined prior distribution to enforce tissue smoothness, the RIM learns the relationship between neighboring voxels directly from the data, making no assumptions about the prior distribution of values. This might improve mapping robustness to acquisition noise.

We evaluated each method in terms of the precision and accuracy of measurements. First, noise robustness was assessed via Monte Carlo experiments with a simulated data set with varying noise levels. Second, we evaluated the quantitative maps' quality concerning each method's ability to retain small structures within the brain. Third, the precision and accuracy in real scans were evaluated via a test-retest experiment using a hardware phantom. Lastly, we used in-vivo scans to evaluate precision in a test-retest experiment with two healthy volunteers.

2. QMRI framework

2.1. Signal modeling

Let $\boldsymbol{\kappa}$ be the parameter maps to be inferred, such that $\boldsymbol{\kappa}(\boldsymbol{x}) \in \mathbb{R}^Q$ is a vector containing Q tissue parameters of a voxel indexed by the spatial coordinate $\boldsymbol{x} \in \mathbb{N}^D$. Then, we assume that the MRI signal in each voxel of a series of N weighted images $\boldsymbol{S} = \{S_1, \dots, S_N\}$ follows a parametric model $f_n(\boldsymbol{\kappa}(\boldsymbol{x})) : \mathbb{R}^Q \mapsto \mathbb{R}$ so

$$S_n(\boldsymbol{x}) = f_n(\boldsymbol{\kappa}(\boldsymbol{x})) + \epsilon_n(\boldsymbol{x}), \tag{1}$$

where $n = \{1, \dots, N\}$ indexes the image in the set and $\epsilon_n(\boldsymbol{x})$ is the noise at voxel $\boldsymbol{x}$.

For images with SNR larger than three, the acquired signal at position $\boldsymbol{x}$ can be well described by a Gaussian distribution (Sijbers et al., 1998; Gudbjartsson and Patz, 1995), with probability density function denoted by $p(S_n(\boldsymbol{x}_m) \mid f_n(\boldsymbol{\kappa}(\boldsymbol{x}_m)),\ \sigma)$, where $m \in \{1, \dots, M\}$ is the voxel index, M the number of voxels within the MR field-of-view and σ is the standard deviation of the noise.

2.2. Quantitative mapping

2.2.1. Regularized maximum likelihood estimator

The Maximum Likelihood Estimator (MLE) is a statistical method that infers parameters of a model by maximizing the likelihood that the model explains the observed data. Because the MLE is asymptotically unbiased and efficient (it reaches the Cramer-Rao lower bound for a large number of weighted images) (Swamy, 1971), it was chosen as the reference method for this study.

Assume $P(\boldsymbol{S}|\boldsymbol{f}(\boldsymbol{\kappa}), \sigma)$ is the joint probability density function (PDF) of all independent voxels in $\boldsymbol{S}$ from which a negative log-likelihood function $L(\boldsymbol{\kappa}, \sigma|\boldsymbol{S})$ is defined. Additionally, let $\Psi(\boldsymbol{\kappa})$ be the log of a prior probability distribution over $\boldsymbol{\kappa}$, introduced to enforce map smoothness. Then the ML estimates $\hat{\boldsymbol{\kappa}}$ are found by solving

$$\hat{\boldsymbol{\kappa}} = \underset{\boldsymbol{\kappa}}{\operatorname{argmin}}\ L(\boldsymbol{\kappa}, \sigma|\boldsymbol{S}) + \Psi(\boldsymbol{\kappa}), \tag{2}$$

in which we assume that σ can be estimated by alternative methods and is, therefore, not optimized.

Note that, although Eq. (2) strictly defines an MAP estimator, we choose to use the term regularized MLE to emphasize that $\Psi(\boldsymbol{\kappa})$ is only applied to promote maps that vary slowly in space. In this work, regularization is used to encourage spatial smoothness of the inversion efficiency map (i.e. B_1 inhomogeneity), while maps linked to proton density and tissue relaxation times are not regularized and their estimation occurs exclusively at the voxel level. Herein, we refer to this method simply as MLE.

2.2.2. ResNet

The Residual Neural Network (ResNet) is a type of feed-forward network that learns to directly map input data to training labels using a concatenation of convolutional layers. It was developed by He et al. (2016) as a solution to the degradation problem that emerges when building deep models (He and Sun, 2014). Skip connections between layers of the network allow the ResNet to fit to the residual of the signal, rather than to the original input, making identity learning simpler, and ensuring that a deeper network will not perform worse than its shallower counterpart in terms of training accuracy (He et al., 2016). For that reason, and because it was shown to be a suitable method for QMRI (Cai et al., 2018), we chose the ResNet as the reference deep learning method for this study.

Let $\Lambda_\phi : \mathbb{R}^N \mapsto \mathbb{R}^Q$ represent a ResNet model for QMRI, parameterized by ϕ, that maps the acquired signal $\boldsymbol{S}$ to tissue parameters $\boldsymbol{\kappa}$, specifically $\hat{\boldsymbol{\kappa}} = \Lambda_\phi(\boldsymbol{S})$. The learning task is to find a model $\Lambda_{\hat{\phi}}$

such that the difference between $\hat{\boldsymbol{\kappa}}$ and $\boldsymbol{\kappa}$ is minimal in the training set, that is

$$\hat{\phi} = \arg\min_{\phi} \left\| \boldsymbol{\kappa} - \Lambda_{\phi}(\boldsymbol{S}) \right\|_2^2. \tag{3}$$

3. The recurrent inference machine: a new framework for QMRI

In the context of inference learning (Chen et al., 2015; Zheng et al., 2015), the Recurrent Inference Machine (RIM) (Putzky and Welling, 2017) framework was conceived to mitigate limitations linked to the choice of priors and optimization strategy. By making them implicit within the network parameters, the RIM jointly learns a prior distribution of parameters and the inference model, unburdening us from selecting them among a myriad of choices.

With this framework, Eq. (2) is solved iteratively, in an analogous way to a regularized gradient-based optimization method. The RIM uses the gradients of the likelihood function to enforce the consistency of the data and to plan efficient parameter updates, speeding up the inference process. Additionally, because this framework is based on a convolutional neural network, it learns and exploits the neighborhood context, providing an advantage over voxel-wise methods. Note that, rather than explicitly evaluating $\Psi(\boldsymbol{\kappa})$, the RIM learns it implicitly from the labels in the training data set.

At a given optimization step $j \in \{0, \ldots, J-1\}$, the RIM receives as input the current estimate of the signal model parameters, $\hat{\boldsymbol{\kappa}}_j$, the gradient of the negative log-likelihood L with respect to $\boldsymbol{\kappa}$, $\nabla_{\boldsymbol{\kappa}}$, and a vector of memory states $\boldsymbol{h}_j$ the RIM can use to keep track of optimization progress and perform more efficient updates. The network outputs an update to the current estimate and the memory state to be used in the next iteration. The update equations for this method are given by

$$\left\{ \Delta\hat{\boldsymbol{\kappa}}_j, \boldsymbol{h}_{j+1} \right\} = \boldsymbol{g}_{\gamma}\left(\hat{\boldsymbol{\kappa}}_j,\ \nabla_{\boldsymbol{\kappa}},\ \boldsymbol{h}_j \right), \tag{4}$$

$$\hat{\boldsymbol{\kappa}}_{j+1} = \hat{\boldsymbol{\kappa}}_j + \Delta\hat{\boldsymbol{\kappa}}_j, \tag{5}$$

where $\Delta\hat{\boldsymbol{\kappa}}_j$ is the output of the network and denotes the incremental update to the estimated maps at optimization step j and $\boldsymbol{g}_{\gamma}$ represents the neural network portion of the framework, called RNNCell, parameterized by γ. A diagram of the RIM is shown on the left of Fig. 1a.

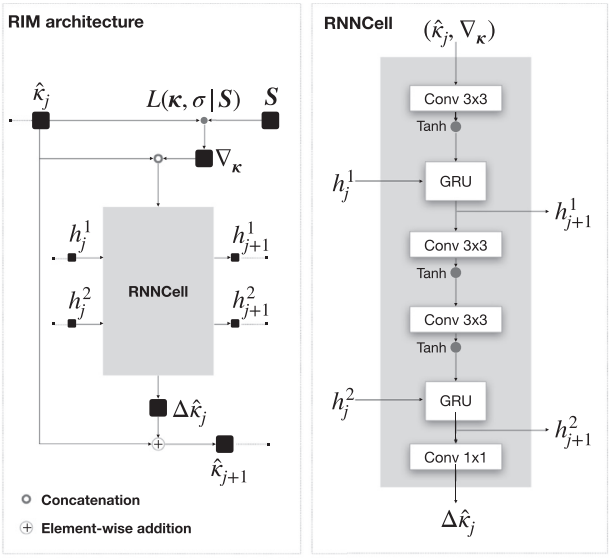
Predictions are compared to a known ground-truth and losses are accumulated at each step, with total loss given by

$$\hat{\gamma} = \arg\min_{\gamma} \frac{1}{J} \sum_{j=0}^{J-1} \left\| \boldsymbol{\kappa} - \hat{\boldsymbol{\kappa}}_{j+1} \right\|_2^2 \tag{6}$$

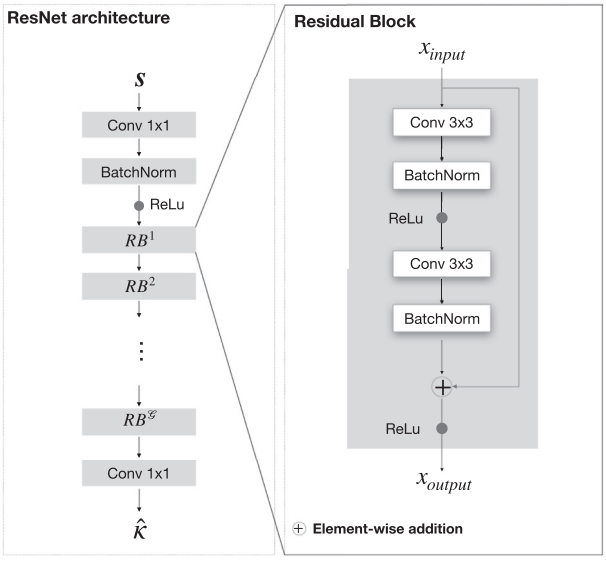
where J is the total number of optimization steps and $\hat{\gamma}$ is the optimal inference model given the training data.

It is important to notice that the RIM uses two distinct loss functions. The likelihood function $L(\boldsymbol{\kappa}|\boldsymbol{S}, \sigma)$ is used to provide the gradient $\nabla_{\boldsymbol{\kappa}}$ to the network and is evaluated in the data input domain (i.e. weighted images). In contrast, Eq. (6) is used to update the network parameters γ, and is evaluated in the parametric map domain (e.g. T_1 or T_2 relaxation maps).

A relevant feature of this framework is that the architecture of the RNNCell, more specifically, the number of input features in the first convolutional layer, only depends on Q, and not on N. This means that RIMs can process series of weighted images $[\boldsymbol{S}_n]$ for $\forall N > 0$.



(a) Recurrent Inference Machine framework



(b) Residual Neural Network architecture

Fig. 1. a) The RIM architecture in detail. The general RIM framework is shown on the left. Dashed lines indicate information passed through iterations. The RNNCell detail is shown on the right of a). Memory states h^*_{j+1} are passed to the next iteration step and used within the Gated Recurrent Units (GRU) to control the relevant information to be used from previous iterations. b) The ResNet architecture is composed of a concatenation of $\mathcal{G}$ residual blocks.

4. Methods

4.1. Sequences and parametric models

The choice of parameters $\boldsymbol{\kappa}$ and the form of the parametric model f_n depend on the pulse sequence used for acquisition.

For the T_1 mapping task in this work, we used the CINE sequence (Atkinson and Edelman, 1991), based on a (popular) fast T_1 quantification method (Look and Locker, 1970). It uses a non-selective adiabatic inversion pulse, applied after the cardiac trigger with zero delay. The heart beat was simulated at a constant

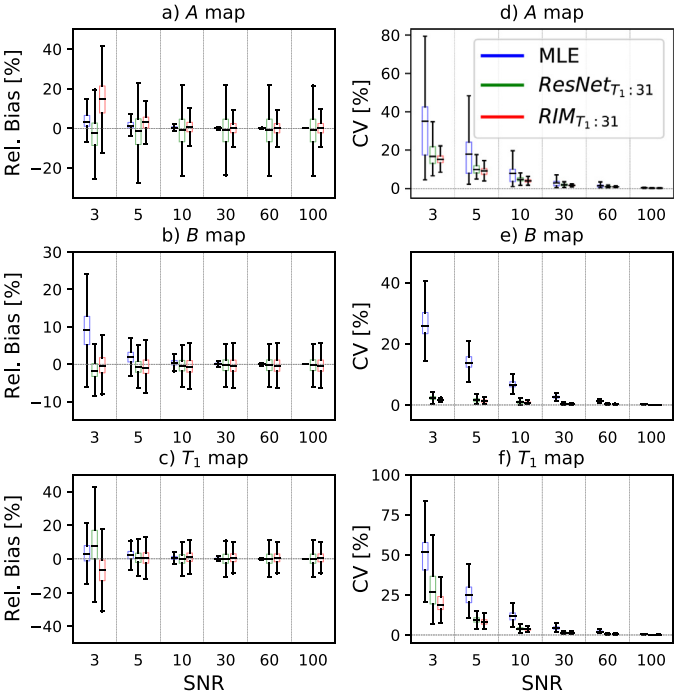


Fig. 2. Results of the Monte Carlo experiment with a T_1w simulated data set for varying *SNR* levels. a), b) and c) show the Relative Bias for the estimated A, B, and T_1 maps compared to simulated ground-truth. Figures d), e) and f) shows the Coefficient of Variation for the same maps. The boxplot represents the distribution of the metric over all pixels in the brain mask. The box extends from the lower to upper quartile values of this data, with a line at the median. The whiskers extend from the box to show the minimum and maximum values for each metric within the brain mask. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

rate of 100 beats per minute using a pulse generator developed in-house. Note that this sequence was originally intended for cardiac imaging, and it was designed to maximize the number of inversion times within a single ECG R-R cycle. Although this might benefit the fitting process (i.e. more points to fit), the scanner operator has limited control over the inversion times used, as well as the number of contrast images acquired.

For this sequence, a common parametric model is given by $f_n(\boldsymbol{\kappa}(\boldsymbol{x}_m)) = \left|A\left(1 - Be^{-\frac{\tau_n}{T_1}}\right)\right|$, where τ_n is the nth inversion time and $\boldsymbol{\kappa}(\boldsymbol{x}_m) = (A, B, T_1)^T$ is the tissue parameter vector at position $\boldsymbol{x}_m$, in which A is a quantity proportional to the proton density and receiver gain, B is linked to the efficiency of the inversion pulse and T_1 is the longitudinal relaxation time. The operator $|\cdot|$ represents the element-wise modulus.

T_2 scans were performed with a T_2 prepared 3D Fast Spin-Echo sequence (Mugler, 2014). In our experiments, we used $f_n(\boldsymbol{\kappa}(\boldsymbol{x}_m)) = \left|Ae^{-\frac{\tau_n}{T_2}}\right|$ as forward model, where τ_n is the nth T_2 preparation time and $\boldsymbol{\kappa}(\boldsymbol{x}_m) = (A, T_2)^T$, with A proportional to the proton density and receiver gain and T_2 the transverse relaxation time.

4.2. Generation of simulated data for training

In this work, we opted to generate training data via model-based simulation pipeline. Training samples were composed of ground truth tissue parameters $\boldsymbol{\kappa}$ and their corresponding set of simulated weighted images $\boldsymbol{S}$. To generate training samples with a spatial distribution that resembles the human brain, ten 3D virtual brain models from the BrainWeb project (Cocosco et al., 1997) were selected. We randomly extract 2D patches from the brain models during training, with patch centers drawn uniformly from the model's brain mask. To introduce the notion of uniform tissue properties within subjects but distinct between subjects, for each patch and tissue separately, the parameters in $\boldsymbol{\kappa}$ were drawn from a normal distribution with values given in Table 1. To enable recovery of intra-tissue variation, voxel-wise Gaussian noise was added to each parameter in $\boldsymbol{\kappa}$, except for B. Because the B value is related to the efficiency of the inversion pulse in inversion recovery (IR) sequences, it is not tissue-specific, and as such, cannot be modeled as above. Its value was simulated as $2 - \Gamma$, where Γ is independently sampled, per patch, from the half-normal distribution (Leone et al., 1961) with standard deviation $\sigma^{\Gamma} = 0.2$.

Using $\boldsymbol{\kappa}, \boldsymbol{S}$ was simulated via Eq. (1), with $\epsilon(\boldsymbol{x})$ an independent zero mean Gaussian noise where, for each patch, standard deviation $\sigma^{\text{acquisition}}$ was drawn from a log-uniform distribution with values in the range $[0.0065, 0.255]$, corresponding to SNR levels in the range of 100 to 3, respectively.

Table 1
Distribution of parameters per tissue and tissue property. T_1 and T_2 values in milliseconds. Values for A are chosen as a fraction of the concentration of protons in the CSF.

Tissue	$\mu^{T_1}_{\text{tissue}}$	$\sigma^{T_1}_{\text{tissue}}$	$\mu^{T_2}_{\text{tissue}}$	$\sigma^{T_2}_{\text{tissue}}$	μ^{A}_{tissue}	$\sigma^{A}_{\text{tissue}}$
CSF	3500†	300	2000†	300	1.0	0.3
GM	1400†	300	110†	30	0.85	
WM	780†	250	80†	20	0.65	
Fat	420†	100	70†	20	0.9	
Muscle	1200†	300	50†	20	0.7	
Muscle skin	1230†	300	50†	20	0.7	
Skull	400‡	100	30±	10	0.9	
Vessels	1980±	300	275±	70	1.0	
Marrow	580∓	100	50†	20	0.8	

† Bojorquez et al., 2017. ‡ Chen et al. (2016). ± Stanisz et al. (2005). ∓ de Bazelaire et al. (2004).

4.3. Evaluation datasets

We performed all scans on a 3T General Electric Discovery MR750 clinical scanner (General Electric Medical Systems, Waukesha, Wisconsin) with a 32-channel head coil.

4.3.1. Hardware phantom

Phantom scans were carried out using the NIST/ISMRM system phantom (Keenan et al., 2017) with parameters for the acquisition of T_1 weighted (T_1w) and T_2 weighted (T_2w) images presented in Table 2 (datasets HP_{T_1} and HP_{T_2}, respectively). The acquisition field-of-view (FOV) contained the phantom's T_1 array for T_1w scans and the T_2 array for T_2w scans. To evaluate the repeatability of each mapping method, $C = 4$ consecutive acquisitions were performed without moving the phantom and with minimal time interval between scans.

4.3.2. In-vivo

Our Institutional Review Board approved the volunteer study and informed consent was obtained from 2 healthy adults. $C = 2$ repeated scans per volunteer were acquired for both T_1 and T_2 experiments to evaluate repeatability with in-vivo data. The FOV used was similar for T_1 and T_2 experiments and was oriented in the axial direction, with the middle slice positioned at the level of the body of the corpus callosum. These datasets, acquired with a slice thickness of 3mm, are referred to as IV_{T_1} and IV_{T_2}, respectively. Details on acquisition settings are given in Table 2. Finally, to evaluate the performance of the estimators under low SNR conditions, we repeated the T_1w acquisition using a slice thickness of 1.5mm (data set called $IV^{\text{noisy}}_{T_1}$), in which a single slice, positioned above the corpus callosum, was acquired. Again, $C = 2$ repeated scans were acquired for each volunteer to assess each method's repeatability.

Table 2
Acquisition settings for the evaluation datasets. *HP* denotes the phantom scans while *IV* are the in-vivo scans.

data set	HP_{T_1}	IV_{T_1}	$IV^{noisy}_{T_1}$	HP_{T_2}	IV_{T_2}
FOV (pixel)	210x210x15	210x210x10	210x210x1	210x210x15	210x210x10
Slice thickness (mm)	1.5	3.0	1.5	1.5	3.0
Spacing (mm)	1.5	1.5	-	1.5	1.5
In-plane voxel size (mm)			0.82		
Repetition Time (ms)		8192		2010, 2020, 2040, 2080, 2160, 2320	
τ (Preparation time) (ms)		4		10, 20, 40, 80, 160, 320	
τ (Inversion Times) (ms)	**23 TIs**: 172, 204, 237, 270, 303, 335, 368, 401, 434, 467, 499, 532, 565, 598, 630, 663, 696, 729, 761, 794, 827, 860, 893	**31 TIs**: 139, 166, 193, 219, 246, 272, 299, 325, 352, 379, 405, 432, 458, 485, 511, 538, 565, 591, 618, 644, 671, 697, 724, 751, 777, 804, 838, 857, 883, 915, 937	**25 TIs**: 172, 204, 237, 270, 303, 335, 368, 401, 434, 467, 499, 532, 565, 598, 630, 663, 696, 729, 761, 794, 827, 860, 893, 925, 958		-
Flip Angle (°)		10			-
Acceleration factor			2		
C (nr. of repeated scans)	4	2	2	4	2
Acq. time/scan (min)	4.3	7.5	1.6	3.2	3.2

4.4. Implementation details

The codes for all methods, trained models and the data used in the experiments are available online[1].

4.4.1. MLE

To promote smoothness of the B map, the prior $\Psi(\boldsymbol{\kappa})$ was set as the Laplace operator, given by the sum of the second spatial derivatives in every dimension. A weighting term λ_B is introduced to control the strength of the regularization. With $\lambda_B = 0$, the field is estimated voxel-wise with high accuracy, at the cost of precision. When $\lambda_B = \infty$, the field is constant, with value equal to some weighted average of the true B value over the FOV. The optimal λ_B value is partly dependent on the noise level of the scans, thus it might vary between datasets.

Here, we selected $\lambda_B = 500$ based on experiments with data set IV_{T_1} (not shown). We tested multiple λ_B values (0, 1, 5, 10, 50, 100, 500, 1000) and chose the one that produced T_1 maps with the lowest number of outliers. This value was used for all T_1 mapping experiments. The remaining A, T_1 and T_2 maps were not regularized.

To prevent the estimator from getting stuck in a local minimum far from the optimal target, we initialize $\boldsymbol{\kappa}$ via an iterative linear search within a pre-specified range of values per parameter. Following initialization, parameters are estimated with a non-linear trust region optimization method. The estimation pipeline was implemented in MATLAB with in-house custom routines (Poot and Klein, 2015).

4.4.2. Network training

To train both neural networks, 7200 2D patches of size 40×40 per brain model were generated during training and arranged in mini-batches of 24 samples, for a total of 3000 training iterations. Training times were approximately 11h for T_1 RIM models, 9h for T_1 ResNet models, 7h for T_2 RIM and 6h for the T_2 ResNet model.

We used the ADAM optimizer with an initial learning rate of 0.001 and set the initial network weights with the Kaiming initialization (He et al., 2015). PyTorch 1.3.1 was used to implement and train the models. The networks were trained on a GPU Nvidia P100, and all experiments (including timing) were performed on an Intel Core i5 2.7 GHz CPU.

4.4.3. ResNet architecture

Our implementation of the ResNet is a modified version of He et al. (2016). Pooling layers were removed to ensure limited influence between distant regions of the brain, effectively enforcing the use of local spatial context during inference. Additionally, our ResNet does not contain fully connected layers to adapt the network for a voxel-wise regression problem. All convolutions are zero-padded to maintain the patch size.

The first convolutional layer has a 1×1 filter, and it is used to increase the number of features from N (the number of weighted images) to 40. This layer is followed by a batch normalization (BatchNorm) layer and a ReLu activation function. The core component of the network, denoted as the residual block (RB), comprises two 3×3 convolutional layers, two BatchNorm layers, and two ReLu activations, arranged as depicted on the right of Fig. 1b. Within a given RB, the number of features in each convolutional layer is the same. The skip connection is characterized by the element-wise addition between the input and the output of the second BatchNorm layer. In total, $\mathcal{G} = 12$ residual blocks are sequentially linked, with the number of feature channels in each block empirically chosen as $[40, 40, 80, 80, 160, 320, 160, 80, 80, 40, 6]$. The network architecture is completed by one 1×1 convolutional filter, used to reduce the number of features to Q. Details on the general architecture are presented on the left of Fig. 1b.

Note that, due to differences in the inversion times used for the acquisition of T_1 weighted datasets (Table 2), we trained three ResNet models for the T_1 mapping task: (1) Training data set generated with $N = 23$ inversion times $\text{ResNet}_{\boldsymbol{T_1}:\mathbf{23}}$), (2) with $N = 25$ inversion times $\text{ResNet}_{\boldsymbol{T_1}:\mathbf{25}}$) and (3) with $N = 31$ inversion times $\text{ResNet}_{\boldsymbol{T_1}:\mathbf{31}}$). Finally, a fourth model was trained on the T_2 mapping task, denoted as $\text{ResNet}_{\boldsymbol{T_2}}$, with $N = 6$ echo times.

4.4.4. RIM architecture

The RNNCell (shown in detail on the right of Fig. 1a) is composed of four convolutional layers and two Gated Recurrent Units (GRUs). The first 3×3 convolutional layer is followed by a hyperbolic tangent (*tanh*) link function, and its output, with 36 feature channels, is passed to the first GRU, which produces 36 output channels. The output of this unit ($\boldsymbol{h}^1_{j+1}$), also used as the first memory state, goes through two 3×3 convolutional layers with 36 output features, each followed by a *tanh* activation. The data then passes through a second GRU, which generates the second memory state $\boldsymbol{h}^2_{j+1}$. The last layer is a 1×1 convolutional layer used to reduce the dimensionality of the feature channels, and it outputs Q features, corresponding to the number of tissue parameters in $\boldsymbol{\kappa}$. All convolutional layers are zero-padded to retain the original image size.

[1] https://gitlab.com/e.ribeirosabidussi/emcqmri_relaxometry

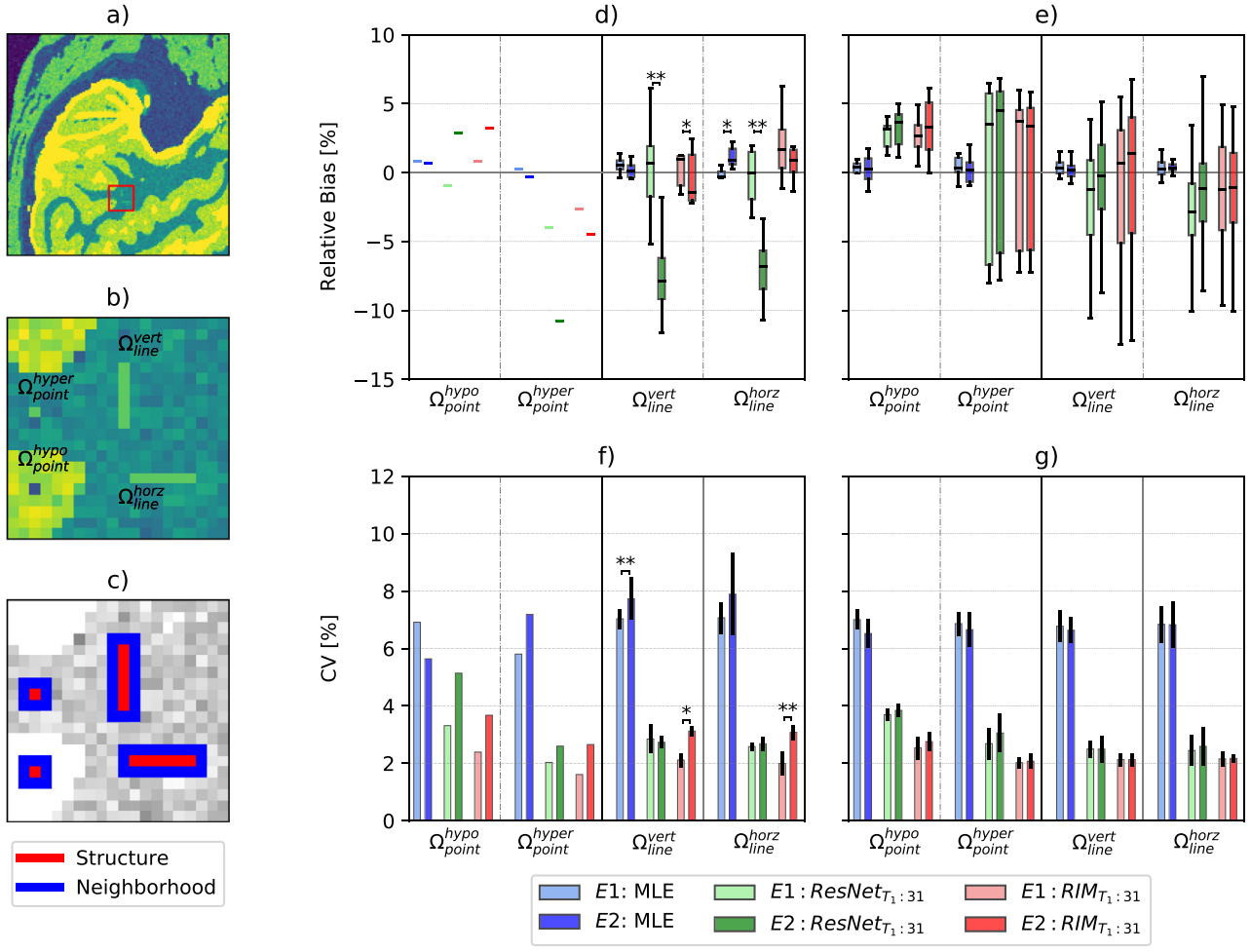


Fig. 3. Evaluation of image blurriness in terms of Relative Bias and CV. a) Ground-truth T_1 map used to generate the weighted images S. The red box indicates the position of the simulated artefacts. b) The four simulated structures. c) Representation of the areas of interest. The blue areas are the structures, and red areas are their immediate neighborhood. d) Relative Bias over one hundred repetitions within the Structure region. e) Relative Bias over one hundred repetitions within the Neighborhood region. f) CV over one hundred repetitions within the Structure region. g) CV over one hundred repetitions within the Neighborhood region. In all plots, the box extends from the lower to upper quartile values of the data, with a line at the median. The whiskers extend from the box to show the range of the data. The vertical black lines at the top of the bars (plots f) and g)) show the standard deviation over the data. Significant differences between scenarios $E1$ and $E2$ are indicated by * and **, representing $p < 0.05$ and $p < 0.01$, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The parameter vector $\hat{\boldsymbol{\kappa}}$ was initialized as $A = \text{MIP}(\boldsymbol{S})$, $B = 2$, $T_1 = 1000$ *ms* and $T_2 = 100$ *ms*, where MIP is the Maximum Intensity Projection per voxel over all weighted images in the set. We used $J = 6$ optimization steps for all RIM models.

Similarly to the ResNet, we trained three RIM models on the T_1 mapping ($\text{RIM}_{\boldsymbol{T_1}:\boldsymbol{23}}$, $\text{RIM}_{\boldsymbol{T_1}:\boldsymbol{25}}$, and $\text{RIM}_{\boldsymbol{T_1}:\boldsymbol{31}}$) and one model on the T_2 task ($\text{RIM}_{\boldsymbol{T_2}}$). Notice that, while all T_1 datasets could be processed by a single RIM model, as the number of input features in the first convolutional layer does not depend on N, slight variations in inversion times might affect estimation error. This aspect will be assessed in Section 5, as it supplies information on the RIM's generalizability.

4.5. Quantitative evaluation

The prediction accuracy was evaluated in terms of the Relative Bias between the reference parameter values $\boldsymbol{\kappa}$ and the estimated parameters $\hat{\boldsymbol{\kappa}}^c \in \{\hat{\boldsymbol{\kappa}}^1, \ldots, \hat{\boldsymbol{\kappa}}^C\}$ for each repeated experiment c, defined as

$$\text{Relative Bias } [\%] = \frac{1}{C}\sum_{c=1}^{C}\left[\left(\hat{\boldsymbol{\kappa}}^c - \boldsymbol{\kappa}\right) \oslash \boldsymbol{\kappa}\right] \times 100\%, \tag{7}$$

where C is the number of repeated experiments and $\oslash$ denotes the element-wise division. The Coefficient of Variation (CV) was used to measure the repeatability of the predictions, and it is given by

$$\text{CV } [\%] = \left(\text{SD}^c\left(\hat{\boldsymbol{\kappa}}^c\right) \oslash \frac{1}{C}\sum_{c=1}^{C}\hat{\boldsymbol{\kappa}}^c\right) \times 100\%, \tag{8}$$

where SD^c denotes the standard deviation over C estimates $\hat{\boldsymbol{\kappa}}$.

5. Experiments

5.1. Simulated data set

5.1.1. Noise robustness

To assess each method's robustness to noise and mapping quality, we generated the simulated T_1w data with the process described in Section 4.2 using a 2D slice of a virtual brain model not included in the training, matrix size 256×256 and inversion times of data set IV_{T_1}.

For the same ground-truth T_1, A and B maps, $C = 100$ realisations of acquisition noise were simulated per $\text{SNR} \in [3, 5, 10, 30, 60, 100]$. The Relative Bias and CV were computed per pixel and their distribution over all pixels within a brain mask is

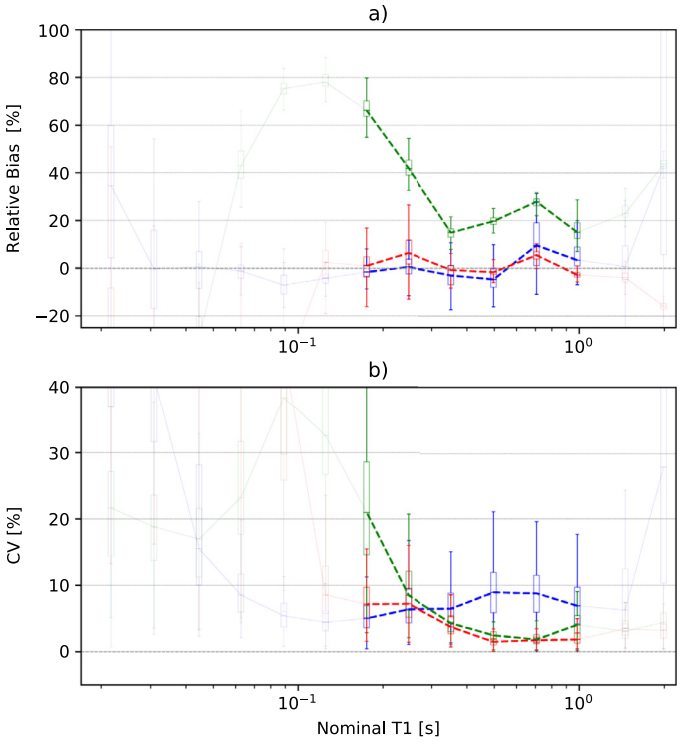


Fig. 4. Estimation of T_1 values in the ISMRM/NIST phantom. RIM results shown in red, MLE in blue and ResNet in green. a) Distribution of Relative Bias over all pixels within a ROI versus nominal T_1 values in the phantom. b) Box plot of the CV in the different spheres/ROIs of the phantom, plotted as a function of their nominal T_1 value. In both figures, the fully-coloured strokes indicate the spheres with T_1 values within the range of inversion times. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

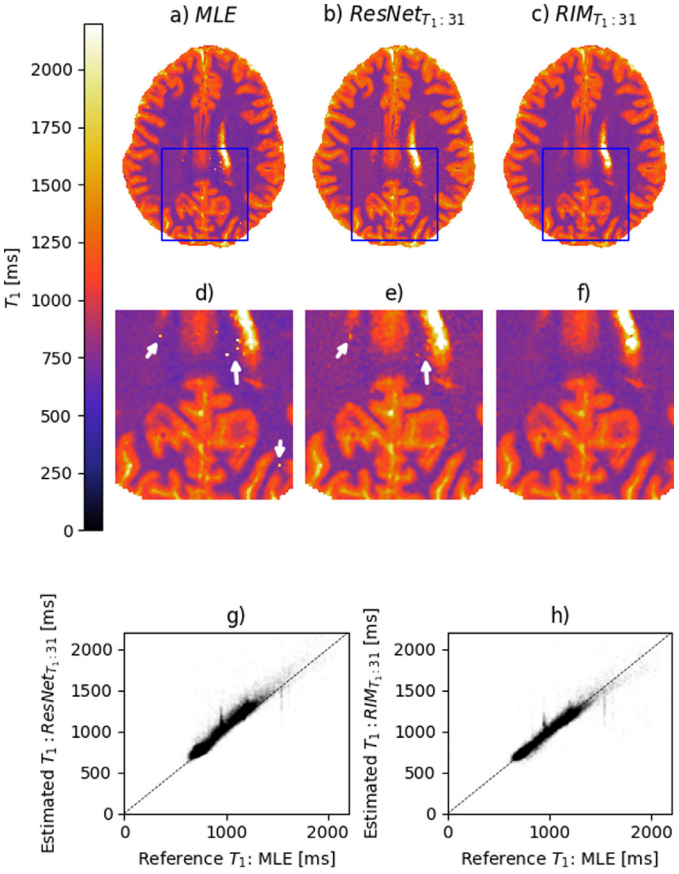


Fig. 5. T_1 maps estimated from the $\boldsymbol{IV}_{T_1}$ data set. Scan 1 of volunteer 1 is shown. a-c) T_1 maps generated by each mapping method and the detail (blue box) shown in figures d-f). The white arrows indicate estimation outliers. g) Agreement between the ResNet and MLE and h) RIM and MLE. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

shown. The models RIM$_{\boldsymbol{T_1}:\mathbf{31}}$ and ResNet$_{\boldsymbol{T_1}:\mathbf{31}}$ were used in this experiment.

5.1.2. Blurriness analysis

We assessed the quality of the quantitative maps in terms of blurriness. Here, we defined blurriness as the amount of error introduced to a pixel, in terms of Relative Bias and CV, due to the influence of its neighbors and vice-versa. In this experiment, our interest lies on how well each mapping method can preserve the true T_1 value in small structures (e.g. one pixel), specifically hypo- and hyper-intense regions that are at risk of being blurred away by the neural networks.

To simulate the presence of these small anatomical structures, we changed the T_1 value of selected pixels in a ground-truth T_1 map (Fig. 3a), described as follows: $\Omega^{\text{hypo}}_{\text{point}}$ is a hypo-intense pixel ($T_1 = 400ms$) within the gray mater of this map (shown in detail in Fig. 3b); $\Omega^{\text{hyper}}_{\text{point}}$ is a hyper-intense pixel ($T_1 = 1200ms$) within the white mater (WM); $\Omega^{\text{vert}}_{\text{line}}$ is a hyper-intense vertical line ($T_1 = 1200ms$) in the WM; and $\Omega^{\text{horz}}_{\text{line}}$ is a hyper-intense horizontal line ($T_1 = 1200ms$) also in the WM.

We measured the Relative Bias and CV per pixel in a Monte Carlo experiment with $C = 100$ noise realizations (SNR=10). Each metric's median and standard deviation are reported for two disjoint regions in the estimated T_1 map, referred to as Structure and Neighborhood (Fig. 3c). This scenario, containing simulated structures, is called $E2$, and was compared to the baseline error in the same regions in the original T_1 map (scenario $E1$). An independent t-test was applied to identify significant differences between $E1$ and $E2$. The models RIMRIM$_{\boldsymbol{T_1}:\mathbf{31}}$ and ResNetResNet$_{\boldsymbol{T_1}:\mathbf{31}}$ were used in this experiment.

5.2. Evaluation with hardware phantom

We manually drew Regions of Interest (ROIs) within every sphere in the phantom and calculated the Relative Bias and CV per pixel within each ROI for T_1 and T_2 tasks. Since nominal parameter values within the spheres, as reported by Keenan et al. (2017) and used as the reference $\boldsymbol{\kappa}$, include relaxation times shorter and longer than values normally found in brain tissues, we calculated the overall accuracy and repeatability as the average Relative Bias and CV over all pixels in spheres with parameter value in between the lowest and highest τ used for training (Table 2).

Because this data set was acquired with 23 inversion times, models RIM$_{\boldsymbol{T_1}:\mathbf{23}}$ and ResNet$_{\boldsymbol{T_1}:\mathbf{23}}$ were used.

5.3. Evaluation with in-vivo scans

To evaluate the precision of estimates from in-vivo data, we compared T_1 and T_2 maps from all methods in terms of pixel-wise CV for all in-vivo scans. We also performed a visual comparison of the maps.

We evaluated the mapping quality in in-vivo scans regarding the sharpness of the boundary between gray mater and white mater. Twenty lines perpendicular to the tissue interface (Fig. 7a) were manually drawn in the measured quantitative maps. For each line, linear interpolation was used to reconstruct the T_1 values along them and a sigmoid model, given by $y(x) = V/(1 + e^{-\upsilon(x-x_0)}) + b$, was fit using the mean squared error (MSE) as objective function. The parameter υ denotes the slope of the fitted

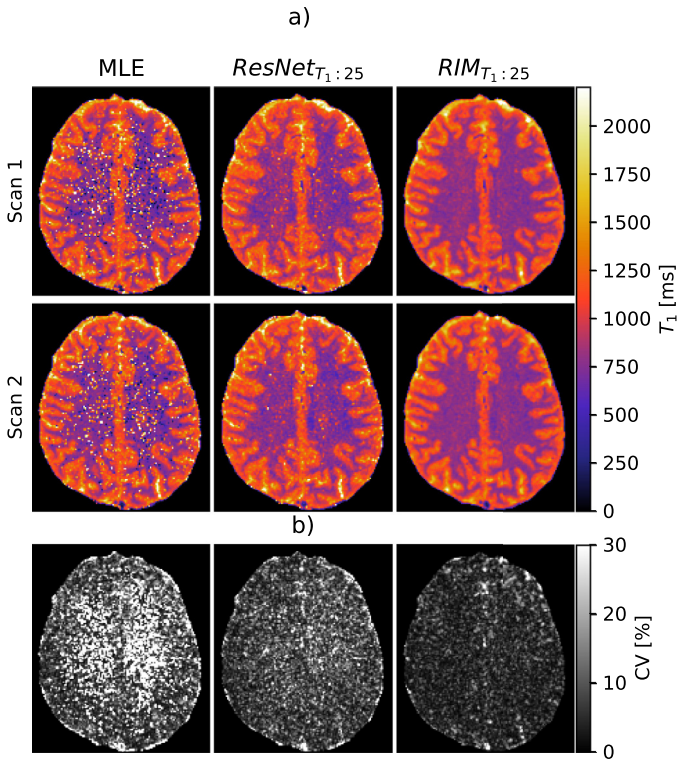


Fig. 6. T_1 maps estimated from the $IV^{noisy}_{T_1}$ data set a) T_1 maps estimated from volunteer 1 for repeated scans 1 and 2. b) Their respective pixel-wise CV map.

sigmoid and was used as a measure of boundary sharpness. A paired *t*-test was performed to evaluate significant differences between mapping methods.

5.4. Model generalizability

In this experiment, we evaluated how well the RIM can generalize to datasets with different acquisition settings, specifically, the variation of the inversion times in the three T_1w datasets. In contrast to the ResNet architecture, which depends on the number of weighted images in the series, the RIM can process inputs of any length.

We used the three RIM$_{T_1}$ models (RIM$_{\boldsymbol{T_1}:\mathbf{23}}$, RIM$_{\boldsymbol{T_1}:\mathbf{25}}$ and RIM$_{\boldsymbol{T_1}:\mathbf{31}}$) to infer T_1 maps from each T_1w data set, and computed the CV for the repeated experiments in each. The results were compared to the MLE and data set-specific ResNet models.

6. Results

6.1. Simulated data set

Fig. 2(a)–(c) show the Relative Bias measured for A, B and T_1 maps in the experiment with simulated T_1w data. For most cases where SNR > 3, all methods produced quantitative maps with comparable median Relative Bias, but both neural networks displayed a larger range of values than the MLE. The CV for all SNR levels is shown in Fig. 2(d)-(f) for the same data. The RIM presented lower CV than the other methods for all SNRs. In comparison, the MLE displayed significantly higher CV compared to RIM and ResNet, accentuated in low SNR. The results of the experiments with simulated T_2w data were similar and are shown in figure A1 of the Supplementary Results.

Fig. 3(d)–(g) show the results of the blurriness analysis. Specifically, Fig. 3(d) and (f) depict the Relative Bias and CV measured per pixel within the Structure area. We observe that both neural networks presented increased Relative Bias compared to scenario $E1$. For the RIM, the highest increase occurred for $\Omega^{\text{hypo}}_{\text{point}}$, with Relative Bias going from 0.68% to 3.43%. This difference represents an average error of 11ms over the ground-truth T_1 value of 400ms, or a loss of 0.81% in T_1 contrast between the pixel and its neighbors, with average T_1 of 1350ms. The ResNet showed considerably higher bias than RIM when small structures were added, while for the MLE, the difference between scenarios $E1$ and $E2$ is not significant (with exception for $\Omega^{\text{horz}}_{\text{line}}$). The RIM showed increased CV for all structures compared to the baseline, but values were still lower than the MLE's and comparable to the ResNet's. Figs. 3(e) and 3(g) show the Relative Bias and CV for the Neighborhood region. We observe higher Relative Bias for RIM and ResNet than the MLE, with a wider range of values, but we found no significant differences between $E1$ and $E2$ for any of the cases.

The average computing time to produce $\hat{\boldsymbol{\kappa}}$ from $N = 31$ weighted images (with size 256 $\times$ 256 pixels) was measured as 3.8s for the RIM$_{T_1:31}$, 27s for ResNet$_{T_1:31}$ and 575s for the MLE.

6.2. Evaluation with hardware phantom

The T_1 quantification results are shown in Fig. 4. In Fig. 4(a) we present the Relative Bias for the different spheres in the phantom. The average Relative Bias was computed over the spheres in the restricted τ domain (full-color lines), in which the RIM$_{\boldsymbol{T_1}:\mathbf{23}}$ model shows lower error (1.34%) compared to the MLE (1.71%) and ResNet$_{\boldsymbol{T_1}:\mathbf{23}}$ (31.06%). The CV as a function of T_1 values is shown in Fig. 4(b). The average CV over the restricted τ domain was measured as 3.21% for RIM$_{\boldsymbol{T_1}:\mathbf{23}}$, 7.56% for MLE and 7.5% for ResNet$_{\boldsymbol{T_1}:\mathbf{23}}$.

The results for the T_2 mapping task with the hardware phantom are shown in figure A2 of the Supplementary Results, where we observed larger Relative Bias for all methods.

6.3. Evaluation with in-vivo scans

The T_1 maps generated by each method for volunteer 1 in the low noise data set IV_{T_1} are shown in Fig. 5(a)–(c). We observe the presence of outliers in the MLE and ResNet$_{\boldsymbol{T_1}:\mathbf{31}}$ (white arrows in Fig. 5(d)–(f)), while the RIM$_{\boldsymbol{T_1}:\mathbf{31}}$ produced a clean T_1 map. The scatter plot in Fig. 5(h) shows that the RIM estimate is nearly unbiased when compared to the MLE's, while the ResNet presented overestimated T_1 values (Fig. 5(g)).

T_1 maps inferred from the noisier data set $IV^{\text{noisy}}_{T_1}$ are shown in Fig. 6(a). The RIM$_{\boldsymbol{T_1}:\mathbf{25}}$ showed increased noise robustness compared to the MLE and ResNet$_{\boldsymbol{T_1}:\mathbf{25}}$, clearly outperforming these methods in terms of outliers. The CV maps, computed per pixel, are presented in Fig. 6(b) and shows that the RIM$_{\boldsymbol{T_1}:\mathbf{25}}$ model produces low-variance quantitative maps, with average CV over all pixels equal to 6.4%, compared to 17.1% from the MLE and 11.06% from the ResNet$_{\boldsymbol{T_1}:\mathbf{25}}$.

Fig. 7(c) shows the result of the image quality analysis for in-vivo scans. The figure depicts the distribution of the sigmoid slope υ for each method across all 20 lines. The whiskers indicate the minimum and maximum υ values, the boxes show the lower and upper quartiles and the solid horizontal line their median. The paired *t*-test shows no significant differences between methods.

The results of the T_2 in-vivo experiments are shown in Fig. 8. Fig. 8(a)–(c) show the T_2 maps estimated by the MLE, ResNet and RIM, respectively. Differences between the ResNet and MLE are shown in Fig. 8(d) and between RIM and MLE in Fig. 8(e). The ResNet overestimated the T_2 values for most of the brain regions, with average difference of 32.2ms across all pixels in a brain mask (including CSF), while the average T_2 difference between the RIM and MLE was 2.36ms. Figures h-m) show the same maps for a wider range of T_2 values. In figures n) and o), the range of the

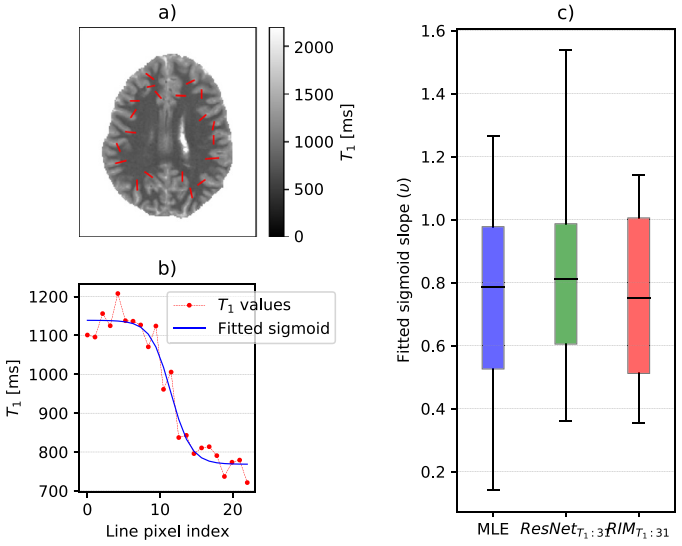


Fig. 7. Evaluation of the integrity of the GM/WM boundaries. a) Detail on the twenty lines were manually drawn perpendicular to the GM/WM interface indicated by the red lines. b) An example of the sigmoid fitting for one of the lines. c) The box plot depicts distribution of the absolute sigmoid slope (υ) for all 20 lines for each mapping method. We found no significant differences between the methods. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

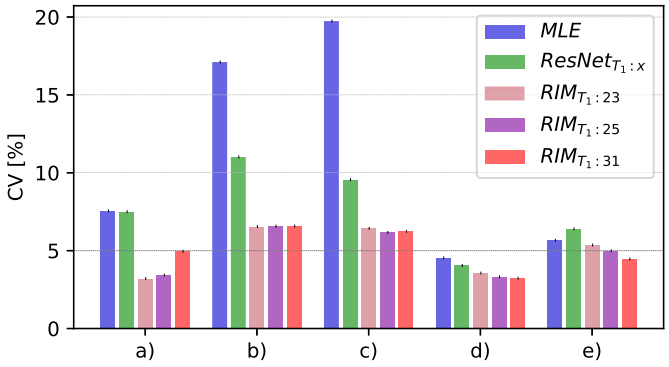


Fig. 9. Results of the model generalisability experiment. The 3 RIM models ($RIM_{T_1:23}$, $RIM_{T_1:25}$ and $RIM_{T_1:31}$) for T_1 mapping were used to estimate data from all datasets and compared to the results from MLE and ResNet. a) data set $\boldsymbol{HP_{T_1}}$ (23 TIs), b) data set $\boldsymbol{IV^{noisy}_{T_1}}$ **:** $\boldsymbol{Volunt.1}$ (25 TIs) c) data set $\boldsymbol{IV^{noisy}_{T_1}}$ **:** $\boldsymbol{Volunt.2}$ (25 TIs) d) data set $\boldsymbol{IV_{T_1}}$ **:** $\boldsymbol{Volunt.1}$ (31 TIs) e) $\boldsymbol{IV_{T_1}}$ **:** $\boldsymbol{Volunt.2}$ (31 TIs). The median CV over all pixels containing tissues of interest (phantom spheres or brain tissue) is shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

scatter plots f) and g) was extended to show T_2 CSF values. Neither neural network was able to correctly estimate T_2 values above 750 ms.

6.4. Model generalizability

Fig. 9 illustrates the CV of the different models evaluated on all T_1w datasets. The graph shows that the RIM produces estimates with lower variance than the MLE and ResNet, regardless of the number of inversion times used to create the training set. Note that, in every case, the RIM trained for the specific data performs slightly better than the other RIM models. However, we found no significant differences in repeatability between these models.

7. Discussion

This work presented a novel approach for MR relaxometry using Recurrent Inference Machines. Previous works showed that RIMs produce state-of-the-art predictions solving linear reconstruction problems. Here, we expanded the framework and demonstrated that it could be successfully applied to non-linear inference problems, outperforming a state-of-the-art Maximum Likelihood Estimator and a ResNet model in T_1 and T_2 mapping tasks.

In simulated experiments, we observed that the RIM reduces the variance of estimates without compromising accuracy, suggesting higher robustness to acquisition noise than the MLE, and attesting to the advantages of using the neighborhood context in the inference process. In addition, for low SNR, the RIM had lower variance than the ResNet, suggesting that the neighborhood context alone is not the sole responsible for the increased quality, and that the data consistency term (likelihood function) in the RIM framework helps to produce more reliable estimates. This showcases a major advantage of the RIM framework over current conventional and deep learning methods for QMRI.

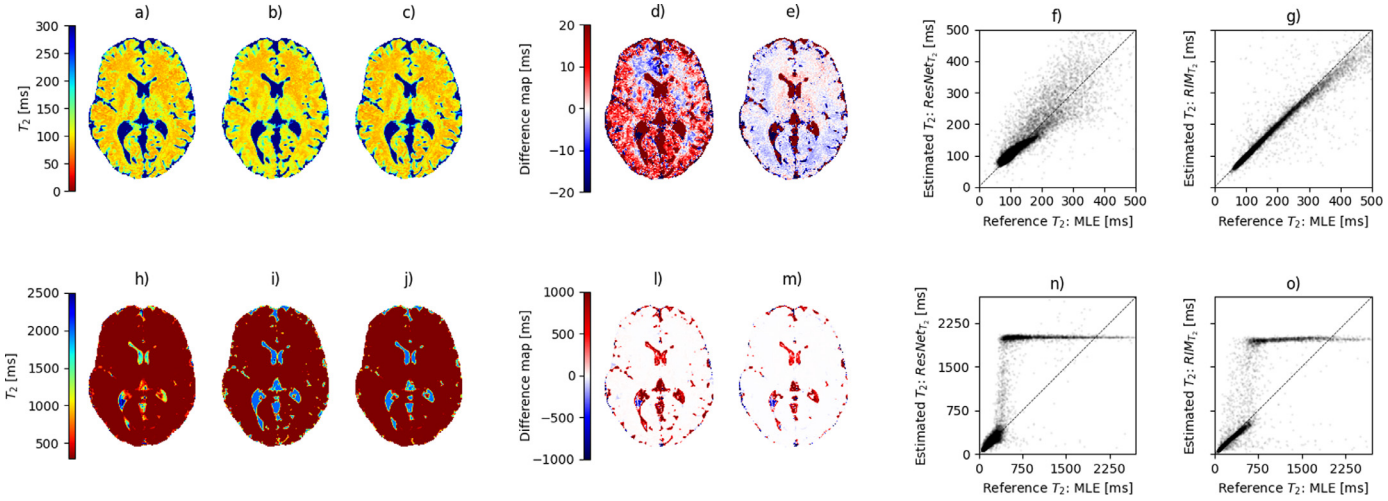


Fig. 8. T_2 mapping. Figures a-c) T_2 maps from data set IV_{T_2} for the MLE, ResNet and RIM, respectively. The same maps are shown in figures h-j) in a different scale to showcase differences in high T_2 values. Difference maps are shown in figures d) $ResNet_{T_2}$ - MLE_{T_2}, and e) RIM_{T_2} - MLE_{T_2}. The same difference maps are shown in figures l) and m), with a larger range of values to display CSF differences. Figures f) and g) show the agreement between $ResNet_{T_2}$ and MLE_{T_2}, and RIM_{T_2} and MLE_{T_2} for T_2 values $< 500ms$, respectively. The same plots are presented in figures n) and o) with extended range of T_2 values.

An important consideration when using learning-based methods is the time used for training. The RIM required 7 to 11 h of training, depending on the data set, until the training loss converged. Although it seem long, this process is performed only once, and the time to learn the models is offset by the inference speed of the neural networks. Running on a CPU, the MLE took approximately 110 h to process the entire simulated data set (700 samples), while the RIM took approx. 45 min.

The phantom experiments performed to assess the Relative Bias and CV in real, controlled scans showed that the RIM has the lowest Relative Bias among the evaluated methods. The ResNet presented significantly higher error, which indicates that it does not generalize well to unseen structures, and the use of simulated training data with this model should be carefully considered. Because the RIM can generalize well, using simulated data for training represents a significant advantage over models trained with real-data when considering data set flexibility, since any combination of parameter values can be simulated and the training data set can be arbitrarily large.

In the in-vivo T_1 mapping experiments, the RIM produces quantitative maps similar to those from the MLE, with higher robustness to noise. Although the ResNet estimates parametric maps consistent with reported T_1 relaxation time of brain tissues, these values are often overestimated compared to the MLE. Experiments with the noisy T_1 data set show that the RIM compensates the acquisition noise without blurring brain structures and tissue edges. This indicates a strong and effective learned prior. The MLE presented a high number of outliers, which drastically reduces precision.

Note, however, that we used the same λ_B for all MLE experiments. As the optimal regularization strength is dependent on the noise level, the chosen λ_B might not be optimal for all datasets, and the MLE precision might improve for different values. Additionally, as our goal was to demonstrate the advantages of the RIM's learned prior over a voxel-wise estimator, we chose not to impose a prior on the remaining quantitative parameters. We believe, however, that regularization of the T_1 and T_2 maps could improve the MLE precision, at the expense of additional bias.

Additionally, the CINE sequence was originally designed for cardiac imaging, and only allows limited control over the inversion times and number of images acquired. The use of this sequence for brain imaging should be cautiously considered, since, by design, the longest inversion times are on the order of one cardiac cycle ($\sim$ 1s), and might not cover the entire range of T_1 values in brain tissues.

The T_2 experiments demonstrated an important consideration of training the networks with simulated data. For $T_2 < 500$ ms, the RIM presented a strong agreement with the MLE, while the ResNet produced, in average, overestimated values. T_2 values above this range, however, were mapped to a narrow distribution around 2000ms, which corresponds to the mean T_2 value of CSF in the training set. We believe this behavior is due to the distribution of values used for training, which contained a gap between CSF and blood vessels ($T_2 = 275$ms). Therefore, we advise caution when generating training datasets, since the parameter distribution should match the data to which the RIM will be applied (e.g. in-vivo scans).

The anatomical integrity of quantitative maps is an essential factor when evaluating the quality of a mapping method. The RIM and the ResNet use the pixel neighborhood's information to infer the parameter value at that pixel, which creates valid concern regarding the amount of blur introduced by the convolutional kernels. We demonstrated in simulation experiments that, although the RIM does introduce a limited amount of blur to the quantitative maps, small structures are still confidently retained, and the error introduced by the pixel neighborhood does not represent a significant change in the relaxation time of those structures. Additionally, in in-vivo experiments, both deep learning methods produce relaxation maps with similar structural characteristics to the maps inferred by the MLE. More concretely, the T_1 relaxation times in the interface between gray and white mater follow a similar transition pattern to the MLE, further suggesting that the RIM does not introduce sufficient blur to alter brain structures, even in in-vivo scans.

8. Conclusion

We proposed a new method for T_1 and T_2 mapping based on the Recurrent Inference Machines framework. We demonstrated that our method has higher precision than, and similar accuracy levels as an Maximum Likelihood Estimator and higher precision and higher accuracy than an implementation of the ResNet. The experimental results show that the proposed RIM can generalize well to unseen data, even when acquisition settings vary slightly. This allows the use of simulated data for training, representing a substantial improvement over previously proposed QMRI methods that depend on alternative mapping methods to generate ground-truth labels. Lastly, the RIM dramatically reduces the time required to infer quantitative maps by 150-fold compared to our implementation of the MLE, showing that our proposed method can be used in large studies with modest computing costs.

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.

CRediT authorship contribution statement

E.R. Sabidussi: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **S. Klein:** Conceptualization, Methodology, Validation, Writing – review & editing, Supervision. **M.W.A. Caan:** Conceptualization, Methodology, Validation, Writing – review & editing. **S. Bazrafkan:** Validation, Writing – review & editing. **A.J. den Dekker:** Validation, Writing – review & editing. **J. Sijbers:** Validation, Writing – review & editing, Funding acquisition. **W.J. Niessen:** Writing – review & editing, Project administration. **D.H.J. Poot:** Conceptualization, Methodology, Validation, Investigation, Writing – review & editing, Supervision.

Acknowledgements

This work is part of the project B-QMINDED which has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 764513.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at10.1016/j.media.2021.102220.

References

Atkinson, D.J., Edelman, R.R., 1991. Cineangiography of the heart in a single breath hold with a segmented turboFLASH sequence. Radiology 178 (2), 357–360. doi:10.1148/radiology.178.2.1987592.

de Bazelaire, C.M.J., Duhamel, G.D., Rofsky, N.M., Alsop, D.C., 2004. MR imaging relaxation times of abdominal and pelvic tissues measured in vivo at 3T: preliminary results. Radiology 230 (3), 652–659. doi:10.1148/radiol.2303021331.

Bojorquez, J.Z., Bricq, S., Acquitter, C., Brunotte, F., Walker, P.M., Lalande, A., 2017. What are normal relaxation times of tissues at 3T? J. Magn. Reson. Imaging 35, 69–80. doi:10.1016/j.mri.2016.08.021.

Cai, C., Wang, C., Zeng, Y., Cai, S., Liang, D., Wu, Y., Chen, Z., Ding, X., Zhong, J., 2018. Single-shot T2 mapping using overlapping-echo detachment planar imaging and a deep convolutional neural network. Magn. Reson. Med. 80 (5), 2202–2214. doi:10.1002/mrm.27205.

Cercignani, M., Dowell, N.G., Tofts, P., 2018. Quantitative MRI of the Brain: Principles of Physical Measurement. CRC Press, Taylor and Francis Group.

Chen, J., Chang, E.Y., Carl, M., Ma, Y., Shao, H., Chen, B., Wu, Z., Du, J., 2016. Measurement of bound and pore water T1 relaxation times in cortical bone using three-dimensional ultrashort echo time cones sequences. Magn. Reson. Med. 77 (6), 2136–2145. doi:10.1002/mrm.26292.

Chen, Y., Yu, W., Pock, T., 2015. On learning optimized reaction diffusion processes for effective image restoration. CoRR. abs/1503.05768, eprint1503.05768, http://arxiv.org/abs/1503.05768

Cheng, H.-L.M., Stikov, N., Ghugre, N.R., Wright, G.A., 2012. Practical medical applications of quantitative MR relaxometry. J. Magn. Reson. Imaging 36 (4), 805–824. doi:10.1002/jmri.23718.

Cocosco, C.A., Kollokian, V., Kwan, R.K.-S., Pike, G.B., Evans, A.C., 1997. BrainWeb: online interface to a 3D MRI simulated brain database. NeuroImage 5, 425.

Conlon, P., Trimble, M., Rogers, D., Callicott, C., 1988. Magnetic resonance imaging in epilepsy: a controlled study. Epilepsy Res. 2 (1), 37–43. doi:10.1016/0920-1211(88)90008-3.

Deoni, S.C., Williams, S.C., Jezzard, P., Suckling, J., Murphy, D.G.M., Jones, D.K., 2008. Standardized structural magnetic resonance imaging in multicentre studies using quantitative T1 and T2 imaging at 1.5T. NeuroImage 40 (2), 662–671. doi:10.1016/j.neuroimage.2007.11.052.

Erkinjuntti, T., Ketonen, L., Sulkava, R., Sipponen, J., Vuorialho, M., Iivanainen, M., 1987. Do white matter changes on MRI and CT differentiate vascular dementia from Alzheimers disease? J. Neurol. Neurosurg. Psychiatry 50 (1), 37–42. doi:10.1136/jnnp.50.1.37.

Gudbjartsson, H., Patz, S., 1995. The rician distribution of noisy MRI data. Magn. Reson. Med. 34 (6), 910–914. doi:10.1002/mrm.1910340618.

He, K., Sun, J., 2014. Convolutional neural networks at constrained time cost. CoRR. abs/1412.1710. eprint: 1412.1710, http://arxiv.org/abs/1412.1710

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. CoRR. abs/1502.01852. eprint: 1502.01852., http://arxiv.org/abs/1502.01852

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) doi:10.1109/cvpr.2016.90.

Jeelani, H., Yang, Y., Zhou, R., Kramer, C.M., Salerno, M., Weller, D.S., 2020. A myocardial T1-mapping framework with recurrent and U-Net convolutional neural networks. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, pp. 1941–1944. doi:10.1109/isbi45749.2020.9098459.

Keenan, K. E., Stupic, K. F., Boss, M. A., Russek, S. E., Chenevert, T. L., Prasad, P. V., Reddick, W. E., Zheng, J., Hu, P., Jackson, E. F., et al., 2017. Comparison of T1 measurement using ISMRM/NIST system phantom.

Larsson, H.B.W., Frederiksen, J., Petersen, J., Nordenbo, A., Zeeberg, I., Henriksen, O., Olesen, J., 1989. Assessment of demyelination, edema, and gliosis by in vivo determination of T1 and T2 in the brain of patients with acute attack of multiple sclerosis. Magn. Reson. Med. 11 (3), 337–348. doi:10.1002/mrm.1910110308.

Leone, F.C., Nelson, L.S., Nottingham, R.B., 1961. The folded normal distribution. Technometrics 3 (4), 543–550. doi:10.1080/00401706.1961.10489974.

Liu, F., Feng, L., Kijowski, R., 2019. MANTIS: model-augmented neural network with incoherent k -space sampling for efficient MR parameter mapping. Magn. Reson. Med. 82 (1), 174–188. doi:10.1002/mrm.27707.

Lønning, K., Putzky, P., Sonke, J.-J., Reneman, L., Caan, M.M.A., Welling, M., 2019. Recurrent inference machines for reconstructing heterogeneous MRI data. Med. Image Anal. 53, 64–78. doi:10.1016/j.media.2019.01.005.

Look, D.C., Locker, D.R., 1970. Time saving in measurement of NMR and EPR relaxation times. Rev. Sci. Instrum. 41 (2), 250–251. doi:10.1063/1.1684482.

Lu, H., 2019. Physiological MRI of the brain: emerging techniques and clinical applications. NeuroImage 187, 1–2. doi:10.1016/j.neuroimage.2018.08.047.

Morningstar, W.R., Levasseur, L.P., Hezaveh, Y.D., Blandford, R., Marshall, P., Putzky, P., Rueter, T.D., Wechsler, R., Welling, M., 2019. Data-driven reconstruction of gravitationally lensed galaxies using recurrent inference machines. Astrophys. J. 883 (1), 14. doi:10.3847/1538-4357/ab35d7.

Mugler, J.P., 2014. Optimized three-dimensional fast-spin-echo MRI. J. Magn. Reson. Imaging 39 (4), 745–767. doi:10.1002/jmri.24542.

Poot, D.H.J., Klein, S., 2015. Detecting statistically significant differences in quantitative MRI experiments, applied to diffusion tensor imaging. IEEE Trans. Med. Imaging 34 (5), 1164–1176. doi:10.1109/tmi.2014.2380830.

Putzky, P., Welling, W., 2017. Recurrent inference machines for solving inverse problems. 1706.04008.

Ramos-Llorden, G., Den Dekker, A.J., Van Steenkiste, G., Jeurissen, B., Vanhevel, F., Van Audekerke, J., Verhoye, M., Sijbers, J., 2017. A unified maximum likelihood framework for simultaneous motion and $T1$ estimation in quantitative mr T1 mapping. IEEE Trans. Med. Imaging 36 (2), 433–446. doi:10.1109/tmi.2016.2611653.

Shao, J., Ghodrati, V., Nguyen, K.-L., Hu, P., 2020. Fast and accurate calculation of myocardial T1 and T2 values using deep learning Bloch equation simulations (DeepBLESS). Magn. Reson. Med. 84 (5), 2831–2845. doi:10.1002/mrm.28321.

Sijbers, J., Den Dekker, A., 2004. Maximum likelihood estimation of signal amplitude and noise variance from MR data. Magn. Reson. Med. 51 (3), 586–594. doi:10.1002/mrm.10728.

Sijbers, J., Den Dekker, A.J., Verhoye, M., Raman, E.R., Van Dyck, D., 1998. Optimal estimation of T2 maps from magnitude MR images. In: Hanson, K.M. (Ed.), Medical Imaging 1998: Image Processing. International Society for Optics and Photonics, SPIE, pp. 384–390. https://doi.org/10.1117/12.310915

Smit, H., Guridi, R.P., Guenoun, J., Poot, D.H.J., Doeswijk, G.N., Milanesi, M., Bernsen, M.R., Krestin, G.P., Klein, S., Kotek, G., 2013. T1 mapping in the rat myocardium at 7T using a modified CINE inversion recovery sequence. J. Magn. Reson. Imaging 39 (4), 901–910. doi:10.1002/jmri.24251.

Stanisz, G.J., Odrobina, E.E., Pun, J., Escaravage, M., Graham, S.J., Bronskill, M.J., Henkelman, R.M., 2005. T1, T2 relaxation and magnetization transfer in tissue at 3T. Magn. Reson. Med. 54 (3), 507–512. doi:10.1002/mrm.20605.

Swamy, P.A.V.B., 1971. Statistical Inference in Random Coefficient Regression Models. Springer doi:10.1007/978-3-642-80653-7.

Weiskopf, N., Suckling, J., Williams, G., Correia, M.M., Inkster, B., Tait, R., Ooi, C., Bullmore, E.T., Lutti, A., 2013. Quantitative multi-parameter mapping of R1, PD*, MT, and R2* at 3T: a multi-center validation. Front. Neurosci. 7. doi:10.3389/fnins.2013.00095.

Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.H.S., 2015. Conditional random fields as recurrent neural networks. CoRR. abs/1502.03240. eprint: 1502.03240., http://arxiv.org/abs/1502.03240