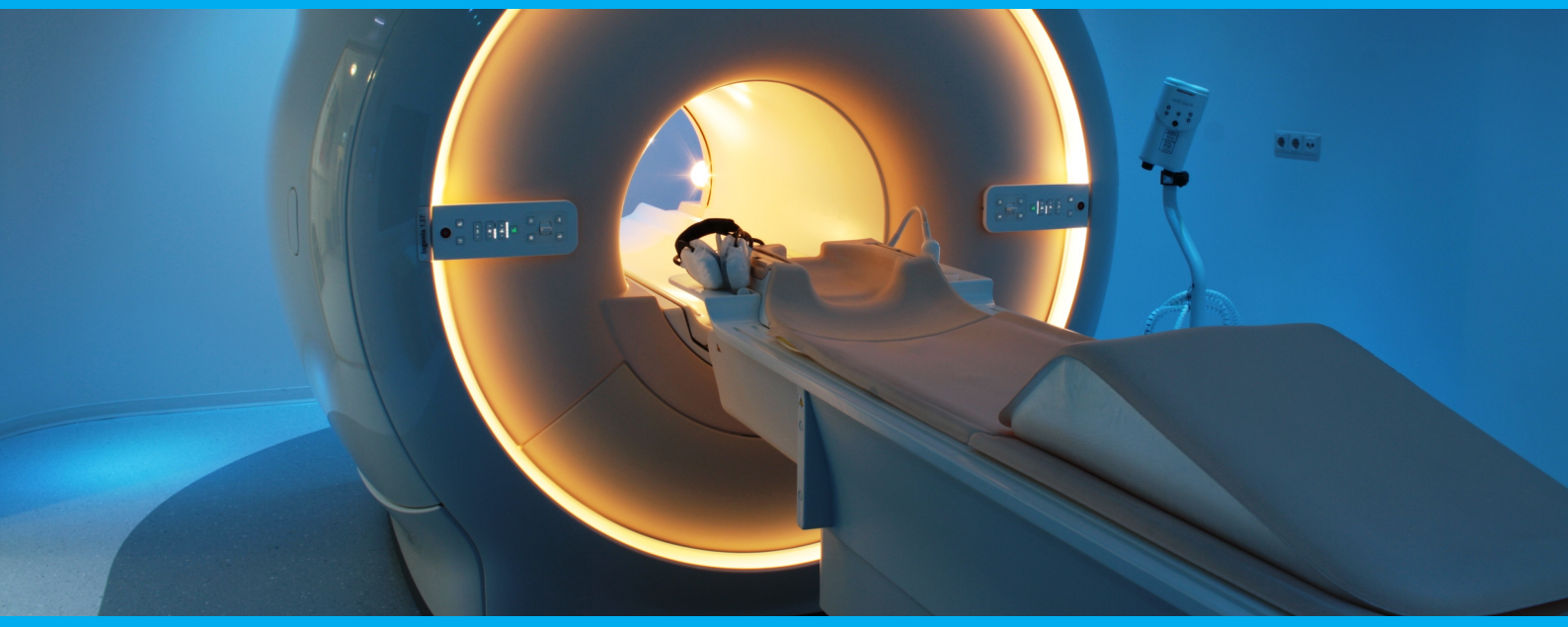
Recurrent Inference Machines for accelerated MRI acquisition trained using simulated undersampled k-space data

M.D. Izarin



TU Delft

Recurrent Inference Machines for accelerated MRI acquisition trained using simulated undersampled k-space data

by

M.D. Izarin

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday, August 26, 2021 at 15:00.

Student number:	4310144	
Project duration:	September 04, 2020 – August 26, 2021	
Thesis committee:	Dr. ir. R. van de Plas	TU Delft
	Prof. (Dr. ir.) W. J. Niessen	Erasmus MC & TU Delft
	Assistant prof. (Dr. ir.) D. H. J. Poot	Erasmus MC, supervisor
	MSc. E. Ribeiro Sabidussi	Erasmus MC, supervisor

TU Delft

Abstract

Recently, many advancements have been made in accelerated MRI reconstruction with the use of neural networks. Such deep learning methods learn a suitable MRI prior distribution from large sets of training data. For MRI images acquired with an uncommon scanning sequence, large datasets required for training are not available. Additionally, deep learning methods do not generalize well to unseen data. Therefore, in this research, a framework is proposed for accelerated MRI reconstruction trained with simulated data. The framework uses a Recurrent Inference Machine (RIM). The RIM is a deep learning framework that learns an iterative inference process. The RIM framework has been chosen as it is designed to learn the inversion process itself rather than the image statistics. Therefore, RIMs have a low tendency to overfit, and a high capacity to generalize to unseen data. The framework is evaluated by reconstructing undersampled data of in-vivo brain MRI images and comparing them with zero-filled reconstructions, reconstructions of an identical framework trained with in-vivo data and ESPIRiT reconstructions. The comparison shows that the framework does partly learn the inference process; however, the reconstructions still contain artefacts and the reconstruction of the framework trained with in-vivo data and the ESPIRiT method are of higher quality. For simulated data to replace in-vivo data for the training of the RIM, the simulated data has to be more similar to the in-vivo data.

Contents

1

Introduction

Magnetic Resonance Imaging (MRI) is a non-invasive imaging technology that uses the concept of nuclear magnetic resonance to produce detailed anatomical images. In clinical environments, MRI allows for the survey and quantification of physiological and metabolic features of tissues. These can be used to extract information about pathological processes and conditions, such as tumour growth (Young & Knopp, 2006), atrophy (Rocca et al., 2017) and neurodegenerative disorders (Koikkalainen et al., 2016). The MRI signal measured by the MRI scanner decays exponentially with time, limiting the time frame in which the data can be acquired with a sufficient signal to noise ratio (Macovski, 1996). Furthermore, the magnetic gradients are constrained by the maximum slew-rate and gradient strength (Boubela et al., 2014). As a consequence of this, MRI scans are slow. Some negative effects of this are discomfort for patients and higher costs of a scan, as the patient throughput is limited (Cohen & Weisskoff, 1991). Additionally, long scanning times makes MRI images more susceptible to motion artefacts caused by physiological processes (respiration, heart beat) and patient movement (Zaitsev et al., 2015).

Motivated by these limitations, many attempts have been made to accelerate MRI scan times. Examples of strategies for shortening MRI scan time are the use of multiple receiver coils (Deshmane et al., 2012), improving the magnetic gradient performance (Heidemann et al., 2003), increasing the magnetic field strength (Schnell et al., 1999) or reconstruction of undersampled data. The focus of this research will be on the latter strategy. The reconstruction of undersampled MRI data is referred to as 'accelerated MRI reconstruction'. The measurements made during an MRI scan are stored in the k-space, an array of numbers representing spatial frequencies in the MRI image. When undersampling the data samples from the k-space are missing. As a result of this aliasing artefacts apear in the reconstructed image.

Many acceleration methods have been developed for reconstructing artefact free MRI images from undersampled k-space data. Parallel imaging (PI) methods aim to do this by integrating the spatial sensitivity of the receiver coils of the MRI scanner (Pruessmann et al., 1999) into the image reconstruction. Commonly, the spatial sensitivity of the coils is introduced in the reconstruction in the from of sensitivity maps. Examples of PI imaging reconstruction methods are SENSE (Pruessmann et al., 1999) and GRAPPA (Griswold et al., 2002). SENSE uses a matrix inversion process that unfolds and combines the aliased images from each coil using sensitivity maps estimated from separate low-resolution images. The GRAPPA method extracts the information needed for the sensitivity maps from a separately acquired fully sampled k-space for each coil, known as the auto-calibration signal (ACS) region. From the ACS of each coil, weighting factors are calculated. These reflect how each coil smears, distorts and displaces frequencies within the k-space data (Elster, 2021). This information is used to fill in the missing k-space points; from the filled-in k-space an image is reconstructed. Ad-

vantages of the GRAPPA method over SENSE are higher quality reconstructions for moving organs. However, if speed is the main concern, SENSE has the advantage, as the acquisition of extra self-calibration k-space lines may increase scan times. An additional advantage of SENSE is that it generally more accurately reconstructs images for high accelerations.

An alternative technique for accelerated MRI acquisition is Compressed Sensing (CS) (Lustig et al., 2007). CS aims to exploit the natural sparsity of medical images in some transform domains. When randomly sampling the data, artefacts appear as incoherent noise-like patterns in the image. These artefacts can be removed by applying a sparsity transform, e.g. wavelets or curvelets (Qu et al., 2010) and then making the assumption that certain components in the transformed image contain the majority of the artefacts. This assumption is then used to reduce the artefacts in the image domain. Mathematically, this entails solving an optimization problem, which is generally solved using an iterative non-linear reconstruction method. Advantages of CS over PI are that it allows for higher acceleration factors and produces reconstructions with higher SNR (Lustig et al., 2007). However, compared to PI, the iterative non-linear reconstruction methods are relatively slow, this significantly increases the reconstruction time.

A method that combines GRAPPA, SENSE and CS is total variation ESPIRiT (TV-ESPIRiT) (Uecker et al., 2014)(Chunli et al., 2017)(Uecker et al., 2015). Similarly to GRAPPA, this method computes sensitivity maps from the ACS. However, ESPIRiT computes the maps using an eigenvalue decomposition, resulting in higher-quality sensitivity maps compared to GRAPPA or SENSE. These maps enable better reconstructions, which are more robust against artefacts, in particular ghosting artefacts. Subsequently, the sensitivity maps are used in a method that combines SENSE with a compressed sensing reconstruction method that includes a total variation regularization. Compared to GRAPPA and SENSE, TV-ESPIRiT is more robust, provides more accurate reconstruction and has better noise behaviour. Additionally, TV-ESPIRiT is more suited to coherent undersampling schemes than conventional CS methods. Disadvantages of TV-ESPIRiT are that it requires a separately acquired fully-sampled ACS, which increases the scan time. Additionally, it cannot be applied directly to coherently sampled data, as an additional gridding step is required to generate calibration data (Wright et al., 2014).

The emerging trend within accelerated MRI reconstruction is the use of deep learning. Accelerated MRI deep learning methods learn information about a typical MRI image from large datasets. This information is used to facilitate the reconstruction of the undersampled MRI images. An example of an accelerated MRI deep learning method is the ADMM-Net (Yang et al., 2017), which uses a neural network to parametrize the alternating direction method of multipliers (ADMM) (Boyd et al., 2011), a method that solves convex optimization problems by breaking them into smaller pieces, each of which are then easier to handle. The ADMM-Net achieved higher reconstruction accuracy than the original ADMM method while keeping the computational efficiency of the ADMM algorithm. Another successful deep learning method was developed by Jin et al., 2017. This method utilizes a U-net architecture, originally designed for image segmentation tasks (Ronneberger et al., 2015). The U-net takes an aliased weighted image as the input and aims to reduce the artefacts of the image by combining high-level features with lower-level global information of the image. The U-net enables faster reconstructions than conventional CS methods; however, it is prone to overfitting on the training data.

Most deep learning accelerated MRI reconstruction methods require large quantities of training data. For common scanning sequences, such as T1, T2 and proton density weighting, training data is widely available, but for new scanning sequences the training data must be acquired. This is a time-intensive and costly operation. This research aims to evaluate the potential of using simulated data to train deep learning reconstruction methods to reduce the dependency on large in-vivo data sets. A major hurdle for using simulated data is the limited capacity of many deep learning methods to generalize well for unseen data (Ahishakiye et al., 2021). An accelerated MRI deep learning method that showed potential to generalize well across different data types is the framework proposed by Lønning et al., 2018. This framework is based on the recurrent inference machine

(RIM) (Putzky & Welling, 2017). A RIM is an iterative method that contains a Recurrent Neural Network cell (RNN). The RIM generalizes well for unseen data because the architecture of the RIM has been designed such that the focus shifts toward learning the optimization procedure itself rather than the image statistics. Because of this property, the RIM framework is chosen as the framework used in this research to evaluate the potential of using simulated training data.

For this research an accelerated MRI reconstruction method was build based on the Recurrent Inference Machine. This method was trained using simulated data. For the evaluation of this method, in-vivo data from the fastMRI challenge (Zbontar et al., 2018) was used. The optimal hyperparameters for the method were determined experimentally. To assess the performance of the method, the normalized mean square error (NMSE), peak signal-to-noise ratio (PSNR) and structural similarity index measures (SSIM) of the reconstructed images was measured and the results were compared to images reconstructed by a baseline RIM (trained with in-vivo images), the TV-ESPIRiT, and the i-RIM (invertible-RIM), an advanced RIM model proposed by Putzky and Welling, 2019.

2

Theory

2.1 The magnetic resonance imaging signal

The information in this section has been acquired from the chapter 'Basic MRI physics' in the book Cardiovascular Magnetic Resonance Imaging (Aletras, 2008) unless another citation is stated. MRI makes use of magnetic resonance properties of hydrogen atoms in water molecules. Hydrogen atoms posses a property known as spin. The spin of a charged particle is associated with a magnetic dipole moment, μ. Naturally, all individual μ in the tissue are oriented in all directions, resulting in a local net magnetization vector M of zero. When applying a strong magnetic field B_0, the μ of all hydrogen atoms align parallel and anti-parallel to the external magnetic field. In practice, μ is not stationary, but is precessing about B_0 with the Larmor frequency, f_L, of the associated atom. When all μ are aligned with B_0, the parallel orientation is slightly favoured over the anti-parallel, resulting in a small excess of μ along the parallel direction. This effect, combined with the fact that the individual spin vector components perpendicular to B_0 cancel each other out, makes that M is stationary and points in the direction of B_0.

The scanner's receiver coils can not measure a stationary M. To enable the measurement of M, a radiofrequency (RF) pulse with the Larmor frequency of hydrogen atoms is applied when B_0 is active. Subsequently, the spins start to resonate with the RF-pulse, 'tipping' M down into the plane perpendicular to B_0, while M rotates around B_0. The rotating M induces a small oscillating current (emf) in the receiver coils. This is the MRI signal, which is measured during MRI scans. The signal decays after the RF pulse is applied; in a process called relaxation. The relaxation can be categorized into three different types: T1 relaxation, which is responsible for returning the 'tipped' M to the original direction of B_0, T2 relaxation, caused by the microscopic changes in the magnetic field that the spins experience and T2*, which is the result of the same mechanism responsible for T2 relaxation, combined with the effect of local inhomogeneities in B_0. The relaxation rates are tissue-specific and are characterized by the T1, T2 and T2* relaxation times. By adjusting specific scanning parameters of the MRI machine, tissue property weighted images can be captured, for which the contrast is mainly dictated by one of the relaxation times.

The measurements made during an MRI scan can be expressed in the k-space. The k-space is an array of numbers representing spatial frequencies and quadrature phases in the MRI image. The k-space is a two or three-dimensional space with one frequency encoding and one or two (in 3D) phase encoding directions. The centre of the k-space contains the lower frequencies, which describe the overall contrast, brightness and larger structures. The periphery contains the higher frequencies which describe the edges and finer details of the MRI image.

2.2 Fully sampled acquisition

From the data in the k-space, a weighted image can be reconstructed. Traditionally, the Nyquist criterion is used to determine if an image can be accurately reconstructed. If the criterion is not met, aliasing occurs, which is the incorrect measurement of the signal's frequency. As a consequence of this the reconstructed images might be affected by artefacts. If the number of samples acquired in k-space satisfies the Nyquist criterion (Shao, 2016), this k-space is referred to as fully-sampled, and an inverse Fourier transform (IFT) can be used to reconstruct the MRI image, without loss of information.

2.3 Accelerated MRI

The total scan time can be reduced by acquiring fewer data points in the k-space. The undersampled k-space can be reconstructed with a zero-filled reconstruction method. This method fills in the missing k-space points with zeros and reconstructs the image from the filled-in k-space using an IFT. However, the zero-filled k-space violates the Nyquist criterion, and the reconstruction process introduces artefacts to the reconstructed MRI image.

Depending on the k-space sampling strategy, different types of artefacts can corrupt the images (Lustig et al., 2007). Some examples are shown in Figure 2.1. Figure 2.1A presents the MRI image reconstructed from a fully sampled k-space. Figure 2.1B shows that, for uniform undersampling (i.e. only some lines of k-space are acquired), coherent, ghost-like artefacts are generated. Figures 2.1C and 2.1D shows the reconstructions for randomly sampled k-spaces, which contain incoherent artefacts that look much like additive random noise. Note that, as the centre of the k-space encodes the larger structures and overall contrast of the image, the artefacts can be reduced if the centre of the k-space has a higher sampling density.

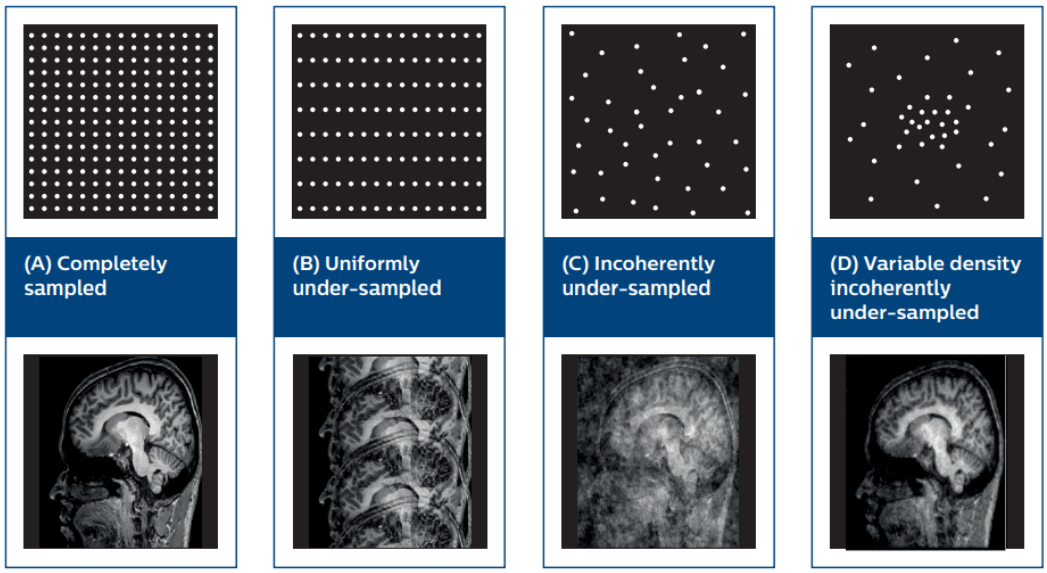


Figure 2.1: Various sampling patterns and the resulting reconstructed weighted images containing different types of artefacts. (Geerts-Ossevoort L, 2018)

2.3.1 Forward model

The forward model describes how the undersampled k-space is formed from the artefact-free weighted image. For the forward model of a 2D weighted image and the corresponding undersampled k-spaces, let $W \in C^{n \times m}$ be an unknown artefact free weighted image with n and m as the size of the image. The mapping from W to a coil weighted image $W_c \in C^{n \times m}$ is performed by multiplying W with the sensitivity map $M_c \in C^{n \times m}$, which scales each voxel according to the spatial sensitivity of one of the scanner's C receiver coils coils. The projection from the image to the k-space can be performed using a Fourier transform $\mathscr{F}$. The k-space of each coil is subject to a zero-mean Gaussian noise $n_c \in C^{n \times m}$, with a standard deviation of n_σ. The trajectory which is traversed in k-space can be modelled by introducing an undersampling mask $U \in \{0,1\}^{n \times m}$ that 'selects' the acquired samples from the fully-sampled k-space. This results in the measured signal in the coil $S_c \in C^{n \times m}$. The forward model which describes the whole operation is the following

$$S_c = U \odot \Big(\mathscr{F}\Big(M_c \odot W \Big) + n_c \Big), \quad c = 1, ..., C, \tag{2.1}$$

where $\odot$ denotes element-wise multiplication

2.3.2 Statistical noise model

The goal of accelerated MRI reconstruction is to find an inverse transform of the forward model described in Equation 2.1. A common approach for this is the use of statistical methods such as the Maximum Likelihood Estimate (Myung, 2003) and the Maximum A Posteriori Estimate (Bassett & Deride, 2019). These methods require a parametric statistical model of the weighted images, which can be derived from the statistical noise model. Take $S = \{S_1, S_2, ... S_C\}$ and let $P_{i,j}(S_{i,j} \mid W, n_\sigma)$ be the Gaussian probability density function (PDF) of observing the undersampled signal for a weighted image W, where i and j denote the position in k-space. Assuming that all k-space points in S are independent and that the measurement noise is Gaussian distributed with a standard deviation n_σ, the joint PDF of all k-space points of all coils is given by

$$P(S \mid W, n_\sigma) = \prod_{c=1}^{C} \prod_{i,j}^{m,n} P\left(S_{c,i,j} \mid W, n_\sigma\right). \tag{2.2}$$

From the joint PDF the log-likelihood function can be obtained. A likelihood function quantifies how well the statistical model parameters fit the given data. The likelihood function can be obtained by regarding Equation 2.2 as a function of W. By taking the log of the likelihood, multiplications are replaced with additions, which helps to stabilize the optimization, while the location of the maximum of the likelihood stays the same. Let $\hat{S}_c := U \odot \mathscr{F}\Big(M_c \odot W \Big)$ be the estimated k-space signal for coil c and $\hat{W}$ an estimated weighted image. The corresponding joint log-likelihood of the Gaussian distribution is then given by

$$\mathcal{L}(\hat{W}, n_\sigma \mid S) = \log P(S \mid \hat{W}, n_\sigma) = -\frac{CN}{2} \log(2\pi n_\sigma^2) - \frac{1}{n_\sigma^2} \sum_{c=1}^{C} \sum_{i,j}^{m,n} (S_{c,i,j} - \hat{S}_{c,i,j})^2, \tag{2.3}$$

where N is equal to the total number of non zero entries in U.

3

Methods

3.1 Recurrent Inference Machine

The Recurrent Inference Machine is a framework that has specifically been developed for inverse problem solving by Putzky and Welling, 2017 and has successfully been implemented for accelerated MRI reconstruction by Lønning et al., 2018. A RIM is an iterative method that contains a Recurrent Neural Network architecture. In the context of accelerated MRI reconstruction a RIM evaluates an estimate of the (artefact-free) weighted image, a hidden memory state, and the gradient of a likelihood term, to reconstruct a weighted image. The likelihood term encodes information about the known forward model and measures how well the current estimate of the model reproduces the observed measurements. The RIM is able to perform the inference process (reconstructing artefact free weighted images from undersampled k-spaces) by learning a optimization model that minimizes the likelihood function and a prior. The prior is learned by calculating a loss between the ground truth image and the final estimate and using this loss to update the network parameters. In this section the implementation of the RIM used in this research for the reconstruction of undersampled k-space data is presented. The framework was implemented in Python using the PyTorch 1.7.0 package.

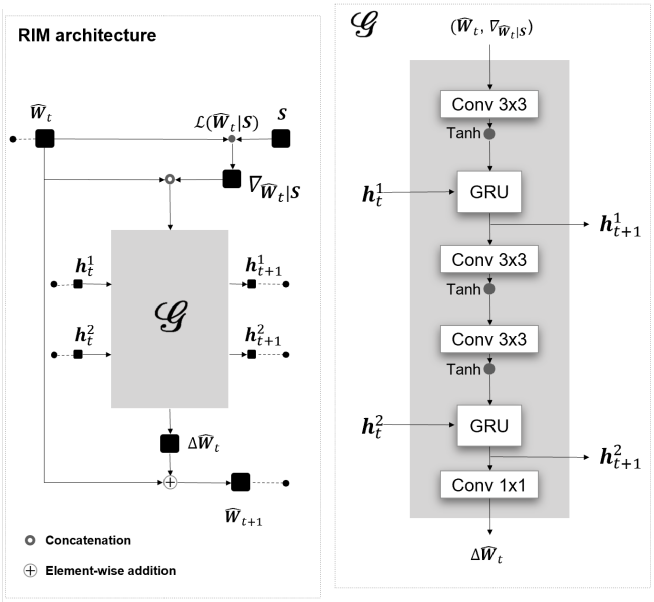
3.1.1 Framework

The framework of the RIM can be seen in Figure 3.1a. The RIM draws many similarities with the common gradient-descent optimization method, a first-order iterative optimization algorithm for finding the local minimum of the negative of the log-likelihood. At a given iteration step $t \in \{0, ...T-1\}$, the neural network part of the RIM ϕ receives as input the current estimate $\hat{W}_t$, the gradient of the negative log-likelihood $\mathcal{L}(\hat{W}, n_\sigma \mid S)$, with respect to $\hat{W}_t$, $\nabla_{S|\hat{W}_t}$, and the hidden states $h_t^* = [h_t^1, h_t^2]$. From these inputs $\mathcal{G}$ generates an update to the current estimate, $\Delta\hat{W}_t$, and the memory states that are used in the next iteration, h_{t+1}^*. The update equations of the RIM at each iteration, are the following

$$\{\Delta\hat{W}_t, h_{t+1}^*\} = \mathcal{G}_\phi\left(\nabla_{S|\hat{W}_t}, \hat{W}_t, h_t^*\right); \tag{3.1}$$

$$\hat{W}_{t+1} = \hat{W}_t + \Delta\hat{W}_t, \tag{3.2}$$

where ϕ represents trainable network parameters. The hidden states are initially set to zero: $h_{t=0}^* = \mathbf{0}$, and $\hat{W}_{t=0}$ is initialized as described in Section 3.1.4.



(a) The framework of the RIM

(b) The neural network part of the RIM $\mathcal{G}$

Figure 3.1: The RIM framework, where lines originating from the dots indicate information passing through iterations. The inputs of the RIM are $\hat{W}_t$, the estimate of the weighted image for iteration t, and S, the undersampled k-spaces. From the inputs of the RIM the gradient of the log-likelihood is computed $\nabla_{S|\hat{W}_t}$ which is concatenated with $\hat{W}_t$ to be the input of the neural network part $\mathcal{G}$. Additional inputs of $\mathcal{G}$ are the memory states h_t^*. The $\mathcal{G}$ computes h_{t+1}^*, the new memory stated for the next iteration, and $\Delta\hat{W}_t$, an update to $\hat{W}_t$. $\hat{W_{t+1}}$ and h_{t+1}^* are passed to the next iteration.

3.1.2 Neural network architecture

The composition of $\mathcal{G}$ can be seen in Figure **??**. The inputs of $\mathcal{G}$ are $\hat{W}_t$, $\nabla_{S|\hat{W}_t}$ and h_{t+1}^*. $\mathcal{G}$ consists of three convolutional layers and two Gated Recurrent Units (GRU). A GRU is a specific type of RNN gating mechanism specifically designed to handle long-term dependencies. Additionally, a GRU resolves the vanishing gradient problem, which regular RNN units suffer from (Hu et al., 2018). For the GRUs, the number of input channels is equal to the number of output channels. The exact number of output channels of the convolutional layers, ψ, is identical for each layer and was determined experimentally. The first 3×3 convolutional layer is followed by a hyperbolic tangent ($\tanh$) activation function. This is followed up by the first GRU, which additionally takes the hidden memory state (h_t^1) as an input. The GRU outputs h_{t+1}^1, and passes this through two successive 3 × 3 convolutional layers, each succeeded by a $\tanh$ activation function. The data then goes through the second GRU, which produces the second hidden state h_{t+1}^2. The last element of the $\mathcal{G}$ is a 1x1 convolutional channel which reduces dimensionality of the channels and outputs $\Delta\hat{W}_t$. the update of the estimated weighted image at each iteration, $\hat{W}_t$. To retain the original image size throughout $\mathcal{G}$, the 3×3 convolutional layers are zero-padded. The inputs and outputs of the $\mathcal{G}$ are complex-valued and are split, such that the real and imaginary values are processed as two separate channels.

3.1.3 Gradient of the log-likelihood

The gradient of the log-likelihood described in Equation 2.3, with respect to W, is given by

$$\nabla_{S|\hat{W}_t} = \frac{1}{n_\sigma^2} \sum_{c=1}^{C} M_c^* \odot \mathscr{F}^{-1}(U \odot (\hat{S}_c - S_c)), \tag{3.3}$$

where M_c^* is the conjugate matrix of M_c. For the implementation in the RIM, the gradient of the likelihood function, can be simplified by assuming n_σ is constant. Because, with n_σ constant, $\nabla_{S|\hat{W}_t}$ does not depend on n_σ, therefore its value has no influence in the optimization of W. Therefore, for the optimization only, n_σ is set to 1.

3.1.4 Initialisation of weighted image estimates

The optimization performed by the RIM improves if the initial guess of the weighted image $\hat{W}_0$ is close to the ground truth. The $\hat{W}_0$ used in this framework is the zero-filled reconstruction of $\{S_c\}_{c=1}^C$, which is estimated using the sensitivity masks of each coil. The method for reconstruction the zero-filled reconstruction is a component of SENSE (Pruessmann et al., 1999), and is described in the following equation

$$\hat{W}_0 = \sum_{c=1}^{C} M_c^* \odot (\mathscr{F}^{-1} S_c), \tag{3.4}$$

where $\mathscr{F}^{-1}$ is the inverse Fourier transform and M_c^* is the complex conjugate of M_c.

3.1.5 Loss function

The loss function quantifies the performance of the RIM by comparing the known ground truth image reconstructed from fully sampled k-spaces, W, with the estimates. For the RIM in this research the loss function is the Normalized Mean Square Error (NMSE) over all iterations t. The loss is computed for every inference step as this encourages the model to produce predictions that improve for each time step. The NMSE is chosen over the MSE as the normalization enables easier comparison of the training losses between datasets with different image intensities. With the NMSE as the loss function, the training of the RIM can be written as the following optimization problem

$$\hat{\phi} = \arg\min_{\phi} \left\{ \frac{1}{T} \sum_{t=1}^{T} \frac{\|\hat{W}_t - W\|_2^2}{\|\hat{W}_t\|_2^2} \right\}, \tag{3.5}$$

where ϕ denotes the network parameters and $\hat{\phi}$ are the optimal network parameters given the training data. The goal of the training is to improve the reconstruction performance by updating ϕ. This is achieved using the principle of backpropagation, which is a method that produces updates to ϕ using the gradients of the loss function in combination with an optimization algorithm. In this research, the backpropagation method used for the training utilizes the Adam optimiser (Kingma & Ba, 2014).

3.1.6 Training

Firstly, two separate RIMs are trained using simulated training data one for an acceleration factor Λ of 4 ($\text{RIM}_{\Lambda=4}^{\text{sim}}$) and another for $\Lambda = 8$ ($\text{RIM}_{\Lambda=8}^{\text{sim}}$). The method for simulating the testing data is described in Section 3.3. These RIMs are used to asses how well the framework is able to reconstruct the simulated data. Two other RIMs are trained with simulated data as well ($\text{RIM}_{\Lambda=4}^{\text{in-vivo}}$, $\text{RIM}_{\Lambda=8}^{\text{in-vivo}}$). Again, one for each acceleration. However, these RIMs are used to reconstruct in-vivo data. Details about the in-vivo data can be found in Section 3.4. Finally two additional RIMs ($\text{RIM}_{\Lambda=4}^{\text{baseline}}$, $\text{RIM}_{\Lambda=8}^{\text{basleline}}$), again, one for each acceleration, are trained using in-vivo training data. These are used as baseline models. So in total, 6 separate RIMs are trained and evaluated.
The hyperparameters of the training are: learning rate l, number of output channels of the convolutional layers,ψ, batch size b, and number of inference steps T. The choice for b and T is limited by the memory of the GPU. Lønning et al., 2018 showed that increasing T results in better reconstruction quality. For the training data used in this research and for the minimum number of ψ that still produced reasonable reconstructions, if $b = 2$, the framework can run, at most, $T = 2$ iterations. Only performing two iteration steps significantly reduces the reconstruction performance. Therefore, for all models $b = 1$. Again, for the training data in this

research and the minimum desired ψ, the GPU limitation only allows for $T = 2$. The other hyperparameters, l and ψ have been determined experimentally in the experiment described in Section **??**.

3.2 Reference reconstruction methods

3.2.1 Total variation ESPIRiT

The total variation ESPIRiT (TV-ESPIRiT) is an accelerated MRI reconstruction method that combines the coil sensitivity map estimation of ESPIRiT, with a CS reconstruction based on a Total Variation regularization and SENSE. This method is a functionality of the BART Toolbox (Uecker et al., 2015).

The ESPIRiT method uses k-space kernel operations to derive a set of eigenvector maps that function as coil sensitivity maps. The eigenvector maps are computed from a fully sampled center of k-space, known as the auto-calibration signal (ACS) region. By moving a GRAPPA (Griswold et al., 2002) kernel across each position in the ACS, a calibration matrix, E, is formed. From E, a set of coil sensitivity maps is derived by performing an eigenvector decomposition. This results in a set of eigenvector maps with corresponding eigenvalues. The maps with an eigenvalue of 1 are used as the sensitivity maps for reconstruction.

The spatial information encoded in the sensitivity maps is used in the SENSE method to enable unfolding of the corrupted weighted image. For SENSE, the reconstruction problem can be described by solving the following linear equation:

$$ZS = W, \tag{3.6}$$

where Z denotes the unfolding matrix. To calculate Z, firstly a sensitivity matrix ξ is constructed from the calculated coil sensitivities. Subsequently Z is calculated with the following equation:

$$Z = (\xi^H \epsilon^{-1} \xi)^{-1} \xi^H \epsilon^{-1} \tag{3.7}$$

where the superscript H indicates the transposed complex conjugate and ϵ the noise matrix which describes the levels and correlation of noise in the receiver channels. Equation 3.6 is solved using a compressed sensing reconstruction method with a total variation (TV) regularization. Take Q to be γW, where γ denotes the sparsifying transform. Then, the TV regularized SENSE reconstruction is described by

$$\hat{W}_{ESPIRiT} = \text{argmin}_W \{\|S - ZW\|_2^2 + \lambda \|Q\|_{TV}\}, \tag{3.8}$$

where λ is the regularization parameter, which weights the relative importance of the first term, representing the data consistency and the second term, representing the prior. The regularization term is the TV norm of Q. The method to calculate the TV is described in the research of Bredies et al., 2010.

In this research the l_1-wavelet (Guerquin-Kern et al., 2009) is used as γ. For the reconstructions, λ is set to 10^{-3} for 4-fold accelerations and 10^{-2} for 8-fold accelerations, which corresponds with the highest quality reconstructions in Zbontar et al., 2018. Equation 3.8 is solved using an the alternating direction method of multipliers (ADMM) (Boyd et al., 2011).

3.2.2 i-RIM

The use of an invertible-RIM (i-RIM) for accelerated MRI reconstruction was proposed bu Putzky and Welling, 2019. The main difference between the RIM and the i-RIM is that the i-RIM uses reversible neural network architectures. This significantly reduces the memory consumption of the i-RIM compared to the original RIM. Thus, the framework of the i-RIM allows for more complex neural networks without overburdening the hardware memory. For this research the RIM is not implemented, however the evaluation metrics for the reconstruction of the fastMRI data, reported in Putzky and Welling, 2019, are used as a reference.

3.3 Simulated data set

3.3.1 Generate simulated weighted images

Synthetic MRI is a method for generating weighted images (Bobman et al., 1985). This method is commonly used to produce images with different tissue contrasts from anatomical parametric maps. The equations for synthetic MRI are simplifications of the Bloch equations described in Bloch, 1946. These are a set of macroscopic equations that are used to calculate the nuclear magnetization as a function of time given the tissue parameters. For Fast Spin Echo sequences, which are commonly used for generating T1-weighted, T2-weighted and PD-weighted images, the following Synthetic MRI equation is used to synthesize weighted images (Hagiwara et al., 2017)

$$|W| = Ae^{-\frac{TE}{T_2}}(1 - e^{-\frac{TR}{T_1}}), \tag{3.9}$$

where $|\bullet|$ is the operator for the magnitude of a matrix, A is a quantity proportional to the proton density, T_1 the longitudinal relaxation time, and T_2 the transverse relaxation time. The scanning parameters that characterise the tissue contrast are the repetition time TR and echo time TE. The method for simulating magnitude weighted images was derived from the method described in Sabidussi et al., 2021. Firstly, axial slices are sampled randomly from anatomical brain phantoms, P, from the BrainWeb project (Cocosco et al., 1997). The dimensions of the slices are 362 $\times$ 434 voxels. Each voxel contains a label that corresponds to a specific tissue type. To each voxel of a slice, T_1, T_2 and A are assigned, corresponding to this tissue type. To simulate the tissue variability between different subjects, the values are drawn from a normal distribution for each slice. The values and standard deviation are given in Table 3.1. To enable intra-tissue variation, voxel-wise Gaussian noise is added for each tissue parameter. The standard deviation of the intra-tissue noise is $1/35$th the value of the tissue parameter. Then the borders between the different tissue parameter maps are smoothed by applying a Guassian filter with a standard deviation of 0.4. From the resulting parameter maps, $|W|$ is reconstructed using Equation 3.9. The scanning sequence parameters are set such that they are identical to the in-vivo data set. The repetition time (TR) is 250ms and the echo time (TE) is 3.4ms.The major differences with the simulation method proposed in Sabidussi et al., 2021, is that the standard deviation of the intra-tissue variation is proportional to the tissue parameters values. Additionally the acquisition noise is omitted as this is added in the k-space, further in the simulation pipeline, rather then the image space. An example of $|W|$ can be seen in Figure 3.3.

$|W|$ is a magnitude image and does not contain any phase information. However, the RIM framework reconstructs complex-valued weighted images. The real-valued weighted images are converted to a complex-valued image, W, by assigning a uniform phase to all the voxels of the weighted image. For each slice, the phase of the phase map is drawn from a uniform distribution with values in the range [0, 2π]. From the phase map and $|W|$, a complex-valued weighted image is generated. W is the ground truth used during the training of the RIMs.

3.3.2 Generating undersampled k-space

From W, undersampled k-space were generated. The method to generate the undersampled k-space is the forward model in Equation 2.1. The value for n_σ was determined empirically, by trying different values of n_σ until the difference between the average signal to noise ration (SNR) of the simulated weighted image reconstructed from the 'noisy' coil-k-spaces, and the SNR of the in-vivo dataset was less then one. The SNR of both data sets was calculated by taking the mean value of the voxels containing tissue and dividing this by the standard deviation of the pixels containing the background. The n_σ that was found and used to simulate the k-space noise was 0.005.

For the undersampling masks, U, the lines are omitted in the phase encoding direction to resemble physically achievable accelerations in 2D data acquisitions. As the centre of the k-space contains the low-frequency

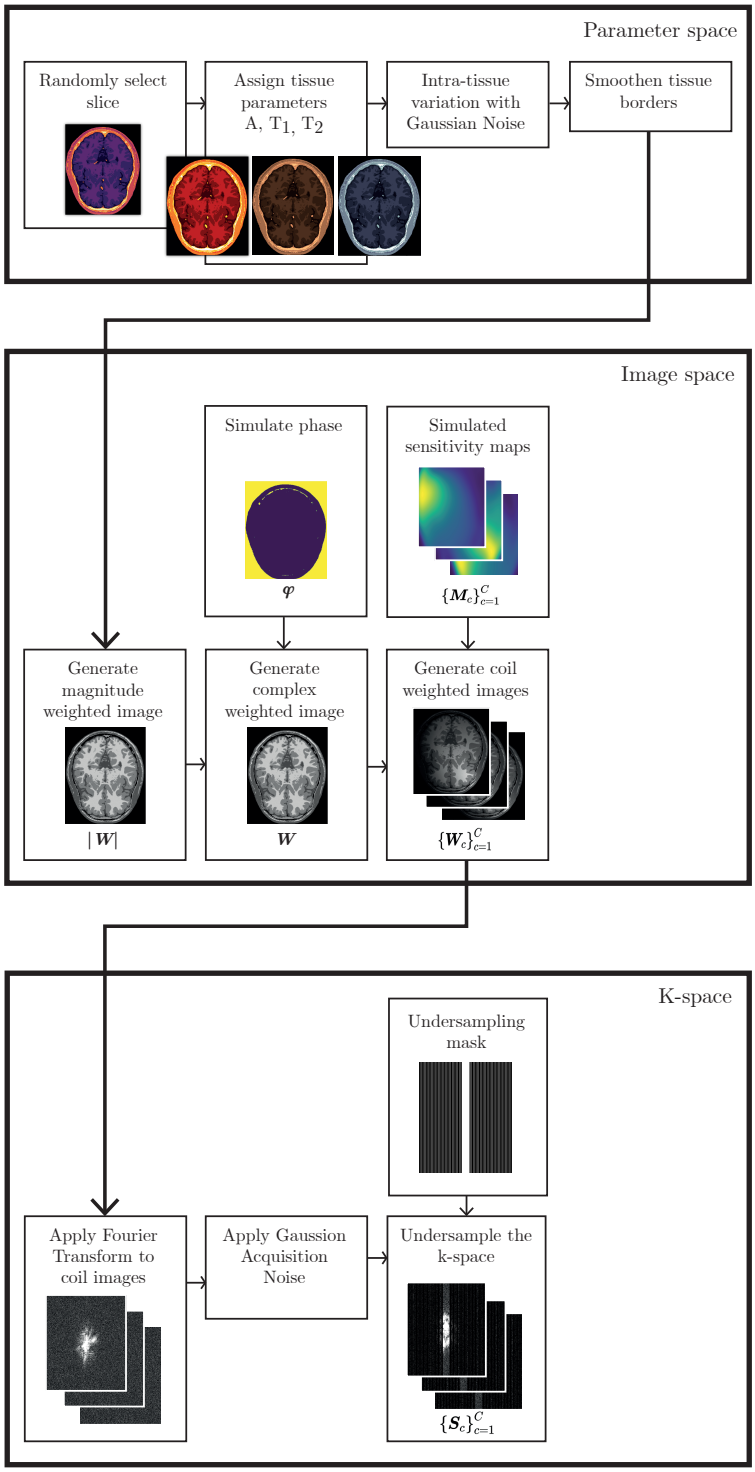


Figure 3.2: The pipeline for generating the simulated training data. In the parameter space, tissue parameters are assigned to a slice of an anatomical brain phantom. Subsequently, tissue variation is introduced, and tissue borders smoothing is applied. A magnitude weighted image, $|W|$ is generated from the parameter maps. $\|W\|$ is converted to a complex weighted image, W, by applying a uniform simulated phase map ϕ. Coil weighted images, $\{W_c\}_{c=1}^{C}$, are generated by multiplication with the individual coil sensitivity maps. The k-spaces are formed by applying a Fourier transform to all coil images. The final steps add acquisition noise and undersample the k-space, the result is a set of undersampled k-spaces $\{S_c\}_{c=1}^{C}$.

Table 3.1: Distribution of parameters per tissue and tissue property. T1 and T2 values in milliseconds. Values for A are chosen as a fraction of the concentration of protons in the CSF.

Tissue	$\mu_{tissue}^{T_1}$	$\sigma_{tissue}^{T_1}$	$\mu_{tissue}^{T_2}$	$\sigma_{tissue}^{T_2}$	μ_{tissue}^{A}	σ_{tissue}^{A}
CSF	3500⋆	300	2000⋆	300	1.0	0.3
Gray Matter	1400⋆	300	110⋆	30	0.85	0.3
White Matter	780⋆	250	80⋆	20	0.65	0.3
Fat	420⋆	100	70⋆	20	0.9	0.3
Muscle	1200⋆	300	50⋆	20	0.7	0.3
Muscle skin	1230⋆	300	50⋆	20	0.7	0.3
Skull	400‡	100	30⋆	10	0	0
Vessels	1980⋆	300	275⋆	70	1.0	0.3
Bone Marrow	580§	100	50⋆	20	0.8	0.3

⋆ Bojorquez et al., 2017
‡ Chen et al., 2017
⋆ Stanisz et al., 2005
§ De Bazelaire et al., 2004

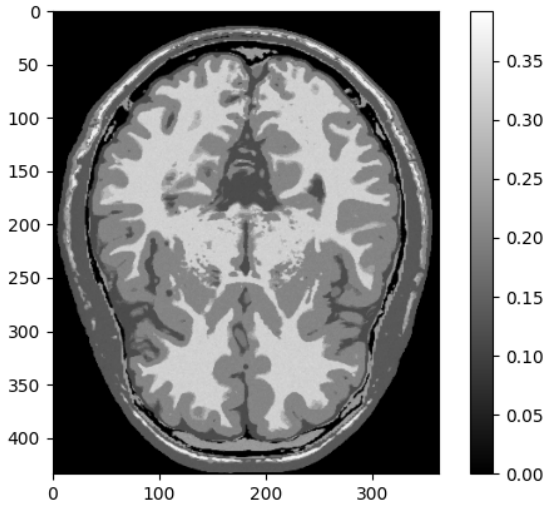


Figure 3.3: An example of a magnitude weighted image generated using the method described in Section 3.3.1

components, which encode the larger structures and overall contrast, the lines adjacent to the centre are fully sampled. When the acceleration factor, Λ, is 4, the fully sampled centre region contains 8% of all the k-space lines; for $\Lambda = 8$, it contains 4% of all lines. The remaining k-space lines are sampled equidistantly from each other, with a distance such that the fraction of fully sampled lines, including the fully sampled centre of k-space, correspond with the desired acceleration factor. Both undersampling masks are shown in Figure 3.4. The complete pipeline for generating the simulated data is presented in Figure 3.2.

The set of simulated sensitivity maps was constructed from the sensitivity maps of the most caudal slice of six different subjects from the in-vivo data set. The in-vivo sensitivity maps were generated using the ESPIRiT method included in the BART toolbox. Subsequently, the sensitivity maps were transformed such that their size matches the size of W and to ensure that the part of the sensitivity maps mapping the brain in the in-vivo weighted image overlaps with the brain in the simulated data. The transformations applied on the in-vivo sensitivity maps were a centre cropping, from the original size of 320 $\times$ 320 to 250 $\times$ 250, followed by a scaling to a size of 434 $\times$ 362, corresponding with the dimensions of the simulated weighted image. The scaling was performed using a spline interpolation of the 3rd order.

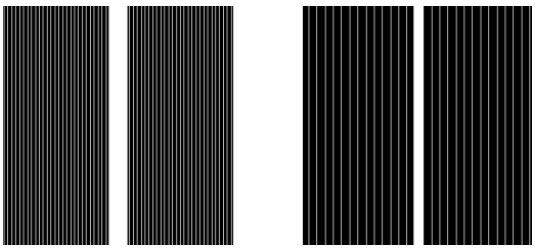


Figure 3.4: The undersampling masks for 4x acceleration (left) and 8x acceleration (right).

3.3.3 Training and testing data set

Two different data sets are defined for the simulated data. The first dataset ($\text{Data}^{\text{sim}}_{\text{train}}$) contains 12 different anatomical brain phantoms and a set of 4 simulated sensitivity maps, $\left\{\{P\}, \{\{M_c\}_{c=1}^{C}\}\right\}$. This data set is used to train the RIMs. The second data set ($\text{Data}^{\text{sim}}_{\text{test}}$) contains 4 different phantoms and two different simulated sensitivity maps and is used to evaluate the reconstruction performance of the RIM for simulated data.

3.4 In-vivo data set

The in-vivo data used in this research is from the the fastMRI project data set (Zbontar et al., 2018). The data set contains raw, fully sampled 2D multi-coil k-space data, $\{F_c\}_{c=1}^{C}$, and corresponding ground truth magnitude weighted images. The data set contains axial T1 weighted, T2 weighted and FLAIR images. However, in this research, only the T1 weighted images are used. The data set contains T1 weighted images of 54 different subjects; of each subject 16 different slices are included. The images are acquired using a FLASH acquisition (Haase, 1990). The MRI scans were performed on a Siemens 'Skyra' 3T scanner with a 16 channel coil array. The data was acquired with a TR of 250ms, a TE of 3.4ms and a flip angle of 70°. The slice thickness is 3mm. For this research, the k-space data is normalized with a factor of 1200, such that the intensity of the in-vivo data approximates the simulated data. This results in a maximum intensity of the reconstructed weighted image around 0.4. The ground-truth images included in the fastMRI data set are magnitude images. However, the RIM framework estimates complex-valued weighted images. Therefore, a complex-valued ground-truth image is constructed from the fully sampled k-space with the method described in Equation 3.4. The sensitivity maps

of each coil were estimated with the ESPIRiT method. This was done once, after which the sensitivity maps were stored on a hard drive to decrease the training and reconstruction time of the RIM.

3.4.1 Training, validation and test data sets

The fastMRI in-vivo data set is split into three different groups: training data ($\text{Data}_{\text{train}}^{\text{in-vivo}}$) containing 43 subjects, validation data ($\text{Data}_{\text{val}}^{\text{in-vivo}}$) containing 5 subjects and testing data ($\text{Data}_{\text{test}}^{\text{in-vivo}}$) containing 5 subjects as well. Each data set contains ground truth weighted images, pre-calculated sensitivity maps and fully sampled coil k-spaces of the subjects, $\left\{\{W\}, \{\{M_c\}_{c=1}^{C}\}, \{\{F_c\}_{c=1}^{C}\}\right\}$. $\text{Data}_{\text{train}}^{\text{in-vivo}}$ is used to train the baseline RIM models. $\text{Data}_{\text{val}}^{\text{in-vivo}}$ is used to provide an unbiased evaluation of a model fit on the in-vivo data while tuning hyperparameters for both $\text{RIM}_{\text{sim}}^{\Lambda=4}$ and $\text{RIM}_{\text{in-vivo}}^{\Lambda=4}$. $\text{Data}_{\text{test}}^{\text{in-vivo}}$ is used to evaluate the reconstruction performance of the proposed methods as well as the reference methods, for in-vivo data. One thing to note is that the amount of subjects in the in-vivo training data is significantly higher than the number of anatomical phantoms for the simulated data. However, this does not result in more unique training samples as the anatomical phantoms contain 180 different slices compared to the 16 of the in-vivo data. Additionally, for the simulated data, the tissue parameters are varied for every slice, resulting in an large amount of unique weighted images.

3.5 Evaluation metrics

All evaluations were performed on the estimated magnitude images. The metrics used for the evaluation are the same as the ones used for the fastMRI challenge. As images in the data sets have varying proportion of background, all metrics are computed for a masked area, $\Upsilon \in W$, that only includes tissue data. For the simulated data, Υ is constructed from the anatomical model used to generate the weighted images. For the in vivo data, the anatomy is not known; therefore Υ is constructed by including all voxels where $W_p \geq 0.015$. Subsequently, background voxels still included in the mask are removed by only retaining the largest connected object in Υ. Finally, missing voxels containing tissue are included in Υ using binary morphological closing (Virtanen et al., 2020).

3.5.1 Normalized mean square error

The normalized mean square error (NMSE) is a widely used measure of reconstruction quality for MRI images. The NMSE is the mean squared error (MSE) normalized by the power of the image. The advantage of using the NMSE is that it incorporates both the variance of the estimation and its bias (Lebanon, 2010). The NMSE between the reconstructed image ($\hat{W}$) and the artefact-free weighted image W is given by

$$\text{NMSE}\left(\hat{W}, W\right) = \sum_{p \in \Upsilon} \frac{\left(|\hat{W}_p| - |W_p|\right)^2}{\sum_{p \in \Upsilon} |W_p|^2}. \tag{3.10}$$

3.5.2 Peak signal-to-noise ratio

The peak signal-to-noise ratio (PSNR) represents the ratio between the power of the maximum possible image intensity across a volume and the power of noise and other errors and is defined as

$$\text{PSNR}\left(\hat{W}, W\right) = 10 \log_{10} \frac{\max(|W|)^2}{\frac{1}{n}\sum_{p \in \Upsilon}(|\hat{W}_p| - |W_p|)^2}, \tag{3.11}$$

where n is the number of voxels included in Υ. Higher values of PSNR indicate a better reconstruction.

3.5.3 Structural similarity index measure

The structural similarity index measure (SSIM) was developed with the idea that the human visual system is highly adapted to extract structures of objects in a scene (Morain-Nicolier, Landré, et al., 2009). The SSIM

considers image degradation as perceived changes in structural information while also incorporating important perceptual phenomena, including luminance masking and contrast masking terms. The local SSIM index of a single voxel is calculated within two local windows ($R \in W$, $\hat{R} \in \hat{W}$). The average of all local SSIM indexes is the global SSIM and is defined as

$$\mathrm{SSIM}\left(\hat{W}, W\right) = \frac{1}{n} \sum_{p \in \Upsilon} \frac{\left(2\mu_{|\hat{R}_p|}\mu_{|R_p|} + c_1\right)\left(2\sigma_{|\hat{R}_p||R_p|} + c_2\right)}{\left(\mu^2_{|\hat{R}_p|} + \mu^2_{|R_p|} + c_1\right)\left(\sigma^2_{|\hat{R}_p|} + \sigma^2_{|R_p|} + c_2\right)}, \tag{3.12}$$

where $\mu_{\hat{R}_p}$ and μ_{R_p} are the average voxel intensities in the windows in $\hat{R}_p$ and R_p, $\sigma^2_{\hat{R}_p}$ and $\sigma^2_{R_p}$ are the variances in the windows and $\sigma_{\hat{R}_p R_p}$ is the covariance between the windows. To stabilize the division two variables are included: $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$ with L as the maximum value of the target volume, $L = \max(|W|)$. For SSIM values reported in this research the following parameters are chosen: a window size of 7 $\times$ 7 and $k_1 = 0.01, k_2 = 0.03$. Higher values of SSIM indicate a better reconstruction.

4

Experiments

4.1 Validation of the forward model

In this experiment, the forward model described in Equation 2.1 was validated. A single slice from $\text{Data}_{\text{train}}^{\text{in-vivo}}$ was selected. From this slice, an undersampled coil k-space, S_c, was reconstructed using two different methods. The first undersampled coil k-space S_c^{data} was from a single fully sampled coil k-space included in $\text{Data}_{\text{train}}^{\text{in-vivo}}$, by multiplying it with an undersampling mask, U, corresponding to $\Lambda = 4$. The second method reconstructs S_c^{forward} from the W and M_c of the same slice and receiver coil as the first method. S_c^{forward} was generated using the forward model in Equation 2.1, where the standard deviation of the noise, n_σ, was set to 0 and a U identical to the U of the first method was used. To compare $S_c^{dataset}$ and $S_c^{forward}$, coil weighted images ($W_c^{dataset}$, $W_c^{forward}$) were reconstructed from both coil k-spaces by applying an inverse Fourier transform. The images were compared visually, and a difference map between the two weighted images was calculated by subtracting them from each other. To evaluate if the difference between the images can be attributed to the acquisition noise, the standard deviation of the difference image is computed. This is compared with the the standard deviation of a portion of the background of $W_c^{dataset}$ that does not contain undersampling artefacts.

4.2 Validation of simulated weighted images

This experiment was performed to quantify how well the simulated weighted images match with the in-vivo dataset. To identify the similarities and differences between the two different data sets, simulated weighted images were generated and compared visually to in-vivo images from the same area of the brain. To assess the differences between the complete datasets, the distribution of the voxel intensities containing tissue for 50 simulated was compared to the distribution of 50 in-vivo weighted images.

4.3 Optimal RIM model: hyperparameter tuning

To ensure that the optimal RIM model was used during the evaluation of the in-vivo data, the hyperparameters learning rate l and number of feature channels between the RNN cell layers ψ of the RIM were tuned. For $\Lambda = 4$ and $\Lambda = 8$, multiple RIMs were trained with $\text{Data}_{\text{train}}^{\text{sim}}$, each RIM with a different combination of hyperparameters. When tuning the RIM for one hyperparameter, the other hyperparameter was kept constant. The different values for the tested hyperparameters are the following

learning rate l	4e-3	**2e-3**	1e-3	5e-4
feature channels ψ	60	80	**95**	120

where the bold value is the constant value used when tuning the other hyperparameter. During the training, 10 slices were reconstructed from each virtual brain phantom included in $\text{Data}^{\text{sim}}_{\text{train}}$. The reconstruction performance of the RIMs was evaluated by reconstructing the undersampled k-spaces in $\text{Data}^{\text{sim}}_{\text{test}}$ for the models after the following training epochs: 25, 50, 100, 200, 400, 600, 800, 1000, 1200 and 1400. The NMSE, PSNR and SSIM of the reconstructed images of each model were evaluated. The models that reconstructs the data with the best SSIM were chosen as $\text{RIM}^{\text{in-vivo}}_{\Lambda=4}$ and $\text{RIM}^{\text{in-vivo}}_{\Lambda=8}$. The SSIM was chosen as the decisive metric as it is specifically designed to take into account the human perception of images, and it is widely used in research to quantify the quality of MRI images. Choosing which reconstructions have the best SSIM is partly subjective, as not necessarily the model with the highest average SSIM is the best performing as the variability of the SSIM over the different reconstructions also has to be accounted for.

4.4 Reconstruction of simulated data

To evaluate how well the RIM framework is able to reconstruct the simulated data, $\text{RIM}^{\text{sim}}_{\Lambda=4}$ and $\text{RIM}^{\text{sim}}_{\Lambda=8}$, were trained with data generated from $\text{Data}^{\text{sim}}_{\text{train}}$ and were used to reconstruct data generated from $\text{Data}^{\text{sim}}_{\text{test}}$. The hyperparameters of the training were set to the optimal hyperparameters determined in Experiment 4.3. During the training, 10 slices were reconstructed from each virtual brain phantom for each training epoch. Both models were trained until the training loss decreased less than 2e-5 over 50 epochs. The models of the final training epoch were used to reconstruct data generated from $\text{Data}^{\text{sim}}_{\text{test}}$. Note that in this experiment different models were used then the ones determined in Experiment 4.3. The reconstruction of a single simulated weighted image was compared visually with zero-filled and TV-ESPIRiT reconstructions. Additionally, the NMSE, PSNR and SSIM of the reconstructed images of each model were evaluated.

4.5 Reconstruction of in-vivo data

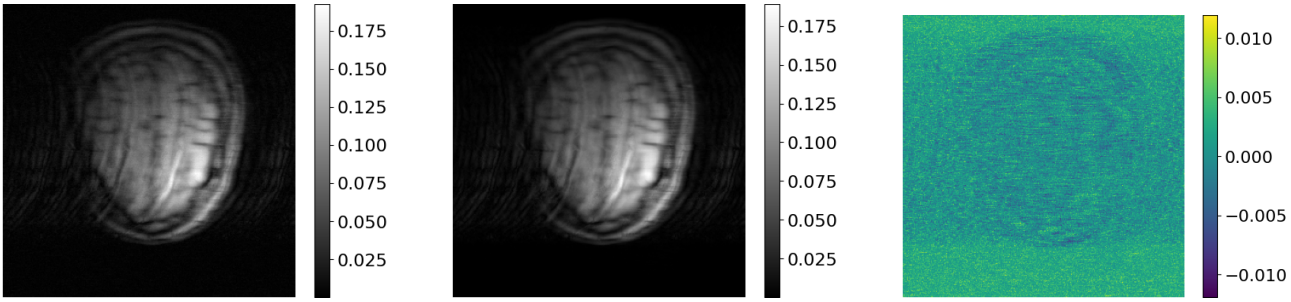
In this experiment the quality of in-vivo reconstructions of the models determined in Experiment 4.3 ($\text{RIM}^{\text{in-vivo}}_{\Lambda=4}$, $\text{RIM}^{\text{in-vivo}}_{\Lambda=8}$) is evaluated. The baseline RIMs which were used as a reference were trained with $\text{Data}^{\text{in-vivo}}_{\text{train}}$. At each training epoch 10 slices were reconstructed from each subject. Both models were trained until the training loss decreased less than 2e-5 over 50 epochs. The models of the final training epoch were used as $\text{RIM}^{\text{baseline}}_{\Lambda=4}$ and $\text{RIM}^{\text{baseline}}_{\Lambda=8}$. The data from $\text{Data}^{\text{in-vivo}}_{\text{test}}$ was reconstructed by the $\text{RIM}^{\text{in-vivo}}$, $\text{RIM}^{\text{baseline}}$ and TV-ESPIRiT methods. The reconstruction of an identical slice from all the different methods was compared visually. Additionally, the NMSE, PSNR and SSIM of all the reconstructions for these methods were compared with each other and the reconstruction metrics of the i-RIM reported in Putzky et al., 2019.

5

Results

5.1 Validation of the forward model

Figure 5.1 shows the results of the experiment for validating the forward model. Figure 5.1a show the coil weighted images reconstructed from the acquired k-space ($W_c^{dataset}$) and Figure 5.1b shows the coil weighted image generated by using the forward model of Equation 2.1 ($W_c^{forward}$) The difference between $W_c^{dataset}$ and $W_c^{forward}$ can be seen in Figure 5.1c. The pattern in Figure 5.1c looks similar to Gaussian noise. The standard deviation of the difference image is 2.4e-3. This is in the same order of magnitude as the standard deviation of the background of $W_c^{dataset}$, which is 1.8e-3.



(a) Acquired undersampled in-vivo coil weighted image

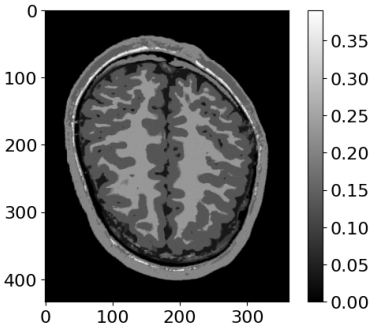
(b) Simulated undersampled in-vivo coil weighted image

(c) Difference image

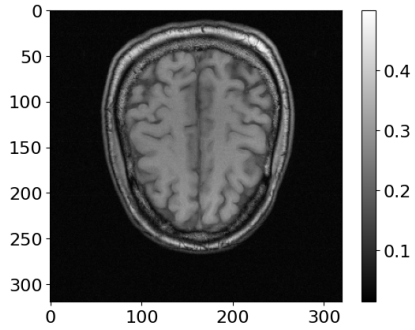
Figure 5.1: (a) Acquired undersampled in-vivo coil weighted image reconstructed from a retrospectively undersampled coil k-space, which was undersampled by multiplying it with a simulated undersampling mask for $\Lambda = 4$. (b) Simulated undersampled in-vivo coil weighted image generated by the forward model in Equation 2.1. The forward model was used to generate an undersampled coil weighted image from the ground truth image. The coil sensitivity maps generated by ESPIRiT were used to create a single coil image from the ground truth image. Subsequently, a undersampled k-space was reconstructed using the Fourier transform and the simulated undersampling mask for $\Lambda = 4$. (c) The difference between the two different coil weighted images, calculated by subtracting Figure 5.1b from Figure 5.1a

5.2 Validation of simulated weighted images

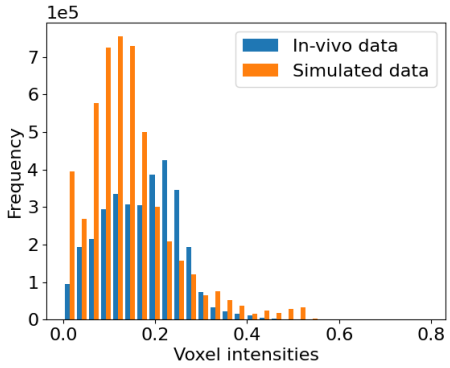
Figure 5.2a shows a magnitude weighted image simulated with the method described in Section 3.3.1. From Figure 5.2b, the observation can be made that, in general, the simulated dataset has a more uniform distribution of intensity values (within a given tissue) than the in-vivo data. Additionally, the interface between white matter and gray matter is sharper in the simulated dataset than in the in-vivo dataset. Figure 5.2c shows the distribution



(a) Magnitude weighted image, simulated from a virtual brain phantom



(b) In-vivo magnitude weighted image from the fastMRI data set



(c) Distribution of the pixel intensities of 50 magnitude weighted images for simulated weighted images and the in-vivo weighted images.

Figure 5.2: Results from the experiment used to evaluate how well the simulated data matches the in-vivo data

of the pixel intensities for 50 different weighted images from the simulated data set and the in-vivo data set. From this distribution it can be observed that most pixel intensities for the simulated data are in the range [0, 0.55], for the in-vivo data this is [0, 0.45]. The peak of the distribution of the simulated data is at 0.15 and for the in-vivo data at 0.25.

5.3 Optimal RIM model: hyperparameter tuning

Figure 5.5-5.6 show the NMSE, PSNR and SSIM of the reconstructed in-vivo images from $\text{Data}_{\text{val}}^{\text{in-vivo}}$ for different training epochs and for acceleration factors $\Lambda = 4$ and $\Lambda = 8$. The reconstruction metrics in between 600 and 1400 epochs are not included in any of the figures, as the reconstruction performance stagnates after 600 epochs and for some metrics even becomes worse. This effect can be observed when comparing the metrics of the first 600 training epochs with the metrics of the models trained for 1400 epochs. A consequence of the deteriorating effect of training for many epochs is that all the best performing models were trained for 600 or fewer epochs. For $\Lambda = 4\times$ (Figure 5.5 and 5.4), the largest improvement in the reconstruction metrics is achieved in the first 400 training epochs. For $\Lambda = 8$ (Figure 5.5 and 5.6), the NMSE does not decrease significantly when training the network. For the PSNR and SSIM, training does increase the performance slightly. However, similarly to $\Lambda = 4$, no big improvements are observed in any of the metrics after 400 epochs. The models with the highest SSIM for each acceleration factor and the corresponding training epoch can be found in Table 5.1. These models were used for the reconstruction of in-vivo data in Experiment 4.5.

Model	learning rate l	number of channels ψ	Epoch
$\text{RIM}^{\text{in-vivo}}_{\Lambda=4}$	2e-3	60	400
$\text{RIM}^{\text{in-vivo}}_{\Lambda=8}$	4e-3	80	200

Table 5.1: Optimal parameters corresponding with the models which reconstruct the images with the highest SSIM

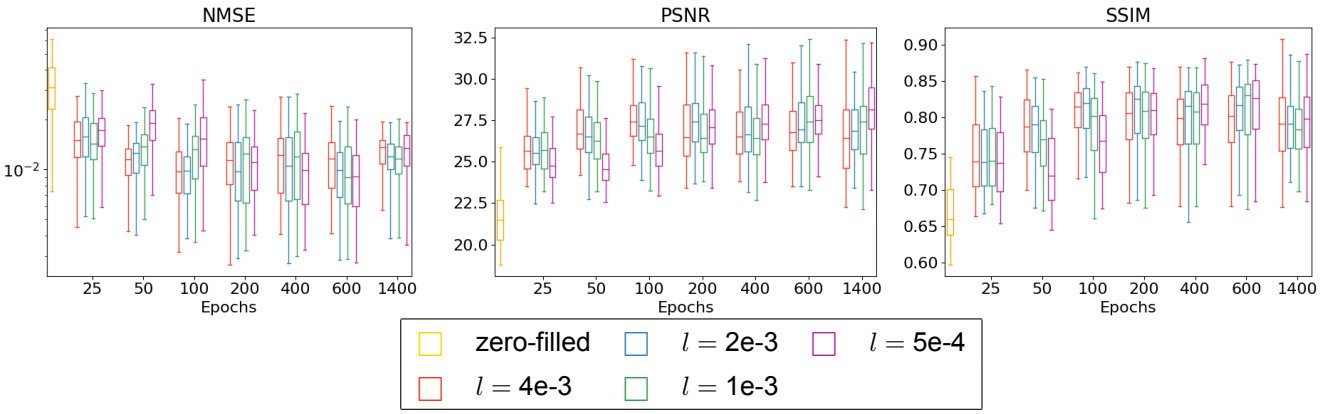


Figure 5.3: Boxplots of the NMSE, PSRN and SSIM of reconstructions of different RIMs for $\Lambda = 4$. The RIMs were trained with $\text{Data}^{\text{sim}}_{\text{train}}$ and reconstructed data from $\text{Data}^{\text{in-vivo}}_{\text{val}}$. The RIMs were evaluated at different training epochs. The RIMs were trained with a constant learning channel size $\psi = 95$ and each RIM was trained with a different l

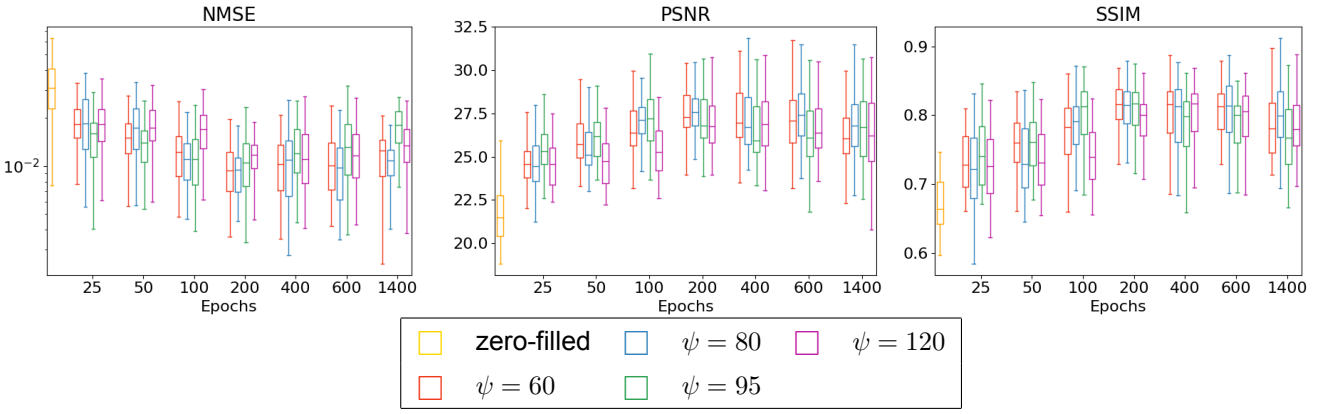


Figure 5.4: Boxplots of the NMSE, PSRN and SSIM of reconstructions of different RIMs for $\Lambda = 4$. The RIMs were trained with $\text{Data}^{\text{sim}}_{\text{train}}$ and reconstructed data from $\text{Data}^{\text{in-vivo}}_{\text{val}}$. The RIMs were evaluated at different training epochs. The RIMs were trained with a constant learning rate $l =$ 2e-3 and each RIM was trained with a different number of channels ψ

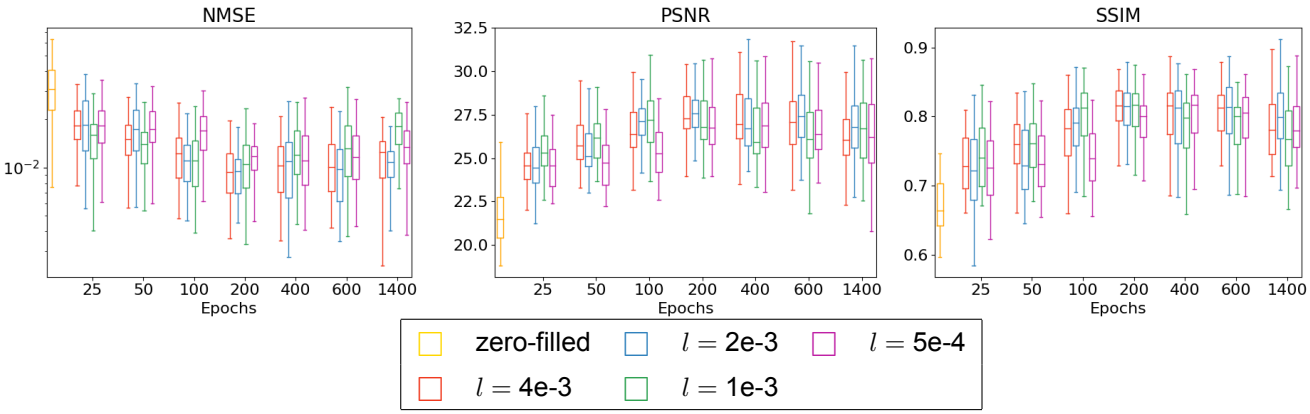


Figure 5.5: Boxplots of the NMSE, PSRN and SSIM of reconstructions of different RIMs for $\Lambda = 8$. The RIMs were trained with $\text{Data}_{\text{train}}^{\text{sim}}$ and reconstructed data from $\text{Data}_{\text{val}}^{\text{in-vivo}}$. The RIMs were evaluated at different training epochs. The RIMs were trained with a constant learning channel size $\psi = 95$ and each RIM was trained with a different l

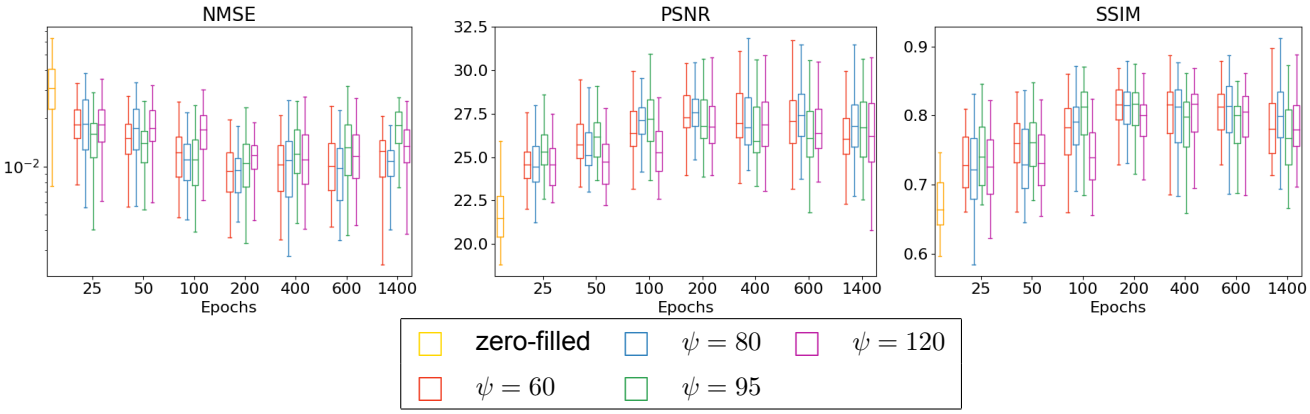


Figure 5.6: Boxplots of the NMSE, PSRN and SSIM of reconstructions of different RIMs for $\Lambda = 8$. The RIMs were trained with $\text{Data}_{\text{train}}^{\text{sim}}$ and reconstructed data from $\text{Data}_{\text{val}}^{\text{in-vivo}}$. The RIMs were evaluated at different training epochs. The RIMs were trained with a constant learning rate $=$ 2e-3 and each RIM was trained with a different number of channels ψ

5.4 Reconstruction of simulated data

Figure 5.7 shows the zero-filled, RIM^{sim} and TV-ESPIRiT reconstructions of $\text{Data}_{\text{test}}^{\text{sim}}$. For $\Lambda = 4$, $\text{RIM}_{\Lambda=4}^{\text{sim}}$ eliminates most ghosting artefacts and the error depicted in the difference image is lower then TV-ESPIRiT. For TV-ESPIRiT reconstruction, ghosting artefacts of the outer tissue of the head can be seen in the brain tissue. These are not present in the RIM^{sim} reconstruction. For $\Lambda = 8$, the reconstruction of the $\text{RIM}_{\Lambda=8}^{\text{sim}}$ still contain many artefacts. Ghosting artefacts are present in the top middle of the image, characterized by the repeating pattern in the difference image. The difference images show that the error of TV-ESPIRiT is larger then $\text{RIM}_{\Lambda=8}^{\text{sim}}$. The evaluation metrics shown in Figure 5.8 support the observations made for both accelerations, where for all metrics the reconstructions of RIM^{sim} are better then the zero-filled and TV-ESPIRiT reconstructions.

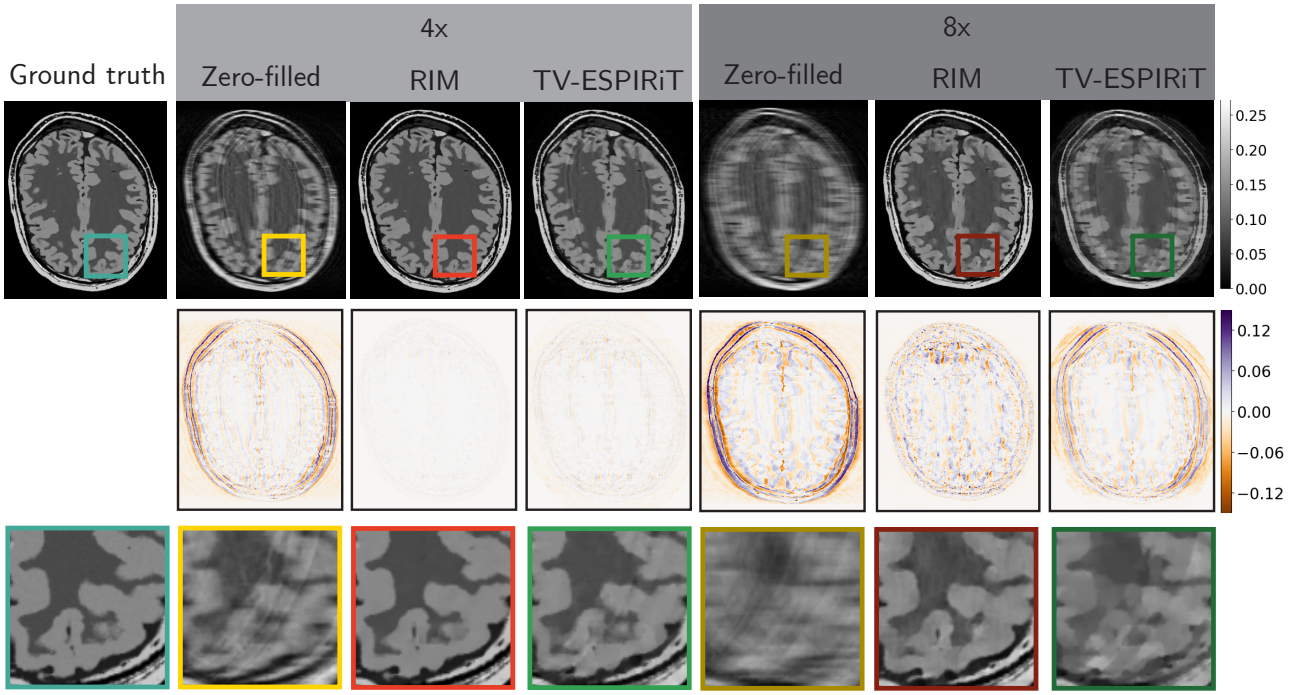


Figure 5.7: Zero-filled and RIM^{sim} reconstructions of simulated data for $\Lambda = 4$ and $\Lambda = 8$. The image directly under the reconstruction shows the ground truth image minus the reconstruction. In the bottom row, a close-up is shown of the highlighted area in the ground truth image and the reconstructions

5.5 Reconstruction of in-vivo data

Figure 5.9, depicts reconstructions for $\Lambda = 4$ and $\Lambda = 8$ of a single slice from $\text{Data}^{\text{in-vivo}}_{\text{test}}$, reconstructed using $\text{RIM}^{\text{in-vivo}}$, $\text{RIM}^{\text{baseline}}$ and TV-ESPIRiT. For $\Lambda = 4$, the reconstructions of $\text{RIM}^{\text{in-vivo}}_{\Lambda=4}$ and TV-ESPIRiT both contain similar ghosting artefacts. These artefacts are not visible in the reconstruction of $\text{RIM}^{\text{baseline}}_{\Lambda=4}$. The difference image of $\text{RIM}^{\text{in-vivo}}_{\Lambda=4}$ shows larger errors when compared to the reconstructions of $\text{RIM}^{\text{baseline}}_{\Lambda=4}$ and TV-ESPIRiT.

For $\Lambda = 8$, the reconstructions of every method show many artefacts. The difference image of $\text{RIM}^{\text{in-vivo}}_{\Lambda=8}$ shows more voxels with high errors then the difference image of $\text{RIM}^{\text{baseline}}_{\Lambda=8}$ and TV-ESPIRiT. The difference image of $\text{RIM}^{\text{baseline}}_{\Lambda=8}$ shows the lowest errors of all the different methods.

Figure 5.10 shows the evaluation metrics for the reconstructions of $\text{Data}^{\text{in-vivo}}_{\text{test}}$ for the different reconstruction methods. The averages and 95% confidence intervals of these metrics can be found in Table 5.2. For $\Lambda = 4$, all metrics of $\text{RIM}^{\text{in-vivo}}_{\Lambda=4}$ are higher then the zero-filled reconstructions. When looking at the SSIM, the reconstructions of $\text{RIM}^{\text{in-vivo}}_{\Lambda=4}$ are similar to TV-ESPIRiT reconstructions, however, for the other metrics (NSME, PSNR) TV-ESPIRiT reconstructions are better. $\text{RIM}^{\text{baseline}}_{\Lambda=4}$ outperforms the other methods for every metric. For $\Lambda = 8$, the PSNR and SSIM of $\text{RIM}^{\text{in-vivo}}_{\Lambda=8}$ is better then the zero-filled reconstruction, however, for the NMSE almost no improvement is observed. TV-ESPIRiT is better then $\text{RIM}^{\text{in-vivo}}_{\Lambda=8}$ in every metric. Again, for every metric the reconstructions of $\text{RIM}^{\text{baseline}}_{\Lambda=8}$ are the best of any method. When looking a the performance metrics of the i-RIM in Table 5.2, for $\Lambda = 8$, the difference between the i-RIM and the baseline RIM is small, the NMSE of the reconstructions of the baseline RIM is even better. However, for $\Lambda = 8$, the difference between i-RIM and $\text{RIM}^{\text{baseline}}_{\Lambda=8}$

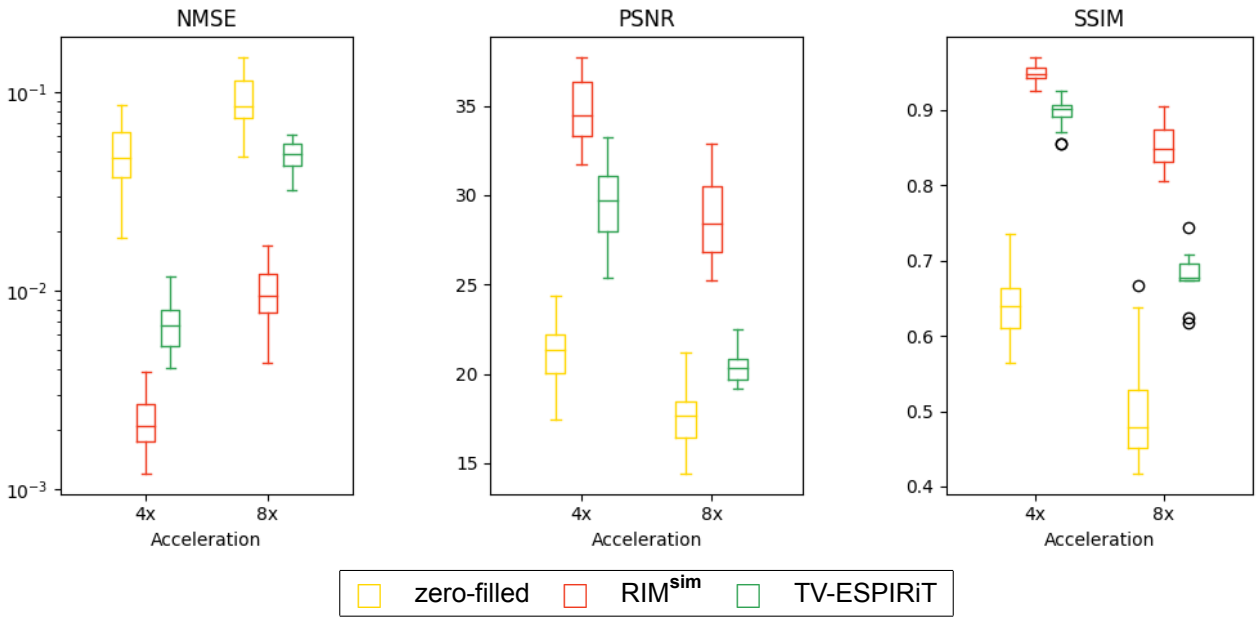


Figure 5.8: Results of the reconstruction of simulated validation data with RIM^{sim}. The distribution of the NMSE, PSRN and SSIM are shown in boxplots for $\Lambda = 4$ and $\Lambda = 8$. These are compared with the zero-filled reconstruction of the same data. The box extends from the lower to upper quartile values of the measured metrics, with a line at the median. The whiskers extend from the box to show the minimum and maximum values for each metric.

is large.

For both acceleration factors, the reconstructed phases of the weighted images shown in Figure 5.9 can be seen in Figure 5.11. The differences between the individual phase reconstructions are not easy to identify. However, the difference maps show that all reconstruction methods produce the highest errors in the areas in which the phase is inconsistent. For this specific weighted image, that area is near the skull. For both acceleration factors the difference is the largest for the reconstructions of RIM$^{in\text{-}vivo}$.

Metric	Λ	zero-filled	RIM$^{in\text{-}vivo}$	RIM$^{in\text{-}vivo}$	TV-ESPIRiT	i-RIM
NMSE ($\times 10^{-2}$)	4	3.1 (2.7, 3.5)	1.1 (8.9, 1.3)	0.38 (0.29, 0.48)	0.62 (0.57, 0.68)	0.62
	8	6.6 (5.8, 7.4)	5.6 (4.7, 6.6)	1.9 (1.6, 2.2)	2.9 (2.5, 3.3)	1.03
PSNR	4	21.7 (21.2, 22.3)	27.3 (26.6, 28.0)	32.1 (31.3, 32.8)	29.3 (28.8, 29.7)	38.84
	8	17.7 (17.3, 18.1)	20.7 (20.0, 21.3)	24.8 (24.2, 25.4)	22.1 (21.5, 22.7)	36.19
SSIM	4	0.66 (0.65, 0.68)	0.81 (0.80, 0.82)	0.89 (0.88, 0.90)	0.82 (0.81, 0.83)	0.91
	8	0.48 (0.46, 0.49)	0.57 (0.56, 0.59)	0.71 (0.69, 0.73)	0.65 (0.63, 0.66)	0.89

Table 5.2: Results of the reconstruction of in-vivo testing data with RIM$^{in\text{-}vivo}$ compared with the reconstructions of RIM^{baseline}, TV-ESPIRiT and i-RIM. The values show are the average and the 95% intervals of the NMSE, PSRN and SSIM for $\Lambda = 4$ and $\Lambda = 8$. For the evaluation metrics of the i-RIM, the 95% intervals were not reported in the paper

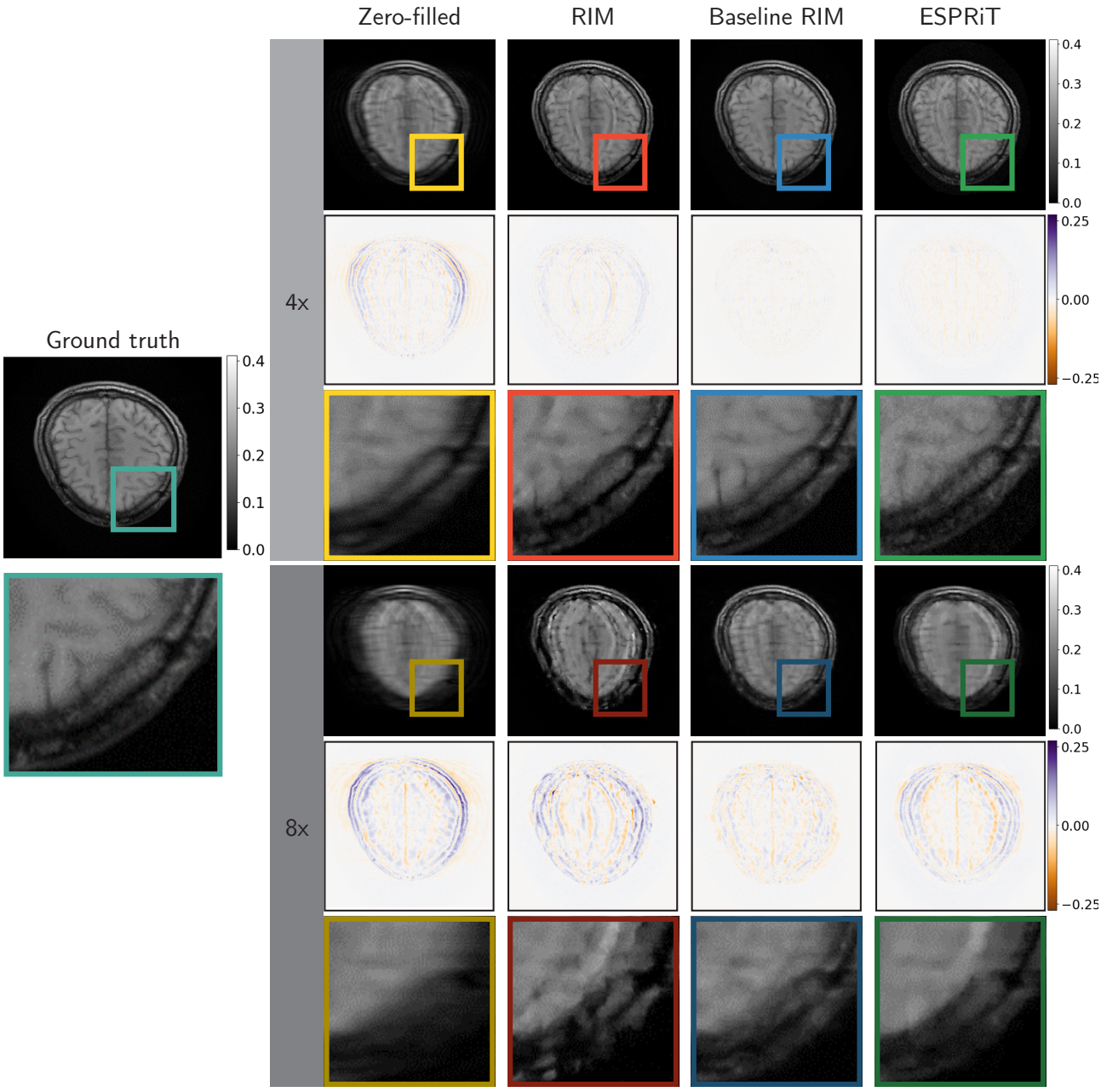


Figure 5.9: Reconstructions of a single slice from $\text{Data}_{\text{test}}^{\text{in-vivo}}$ using different reconstruction methods for $\Lambda = 4$ and $\Lambda = 8$. The reconstruction methods used are: zero-filled, $\text{RIM}^{\text{in-vivo}}$, $\text{RIM}^{\text{baseline}}$ and TV-ESPIRiT. Below each reconstruction a close-up is shown of the highlighted area with the corresponding color.

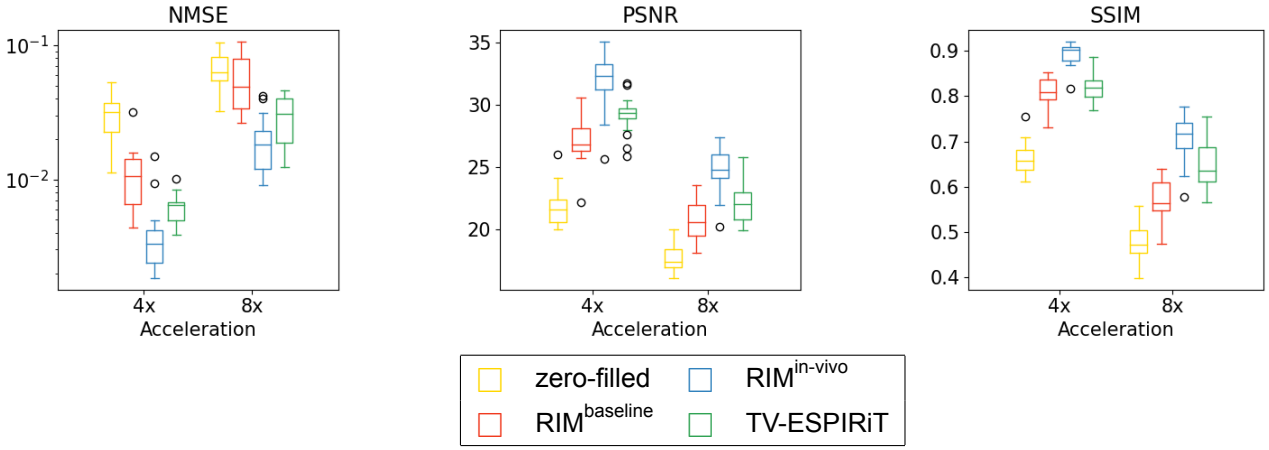


Figure 5.10: Results of the reconstruction of in-vivo testing data with RIM$^{\text{in-vivo}}$. The NMSE, PSRN and SSIM are shown in boxplots for $\Lambda = 4$ and $\Lambda = 8$. These are compared with the zero-filled, RIM$^{\text{baseline}}$ an TV-ESPIRiT reconstructions of the same data. The box extends from the lower to upper quartile values of the measured metrics, with a line at the median. The whiskers that extend from the box to show the minimum and maximum values for each metric.

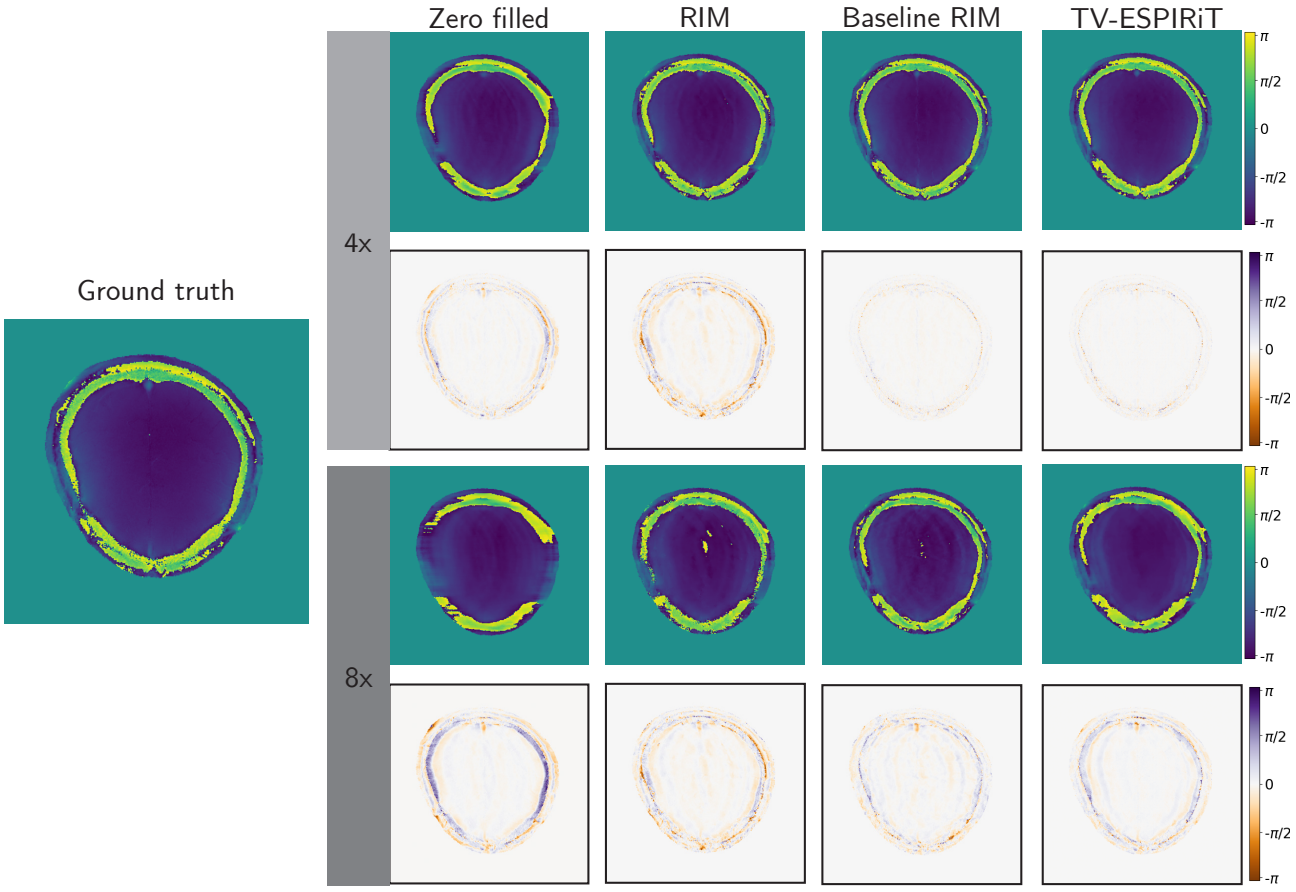


Figure 5.11: Phase maps which were reconstructed from the reconstructed weighted images depicted in Figure 5.9. The phase maps of the reconstruction of different methods is shown in the first and second row. Under each phase map the difference with the ground truth phase is depicted.

6

Discussion

6.1 Validation of forward model

The standard deviation of the difference image shown in Figure 5.1 was in the same order of magnitude as the standard deviation of the background $W_c^{dataset}$. From this the conclusion can be made that the forward model accurately represents the real-life measurements made by a single receiver coil. This is important as the RIM uses this forward model for the optimization, thus, if it is inaccurate, the RIM reconstruction performance is limited.

6.2 Reconstruction of simulated data

The results of Experiment 4.4 show that, for the simulated data used in this research and $\Lambda = 4$, the RIM framework is a good reconstruction method. The reconstructions of $\text{RIM}_{\Lambda=4}^{\text{sim}}$ were better then TV-ESPIRiT and had almost no artefacts. However, for $\Lambda = 8$ the RIM did not perform well. The reconstructions still showed many large artefacts. A reason for the worse performance of the RIM for $\Lambda = 8$ is that the likelihood function evaluates less k-space points, which makes the optimization performed by the RIM harder. Additionally, the initial weighted image computed from the undersampled k-space is further from the ground truth, again making the optimization harder. The performance of the RIM could be improved by increasing the network complexity.

6.3 Reconstruction of in-vivo data

From the results of Experiment 4.5 it can be concluded that for $\Lambda = 4$, the RIM learns an optimization method from the training with simulated data, that can also be used to reconstruct $\text{Data}_{\text{val}}^{\text{in-vivo}}$. However, for $\Lambda = 8$, the RIM did not learn an optimization that improved the in-vivo reconstructions. The results of Experiment 4.5 also show that the training of the RIM with simulated data does not successfully substitute the training with in-vivo data. Even though the RIM should be able to generalize well across different types of data, this does not seem to be the case for the method proposed in this research. This behaviour could be caused by multiple factors.

6.4 Limiting factors

6.4.1 Simulated acquisition noise

One factor that could have negatively impacted the generalisation of the proposed method is the the standard deviation of the acquisition noise, which was constant during the training. Changing this to pick an acquisition noise from a range of values could improve the generalisation.

6.4.2 Scanning sequence

A limitation of the proposed method could be the discrepancies between the sequence model used to generate the simulated data and the pulse sequence used to acquire the in-vivo dataset. To simulate the training data, a Fast Spin Echo sequence was used. However, the in-vivo data is acquired using a FLASH sequence, for which a more accurate model is described in Winkler et al., 1988. The models are identical for a flip angle of 90°; however, the in-vivo data set is acquired using a flip angle of 70°. As a result of this, the images simulated in this research are more T1 weighted than the images in the in-vivo dataset. Changing the sequence model to the one described for FLASH sequences would result in more accurate simulated images that might improve the the performance of the proposed method.

6.4.3 Simulated weighted image

Even though the RIM should generalize well for unseen data, differences between the appearance of the weighted images could still play an important role in the quality of reconstructions. One such difference, observed in Experiment 4.2, is the mismatch of the distributions of the pixel intensities shown in Figure 5.2c. Another difference which could influence the performance, is the uniform distribution of intensity values (within a given tissue) of the simulated weighted image when compared to the in-vivo data. The cause of this discrepancy is that, in real tissue, tissue properties, such as the relaxation times and protons density, are not uniform throughout one tissue type. Additionally, magnetic homogeneity also plays a role in changing the contrast across the image. Even if the tissue properties are uniform throughout the tissue, the corresponding tissue in the weighted image might have varying intensities due to these magnetic imperfections. Another difference between the datasets which could have an influence, is the sharper border between the different type of tissues observed in the simulated data.

6.4.4 Phase simulation

The phase simulation method is another property of the simulated test data that could be improved to match the in-vivo data better. Currently, the phase is uniform throughout the simulated weighted image; however, in the in-vivo data, the phase varies. The results of Experiment 4.5 showed that, when compared with the other methods, the phase error was the largest for RIM^{sim}. Thus, strategies for a more realistic phase map, such as varying phase maps or using phase maps extracted from in-vivo data, could increase the performance of the RIM.

6.4.5 Network complexity

The bad generalization of the proposed method could be caused by lacking network complexity of the RIM. A more complex neural network could increase performance of the proposed methods as more network parameters are available that could learn the generalization. Thus, increasing the network complexity should yield better performance. However, the network complexity is limited by the high load on the hardware memory. Strategies that could be explored to increase the RIM complexity include training the RIM on GPUs with higher memory capacity or replacing the current RIM framework with the framework of the i-RIM, as the i-RIM allows for more complex network architectures.

6.5 Initialisation of the weighted image

If the initialisation of the weighted image is closer to the ground truth image, the optimisation becomes easier for the RIM, as the initial image is closer to the global minimum. Currently, the initial weighted image is the zero-filled reconstruction. This could be replaced by reconstructions from a conventional reconstruction method, such as

the GRAPPA, SENSE or TV-ESPIRiT. However, this will come with a disadvantage. As these reconstructions are slower than zero-filled reconstructions, especially for TV-ESPIRiT, initialising the weighted images with these methods would significantly increase the training and reconstruction time of the RIM.

6.6 Transfer learning

The problem of limited training data could be resolved by pre-training the network with simulated data and fine-tuning the RIM by training it with a smaller set of in-vivo training data. Acquiring a smaller dataset is cheaper and faster and the chance is higher that the a large enough data set has already been acquired.

7

Conclusion

In this research, the viability of using simulated data to train a Recurrent Inference Machine for accelerated MRI, was evaluated. The motivation for the use of simulated training data is the limited amount of in-vivo training data for uncommon scanning sequences. The RIM framework was chosen because it was shown to reconstruct high quality MRI images from accelerated scans. Additionally, this deep learning framework shows good generalisability to different image contrasts. For low acceleration factors (4x), the proposed method improved the reconstruction quality over zero-filled reconstructions and for performed on par with the TV-ESPIRiT method. However, the reconstruction quality of the proposed method was inferior to the baseline RIM trained with in-vivo data. The reconstructions from the framework proposed in this research show that for acceleration factors of 4, the RIM does generate better reconstructions when compared to the zero-filled reconstructions. However, when comparing the reconstructions with reconstructions from a baseline RIM trained with in-vivo data and the TV-ESPIRiT method, the quality of reconstructions of the RIM trained with simulated data is worse. A possible reason for this is that the data simulation method does not produce weighted images that accurately represent the in-vivo data. Another possible cause is that the RIM has too low complexity. To properly assess if simulated training data can replace in-vivo training data, further research is required that focuses on improving the simulated data and/or increasing the RIM complexity. Another strategy could be replacing the current framework with a framework that is more suited for accelerated MRI reconstruction and that still is able to generalize well for for unseen data, such as the i-RIM.

Bibliography

Ahishakiye, E., Van Gijzen, M. B., Tumwiine, J., Wario, R., & Obungoloch, J. (2021). A survey on deep learning in medical image reconstruction. *Intelligent Medicine*.

Aletras, A. H. Basic MRI physics. In: *Cardiovascular magnetic resonance imaging*. Springer, 2008, pp. 1–31.

Bassett, R., & Deride, J. (2019). Maximum a posteriori estimators as a limit of bayes estimators. *Mathematical Programming*, *174*(1), 129–144.

Bloch, F. (1946). Nuclear induction. *Physical review*, *70*(7-8), 460.

Bobman, S. A., Riederer, S. J., Lee, J. N., Suddarth, S. A., Wang, H. Z., Drayer, B. P., & MacFall, J. R. (1985). Cerebral magnetic resonance image synthesis. *American Journal of Neuroradiology*, *6*(2), 265–269.

Bojorquez, J. Z., Bricq, S., Acquitter, C., Brunotte, F., Walker, P. M., & Lalande, A. (2017). What are normal relaxation times of tissues at 3 T? *Magnetic resonance imaging*, *35*, 69–80.

Boubela, R. N., Kalcher, K., Nasel, C., & Moser, E. (2014). Scanning fast and slow: Current limitations of 3 tesla functional MRI and future potential. *Frontiers in physics*, *2*, 1.

Boyd, S., Parikh, N., & Chu, E. (2011). *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc.

Bredies, K., Kunisch, K., & Pock, T. (2010). Total generalized variation. *SIAM Journal on Imaging Sciences*, *3*(3), 492–526.

Chen, J., Chang, E. Y., Carl, M., Ma, Y., Shao, H., Chen, B., Wu, Z., & Du, J. (2017). Measurement of bound and pore water T1 relaxation times in cortical bone using three-dimensional ultrashort echo time cones sequences. *Magnetic resonance in medicine*, *77*(6), 2136–2145.

Chunli, W., Xiaowan, L., Cuili, L., & Shuo, L. An improved total variation regularized sense reconstruction for mri images. In: *2017 29th chinese control and decision conference (ccdc)*. IEEE. 2017, 5005–5009.

Cocosco, C. A., Kollokian, V., Kwan, R. K.-S., Pike, G. B., & Evans, A. C. Brainweb: Online interface to a 3D MRI simulated brain database. In: *Neuroimage*. Citeseer. 1997.

Cohen, M. S., & Weisskoff, R. M. (1991). Ultra-fast imaging. *Magnetic resonance imaging*, *9*(1), 1–37.

De Bazelaire, C. M., Duhamel, G. D., Rofsky, N. M., & Alsop, D. C. (2004). MR imaging relaxation times of abdominal and pelvic tissues measured in vivo at 3.0 t: Preliminary results. *Radiology*, *230*(3), 652–659.

Deshmane, A., Gulani, V., Griswold, M. A., & Seiberlich, N. (2012). Parallel MR imaging. *Journal of Magnetic Resonance Imaging*, *36*(1), 55–72.

Elster, A. D. (2021). *Questions and answers in MRI*. https://mri-q.com/index.html

Geerts-Ossevoort L, D. A., de Weerdt E. (2018). Compressed SENSE: Speed done right every time.

Griswold, M. A., Jakob, P. M., Heidemann, R. M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., & Haase, A. (2002). Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *47*(6), 1202–1210.

Guerquin-Kern, M., Van De Ville, D., Vonesch, C., Baritaux, J.-C., Pruessmann, K. P., & Unser, M. Wavelet-regularized reconstruction for rapid MRI. In: *2009 ieee international symposium on biomedical imaging: From nano to macro*. IEEE. 2009, 193–196.

Haase, A. (1990). Snapshot flash mri. applications to t1, t2, and chemical-shift imaging. *Magnetic Resonance in Medicine*, *13*(1), 77–89.

Hagiwara, A., Warntjes, M., Hori, M., Andica, C., Nakazawa, M., Kumamaru, K. K., Abe, O., & Aoki, S. (2017). SyMRI of the brain: rapid quantification of relaxation rates and proton density, with synthetic MRI, automatic brain segmentation, and myelin measurement. *Investigative radiology*, *52*(10), 647.

Heidemann, R. M., Özsarlak, Ö., Parizel, P. M., Michiels, J., Kiefer, B., Jellus, V., Müller, M., Breuer, F., Blaimer, M., Griswold, M. A., et al. (2003). A brief review of parallel magnetic resonance imaging. *European radiology*, *13*(10), 2323–2337.

Hu, Y., Huber, A., Anumula, J., & Liu, S.-C. (2018). Overcoming the vanishing gradient problem in plain recurrent networks. *arXiv preprint arXiv:1801.06105*.

Jin, K. H., McCann, M. T., Froustey, E., & Unser, M. (2017). Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, *26*(9), 4509–4522.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Koikkalainen, J., Rhodius-Meester, H., Tolonen, A., Barkhof, F., Tijms, B., Lemstra, A. W., Tong, T., Guerrero, R., Schuh, A., Ledig, C., et al. (2016). Differential diagnosis of neurodegenerative diseases using structural MRI data. *NeuroImage: Clinical*, *11*, 435–449.

Lebanon, G. (2010). Bias, variance, and mse of estimators. *Georgia Institute of Technology. Atlanta*.

Lønning, K., Putzky, P., Caan, M. W., & Welling, M. (2018). Recurrent inference machines for accelerated MRI reconstruction.

Lustig, M., Donoho, D., & Pauly, J. M. (2007). Sparse mri: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *58*(6), 1182–1195.

Macovski, A. (1996). Noise in MRI. *Magnetic resonance in medicine*, *36*(3), 494–497.

Morain-Nicolier, F., Landré, J. et al. (2009). Gray level local dissimilarity map and global dissimilarity index for quality of medical images. *IFAC Proceedings Volumes*, *42*(12), 281–286.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology*, *47*(1), 90–100.

Pruessmann, K. P., Weiger, M., Scheidegger, M. B., & Boesiger, P. (1999). SENSE: sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *42*(5), 952–962.

Putzky, P., Karkalousos, D., Teuwen, J., Miriakov, N., Bakker, B., Caan, M., & Welling, M. (2019). i-RIM applied to the fastMRI challenge. *arXiv preprint arXiv:1910.08952*.

Putzky, P., & Welling, M. (2017). Recurrent inference machines for solving inverse problems. *arXiv preprint arXiv:1706.04008*.

Putzky, P., & Welling, M. (2019). Invert to learn to invert. *Advances in Neural Information Processing Systems*, *32*, 446–456.

Qu, X., Zhang, W., Guo, D., Cai, C., Cai, S., & Chen, Z. (2010). Iterative thresholding compressed sensing MRI based on contourlet transform. *Inverse Problems in Science and Engineering*, *18*(6), 737–758.

Rocca, M. A., Battaglini, M., Benedict, R. H., De Stefano, N., Geurts, J. J., Henry, R. G., Horsfield, M. A., Jenkinson, M., Pagani, E., & Filippi, M. (2017). Brain MRI atrophy quantification in MS: from methods to clinical application. *Neurology*, *88*(4), 403–413.

Ronneberger, O., Fischer, P., & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2015, 234–241.

Sabidussi, E. R., Klein, S., Caan, M. W., Bazrafkan, S., den Dekker, A. J., Sijbers, J., Niessen, W. J., & Poot, D. H. (2021). Recurrent inference machines as inverse problem solvers for MR relaxometry. *arXiv preprint arXiv:2106.07379*.

Schnell, R., Hill, P. B., Esser, E., et al. (1999). Methoden der empirischen sozialforschung.

Shao, Y. (2016). PAPR reduction and receiver sensitivity improvement in 16QAM-OFDM RoF system using DMT modulation and BTN-PS technique. *Optics & Laser Technology*, *77*, 59–63.

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. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *54*(3), 507–512.

Uecker, M., Lai, P., Murphy, M. J., Virtue, P., Elad, M., Pauly, J. M., Vasanawala, S. S., & Lustig, M. (2014). ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. *Magnetic resonance in medicine*, *71*(3), 990–1001.

Uecker, M., Ong, F., Tamir, J. I., Bahri, D., Virtue, P., Cheng, J. Y., Zhang, T., & Lustig, M. Berkeley advanced reconstruction toolbox. In: *Proc. intl. soc. mag. reson. med*. *23*. (2486). 2015.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Winkler, M. L., Ortendahl, D., Mills, T., Crooks, L., Sheldon, P., Kaufman, L, & Kramer, D. (1988). Characteristics of partial flip angle and gradient reversal MR imaging. *Radiology*, *166*(1), 17–26.

Wright, K. L., Hamilton, J. I., Griswold, M. A., Gulani, V., & Seiberlich, N. (2014). Non-cartesian parallel imaging reconstruction. *Journal of Magnetic Resonance Imaging*, *40*(5), 1022–1040.

Yang, Y., Sun, J., Li, H., & Xu, Z. (2017). ADMM-Net: A deep learning approach for compressive sensing MRI. *arXiv preprint arXiv:1705.06869*.

Young, R. J., & Knopp, E. A. (2006). Brain mri: Tumor evaluation. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *24*(4), 709–724.

Zaitsev, M., Maclaren, J., & Herbst, M. (2015). Motion artifacts in MRI: A complex problem with many partial solutions. *Journal of Magnetic Resonance Imaging*, *42*(4), 887–901.

Zbontar, J., Knoll, F., Sriram, A., Murrell, T., Huang, Z., Muckley, M. J., Defazio, A., Stern, R., Johnson, P., Bruno, M., et al. (2018). fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv preprint arXiv:1811.08839*.