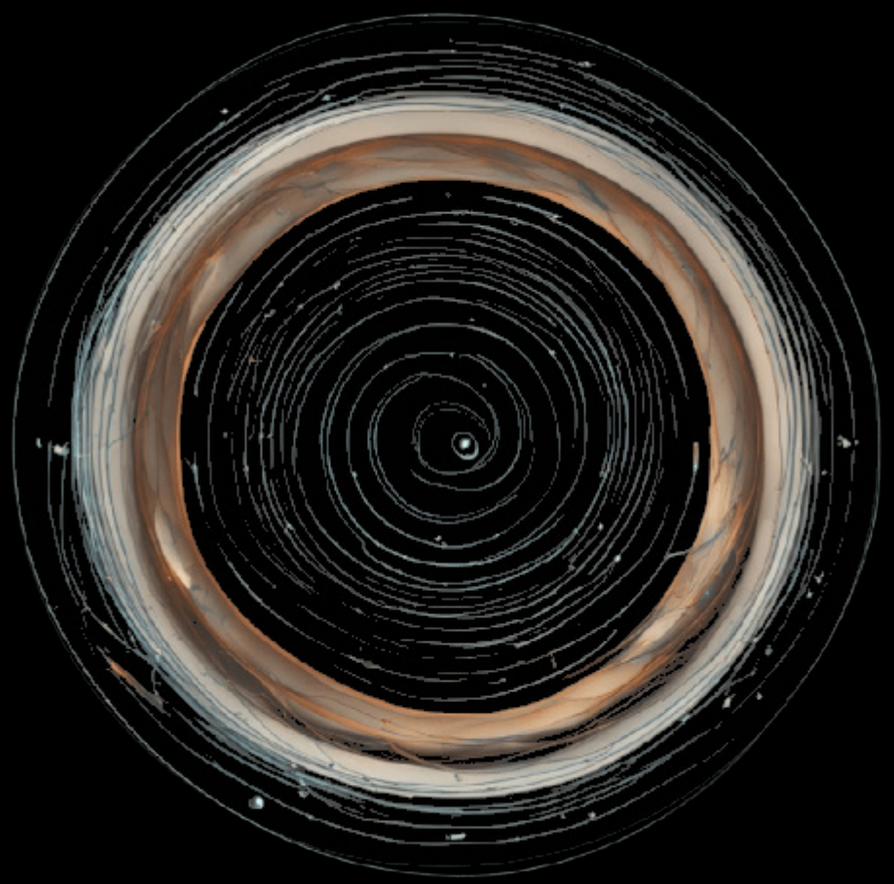
Deep multitask learning for accelerating Magnetic Resonance Imaging

Karkalousos, D.

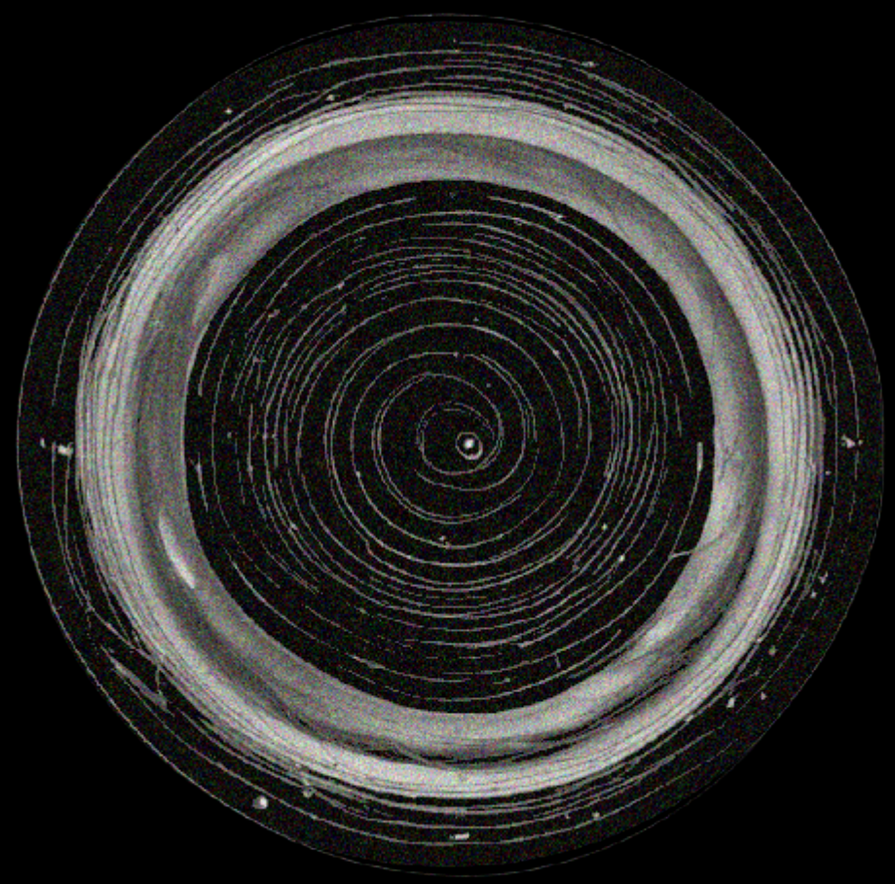
Publication date
2025
Document Version
Final published version



DEEP MULTITASK LEARNING FOR ACCELERATING MAGNETIC RESONANCE IMAGING

Dimitrios Karkalousos

DEEP MULTITASK LEARNING FOR ACCELERATING MAGNETIC RESONANCE IMAGING



Dimitrios Karkalousos

Deep multitask learning for accelerating Magnetic Resonance Imaging

Dimitrios Karkalousos

Deep multitask learning for accelerating Magnetic Resonance Imaging

The research in this thesis was funded by the STAIRS project under the TKI-PPP program. The collaboration project was co-funded by the PPP Allowance made available by Health Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships.

Cover: Dimitrios Karkalousos

Layout: Dimitrios Karkalousos

Deep multitask learning for accelerating Magnetic Resonance Imaging

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op woensdag 26 februari 2025, te 13.00 uur

door Dimitrios Karkalousos
geboren te Athens

Promotiecommissie

Promotores:	prof. dr. H.A. Marquering	AMC-UvA
	prof. dr. I. Išgum	AMC-UvA
Copromotores:	dr. ir. M.W.A. Caan	AMC-UvA
Overige leden:	prof. dr. ir. M. Staring	Universiteit Leiden
	prof. dr. L. Reneman	AMC-UvA
	prof. dr. ir. J.J. Sonke	NKI-AvL
	prof. dr. ir. G.J. Strijkers	AMC-UvA
	dr. E. Gavves	Universiteit van Amsterdam
	dr. ir. D.H.J. Poot	Erasmus MC

Faculteit der Geneeskunde

Contents

1

Introduction

Magnetic Resonance Imaging (MRI) has transformed medical imaging by offering unparalleled non-invasive visualization of soft tissues, contrasting with techniques such as Computed Tomography (CT) that rely on ionizing radiation. By employing strong magnetic fields, typically ranging from 1.5 to 7 Tesla, MRI facilitates the acquisition of images with a high signal-to-noise ratio (SNR), providing detailed spatial resolution of anatomical and pathological structures essential for clinical assessment. The versatility of MRI lies in the employment of a variety of imaging sequences, where parameters such as sequence timing, including the echo and repetition time, flip angle of radio frequency pulses, and the use of inversion pulses determine the contrast of the image. $T_1/T_2/T_2^*$-weighted and fluid-attenuated inversion recovery (FLAIR) contrasts are among the most commonly used modalities, each offering a different representation of anatomy and pathology in the human body.

MRI transcends conventional anatomical imaging, which focuses mainly on structural visualization of tissues and organs, including functional and quantitative assessments for applications in neurological diagnostics, therapeutic monitoring, and biomarker creation. For example, in acute settings, such as stroke diagnosis, MRI can reveal early ischemic changes and penumbral tissue [1], enabling dynamic assessment of tissue viability and guiding time-critical intervention decisions. In chronic conditions such as multiple sclerosis (MS), MRI allows precise quantification of lesion load and distribution [2], facilitating both diagnosis and therapeutic monitoring through longitudinal assessment of disease progression. Furthermore, advanced MRI techniques have emerged as valuable tools for providing insights into pathophysiological mechanisms that cannot be assessed with conventional structural imaging, such as iron quantification for studying neurodegenerative processes and age-related changes in the brain [3],

Image acquisition fundamentally depends on the spatial frequency information in the corresponding domain, known as k-space. During acquisition, the frequency and phase of the measured signal are modulated repetitively by varying readout gradients. These gradients establish a linear relationship between spatial position and resonance frequency. The MR signal is sampled along the frequency-encoding direction. Among the broad spectrum of MRI readout strategies, 2D and 3D Euclidean imaging are the most widely used

in the clinic. In 2D imaging, multiple slices of a single view are acquired sequentially, using one readout direction and one phase-encoding direction. In contrast, 3D imaging acquires a volumetric view across all three anatomical planes (axial, coronal, and sagittal), with a single readout direction and two phase-encoding directions simultaneously acquiring data from an entire volume. The objective of the reconstruction task is to transform the sampled k-space data into a visually interpretable image. This transformation is realized through the inverse Fourier transform.

Nevertheless, the need for continuous sampling in k-space introduces an inherent temporal constraint. MRI acquisition can be accelerated by strategically undersampling the k-space along the phase-encoding dimensions. However, acquiring a sparse representation of the k-space introduces complex aliasing artifacts, as it violates the Nyquist-Shannon sampling theorem, which establishes a minimum sampling rate required to accurately represent a continuous-time signal without aliasing [4]. Balancing this trade-off between acquisition speed and image fidelity presents significant challenges, where the presence of system noise, the noise produced by the scanner during acquisition, further amplifies these fundamental challenges of the reconstruction task.

These combined factors of prolonged acquisition times and image reconstruction challenges can potentially delay treatment decisions and compromise clinical efficiency, crucial in time-critical settings such as acute ischemic stroke. In addition, the requirement for patients to remain still for extended periods to minimize motion artifacts can be particularly challenging for those who are uncomfortable or claustrophobic. The cumulative impact of these technical and practical constraints affects individual patient care while limiting the number of patients that can be examined in a given period, potentially increasing waiting times and reducing overall accessibility to this valuable diagnostic tool.

The overarching aim of this thesis is to accelerate MR imaging while maintaining or improving diagnostic image quality. Central to our research is the formulation of the reconstruction process through a linear forward model, which mathematically describes the relationship between the sampled k-space measurements and the underlying image. This theoretical framework serves as the foundation for understanding the challenges and opportunities inherent in accelerated MRI reconstruction.

1.1 Accelerated MRI Reconstruction

The linear forward measurement model that describes the process of acquiring, undersampling, and transforming k-space signals is mathematically expressed as $y_i = A(x) + \sigma, i = 1, ..., c$, where i denotes a single receiver coil for a total number of c coils. $A : \mathbb{C}^n \mapsto \mathbb{C}^{n \times n_c}$ models the linear forward operator of accelerating an MRI acquisition, with n representing the total number of pixels in the true image x and $\sigma \in \mathbb{C}^n$ denotes the measured white noise from the scanner, assumed to be constant across all coil channels. Consequently, the observed measurement vector $y \mapsto \mathbb{C}^m$ emerges as the result of applying the linear transformation A to the signal vector $x \mapsto \mathbb{C}^n$, with $m \ll n$, further perturbed by an additive noise vector σ.

The forward operator A is given by $A = P \odot \mathcal{F} \odot \epsilon$, with P denoting the undersampling scheme, where the type of imaging sequence (2D or 3D) dictates the specific undersampling scheme used. 2D imaging utilizes 1D undersampling along one phase-encoding direction, while 3D imaging employs undersampling in two phase-encoding directions. The

undersampling patterns, often characterized by random, equidistant, Gaussian, or Poisson distributions, introduce varying degrees of incoherence into the acquired data, influencing the complexity of the subsequent reconstruction task.

Next, in A, $\mathcal{F}$ denotes the Fourier transform and $\epsilon : \mathbb{C}^n \times \mathbb{C}^{n \times n_c} \mapsto \mathbb{C}^{n \times n_c}$ denotes the expand operator, which transforms a single-coil image x into x_c multi-coil images, given by $\epsilon(x) = (S_0 \odot x, ..., S_c \odot x) = (x_0, ..., x_c)$. S denotes the coil sensitivity maps, a matrix representing the spatial sensitivities that scale every voxel in the multi-coil images by a complex number.

The concept of using multiple receiver coils positioned around the target area to simultaneously acquire data from different spatial locations is known as Parallel Imaging (PI) [5]. The introduction of PI marked a significant milestone in accelerating MRI acquisition, effectively reducing scanning times and increasing SNR. However, increased acceleration factors lead to decreased PI efficacy, due to the ill-posed nature of the inverse problem mapping $y \mapsto x$.

An approach to approximate a solution to this inverse problem is through Maximum A Posteriori (MAP) estimation, which leverages prior knowledge of MRI to constrain the solution space. MAP estimation can be formulated as an optimization problem: $x_{MAP} = \arg\max_x (\log p(y|x) + \log p(x))$, where $p(y|x)$ denotes the likelihood function and $p(x)$ represents the prior probability distribution of the signal. This formulation embodies the essence of Bayesian inference, seeking to balance the information in the measured data and the a priori assumptions about the signal's characteristics. The solution is typically obtained through iterative optimization algorithms by following the gradient of the objective function.

1.1.1 Compressed Sensing

The emergence of Compressed Sensing (CS) marked another significant advancement in accelerating MRI [6]. CS leverages the inherent sparsity of MRI data in certain transform domains to achieve significant reductions in scanning times. These transform domains include the wavelet domain, characterized by the inherent sparsity of medical images due to their hierarchical structure [7]; the domain of finite differences [8], which captures smooth features within anatomical structures; the discrete cosine transform domain [9], which identifies periodic patterns in images; and the gradient domain [10], which represents sharp edges and boundaries between various tissue types. When coupled with incoherent k-space undersampling, these sparsifying transforms can enable robust image reconstruction from substantially fewer k-space measurements. The wavelet domain, in particular, emerged as a cornerstone in CS-MRI applications, as it effectively captures both localized features and global image characteristics [11].

CS reconstruction can be mathematically expressed as a constrained optimization problem: $\min |\Psi x|_1$ subject to $|A(x) - y|_2^2 \leq \sigma$, where Ψ denotes the sparsifying wavelet transform and σ represents a small constant related to the noise level [6]. In CS, undersampling is typically modeled using pseudo-random variable-density sampling, with a denser sampling of the low-frequencies and a more sparse sampling of the higher frequencies. This strategy preserves crucial structural information while introducing incoherent aliasing artifacts that CS algorithms can effectively suppress. The integration of PI with CS, known as PICS, has further advanced the field by enabling image reconstruction at higher acceleration

factors.

However, implementing CS reconstruction in clinical practice faces several significant challenges. The undersampling patterns must be optimized to balance the theoretical requirements of incoherent sampling against practical hardware constraints and physiological limitations. CS performance strongly depends on image contrast characteristics, with high-contrast features, such as white and gray matter, being more easily recoverable than subtle low-contrast structures, such as deep gray matter (e.g., basal ganglia and thalamus) in T_1-weighted brain MRI. This contrast-dependent behavior can impact diagnostic accuracy, particularly at higher acceleration factors. Furthermore, the computational demands of CS reconstruction present practical challenges in clinical workflows where timely results are crucial, necessitating a delicate balance between reconstruction quality and imaging times.

1.1.2 Deep Learning

Deep learning (DL) has ushered in a transformative era in medical image processing, offering novel approaches to long-standing challenges in acquisition, reconstruction, and analysis. The progression from Ciresan et al.'s pioneering work in neuronal structure segmentation [12] to the revolutionary introduction of the U-Net [13] marked a significant advancement in the analysis and interpretation of medical images on a large scale. The U-Net has become a cornerstone in medical image analysis, spawning numerous adaptations across various imaging modalities [14] and tasks, including MRI reconstruction [15].

The application of DL to MRI reconstruction has evolved through several stages. Initial approaches utilized convolutional neural networks (CNNs) [16], but were limited by their inability to incorporate domain-specific knowledge. A pivotal advancement in DL-based accelerated MRI reconstruction emerged with the development of physics-informed DL networks. Physics-informed DL networks leverage the prior knowledge about MR imaging by integrating the forward model of MRI reconstruction into their learning scheme through the enforcement of data consistency [17, 18]. These advanced networks have been shown to effectively generalize to diverse data, exhibiting robustness to variations in acquisition parameters, anatomical regions, and diseases [19–21], while allowing for rapid reconstruction times, overcoming the limitations of CS reconstruction.

1.2 Deep Multitask Learning

Although DL has demonstrated remarkable success in various medical imaging tasks, from image reconstruction to analysis, those tasks are usually performed separately of each other. This independent task execution fails to leverage the rich interconnections inherent in medical imaging, where the quality of the downstream analysis tasks is fundamentally dependent on the quality of the prior tasks. For example, raw k-space data must first be reconstructed into high-quality images, which then can serve as input for estimating quantitative parameter maps, ultimately enabling accurate tissue segmentation. Each stage presents unique learning challenges: the computational complexity of sophisticated reconstruction algorithms, the inherent difficulty of learning accurate quantitative parameter mappings from reconstructed images, and the challenge of learning robust segmentation models from limited annotated training data. This cascade of interdependent tasks highlights the broader challenge of efficient medical image analysis.

Recent work by Adler et al. [22] explored the importance of task-relatedness in medical imaging, by investigating the fundamental relationships between image acquisition, reconstruction, and analysis tasks. Building on task-relatedness in medical imaging tasks and the principles of multitask learning (MTL) [23], this thesis introduces the concept of Deep Multitask Learning (DMTL). In MTL, simultaneous learning of related tasks can enhance generalization and performance across all tasks involved through shared representations. DMTL extends this concept to medical imaging applications, not merely by identifying task relationships, but actively leveraging the intricate relationships between imaging tasks, using DL networks, to improve overall performance. An abstract representation of DMTL can be seen in Fig. 1.1.

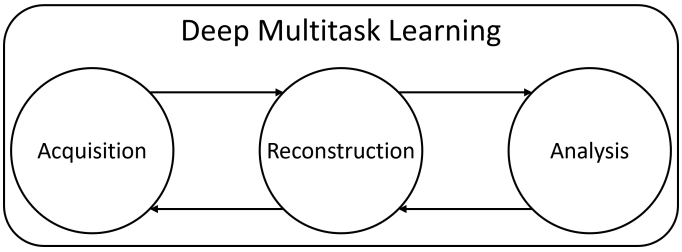


Figure 1.1: Abstract representation of the deep multitask learning (DMTL) concept applied in medical imaging.

1.3 Thesis Aim & Outline

This thesis aims to accelerate MR imaging by developing and validating novel methodological frameworks that bridge physics-informed DL reconstruction with MTL approaches through the concept of DMTL. The chapters follow a structured path from methodological developments in accelerated MRI reconstruction to large-scale and practical clinical evaluations, then to the introduction and application of MTL for joint reconstruction and segmentation, and finally to the practical implementation of DMTL.

In **Chapter 2**, we assess the importance of data consistency in physics-informed DL reconstruction networks. At the same time, we introduce a novel reconstruction network, the Cascades of Independently Recurrent Inference Machines (CIRIM). The CIRIM sequentially connects multiple Recurrent Inference Machines (RIM) [18] to achieve an optimal trade-off between fast reconstruction times and high image quality. The integration of cascades of RIMs addresses the critical issue of vanishing and exploding gradients commonly encountered in Recurrent Neural Networks (RNNs) [24]. The performance of the CIRIM and several state-of-the-art DL-based reconstruction networks and CS is evaluated on multiple heterogeneous datasets, including brain and knee imaging. Furthermore, the generalizability of these models is tested in a real-world scenario, including the reconstruction of unseen during training data of patients with MS lesions.

The robustness and clinical applicability of DL-based reconstruction is extensively evaluated in **Chapters 3** and **4**. **Chapter 3** presents our contributions to the Multi-Coil MRI (MC-MRI) reconstruction challenge, a pivotal initiative alongside the fastMRI challenges [19, 20], which establishes standardized evaluation frameworks for DL-based reconstruction. **Chapter 4** evaluates the performance of the CIRIM in a challenging clinical context,

comprehensively comparing it with the clinical standard PICS in reconstructing highly accelerated (twelve-fold) 3D FLAIR data encompassing various neurological conditions, such as stroke, MS, tumors, and Meniere's disease. This evaluation, supported by both quantitative metrics and an expert radiologists' assessment, provides crucial insights into the practical utility of DL-based reconstruction.

In **Chapter 5**, we argue that the reconstruction task should be perceived as a standalone task but rather consider the task-relatedness of subsequent tasks, such as segmentation, to improve overall performance. By connecting a segmentation network to the CIRIM, we propose a novel MTL approach for joint reconstruction and segmentation (MTLRS). Building upon the developments of the previous chapters, in **Chapter 6**, we present the Advanced Toolbox for Multitask Medical Imaging Consistency (ATOMMIC). ATOMMIC embodies the conceptualization and practical implementation of DMTL and, more importantly, provides full reproducibility of the research presented in this thesis. Finally, a discussion, limitations, and future directions of DMTL for accelerating MRI are discussed in **Chapter 7**.

REFERENCES

[1] Pierre Seners, Nicole Yuen, Michael Mlynash, Sarah J Snyder, Jeremy J Heit, Maarten G Lansberg, Soren Christensen, Jean-François Albucher, Christophe Cognard, Igor Sibon, et al. Quantification of penumbral volume in association with time from stroke onset in acute ischemic stroke with large vessel occlusion. *JAMA neurology*, 80(5):523–528, 2023.

[2] D Kidd, J W Thorpe, B E Kendall, G J Barker, D H Miller, W I McDonald, and A J Thompson. MRI dynamics of brain and spinal cord in progressive multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, 60(1):15–19, 1996.

[3] Steven D Brass, Nan-kuei Chen, Robert V Mulkern, and Rohit Bakshi. Magnetic resonance imaging of iron deposition in neurological disorders. *Topics in Magnetic Resonance Imaging*, 17(1):31–40, 2006.

[4] H. Nyquist. Certain Topics in Telegraph Transmission Theory. *Transactions of the American Institute of Electrical Engineers*, 47(2):617–644, 1928.

[5] Klaas P. Pruessmann, Markus Weiger, Markus B. Scheidegger, and Peter Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–962, 1999.

[6] Michael Lustig, David Donoho, and John M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[7] Ingrid Daubechies. Ten lectures on wavelets. *Society for industrial and applied mathematics*, 1992.

[8] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[9] Nasir Ahmed, T_ Natarajan, and Kamisetty R Rao. Discrete cosine transform. *IEEE transactions on Computers*, 100(1):90–93, 1974.

[10] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

[11] Michael Lustig, David L. Donoho, Juan M. Santos, and John M. Pauly. Compressed Sensing MRI. *IEEE Signal Processing Magazine*, 25(2):72–82, March 2008. 1498 citations (Crossref) [2024-03-13] 2042 citations (Semantic Scholar/DOI) [2024-03-13] Conference Name: IEEE Signal Processing Magazine.

[12] Dan Ciresan, Alessandro Giusti, Luca Gambardella, and Jürgen Schmidhuber. Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[14] Nahian Siddique, Sidike Paheding, Colin P Elkin, and Vijay Devabhaktuni. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE access*, 9:82031–82057, 2021.

[15] Chang Min Hyun, Hwa Pyung Kim, Sung Min Lee, Sungchul Lee, and Jin Keun Seo. Deep learning for undersampled MRI reconstruction. *Physics in Medicine & Biology*, 63(13):135007, 2018.

[16] Shanshan Wang, Zhenghang Su, Leslie Ying, Xi Peng, Shun Zhu, Feng Liang, Dagan Feng, and Dong Liang. Accelerating magnetic resonance imaging via deep learning. In *2016 IEEE 13th international symposium on biomedical imaging (ISBI)*, pages 514–517. IEEE, 2016.

[17] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P. Recht, Daniel K. Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, 2018.

[18] Kai Lønning, Patrick Putzky, Jan-Jakob Sonke, Liesbeth Reneman, Matthan WA Caan, and Max Welling. Recurrent inference machines for reconstructing heterogeneous MRI data. *Medical image analysis*, 53:64–78, 2019.

[19] Florian Knoll, Tullie Murrell, Anuroop Sriram, Nafissa Yakubova, Jure Zbontar, Michael Rabbat, Aaron Defazio, Matthew J. Muckley, Daniel K. Sodickson, C. Lawrence Zitnick, and Michael P. Recht. Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge. *Magnetic Resonance in Med*, 84(6):3054–3070, December 2020.

[20] Matthew J. Muckley, Bruno Riemenschneider, Alireza Radmanesh, Sunwoo Kim, Geunu Jeong, Jingyu Ko, Yohan Jun, Hyungseob Shin, Dosik Hwang, Mahmoud Mostapha, Simon Arberet, Dominik Nickel, Zaccharie Ramzi, Philippe Ciuciu, Jean-Luc Starck, Jonas Teuwen, Dimitrios Karkalousos, Chaoping Zhang, Anuroop Sriram, Zhengnan Huang, Nafissa Yakubova, Yvonne W. Lui, and Florian Knoll. Results of the 2020 fastMRI Challenge for Machine Learning MR Image Reconstruction. *IEEE Transactions on Medical Imaging*, 40(9):2306–2317, 2021.

[21] Youssef Beauferris, Jonas Teuwen, Dimitrios Karkalousos, Nikita Moriakov, Mattha Caan, Lívia Rodrigues, Alexandre Lopes, Hélio Pedrini, Letícia Rittner, Maik Dannecker, Viktor Studenyak, Fabian Gröger, Devendra Vyas, Shahrooz Faghih-Roohi, Amrit Kumar A.K. Jethi, Jaya Chandra J.C. Raju, Mohanasankar Sivaprakasam, Wallace Loos, Richard Frayne, and Roberto Souza. Multi-channel MR Reconstruction (MC-MRRec) Challenge – Comparing Accelerated MR Reconstruction Models and Assessing Their Genereralizability to Datasets Collected with Different Coils. *arXiv*, pages 1–31, 11 2020.

[22] Jonas Adler, Sebastian Lunz, Olivier Verdier, Carola-Bibiane Schönlieb, and Ozan Öktem. Task adapted reconstruction for inverse problems. *Inverse Problems*, 38(7):075006, may 2022.

[23] Rich Caruana. Multitask Learning. *Machine Learning*, 28:41–75, 1997.

[24] S Hochreiter. Long Short-term Memory. *Neural Computation MIT-Press*, 1997.

2

Assessment of data consistency through cascades of independently recurrent inference machines for fast and robust accelerated MRI reconstruction

Karkalousos D., Noteboom S., Hulst H.E., Vos F.M., & Caan M.W.A. (2022).
Physics in Medicine & Biology (2022), 67(12).
DOI: 10.1088/1361-6560/ac6cc2

Abstract

Objective. *Machine Learning methods can learn how to reconstruct magnetic resonance images (MRI) and thereby accelerate acquisition, which is of paramount importance to the clinical workflow. Physics-informed networks incorporate the forward model of accelerated MRI reconstruction in the learning process. With increasing network complexity, robustness is not ensured when reconstructing data unseen during training. We aim to embed data consistency (DC) in deep networks while balancing the degree of network complexity. While doing so, we will assess whether either explicit or implicit enforcement of DC in varying network architectures is preferred to optimize performance.* ***Approach.*** *We propose a scheme called Cascades of Independently Recurrent Inference Machines (CIRIM) to assess DC through unrolled optimization. Herein we assess DC both implicitly by gradient descent and explicitly by a designed term. Extensive comparison of the CIRIM to compressed sensing as well as other Machine Learning methods is performed: the End-to-End Variational Network (E2EVN), CascadeNet, KIKINet, LPDNet, RIM, IRIM, and UNet. Models were trained and evaluated on T_1-weighted and FLAIR contrast brain data, and T_2-weighted knee data. Both 1D and 2D undersampling patterns were evaluated. Robustness was tested by reconstructing 7.5× prospectively undersampled 3D FLAIR MRI data of multiple sclerosis (MS) patients with white matter lesions.* ***Main results.*** *The CIRIM performed best when implicitly enforcing DC, while the E2EVN required an explicit DC formulation. Through its cascades, the CIRIM was able to score higher on structural similarity and PSNR compared to other methods, in particular under heterogeneous imaging conditions. In reconstructing MS patient data, prospectively acquired with a sampling pattern unseen during model training, the CIRIM maintained lesion contrast while efficiently denoising the images.* ***Significance.*** *The CIRIM showed highly promising generalization capabilities maintaining a very fair trade-off between reconstructed image quality and fast reconstruction times, which is crucial in the clinical workflow.*

2.1 Introduction

Magnetic resonance imaging (MRI) non-invasively images the anatomy of the human body. It is important to note that data are acquired in the frequency domain, known as k-space. Conventionally, the measured signals need to adhere to the Nyquist-criterion to allow for inverse Fourier transforming them to the image domain without aliasing. Due to hardware limitations and physical constraints, however, sampling the full k-space leads to long scanning times. Almost 25 years ago, parallel-imaging (PI) [1] was introduced to reduce acquisition times, overcoming hardware and software limitations by applying multiple receiver coil arrays. Each coil has a distinct sensitivity profile which can be exploited in reconstructing undersampled data. With sensitivity encoding (SENSE) the multicoil data are transformed to the image domain through the inverse Fourier Transform, after which a reconstruction algorithm dealiases the images based on the coil sensitivity maps [2]. The combination of PI with compressed sensing (CS) [3, 4] is now standardly applied in clinical settings, allowing for high acceleration factors by utilizing the constrained reconstruction through a sparsifying transform.

Machine Learning (ML) methods can learn how to reconstruct images by training them on acquired data for which a reference reconstruction is available. As such the reconstruction times can be reduced, which is of paramount importance to the clinical

workflow. The UNet [5] may be the most popular network in the field and the base for numerous other methods, as elaborated upon below. Its unique architecture, with the down- and up-sampling operators and the large number of features on the output, has made it a cornerstone approach in image reconstruction today. Although such a network architecture can perform well, its performance is still limited due to operating only in image space without any MR physics knowledge incorporated.

Physics-informed networks were therefore introduced, incorporating the forward model of accelerated MRI reconstruction in the learning process. The variational network (VN) [6] and the recurrent inference machines (RIM) [7–9] proposed to solve the inverse problem of accelerated MRI reconstruction through a Bayesian estimation. Alternatively, scan-specific techniques were used to restore missing k-space from fully-sampled autocalibration data [10–12]. Furthermore, dual-domain networks were proposed to leverage the k-space information and perform corrections both in the frequency domain and the image domain. The Learned Primal-Dual reconstruction technique (LPDNet) [13] replaced the proximal operators in the Primal-Dual Hybrid Gradient algorithm [14] with learned operators, yielding a learning scheme combined with model-based reconstruction. The KIKI-net [15] introduced a sequence of convolutional neural networks (CNN) performed in k-space (K) and image space (I). Later, concatenations of UNets were applied to replace the sequence of CNNs in the KIKI-net [16]. Finally, the Model-Based Deep Learning technique [17] proposed a learned model-based reconstruction scheme involving a data consistency term.

With increasing network complexity, however, robustness is not ensured when reconstructing data unseen during training. This especially concerns clinical data with pathology for which fully sampled reference data cannot be obtained. This was understood in recent MRI reconstruction challenges [18–20], in which deep end-to-end schemes, such as the End-to-End Variational Network (E2EVN) [21], the XPDNet [22], and the Joint-ICNet [23] allowed for higher image quality at increased acceleration factors but not necessarily for generalization to out-of-distribution data containing pathologies. Recurrent neural networks (RNNs), i.e. the RIM and the pyramid convolutional RNN [24], appeared to generalize well on out-of-distribution data due to their nature of maintaining a notion of memory [25]. However, they scored lower on the trained data compared to the previously mentioned networks, possibly due to a limited number of iterations required to avoid gradient instabilities. Such methods would potentially benefit of increased network complexity as can be achieved using a number of cascades of networks [26–28]. The cascades can be considered as stacked networks targeting to resolve aliasing artifacts and to enhance denoising by iteratively evaluating the reconstruction, but without sharing parameters through backpropagation. Unfortunately, a solution may no longer be consistent with the acquired data with increasing network complexity. This raises a need for embedding data consistency in deep networks while balancing the degree of network complexity.

Data consistency (DC) can be embedded into the learning scheme in several ways, such as through gradient descent [6, 8, 21, 29], an iterative energy minimization process, namely variable-splitting [30], generative adversarial networks [31–34], adversarial transformers [35], complex-valued networks [32, 36, 37], transfer learning [38], manifold approximation [39], or through sparsity [40–44]. Recent work evaluated enforcing DC in three ways, by gradient descent, by proximal mapping, and by variable-splitting [45]. It was shown that the training set could be reduced in size by doing so. The best results were obtained when

train and test domains were aligned. However, it remains unknown whether either explicit or implicit enforcement of DC in varying network architectures is the best approach to optimize performance.

This work proposes a scheme called Cascades of Independently Recurrent Inference Machines (CIRIM). The CIRIM comprise RIM blocks sequentially connected through cascades and the efficient Independently Recurrent Neural Network (IndRNN) [46] as recurrent unit. The cascades allow us to train a deep but balanced RNN for improved de-aliasing and denoising, while maintaining stable gradient calculations. The enforcement of DC in an implicit or explicit manner will be assessed by comparison to the E2EVN. The networks are further compared to the CascadeNet [26], the KIKINet [15], the LPDNet [13], the RIM [8], the RIM built with the IndRNN, the UNet [5], and conventional Compressed Sensing reconstruction [4]. The performance is evaluated on multi-modal MRI datasets applying different undersampling strategies. As a clinical application, we focused on reconstructing (out-of-training distribution) FLAIR data of multiple sclerosis patients. Finally, reconstruction times are also assessed as a critical aspect of improving clinical workflow.

2.2 Methods

In this section, first in 2.2.1, the MRI acquisition process is introduced. In 2.2.2, the background on solving the inverse problem of accelerated MRI reconstruction through a Bayesian approach is set. In 2.2.2 and 2.2.2, unrolled optimization by gradient descent is reviewed via the Recurrent Inference Machines (RIM) and the End-to-End Variational Network (E2EVN). The Cascades of Independently Recurrent Inference Machines (CIRIM) is then proposed in 2.2.2, to expand further de-aliasing capabilities of a deep trainable RNN. Furthermore, assessment of data consistency (DC) is performed in 2.2.2 and 2.2.2 to evaluate to what extent the performance of networks depends on the cascades or the DC formulation, or both. In 2.2.2, the loss function is explained with respect to the network's architecture. In 2.2.3, the experiments are described, i.e. the used datasets, the computed evaluation metrics, and the hyperparameters to be optimized.

2.2.1 Accelerated MRI acquisition

The process of accelerating the MRI acquisition can be described through a forward model. Let the true image be denoted by $\mathbf{x} \in \mathbb{C}^n$, with $n = n_x \times n_y$, and let $\mathbf{y} \in \mathbb{C}^m$, with $m \ll n$, be the set of the sparsely sampled data in k-space. The forward model describes how the measured data are obtained from an underlying reference image. For the $\mathbf{i}$th coil of $\mathbf{c}$ receiver coils, the forward model is formulated as:

$$y_i = A(x) + \sigma_i, \quad i = 1, \ldots, c, \tag{2.1}$$

in which $\mathbf{A} : \mathbb{C}^{\mathbf{n}} \mapsto \mathbb{C}^{\mathbf{n} \times \mathbf{n_c}}$ denotes the linear forward operator modeling the sub-sampling process of multicoil data and $\sigma_{\boldsymbol{i}} \in \mathbb{C}^{\mathbf{n}}$ denotes the measured noise for the $\mathbf{i}$th coil. $\mathbf{A}$ is given by

$$A = P \circ F \circ \varepsilon. \tag{2.2}$$

Here, $\mathbf{P}$ is a sub-sampling mask selecting a fraction of samples to reduce scan time. $\boldsymbol{F}$ denotes the Fourier transform, projecting the image onto k-space. $\varepsilon : \mathbb{C}^n \times \mathbb{C}^{n \times n_c} \mapsto \mathbb{C}^{n \times n_c}$

denotes the expand operator, transforming $\mathbf{x}$ into $\mathbf{x_c}$ multicoil images and is given by

$$\varepsilon(x) = (S_0 x, \ldots, S_c x) = (x_0, \ldots, x_c), \tag{2.3}$$

where $\mathbf{S_i}$ are the coil sensitivity maps, a diagonal matrix representing the spatial sensitivities that scale every pixel of the reference image by a complex number.

The adjoint backward operator of $\mathbf{A}$ in (2.2), projecting $\mathbf{y}$ onto image space, is given by

$$A^* = \rho \circ F^{-1} \circ P^T, \tag{2.4}$$

where $\boldsymbol{F}^{-1}$ denotes the inverse Fourier transform, and $\rho : \mathbb{C}^{n \times n_c} \times \mathbb{C}^{n \times n_c} \mapsto \mathbb{C}^n$ denotes the reduce operator that serves for combining the multicoil images $\mathbf{x_c}$ into $\mathbf{x}$. ρ is given by

$$\rho(\mathbf{x_0}, \ldots, \mathbf{x_c}) = \sum_{i=1}^{c} S_i^H \mathbf{x_i}, \tag{2.5}$$

with $\mathbf{H}$ representing Hermitian complex conjugation.

2.2.2 The inverse problem of accelerated MRI reconstruction

The objective when solving the inverse problem of accelerated MRI reconstruction (figure 2.1) is to map the sparsely sampled k-space measurements to an unaliased, highly accurate image. The inverse transformation of restoring the true image from the set of the sparsely sampled measurements can be found through the *Maximum A Posteriori* (MAP) estimator, given by

$$\hat{x}_{MAP} = argmax_x \{\log p(y|x) + \log p(x)\}, \tag{2.6}$$

which is the maximization of the sum of the log-likelihood and the log-prior distribution of $\mathbf{y}$ and $\mathbf{x}$, respectively. The log-likelihood expresses the log probability that k-space data $\mathbf{y}$ are obtained given an image $\mathbf{x}$, yielding a data fidelity term derived from the posterior $p(y|x)$. The log-prior distribution regularizes the solution by representing an MR-image's most likely appearance.

Conventionally, equation (2.6) is reformulated as the following optimization problem

$$\hat{x}_{MAP} = argmin_x \left\{ \sum_{i=1}^{c} d(y_i, A(x)) + R(x) \right\}, \tag{2.7}$$

where d ensures data consistency between the reconstruction and the measurements, reflecting the error distribution given by the log-likelihood distribution in equation (2.6). R is a regularizer weighted by λ, which constrains the solution space by incorporating prior knowledge over x.

Assuming Gaussian distributed data and ignoring the regularization term in equation (2.7), the negative log-likelihood is:

$$\log p(y|x) = -\frac{1}{\sigma^2} \sum_{i=1}^{c} \|A(x) - y_i\|_2^2. \tag{2.8}$$

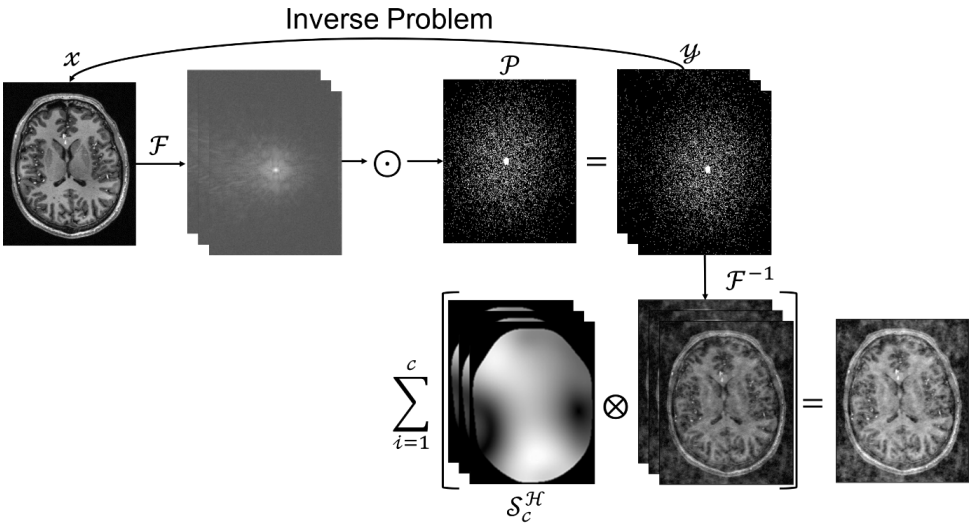


Figure 2.1: The objective in accelerated MRI reconstruction is to solve the inverse problem of recovering an unaliased image (x) from a set of sparsely sampled measurements (y). A forward model starts from the true image representation (x) (top-first), measured over multiple receiver coils (S) (bottom-first image). It is Fourier transformed to k-space (top-second) and sub-sampled using a mask (P) (top-third) to obtain sparsely sampled measurements (y) (top-fourth). Through the inverse Fourier transform (bottom-second) and after combining with coil sensitivity maps (bottom-first), an aliased image is obtained (bottom-third).

Recurrent inference machines (RIM)

The RIM [8] were originally proposed as a general inverse problem solver. The RIM targets iterative optimization of a model with a complex-valued parametrization, requiring taking derivatives with respect to a complex variable. This can be achieved using the Wirtinger- or $\mathbb{CR}$-calculus [47–49]. Gradient descent is performed by us using the Wirtinger derivative, to yield a real-valued cost function of complex values. The unrolled scheme for generating updates is presented in figure 2.2.

Non-convex optimization can be performed based on the approach by [50]. The update rules are learned by the optimizer $\mathbf{h}$, which has its own set of parameters ϕ. Formulating equation (2.6) accordingly, resulting updates are of the form

$$x_{\tau+1} = x_\tau + h_\phi\left(\nabla_{y|x_\tau}, x_\tau\right), \tag{2.9}$$

at iteration τ and for a (a priori set) total number of iterations T.

The gradient of the log-likelihood function is given by

$$\nabla_{y|x} : = \frac{1}{\sigma^2} A^*(A(x) - y). \tag{2.10}$$

The advantage of the RIM is the explicit modeling of the update rule $\mathbf{h}_\phi$ using a recurrent neural network (RNN). In addition to the gradient information, the model is aware of the position of the estimation in variable space equation (2.9).

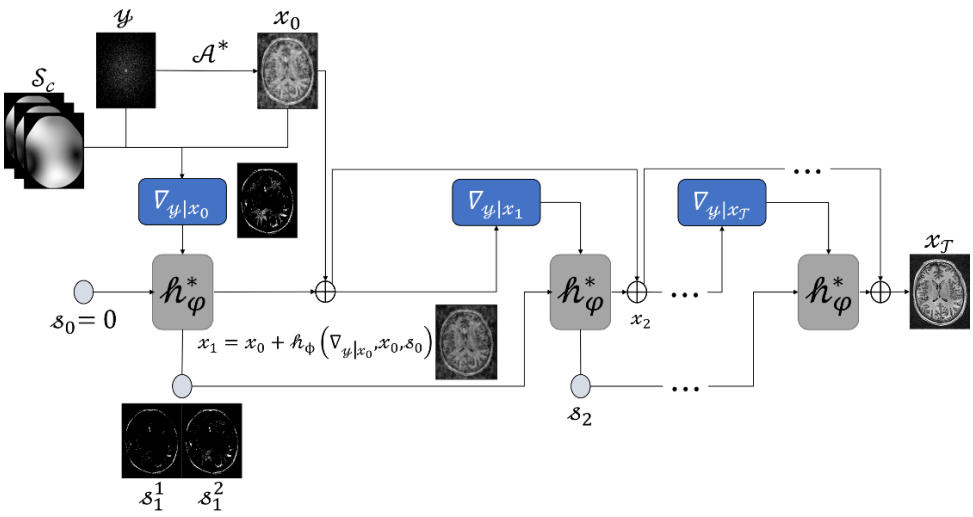


Figure 2.2: The recurrent inference machines unrolled over two iterations. The inputs to the model are the set of sparsely sampled measurements (y) (top, second image), the coil sensitivities maps (S_c) (top, first image), and the initial estimation (x_0) (top, third image) for the estimation of the log-likelihood gradient (llg) ($\nabla_{y|x_0}$). The llg is passed through a network to produce updates; the network maintains hidden states initialized by $s_0 = \mathbf{0}$, $s_1 = \mathbf{0}$. At each iteration (τ) the network updates itself and after total (T) iterations produces the final estimation (x_T) (rightmost).

By inserting equation (2.10) into (2.9), the update equations are obtained, given by

$$\begin{aligned} s_0 &= 0, & x_0 &= A^*(y), \\ s_{\tau+1} &= h^*_\phi\left(\nabla_{y|x_\tau}, x_\tau, s_\tau\right), & x_{\tau+1} &= x_\tau + h_\phi\left(\nabla_{y|x_\tau}, x_\tau, s_{\tau+1}\right). \end{aligned} \tag{2.11}$$

where $\mathbf{h}_{\phi^*}$ is the updated model for state variable $\mathbf{s}$. Equation (2.11) reflects that not the prior is explicitly evaluated, but instead its gradient when performing updates. The step size is learned implicitly in combination with the prior. Therefore, $\mathbf{h}_\phi$ also acts as regularizer $\mathbf{R}$ in equation (2.7). Observe that the RIM contains latent (hidden) states, representing the recurrent aspects of the network.

End-to-end variational network (E2EVN)

The variational network (VN) [6] introduces a mapping to real-valued numbers, going from mapping $\mathbb{C}^n \mapsto \mathbb{C}^m$ to mapping $\mathbb{R}^{2n} \mapsto \mathbb{R}^{2m}$. $\mathbf{x}$ can be computed by least-squares minimization in equation (2.8). As originally proposed in [51] and adapted by the VN, the idea is to perform gradient descent through the iterative Landweber algorithm. By defining a regularizer $\mathbf{R}$, equation (2.7) can be formulated as a trainable gradient scheme with time-varying parameters.

The End-to-End Variational Network (E2EVN) [21] uses a UNet as regularizer (R_{UNet}), whose parameters are learned from the data. Unrolled optimization of the regularized problem in equation (2.7) is performed through cascades, given by

$$\hat{x}_{k+1} = \mathcal{R}_{UNet_k}\left(\mathcal{A}^*(y_k)\right), \tag{2.12}$$

for cascade k, with $1 \leqslant k \leqslant K$ for a total number of K cascades. Next, an explicitly formulated data consistency step applies k-space corrections. This step is given by

$$y_{k+1} = y_k - d(y_k - y) - A(\hat{x}_{k+1}), \tag{2.13}$$

where $d(y_k - y)$ is a soft DC term, with a weighting factor d. The optimization is initialized with the (sparsely sampled) measurement data, $y_{k=1} = y$. The eventual image is obtained via the adjoint operator $x_K = A^*(y_K)$.

In this paper, we test omitting the DC step, in equation (2.13), and evaluate if the network's performance is more dependent on the cascades or the gradient step. In that case, updates are given by equation (2.12). Note that the cascades effectively yield sequentially connected VN blocks, targeting de-aliasing (figure 2.3).

The complex-valued image to complex-valued image mapping is performed in image space by concatenating the real and imaginary parts along the coil dimension. After the regularizer's update in equation (2.12), the image is reshaped to have the real and imaginary parts stacked to a complex (last) dimension.

Cascades of independently recurrent inference machines (CIRIM)

We now propose Cascades of Independently Recurrent Inference Machines (CIRIM), consisting of sequentially connected RIM blocks (Figure 2.3). The cascades allow building a deep RNN without vanishing or exploding gradients issues and further evaluate Eq. (2.7) through K cascades. As such the RIM acts as regularizer (R_{RIM}), while the updates to the CIRIM are given by

$$\hat{x}_{k+1} = x_k + \lambda R_{RIM_k}(x_k), \tag{2.14}$$

for cascade k, with $1 \leqslant k \leqslant K$.

In previous work [7, 8], the gated recurrent unit (GRU) [52] was used as recurrent unit for the RIM. A key novelty of our approach is to include the Independently Recurrent Neural Networks (IndRNN) [46] as a more efficient unit for balancing the network's complexity while increasing the number of trainable parameters through the cascades.

Through the cascades the network's size has increased, but it is unclear whether either implicitly evaluating data consistency through the log-likelihood gradient in equation (2.8) is adequate, or an additional learned gradient step is needed to constrain the solution space further. In a similar manner as in equation (2.13), we assess enforcing DC explicitly and interleaved between the cascades. By doing so, we aim to understand to what extent the network's performance and de-aliasing capabilities depend on the cascades or the formulation of the DC.

Then, the updated prediction of the model is given by

$$\hat{x}_{k+1} = A^*(y_k - d(y_k - y) - A(\hat{x}_{k+1})), \tag{2.15}$$

with $x_K = A^*(y_K)$. If this DC step is omitted, updates to the CIRIM are given by equation (14). Implementation notation for the recurrent units can be found in the appendix.

Loss function

For calculating the loss, we compare magnitude images derived from the complex-valued estimations $\hat{x}$ against the fully sampled reference x. As a loss function, we choose the

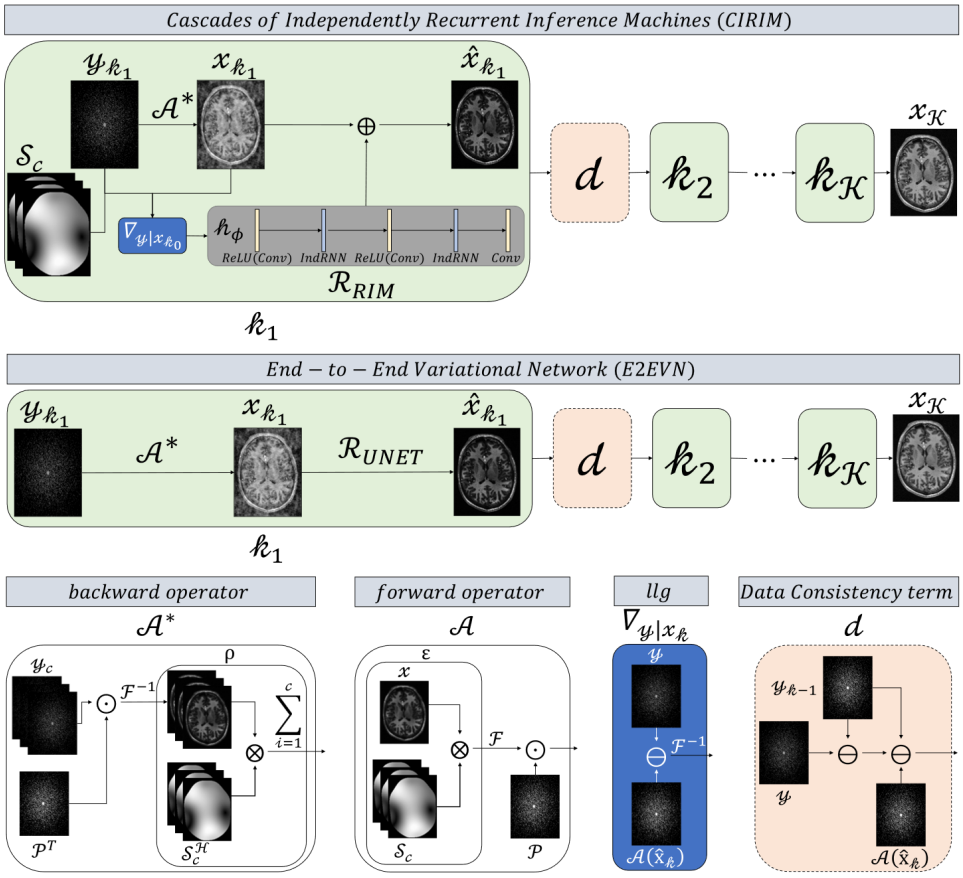


Figure 2.3: Overview scheme for performing unrolled optimization through cascades. The first row represents the Cascades of Independently Recurrent Inference Machines (**CIRIM**), in which a RIM is used as a regularizer ($\mathbf{R_{RIM}}$). The prediction ($\mathbf{\hat{x}_{k_1}}$) of the first cascade ($\mathbf{k_1}$) is given as input to the subsequent cascade ($\mathbf{k_2}$), while an (optional) additional data consistency step can be performed through an explicitly formulated term (**d**). After (**K**) cascades the network returns the final prediction ($\mathbf{x_K}$). The second row depicts the End-to-end Variational Network (E2EVN), where a UNet is used as a regularizer ($\mathbf{R_{UNET}}$). Similarly, as for the CIRIM, the updates are passed through the cascades and the data consistency step. In the third row, first, the backward operator ($\mathbf{A^*}$) is shown, transforming multicoil k-space measurements onto a coil-combined image; second, the forward operator (**A**) is depicted, transforming a coil-combined image into multicoil k-space measurements; third, the log-likelihood gradient ($\nabla_{\mathbf{y}|\mathbf{x_k}}$) reflects the implicit gradient step of the RIM and fourth, the (optional) interleaved between the cascades DC term (**d**) is presented, enforcing an explicit gradient step to the CIRIM and the E2EVN.

$\ell 1 - norm$. The $\ell 1 - norm$ represents the sum of the absolute difference, given by

$$L^{(l_1)}(\hat{x}) = |\hat{x} - x|. \tag{2.16}$$

For the E2EVN and other trained models, the loss is given by equation (16). For the CIRIM, the loss is weighted depending on the number of iterations and averaged over the K cascades, to emphasize the predictions of the later iterations. The loss is then formulated as

Table 2.1: Scan parameters of each dataset used for different experiments. Target anatomy, contrast, scan, and field strength are given, with resolution (res), field-of-view (FOV), time in minutes (with acceleration factor), number of coils (ncoils) and other scan parameters.

Scan/sequence	Field strength	Res (mm)	FOV (mm)	Time (acc)	Ncoils	Parameters
T_1-Brain/3D MPRAGE	3 T	1.0 x 1.0 x 1.0	256 x 256 x 240	10.8 (1x)	32	FA 9°, TFE-factor 150, TI = 900 ms
T_2-Knee/7 TSE	3 T	0.5 x 0.5 x 0.6	160 x 160 x 154	15.3 (1x)	8	FA 90°, TR = 1550 ms, TE = 25 ms
FLAIR-Brain/2D FLAIR	1.5 T/3 T	0.7 x 0.7 x 5	220 x 220	- (1x)	2–24	FA 150°, TR = 9000 ms, TE = 78–126 ms
Pathology study						
MS FLAIR-Brain/3D FLAIR	3 T	1.0 x 1.0 x 1.1	224 x 224 x 190	4.5 (7.5x)	32	TR = 4800 ms, TE = 350 ms, TI = 16500 ms

$$L^{(l_1)}(\hat{x}) = \frac{\sum_{i=1}^{c}\left(\frac{1}{qT}\sum_{\tau=1}^{T} w_\tau |\hat{x}_\tau - x|\right)}{K}, \tag{2.17}$$

where q is the total number of pixels of the image and w_τ is a weighting vector of length T prioritizing the loss at later time-steps. The weights are calculated by setting $w_\tau = 10^{-\frac{T-\tau}{T-1}}$.

2.2.3 Experiments

For our experiments, we used multiple datasets as described in 2.3.1. Scanning parameters of these datasets are given in table 2.1.

Our experiments focused on assessment of the following aspects:

A. Training and validation in fully sampled and retrospectively undersampled data. The undersampling strategy is described in Section 2.2.3.

B. Independent evaluation in prospectively undersampled data of Multiple Sclerosis patients containing white matter lesions.

We trained and compared the CIRIM and the E2VN to the LPDNet, the KIKINet, and the CascadeNet [13, 15, 26], the hyperparameters of which are described in 2.3.3. For comparing the performance of the methods regarding assessment (A) we chose the structural similarity index measure (SSIM) [53] and the peak signal-to-noise ratio (PSNR). For assessment (B), we calculated the contrast resolution (CR), the noise in the white matter (WMN), the noise in the background (BGN), and a resulted weighted average (WA). The metrics are described in 2.3.4.

Datasets

For assessment (A), three fully sampled raw complex-valued multi-coil datasets were obtained. The first dataset was acquired in-house. To this end, eleven healthy subjects were included, from whom written informed consent (under an institutionally approved protocol) was obtained beforehand. The ethics board of Amsterdam UMC declared that this study was exempt from IRB approval. All eleven subjects were scanned by performing 3D T_1-weighted brain imaging on a 3.0 T Philips Ingenia scanner (Philips Healthcare, Best, The Netherlands) in Amsterdam UMC. The data were visually checked to ascertain that they were not affected by motion artifacts. After scanning, raw data were exported and stored for offline reconstruction experiments. The training set was composed of ten subjects (approximately 3000 slices) and the validation set of one subject (approximately 300 slices).

The second dataset consisted of 451 2D multislice FLAIR scans, publicly available through the fastMRI brains dataset [54]. The training set consisted of 344 scans (approximately 5000 slices) and the validation set of 107 scans (approximately 500 slices). The number of coils varied from 2 to 24. The data were cropped in the image domain to 320 for the readout direction by the size of the phase encoding direction (varied from 213 to 320). The cropped images were visually evaluated to not crop any tissue (only air).

The third dataset was composed of 3D knee scans of 20 subjects, available on a public repository [55]. From these data, two subjects were discarded due to observed motion

artifacts. The training set consisted of 17 subjects (approximately 12 000 slices) and the validation set of one subject (approximately 700 slices).

For all datasets, coil sensitivities were estimated using an autocalibration procedure called ecalib from the BART toolbox [56], which leverages the ESPIRiT algorithm [57]. For training and validation, slices were randomly selected by setting a random seed to enable deterministic behavior for all methods and ensure reproducibility. Note that the validation set was only used to calculate the loss at the end of each epoch and not included into the training set. Finally, all volumes were normalized to the maximum magnitude.

For assessment (B), testing the methods' ability to reconstruct unseen pathology, a dataset of 3D FLAIR data of multiple sclerosis patients with known white matter lesions was obtained. Data were prospectively undersampled with a factor of 7.5x based on a Variable-Density Poisson disk distribution. Originally these data were acquired on a 3.0 T Philips Ingenia scanner (Philips Healthcare, Best, The Netherlands) in Amsterdam UMC, within the scope of a larger, ongoing study. The local ethics review board approved this study and patients provided informed consent prior to imaging. A fully-sampled reference scan was also acquired and used to estimate coil sensitivity maps using the caldir method of the BART toolbox [56]. The data were visually checked after which all subjects with motion artifacts were discarded, ending up including 18 patients (approximately 4000 slices).

Undersampling

The masks for retrospective undersampling in assessment (A) were initially defined in 2D. As such the models trained on all modalities could also be used later for reconstructing high-resolution isotropic FLAIR data for assessment (B). The 3D datasets were first Fourier transformed along the frequency encoding axis and used as separate slices along the two-phase encoding axes. The 2D multislice FLAIR dataset was initially Fourier transformed along the frequency encoding axis and undersampled per slice in 2D, to train a model on an identical contrast as in assessment (B), while also having pathology present in the data.

All data were retrospectively undersampled in 2D by sampling k-space points from a Gaussian distribution with a full width at half maximum (FWHM) of 0.7, relative to the k-space dimensions. Hereby the sampling of low frequencies is prioritized whereas incoherent noise is created due to the random sampling. Note that in this way, we abide by the compressed sensing (CS) requirement of processing incoherently sampled data [3]. For autocalibration purposes, data points near the k-space center were fully sampled within an ellipse of which the half-axes were set to 2% of the fully sampled region. Acceleration factors of 4×, 6×, 8×, and 10× were used by randomly generating a sampling mask (P) with according sampling density (both during training and validation).

To abide to the underlying sampling protocol, and to test the model's ability to reconstruct undersampled data in 1D, we performed an additional experiment with retrospective undersampling in just one dimension. Equidistant k-space points were sampled in the phase encoding direction [57]. The acceleration factor was set to four, while the central region was densely sampled retaining eight percent of the fully-sampled k-space.

Hyperparameters

For the CIRIM models, hyperparameters were selected as follows. The number of cascades K was set to 5, the number of channels to 64 for the recurrent and convolutional layers, and the number of iterations T to 8. The hyperparameter search for finding the

optimal number of cascades is shown in the Supplementary Material (available online at stacks.iop.org/PMB/67/124001/mmedia). The kernel size of the first convolutional layer was set to 5×5 and to 3×3 for the second and third layers. The optimization of these hyperparameters is described elsewhere [8]. Next, we trained models on the T_1-Brain dataset, the T_2-Knee dataset, and the FLAIR-Brain dataset to realize the DC step from equation (15).

For the E2EVN models, we chose 8 cascades, 4 pooling layers, 18 channels for the convolutional layers, and included the DC step from equation (13). The hyperparameter search for finding the optimal number of cascades, pooling layers, and number of channels, is again shown in the supplementary material. Then, for further optimization, we experimented with training models on the T_1-Brain dataset, the T_2-Knee dataset, and the FLAIR-Brain dataset while omitting the DC step. The inputs to the UNet regularizer were padded for making the inputs square, setting the padding size to 11, and the outputs were unpadded for restoring the original input size.

For the baseline UNet, the number of input and output channels was set to 2. The number of channels for the convolutional layers was set to 64, and we chose 2 pooling layers. Similar to the E2EVN, the padding size was set to 11, while no dropout was applied. The selected hyperparameters for the UNet were motivated by the configuration in [58].

For the LPDNet, the KIKINet, and the CascadeNet, the choice of the hyperparameters was motivated from the baseline proposed models. For the LPDNet we used the same network architecture for both the primal and the dual part, being a UNet with 16 channels, 2 pooling layers, and padding size of 11, while no dropout was applied. The number of the primals, the duals, and the number of unrolled iterations was set to 5. Similarly, for the KIKINet, we used the UNet architecture for the k-space and the image space networks. The number of channels was set to 64, the number of pooling layers to 2, and the padding size to 11, without applying any dropout. Finally, for the CascadeNet the number of cascades set to 10, using a sequence of CNNs with 64 channels and depth size of 5, without applying batch normalization.

For all models, we applied the ADAM optimizer (Kingma and Ba 2015), setting the learning rate to 1e-3, except for the CascadeNet where the learning rate was set to 1e-5. The batch size was set to 1, allowing training on various input sizes. The data type was set to complex64 for complex-valued data and to float16 for real-valued data. For training models with 2D undersampling, the loss function for the CIRIM is given by equation (17) and for all models by equation (16). For training models with 1D undersampling, we used the SSIM as loss function, motivated by [20], as a better option for resolving artifacts introduced by equidistant undersampling.

CS reconstructions were performed using the BART toolbox [56]. Here we used parallel-imaging compressed sensing (PICS) with a $\ell 1$-wavelet sparsity transform. The regularization parameter was set to $\alpha = 0.005$, which was heuristically determined as a trade-off between aliasing noise and blurring. The maximum number of iterations was set to 60. We tested the reconstruction times of CS on the GPU (turning the -g flag on).

All experiments were performed on an Nvidia Tesla V100 with 32GB of memory. The code was implemented in PyTorch 1.9 [59] and PyTorch-Lighting 1.5.5 [60], on top of novel frameworks [61, 62], and can be found at https://github.com/wdika/mridc.

Evaluation metrics

For quantitative evaluation of the fully-sampled measurements, we compared normalized magnitude images derived from the complex-valued estimations x_τ against the reference x and calculated SSIM and PSNR metrics.

For evaluating robustness on the 3D FLAIR MS data, we computed the contrast resolution (CR), the noise in the white matter (WMN), the noise in the background (BGN), and a resulted weighted average (WA) of those three metrics.

Since the data are not fully-sampled, the CR is an efficient metric to evaluate the signal level between the white matter and the lesions. To compute CR, lesion segmentations were performed using a pretrained multi-view convolutional neural network (MV-CNN). The MV-CNN was previously trained on combined Fast Imaging Employing Steady-state Acquisition (FIESTA), T_2-weighted and contrast-enhanced T_1-weighted data, for eye and tumor segmentation of retinoblastoma patients [63]. For the segmentation of the white matter, the statistical parametric mapping (SPM) toolbox was used [64]. The mean lesion intensity was compared to that of presumed homogeneous surrounding white matter. To that end, the lesion masks were dilated by four voxels and intersected with the whole brain white matter mask. The CR is then defined as the difference between the lesion signal and the signal in the surrounding white matter, divided by the summation of them, given by

$$CR = \frac{s_{\text{lesion}} - s_{\text{WM Surrounding Lesion}}}{s_{\text{lesion}} + s_{\text{WM Surrounding Lesion}}}. \tag{2.18}$$

The WMN is defined as the mode of the gradient magnitude image x, given by

$$WMN = \text{mode}\left(\nabla\left|\frac{x}{\bar{x}_{\text{WM}}}\right|\right), \tag{2.19}$$

where $\bar{x}_{\text{WM}}$ is the mean WM intensity. The background noise (BGN) is computed as the 99-percentile value in the background region, being the complement of a tissue mask.

A weighted average (WA) was eventually defined as the combination of the CR, the WMN, and the BGN after scaling them to maximum value.

Finally, for every scan, the signal-to-noise ratio (SNR) was calculated as follows,

$$SNR = \frac{\overline{t|x|}}{\widetilde{|Y|}}, \tag{2.20}$$

where $\overline{t|x|}$ is the mean value after thresholding the magnitude image x to discard the background, and $\widetilde{|y|}$ the median magnitude value within a square region in the periphery of k-space, which was assumed to be dominated by imaging noise. The threshold t was set using Otsu's method (Otsu 1979).

2.3 Results

Figure 2.4 shows SSIM and PSNR scores upon assessing DC explicitly and implicitly for the CIRIM (figure 2.4(a)) and the E2EVN (figure 2.4(b)). The models were trained on the T_1-Brain dataset, the T_2-Knee dataset, and the FLAIR-Brain dataset.

A qualitative evaluation of the CIRIM's and the E2EVN's performance on the trained datasets, accelerated with ten-times Gaussian 2D undersampling, is presented in figure 2.5.

The CIRIM performed significantly better than the E2EVN on the T_1-Brain and the FLAIR-Brain dataset. On the FLAIR-Brain dataset, the E2EVN failed to accurately reconstruct the center of brain, as well as to resolve noise in the White Matter lesion. On the T_2-Knee dataset, the two models performed comparably in terms of SSIM, while the CIRIM showed a slight improvement in PSNR.

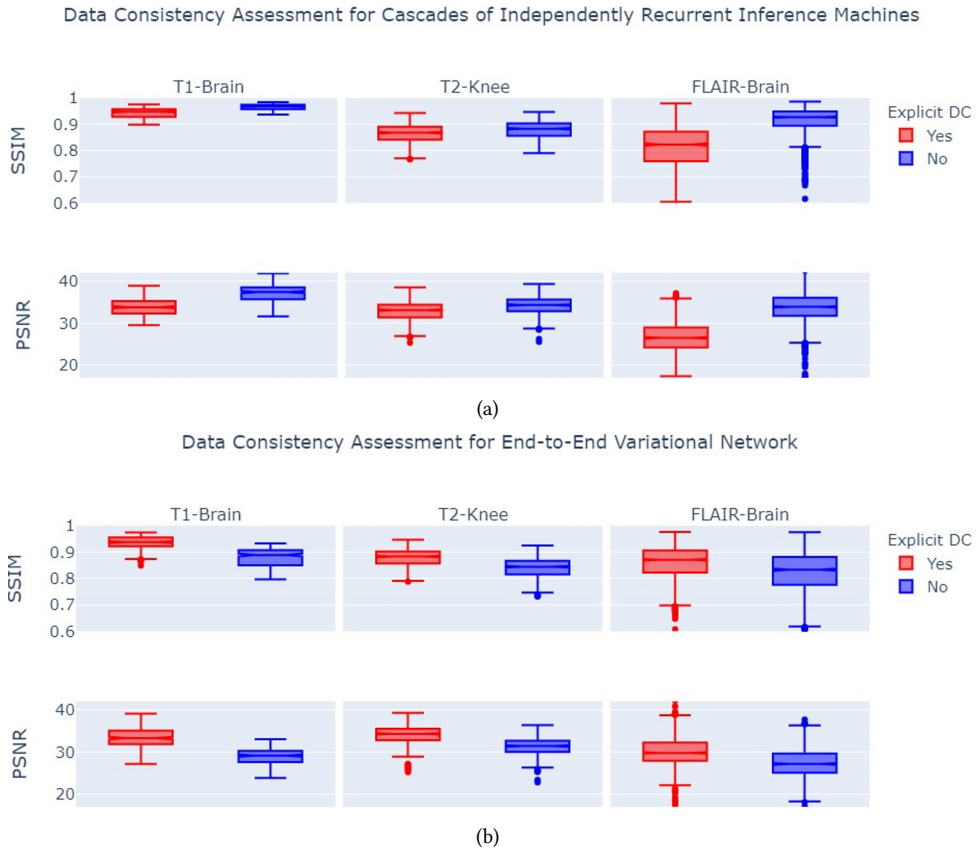


Figure 2.4: Data consistency (DC) assessment for (a) Cascades of Independently Recurrent Inference Machines and (b) End-to-End Variational Network. DC is enforced both explicitly (red) and implicitly (blue). The first row represents SSIM scores and the second row PSNR scores. Performance is reported for models trained on the T_1-Brain dataset (first column), the T_2-Knee dataset (second column), and the FLAIR-Brain dataset (third column).

In figure 2.6, the CIRIM is compared to the RIM and the IRIM. SSIM and PSNR scores are reported for each model trained on the T_1-Brain dataset, the T_2-Knee dataset, and the FLAIR-Brain dataset. The IRIM performed slightly worse compared to the RIM, while the CIRIM performed best.

Table 2.2 collates overall performance of the methods on all training datasets (T_1-Brain, T_2-Knee, FLAIR-Brain). The methods were evaluated with ten-times accelerated data using Gaussian 2D masking, and four times accelerated equidistant 1D masking. For the FLAIR-Brain dataset we dropped the slices outside the head, containing no signal. The

CIRIM performed best in all settings in terms of SSIM and PSNR, while only the E2EVN had comparable performance for the evaluation on the T_2-Knee dataset. Representative reconstructions can be found in the supplementary material, as well as further evaluation for four-, six-, and eight-times acceleration for Gaussian 2D undersampling.

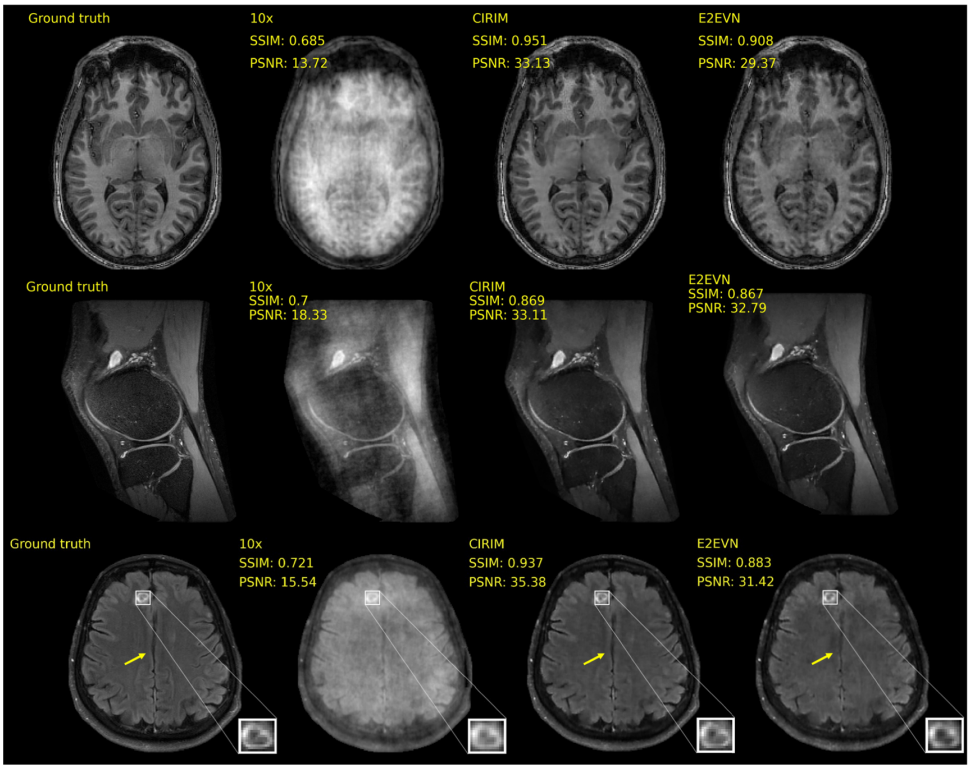


Figure 2.5: Comparison of the CIRIM (third column) to the E2EVN (fourth column) for reconstructing ten-times accelerated slices from the T_1-Brain dataset (first row, first and second image), the T_2-Knee dataset (second row, first and second image), and the FLAIR-Brain dataset (third row, first and second images). For the FLAIR-Brain dataset, the inset focuses on a reconstructed White Matter lesion; obtained through the fastMRI + annotations [65]. The arrow points out to a region of interested.

The trained models on each dataset and undersampling scheme were used to evaluate performance on out-of-training distribution data, containing MS lesions. As summarized in table 2.3, the performance is evaluated quantitatively by measuring the CR of the reconstructed lesions, the WMN, the background noise (BGN) and a WA. A combination of high CR, low WMN and low BGN yields a low WA and reflects highly accurate reconstruction (figures 2.6, S.2.7, S.2.8), such as in the CIRIM FLAIR-Brain model and PICS. The models trained on the FLAIR-Brain, the FLAIR-Brain 1D, and the T_1-Brain datasets scored high on CR and low on WMN compared to the T_2-Knee trained models. The CIRIM and the RIM achieved the lowest BGN. The CascadeNet, the E2EVN, the KIKINet, and the UNet models reported high BGN, in general corresponding to more aliased reconstruction. The LPDNet

Table 2.2: SSIM and PSNR scores of all methods evaluated on the T_1-Brain, T_2-Knee, and FLAIR-Brain datasets. For all datasets performance is reported for ten times acceleration using Gaussian 2D undersampling. For the FLAIR-Brain dataset performance is also reported for four times acceleration using equidistant 1D undersampling. The first column reports the method's name. The second column reports the total number of trainable parameters for each model. Best performing models are highlighted in bold. Methods are sorted in alphabetical order.

Method	Params	T_1-Brain Gaussian2D10x		T_2-Knee Gaussian2D10x		FLAIR-Brain Gaussian2D10x		FLAIR-Brain Equidistant1D4x	
		SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑
CascadeNet	1.1 M	0.922 ± 0.042	30.2 ± 1.5	0.859 ± 0.045	32.2 ± 2.6	0.872 ± 0.106	29.9 ± 5.0	0.913 ± 0.038	30.3 ± 4.6
CIRIM	264k	0.966 ± 0.015	35.8 ± 0.5	0.877 ± 0.039	33.7 ± 2.3	0.906 ± 0.101	32.8 ± 6.2	0.942 ± 0.065	34.3 ± 3.2
E2EVN	19.6 M	0.940 ± 0.023	31.4 ± 1.4	0.877 ± 0.039	33.5 ± 2.5	0.855 ± 0.108	29.1 ± 4.8	0.930 ± 0.062	31.3 ± 5.1
IRIM	53k	0.963 ± 0.017	35.3 ± 0.6	0.870 ± 0.041	33.3 ± 2.3	0.892 ± 0.107	32.0 ± 6.0	0.908 ± 0.093	32.0 ± 5.4
KIKINet	1.9 M	0.925 ± 0.040	31.1 ± 1.3	0.842 ± 0.045	32.1 ± 2.0	0.829 ± 0.113	28.4 ± 4.8	0.919 ± 0.065	30.5 ± 4.6
LPDNet	118k	0.960 ± 0.016	35.0 ± 0.4	0.873 ± 0.038	33.5 ± 2.0	0.858 ± 0.111	29.7 ± 4.7	0.938 ± 0.066	32.3 ± 5.3
PICS		0.866 ± 0.032	30.9 ± 0.7	0.729 ± 0.041	29.7 ± 4.3	0.816 ± 0.074	29.2 ± 7.6	0.876 ± 0.068	30.0 ± 4.2
RIM	94k	0.963 ± 0.017	35.3 ± 0.4	0.872 ± 0.040	33.5 ± 2.3	0.898 ± 0.103	32.3 ± 6.1	0.934 ± 0.069	33.4 ± 3.2
UNet	1.9 M	0.874 ± 0.049	26.6 ± 3.1	0.846 ± 0.048	31.4 ± 3.4	0.795 ± 0.095	26.9 ± 4.1	0.909 ± 0.064	29.4 ± 4.3
Zero-Filled		0.766 ± 0.084	17.3 ± 2.0	0.674 ± 0.031	17.3 ± 1.1	0.703 ± 0.101	16.8 ± 4.2	0.806 ± 0.062	21.5 ± 3.8

achieved high CR and relatively low WMN and BGN, but the observed reconstruction quality was poor. This also highlighted the need for combined metrics and qualitative evaluation to evaluate performance.

Figure 2.7 shows reconstructions of a coronal slice from the MS FLAIR-Brain dataset. Visually, the CIRIM, PICS, RIM, and IRIM reconstructions appear similar. The E2EVN and the CascadeNet showed inhomogeneous intensities and high contrast deviations. The LPDNet showed more aliased reconstructions, with lower contrast levels. The KIKINet and the UNet seemed in our experiments not able to resolve background noise and in general resulted in more distorted images. Example reconstructions of two more subjects including axial and sagittal plane reconstructions can be found in the supplementary material.

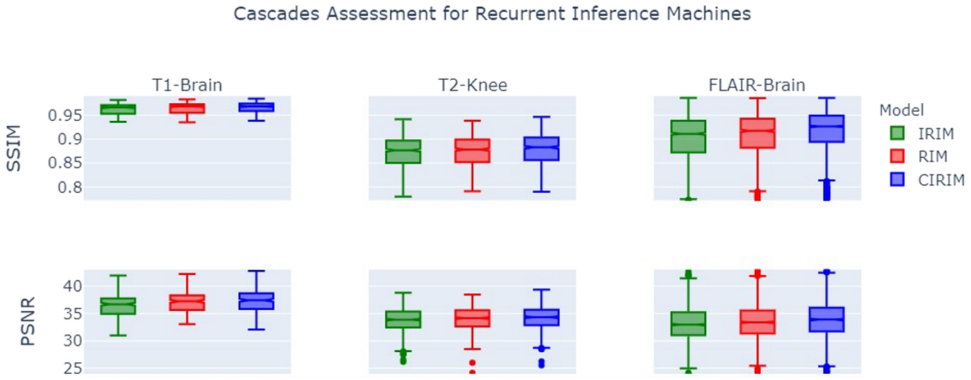


Figure 2.6: Comparison of the Cascades of Independently Recurrent Inference Machines (CIRIM) (blue color), to the recurrent inference machines (RIM) (red color), and the independently recurrent inference machines (IRIM) (green color). Performance is reported for SSIM (first row) and PSNR (second row), on the T_1-Brain dataset (first column), the T_2-Knee dataset (second column), and the FLAIR-Brain dataset (third column).

Finally, in figure 2.8, the reconstruction times of all methods are reported. As input, one volume from the trained fastMRI FLAIR brains dataset was taken, consisting of fifteen slices cropped to a matrix size of 320 × 320. The KIKINet, PICS, and the LPDNet were the slowest methods, requiring 247 ms, 245 ms, and 237 ms respectively to reconstruct the volume. The CIRIM needed 139 ms, the RIM 48 ms, the E2EVN 44 ms, the CascadeNet 42 ms, the IRIM 28 ms, and the UNet 8 ms.

2.4 Discussion

In this paper, we proposed the CIRIM, for a balanced increase in model complexity while maintaining generalization capabilities. We assessed DC both implicitly through unrolled optimization by gradient descent and explicitly by a formulated term. Robustness was evaluated by reconstructing sparsely sampled MRI data containing unseen pathology. The CIRIM was extensively compared to another unrolled network, the E2EVN, and a range of other methods.

In experiments reconstructing brain and knee data containing different contrasts, the proposed CIRIM performed best, with promising generalization capabilities. On the T_2-knee dataset, the E2EVN performed equivalently to the CIRIM, while on the T_1-brain and the

FLAIR-brain datasets for eight- and ten-times acceleration, the measured PSNR dropped by approximately 5% of what compared to what. Visually, this reflected in missing anatomical details such as vessels. The LPDNet, the RIM, and the IRIM performed comparable but slightly worse than the CIRIM. The CascadeNet and the KIKINet, dropped further in SSIM and PSNR on all trained datasets, resulting in more noisy reconstructions. PICS and the UNet showed most of the time overly smoothed results. Interestingly, for 1D undersampling the CascadeNet showed comparable performance to the CIRIM, but it was more sensitive to banding artifacts.

Table 2.3: Independent evaluation of model performance on the 3D FLAIR MS Brains dataset for different training datasets. The reported figures collate: contrast resolution (CR), gradient magnitude with white matter noise (WMN), background noise (BGN), and weighted average (WA). The mean and standard deviation on each metric is given. The best scores are underlined and model highlights in bold. Methods are sorted in alphabetical order.

Method	Trained Dataset	CR↑	WMN↓	BGN↓	WA↓
CascadeNet	T_1-Brain	0.128 ± 0.028	0.135 ± 0.022	0.292 ± 0.078	1.08
	T_2-Knee	0.087 ± 0.040	0.290 ± 0.059	0.302 ± 0.083	1.43
	FLAIR-Brain	0.145 ± 0.030	0.126 ± 0.016	0.265 ± 0.071	0.96
	FLAIR-Brain 1D	0.139 ± 0.025	0.121 ± 0.016	0.309 ± 0.068	1.05
CIRIM	T_1-Brain	0.179 ± 0.025	0.145 ± 0.030	0.172 ± 0.092	1.69
	T_2-Knee	0.097 ± 0.020	0.285 ± 0.034	0.322 ± 0.053	0.62
	FLAIR-Brain	**0.183 ± 0.025**	0.131 ± 0.029	**0.104 ± 0.085**	**0.55**
E2EVN	T_1-Brain	0.173 ± 0.030	0.110 ± 0.017	0.137 ± 0.074	0.62
	T_2-Knee	0.145 ± 0.034	0.144 ± 0.010	0.359 ± 0.095	1.13
	FLAIR-Brain	0.159 ± 0.041	0.116 ± 0.014	0.358 ± 0.064	1.03
	FLAIR-Brain 1D	0.134 ± 0.035	0.141 ± 0.020	0.356 ± 0.052	1.77
IRIM	T_1-Brain	0.159 ± 0.025	0.128 ± 0.027	0.200 ± 0.070	0.80
	T_2-Knee	0.078 ± 0.021	0.260 ± 0.122	0.348 ± 0.118	1.51
	FLAIR-Brain	0.169 ± 0.027	0.145 ± 0.020	0.213 ± 0.075	0.77
	FLAIR-Brain 1D	0.176 ± 0.025	0.151 ± 0.020	0.432 ± 0.075	1.40
KIKINet	T_1-Brain	0.117 ± 0.032	0.184 ± 0.042	0.423 ± 0.075	0.77
	T_2-Knee	0.149 ± 0.026	0.235 ± 0.032	0.294 ± 0.087	1.10
	FLAIR-Brain	0.105 ± 0.077	0.175 ± 0.040	0.626 ± 0.096	1.75
	FLAIR-Brain 1D	0.103 ± 0.026	0.144 ± 0.035	0.352 ± 0.052	1.29
LPDNet	T_1-Brain	0.240 ± 0.046	0.126 ± 0.029	0.210 ± 0.070	0.56
	T_2-Knee	0.030 ± 0.051	0.206 ± 0.031	0.204 ± 0.040	1.34
	FLAIR-Brain	0.117 ± 0.024	0.099 ± 0.012	0.332 ± 0.075	1.15
	FLAIR-Brain 1D	0.066 ± 0.029	0.129 ± 0.024	0.338 ± 0.070	1.40
RIM	T_1-Brain	0.178 ± 0.025	0.168 ± 0.026	0.170 ± 0.093	0.71
	T_2-Knee	0.091 ± 0.036	0.149 ± 0.030	0.251 ± 0.091	1.18
	FLAIR-Brain	0.197 ± 0.029	0.175 ± 0.025	0.134 ± 0.088	0.58
	FLAIR-Brain 1D	**0.183 ± 0.027**	0.158 ± 0.026	0.165 ± 0.074	0.67
UNet	T_1-Brain	0.182 ± 0.034	0.174 ± 0.022	0.276 ± 0.069	0.87
	T_2-Knee	0.125 ± 0.040	0.324 ± 0.084	0.285 ± 0.089	1.93
	FLAIR-Brain	0.085 ± 0.027	0.079 ± 0.010	0.625 ± 0.137	1.72
	FLAIR-Brain 1D	0.087 ± 0.027	0.105 ± 0.023	0.345 ± 0.070	1.40
PICS		0.178 ± 0.025	0.140 ± 0.018	0.147 ± 0.092	0.64
Zero-Filled		0.072 ± 0.023	0.092 ± 0.017	0.372 ± 0.064	1.39

The RIM-based models (RIM, IRIM, CIRIM), trained on FLAIR and T_1-weighted brain data, and PICS, could accurately reconstruct Multiple Sclerosis lesions unseen during

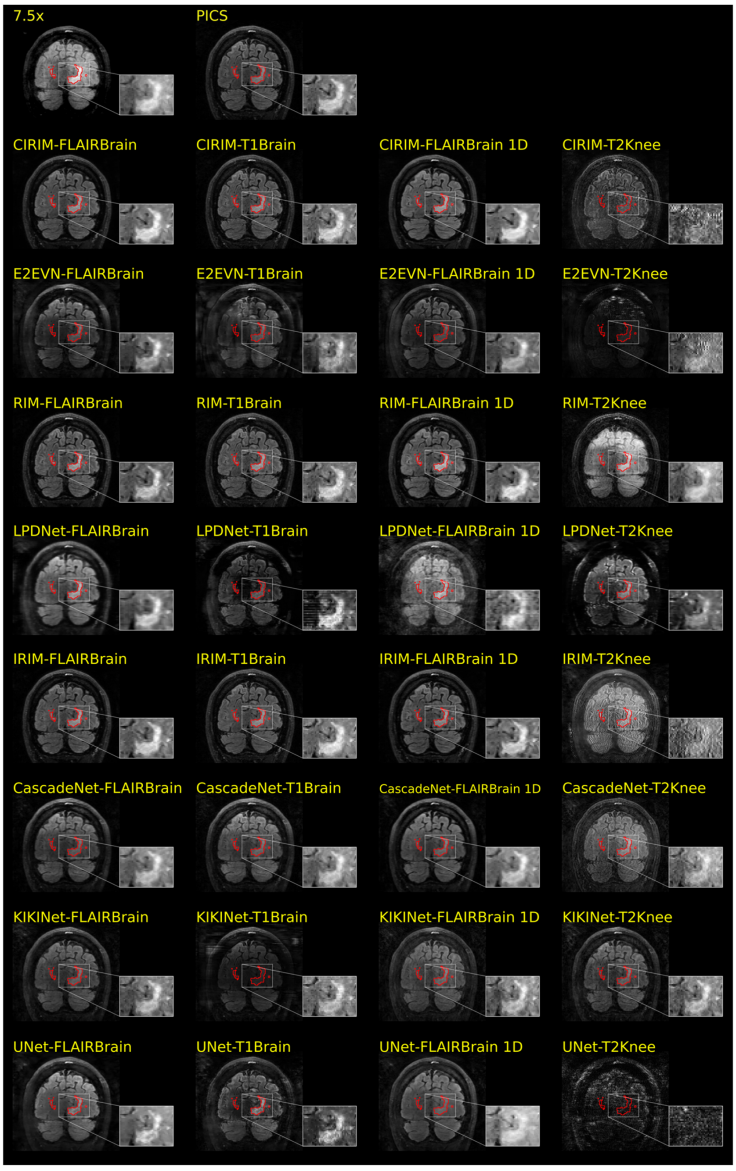


Figure 2.7: Reconstructions of a representative coronal slice of a 7.5x accelerated 3D FLAIR scan of a MS patient. Segmented MS lesions are depicted with red colored contours. Shown is the aliased linear reconstruction (first row-first image), PICS (first row-second image), and models' reconstructions on each trained scheme: the FLAIR-Brain dataset with Gaussian 2D undersampling (second-last row, first column), the T_1-Brain dataset with Gaussian 2D undersampling (second-last row, second column), the FLAIR-Brain dataset with equidistant 1D undersampling (second-last row, third column), and the T_2-Knee dataset with Gaussian 2D undersampling (second-last row, fourth column). The inset on the right bottom of each reconstruction focuses on a lesion region with high spatial detail.

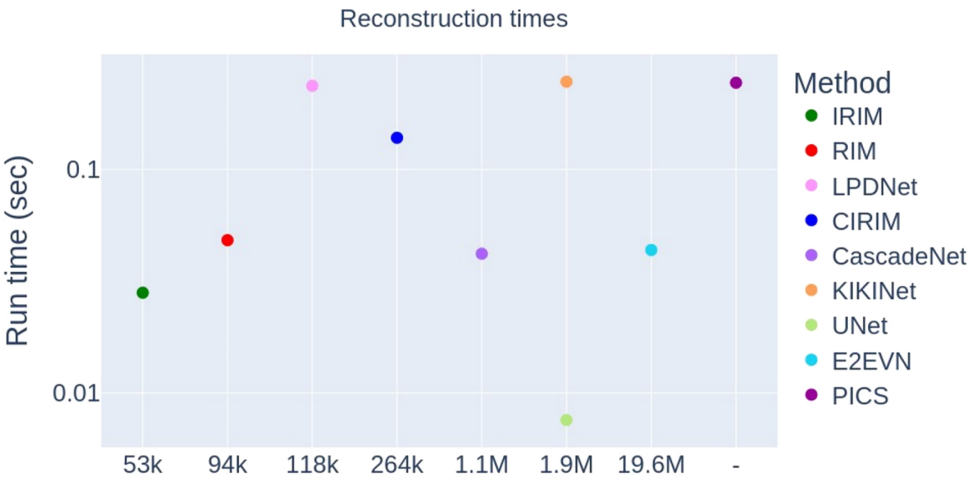


Figure 2.8: Inference times for reconstructing one volume from the FLAIR brains dataset using different methods. The x-axis represents methods' number of trainable parameters. The y-axis shows the run time in seconds.

training. Spatial detail when reconstructing MS lesions was preserved with better denoised images, compared to, e.g. the CascadeNet. The E2EVN and the LPDNet did not show any significant improvement in this respect. The reason for such behavior might be that these scans, in contrast to the training data, came without a fully sampled center since a separate reference scan was acquired. This deviation from the training data could explain the lower performance of some of the models. Conditional deep priors tend to learn dealiasing of undersampled acquisitions on images that they have trained on. In such a situation, learning k-space corrections might be disadvantageous. The KIKINet and the UNet performed significantly worse than the other methods, thereby appearing to be sensitive to noisy inputs. Furthermore, the models trained on knees were inadequate in reconstructing MS lesions, indicating training anatomy preference rather than generalization. Remarkably, this was also realized by the performance of the networks trained on the FLAIR-Brain datasets with equidistant 1D undersampling. All models generalized well on reconstructing the MS data, despite the deviating undersampling scheme (variable density poisson sampling in 2D).

The SNR, the number of coils, and the size of the training dataset appeared to be important parameters that influenced performance. This is to be seen in the reported SSIM and PSNR scores. Here, the E2EVN models performed highest on the largest dataset, i.e. the T_2-weighted knee dataset, which contained approximately 12 000 slices. However, all models experienced lower performance due to lower SNR (17.1 ± 4.5) and number of coils (8), compared to the T_1-weighted brain dataset (3000 slices, SNR of 25.7 ± 5.4, and 32 coils). The FLAIR brain dataset, despite its relatively high SNR (5000 slices and SNR of 23.6 ± 4.8), did not necessarily yield high quality in reconstructed images. The deviating number of coils (from 2 to 24), field strength (1.5 T an 3 T), and matrix sizes, resulted in a challenging dataset to converge with when training a model. In this situation, the advantage of implementing cascades was most apparent, making the CIRIM being robust

with all tested acceleration factors (4x, 6x, 8x—supplementary material and 10x—table 2.2). PICS and the UNet scored overall lower, illustrating that learning a prior with an efficient model is advantageous.

Importantly, our results show that the RIM-based models can reconstruct image details unseen during training. The RIM explicitly contains a formulation of the prior information of an MR image and acts as optimizer itself. Unrolled optimization is performed by gradient descent (Putzky and Welling 2017), such that DC is enforced implicitly. The CIRIM allows to further denoise the reconstructed images through the cascades without sharing parameters, similar to previously proposed deep cascading networks [26, 66, 67]. The cascades thereby allowed us to train an overall deep network of multiple connected RNNs that captures long-range dependencies while avoiding vanishing or exploding gradients. The E2EVN also performs unrolled optimization through cascades, but explicitly enforces DC with a formulated term.

Recent work has pointed out the importance of benchmarking and quantifying the performance of deep networks regarding the GPU memory required for training, the inference times, the applications, and the optimization [45, 68, 69]. With regard to inference times, methods such as the LPDNet and the KIKINet did not seem to improve in speed over the conventional CS algorithm, implemented on the GPU. The reason for these methods being slower is that they consist of deep feed-forward large convolutional layers. The RIM, the E2EVN, and the CascadeNet reduce reconstruction times by a factor of six compared to CS. Here, inference is performed over an iterative scheme, in which sharing of parameters is optimized either through time-steps or cascades. The IRIM and the UNet even further reduce the time by a factor of two and six, respectively. The CIRIM serves as a balanced deep network, being two times faster than the slowest methods and two times slower than the other cascading networks. The performance gain in further denoising and generalization capabilities may counterbalance the need for longer inference times.

2.5 Conclusion

The CIRIM implicitly enforces DC when targeting unrolled optimization through gradient descent. The comparable E2EVN performed best when DC was explicitly enforced, performing well on the training distributions. However, it appeared to be inadequate on reconstructing out-of-training distribution data without a fully sampled center. The CIRIM performed best on all training datasets, tested undersampling schemes and acceleration factors. Also, it showed robust performance on reconstructing accelerated FLAIR data containing MS lesions, achieving good lesion contrast and efficient denoising compared to PICS, the baseline RIM and the IRIM. In contrast, methods such as the CascadeNet and the LPDNet were sensitive to highly noisy untrained data, showing limited generalization capabilities. The KIKINet and the UNet tended to oversimplify the reconstructed images, performing markedly worse than rest methods. To that extent, the impression is that evaluating the forward process of accelerated MRI reconstruction, frequently through time, is of great importance for generalization in other settings. The implemented cascades and the application of the RIM to a deeper network allowed backpropagation on a smaller number of time-steps but on higher frequency for each iteration. Thus, a key advantage of the CIRIM is that it maintains a very fair trade-off between reconstructed image quality and fast reconstruction times, which is crucial in the clinical workflow.

2.6 Acknowledgments

This publication is based on the STAIRS project under the TKI-PPP program. The collaboration project is co-funded by the PPP Allowance made available by Health Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships.

M W A Caan is shareholder of Nico.Lab International Ltd.

2.7 Appendix

Appendix. Gated recurrent unit (GRU) The GRU has two gating units, the reset gate and the update gate. These gates control how the information flows in the network. The update gate regulates the update to a new hidden state, whereas the reset gate controls the information to forget. Both gates act in a probabilistic manner.

The activation of the reset gate $\mathbf{r}$ at iteration τ, for updating equation (9), is computed by

$$r_\tau = \sigma(W_r[s_{\tau-1}, x_\tau] + b_r), \tag{2.21}$$

Similarly, the update gate $\mathbf{z}$ is computed by

$$z_\tau = \sigma(W_z[s_{\tau-1}, x_\tau] + b_z). \tag{2.22}$$

The actual activation of the next hidden state $\mathbf{s}_\tau$ is then computed by

$$s_\tau = (1 - z_\tau) \odot s_{\tau-1} + z_\tau \odot \tilde{s}_\tau, \tag{2.23}$$

where $\odot$ represents the Hadamard product and $\tilde{s}_\tau$ is given by

$$\tilde{s}_\tau = \tanh(W_s[r_\tau \odot s_{\tau-1}, x_\tau] + b_s). \tag{2.24}$$

Appendix. Independently Recurrent Neural Network (IndRNN) The IndRNN addresses gradient decay over iterations, following an independent neuron connectivity within a recurrent layer. The update on equation (9) and at iteration τ is given by

$$s_\tau = \sigma\left(W_{x_\tau} + u \odot s_{\tau-1} + b\right), \tag{2.25}$$

where $\mathbf{W}$ is the weight for the current input, $\mathbf{u}$ is the weight for the recurrent input, and $\mathbf{b}$ is the bias vector.

References

[1] Daniel K. Sodickson and Warren J. Manning. Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays. *Magnetic Resonance in Medicine*, 38:591–603, 10 1997.

[2] Klaas P. Pruessmann, Markus Weiger, Markus B. Scheidegger, and Peter Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–962, 1999.

[3] Michael Lustig, David Donoho, and John M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[4] Michael Lustig, David L. Donoho, Juan M. Santos, and John M. Pauly. Compressed Sensing MRI. *IEEE Signal Processing Magazine*, 25(2):72–82, 2008.

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[6] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P. Recht, Daniel K. Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, 2018.

[7] Patrick Putzky and Max Welling. Recurrent Inference Machines for Solving Inverse Problems. *arXiv*, 2017.

[8] Kai Lønning, Patrick Putzky, Jan Jakob Sonke, Liesbeth Reneman, Matthan W.A. Caan, and Max Welling. Recurrent inference machines for reconstructing heterogeneous MRI data. *Medical Image Analysis*, 53:64–78, 2019.

[9] Patrick Putzky, Dimitrios Karkalousos, Jonas Teuwen, Nikita Miriakov, Bart Bakker, Matthan Caan, and Max Welling. i-RIM applied to the fastMRI challenge. *arXiv*, 2019.

[10] Mehmet Akçakaya, Steen Moeller, Sebastian Weingärtner, and Kâmil Uğurbil. Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging. *Magnetic Resonance in Medicine*, 81(1):439–453, 2019.

[11] Tae Hyung Kim, Pratyush Garg, and Jusin Haldar. LORAKI: Autocalibrated Recurrent Neural Networks for Autoregressive MRI Reconstruction in k-Space. 04 2019.

[12] Yamin Arefeen, Onur Beker, Jaejin Cho, Heng Yu, Elfar Adalsteinsson, and Berkin Bilgic. Scan-specific artifact reduction in k-space (SPARK) neural networks synergize with physics-based reconstruction to accelerate MRI. *Magnetic Resonance in Medicine*, 87(2):764–780, 2022.

[13] Jonas Adler and Ozan Öktem. Learned Primal-Dual Reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1322–1332, 2018.

[14] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40:120–145, 2011.

[15] Taejoon Eo, Yohan Jun, Taeseong Kim, Jinseong Jang, Ho Joon Lee, and Dosik Hwang. KIKI-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. *Magnetic Resonance in Medicine*, 80:2188–2201, 11 2018.

[16] Roberto Souza, Mariana Bento, Nikita Nogovitsyn, Kevin J. Chung, Wallace Loos, R. Marc Lebel, and Richard Frayne. Dual-domain cascade of U-nets for multi-channel magnetic resonance image reconstruction. *Magnetic Resonance Imaging*, 71:140–153, 9 2020.

[17] Hemant K. Aggarwal, Merry P. Mani, and Mathews Jacob. MoDL: Model-Based Deep Learning Architecture for Inverse Problems. *IEEE Transactions on Medical Imaging*, 38(2):394–405, 2019.

[18] Youssef Beauferris, Jonas Teuwen, Dimitrios Karkalousos, Nikita Moriakov, Mattha Caan, Lívia Rodrigues, Alexandre Lopes, Hélio Pedrini, Letícia Rittner, Maik Dannecker, Viktor Studenyak, Fabian Gröger, Devendra Vyas, Shahrooz Faghih-Roohi, Amrit Kumar A.K. Jethi, Jaya Chandra J.C. Raju, Mohanasankar Sivaprakasam, Wallace Loos, Richard Frayne, and Roberto Souza. Multi-channel MR Reconstruction (MC-MRRec) Challenge – Comparing Accelerated MR Reconstruction Models and Assessing Their Genereralizability to Datasets Collected with Different Coils. *arXiv*, pages 1–31, 11 2020.

[19] Florian Knoll, Tullie Murrell, Anuroop Sriram, Nafissa Yakubova, Jure Zbontar, Michael Rabbat, Aaron Defazio, Matthew J. Muckley, Daniel K. Sodickson, C. Lawrence Zitnick, and Michael P. Recht. Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge. *Magnetic Resonance in Medicine*, 84:3054–3070, 2020.

[20] Matthew J. Muckley, Bruno Riemenschneider, Alireza Radmanesh, Sunwoo Kim, Geunu Jeong, Jingyu Ko, Yohan Jun, Hyungseob Shin, Dosik Hwang, Mahmoud Mostapha, Simon Arberet, Dominik Nickel, Zaccharie Ramzi, Philippe Ciuciu, Jean-Luc Starck, Jonas Teuwen, Dimitrios Karkalousos, Chaoping Zhang, Anuroop Sriram, Zhengnan Huang, Nafissa Yakubova, Yvonne W. Lui, and Florian Knoll. Results of the 2020 fastMRI Challenge for Machine Learning MR Image Reconstruction, 2021.

[21] Anuroop Sriram, Jure Zbontar, Tullie Murrell, Aaron Defazio, C. Lawrence Zitnick, Nafissa Yakubova, Florian Knoll, and Patricia Johnson. End-to-End Variational Networks for Accelerated MRI Reconstruction. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12262 LNCS:64–73, 2020.

[22] Zaccharie Ramzi, Philippe Ciuciu, and Jean-Luc Starck. XPDNet for MRI Reconstruction: an Application to the fastMRI 2020 Brain Challenge. *arXiv*, 12 2020.

[23] Yohan Jun, Hyungseob Shin, Taejoon Eo, and Dosik Hwang. Joint Deep Model-based MR Image and Coil Sensitivity Reconstruction Network (Joint-ICNet) for Fast MRI. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 5266–5275, 2021.

[24] Eric Z. Chen, Puyang Wang, Xiao Chen, Terrence Chen, and Shanhui Sun. Pyramid Convolutional RNN for MRI Image Reconstruction. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, XX:1, 12 2019.

[25] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training recurrent neural networks. *30th International Conference on Machine Learning, ICML 2013*, pages 2347–2355, 2013.

[26] Jo Schlemper, Jose Caballero, Joseph V. Hajnal, Anthony N. Price, and Daniel Rueckert. A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction. *IEEE Transactions on Medical Imaging*, 37:491–503, 2018.

[27] Chen Qin, Jo Schlemper, Jose Caballero, Anthony N. Price, Joseph V. Hajnal, and Daniel Rueckert. Convolutional recurrent neural networks for dynamic MR image reconstruction. *IEEE Transactions on Medical Imaging*, 38:280–290, 2019.

[28] Roberto Souza, R Marc Lebel, Richard Frayne, and Rfrayne@ucalgary Ca. A Hybrid, Dual Domain, Cascade of Convolutional Neural Networks for Magnetic Resonance Image Reconstruction. *Proceedings of Machine Learning Research 102:437–446*, 102:1–11, 2019.

[29] Jo Schlemper, Chen Qin, Jinming Duan, Ronald M Summers, and Kerstin Hammernik. Σ-net: Ensembled Iterative Deep Neural Networks for Accelerated Parallel MR Image Reconstruction. pages 1–4, 12 2019.

[30] Jinming Duan, Jo Schlemper, Chen Qin, Cheng Ouyang, Wenjia Bai, Carlo Biffi, Ghalib Bello, Ben Statton, Declan P. O'Regan, and Daniel Rueckert. Vs-net: Variable splitting network for accelerated parallel MRI reconstruction. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11767 LNCS:713–722, 2019.

[31] Guang Yang, Simiao Yu, Hao Dong, Greg Slabaugh, Pier Luigi Dragotti, Xujiong Ye, Fangde Liu, Simon Arridge, Jennifer Keegan, Yike Guo, and David Firmin. DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction. *IEEE Transactions on Medical Imaging*, 37:1310–1321, 2018.

[32] Elizabeth K. Cole, John M. Pauly, Shreyas S. Vasanawala, and Frank Ong. Unsupervised MRI Reconstruction with Generative Adversarial Networks, 2020.

[33] Salman U.H. Dar, Mahmut Yurt, Mohammad Shahdloo, Muhammed Emrullah Ildiz, Berk Tinaz, and Tolga Cukur. Prior-guided image reconstruction for accelerated multi-contrast mri via generative adversarial networks. *IEEE Journal on Selected Topics in Signal Processing*, 14:1072–1087, 10 2020.

[34] Byeongsu Sim, Gyutaek Oh, Jeongsol Kim, Chanyong Jung, and Jong Chul Ye. Optimal Transport Driven CycleGAN for Unsupervised Learning in Inverse Problems. *SIAM Journal on Imaging Sciences*, 13:2281–2306, 1 2020.

[35] Yilmaz Korkmaz, Salman UH Dar, and Mahmut Yurt. Unsupervised MRI Reconstruction via Zero-Shot Learned Adversarial Transformers.

[36] Shanshan Wang, Huitao Cheng, Leslie Ying, Taohui Xiao, Ziwen Ke, Hairong Zheng, and Dong Liang. DeepcomplexMRI: Exploiting deep residual network for fast parallel

MR imaging with complex convolution. *Magnetic Resonance Imaging*, 68:136–147, 2020.

[37] Linfang Xiao, Yilong Liu, Zheyuan Yi, Yujiao Zhao, Linshan Xie, Peibei Cao, Alex T L Leong, and Ed X Wu. Partial Fourier reconstruction of complex MR images using complex-valued convolutional neural networks. *Magnetic Resonance in Medicine*, 87:999–1014, 2 2022.

[38] Salman Ul Hassan Dar, Muzaffer Özbey, Ahmet Burak Çatlı, and Tolga Çukur. A Transfer-Learning Approach for Accelerated MRI Using Deep Neural Networks. *Magnetic Resonance in Medicine*, 84:663–685, 8 2020.

[39] Bo Zhu, Jeremiah Z. Liu, Stephen F. Cauley, Bruce R. Rosen, and Matthew S. Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555:487–492, 2018.

[40] Yan Yang, Jian Sun, Huibin Li, and Zongben Xu. ADMM-net: A deep learning approach for compressive sensing MRI. *arXiv*, pages 1–14, 2017.

[41] Tran Minh Quan, Thanh Nguyen-Duc, and Won Ki Jeong. Compressed Sensing MRI Reconstruction Using a Generative Adversarial Network With a Cyclic Loss. *IEEE Transactions on Medical Imaging*, 37:1488–1497, 2018.

[42] Anuroop Sriram, Jure Zbontar, Tullie Murrell, C. Lawrence Zitnick, Aaron Defazio, and Daniel K. Sodickson. GrappaNet: Combining Parallel Imaging with Deep Learning for Multi-Coil MRI Reconstruction. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 14303–14310, 10 2020.

[43] Xiaohua Zhang, Qiusheng Lian, Yuchi Yang, and Yueming Su. A deep unrolling network inspired by total variation for compressed sensing MRI. *Digital Signal Processing: A Review Journal*, 107:102856, 12 2020.

[44] Nicola Pezzotti, Sahar Yousefi, Mohamed S. Elmahdy, Jeroen Hendrikus Fransiscus van Gemert, Christophe Schuelke, Mariya Doneva, Tim Nielsen, Sergey Kastryulin, Boudewijn PF P.F. F. Lelieveldt, Matthias J. P. JP J.P. van Osch, Elwin de Weerdt, Marius Staring, Christophe Schülke, Mariya Doneva, Tim Nielsen, Sergey Kastryulin, Boudewijn PF P.F. F. Lelieveldt, Matthias J. P. JP J.P. van Osch, Elwin DE Weerdt, Marius Staring, Elwin de Weerdt, and Marius Staring. An Adaptive Intelligence Algorithm for Undersampled Knee MRI Reconstruction. *IEEE Access*, 8:204825–204838, 2020.

[45] Kerstin Hammernik, Jo Schlemper, Chen Qin, Jinming Duan, Ronald M. Summers, and Daniel Rueckert. Systematic evaluation of iterative deep neural networks for fast parallel MRI reconstruction with sensitivity-weighted coil combination. *Magnetic Resonance in Medicine*, pages 1859–1872, 2021.

[46] Shuai Li, Wanqing Li, Chris Cook, Ce Zhu, and Yanbo Gao. Independently Recurrent Neural Network (IndRNN): Building A Longer and Deeper RNN. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 5457–5466, 2018.

[47] Md Faijul Amin, Muhammad Ilias Amin, A. Y.H. Al-Nuaimi, and Kazuyuki Murase. Wirtinger calculus based gradient descent and Levenberg-Marquardt learning algorithms in complex-valued neural networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7062 LNCS:550–559, 2011.

[48] Andy M Sarroff, Victor Shepardson, and Michael A Casey. Learning Representations Using Complex-Valued Nets. 11 2015.

[49] Huisheng Zhang and Danilo P. Mandic. Is a complex-valued stepsize advantageous in complex-valued gradient learning algorithms? *IEEE Transactions on Neural Networks and Learning Systems*, 27:2730–2735, 2016.

[50] Marcin Andrychowicz, Misha Denil, Sergio Gómez Colmenarejo, Matthew W. Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando De Freitas. Learning to learn by gradient descent by gradient descent. *Advances in Neural Information Processing Systems*, pages 3988–3996, 2016.

[51] Yunjin Chen, Wei Yu, and Thomas Pock. On learning optimized reaction diffusion processes for effective image restoration. pages 5261–5269, 2015.

[52] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pages 1724–1734, 2014.

[53] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13:600–612, 2004.

[54] Matthew J. Muckley, Bruno Riemenschneider, Alireza Radmanesh, Sunwoo Kim, Geunu Jeong, Jingyu Ko, Yohan Jun, Hyungseob Shin, Dosik Hwang, Mahmoud Mostapha, Simon Arberet, Dominik Nickel, Zaccharie Ramzi, Philippe Ciuciu, Jean-Luc Starck, Jonas Teuwen, Dimitrios Karkalousos, Chaoping Zhang, Anuroop Sriram, Zhengnan Huang, Nafissa Yakubova, Yvonne Lui, and Florian Knoll. State-of-the-Art Machine Learning MRI Reconstruction in 2020: Results of the Second fastMRI Challenge. 12 2020.

[55] Kevin Epperson, R Rt, Anne Marie Sawyer, R Rt, Michael Lustig, Marcus Alley, Martin Uecker, Patrick Virtue, Peng Lai, and Shreyas Vasanawala. Creation of Fully Sampled MR Data Repository for Compressed Sensing of the Knee. *SMRT Conference*, 2013:1, 2013.

[56] Martin Uecker, Frank Ong, Jonathan I. Tamir, Dara Bahri, Patrick Virtue, Joseph Y Cheng, Tao Zhang, and Michael Lustig. Berkeley Advanced Reconstruction Toolbox. *Proc. Intl. Soc. Mag. Reson. Med.*, 23, 2015.

[57] Martin Uecker, Peng Lai, Mark J. Murphy, Patrick Virtue, Michael Elad, John M. Pauly, Shreyas S. Vasanawala, and Michael Lustig. ESPIRiT-an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. *Magnetic Resonance in Medicine*, 71:990–1001, 3 2014.

[58] Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J. Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, Hersh Chandarana, Zizhao Zhang, Michal Drozdzal, Adriana Romero, Michael Rabbat, Pascal Vincent, Nafissa Yakubova, James Pinkerton, Duo Wang, Erich Owens, C. Lawrence Zitnick, Michael P. Recht, Daniel K. Sodickson, and Yvonne W. Lui. fastMRI: An Open Dataset and Benchmarks for Accelerated MRI. *arXiv*, pages 1–35, 11 2018.

[59] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*, 32, 12 2019.

[60] W. Falcon and The PyTorch Lightning Team. PyTorch Lightning, 2019.

[61] Oleksii Kuchaiev, Jason Li, Huyen Nguyen, Oleksii Hrinchuk, Ryan Leary, Boris Ginsburg, Samuel Kriman, Stanislav Beliaev, Vitaly Lavrukhin, Jack Cook, Patrice Castonguay, Jocelyn Huang, and Jonathan Cohen. NeMo: a toolkit for building AI applications using Neural Modules. 09 2019.

[62] George Yiasemis, Nikita Moriakov, Dimitrios Karkalousos, Matthan Caan, and Jonas Teuwen. DIRECT: Deep Image REConstruction Toolkit. *Journal of Open Source Software*, 7:4278, 5 2022.

[63] Victor I. J. Strijbis, Christiaan M. de Bloeme, Robin W. Jansen, Hamza Kebiri, Huu-Giao Nguyen, Marcus C. de Jong, Annette C. Moll, Merixtell Bach-Cuadra, Pim de Graaf, and Martijn D. Steenwijk. Multi-view convolutional neural networks for automated ocular structure and tumor segmentation in retinoblastoma. *Scientific Reports*, 11:14590, 12 2021.

[64] William Penny, Karl Friston, John Ashburner, Stefan Kiebel, and Thomas Nichols. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier, 2007.

[65] Ruiyang Zhao, Burhaneddin Yaman, Yuxin Zhang, Russell Stewart, Austin Dixon, Florian Knoll, Zhengnan Huang, Yvonne W Lui, Michael S Hansen, and Matthew P Lungren. fastMRI+: Clinical Pathology Annotations for Knee and Brain Fully Sampled Multi-Coil MRI Data. pages 1–7, 9 2021.

[66] Qiaoying Huang, Dong Yang, Pengxiang Wu, Hui Qu, Jingru Yi, and Dimitris Metaxas. MRI Reconstruction Via Cascaded Channel-Wise Attention Network. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1622–1626, 2019.

[67] Jiaqi Chen, Pengfei Zhang, Huafeng Liu, Lei Xu, and Heye Zhang. Spatio-temporal multi-task network cascade for accurate assessment of cardiac CT perfusion. *Medical Image Analysis*, 74:102207, 2021.

[68] Ke Wang, Michael Kellman, Christopher M Sandino, Kevin Zhang, Shreyas S Vasanawala, Jonathan I Tamir, Stella X Yu, and Michael Lustig. Memory-efficient Learning for High-Dimensional MRI Reconstruction.

[69] Zaccharie Ramzi, Philippe Ciuciu, and Jean Luc Starck. Benchmarking MRI reconstruction neural networks on large public datasets. *Applied Sciences (Switzerland)*, 10, 3 2020.

Supplementary Material

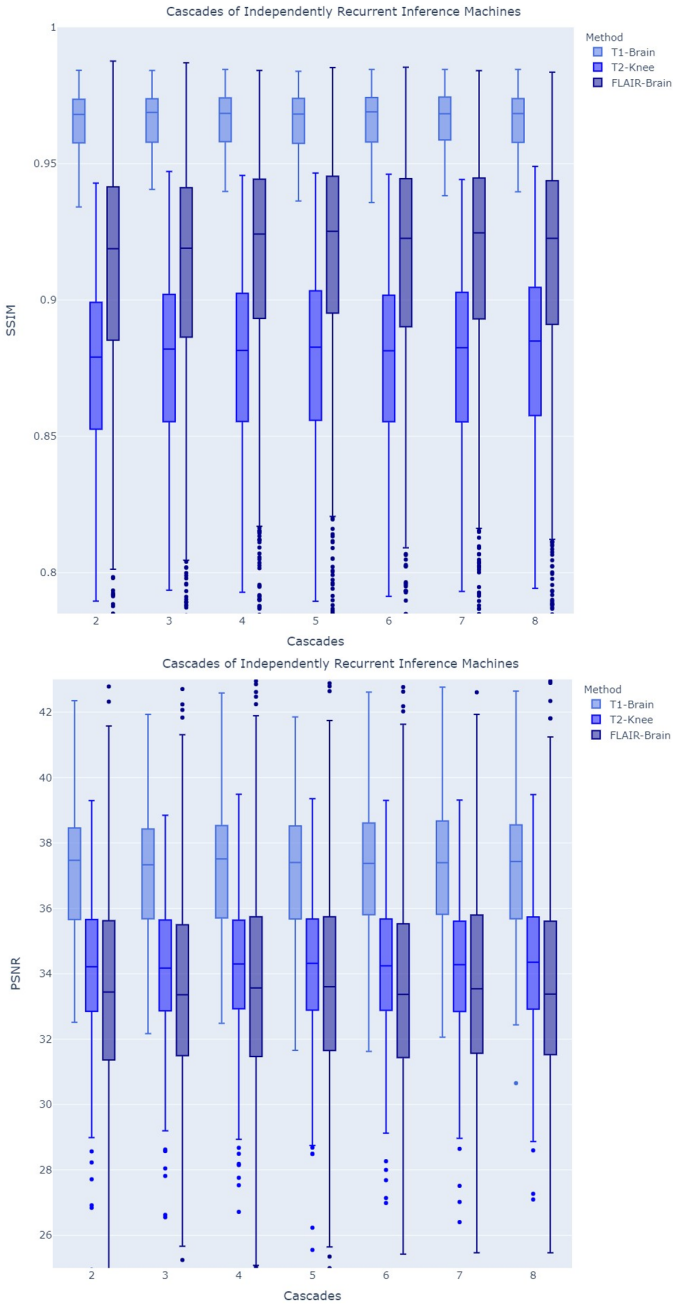


Figure S.2.1: Comparison of varying number of cascades for the Cascades of Independently Recurrent Inference Machines, on the trained datasets (T_1-Brain, T_2-Knee, FLAIR-Brain) using Gaussian 2D 10x undersampling. Top figure reports SSIM scores and bottom figure PSNR scores.

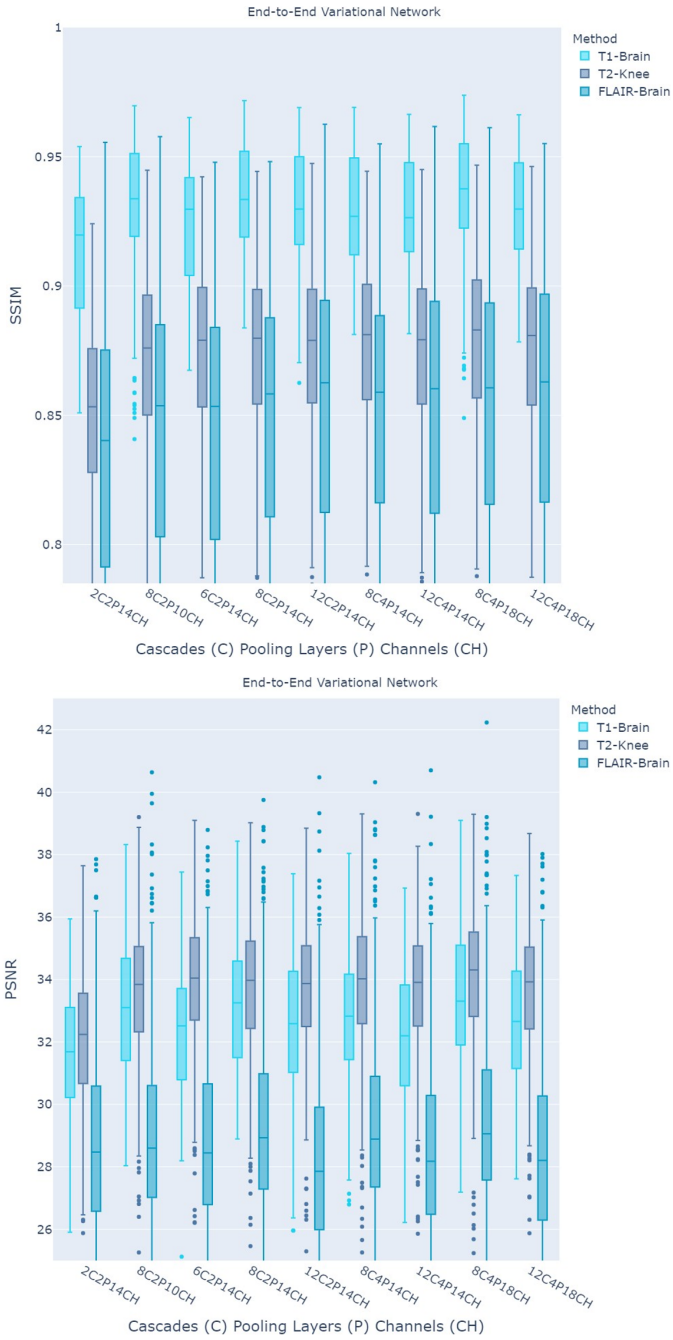


Figure S.2.2: Comparison of varying number of cascades, pooling layers, and channels for the End-to-End Variational Network, on the trained datasets (T_1-Brain, T_2-Knee, FLAIR-Brain) using Gaussian 2D 10x undersampling. Top figure reports SSIM scores and bottom figure PSNR scores.

Table S.2.1: SSIM & PSNR scores of all methods evaluated on the T_1-Brain dataset (third and fourth column), the T_2-Knee dataset (fifth and sixth column), and the FLAIR-Brain dataset (seventh, eighth, ninth and tenth column). For all datasets performance is reported for four times acceleration using Gaussian 2D undersampling. The second column reports the total number of trainable parameters for each model. Best performing models are highlighted in bold. Methods are sorted in alphabetical order.

Method	Params	T_1-Brain Gaussian2D4x		T_2-Knee Gaussian2D4x		FLAIR-Brain Gaussian2D4x	
		SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑
CascadeNet	1.1 M	0.962 ± 0.016	33.3 ± 0.9	0.908 ± 0.030	35.2 ± 2.3	0.925 ± 0.068	33.6 ± 4.7
CIRIM	264k	0.981 ± 0.007	39.2 ± 0.6	0.919 ± 0.027	36.3 ± 2.3	0.945 ± 0.061	36.5 ± 5.3
E2EVN	19.6 M	0.972 ± 0.011	35.6 ± 0.6	0.919 ± 0.027	36.4 ± 2.3	0.912 ± 0.071	32.6 ± 5.3
IRIM	53k	0.980 ± 0.008	38.9 ± 0.6	0.912 ± 0.029	35.8 ± 2.2	0.937 ± 0.066	35.9 ± 5.2
KIKINet	1.9 M	0.960 ± 0.020	34.9 ± 0.2	0.891 ± 0.033	34.6 ± 1.9	0.889 ± 0.074	32.1 ± 4.4
LPDNet	118k	0.976 ± 0.007	37.2 ± 0.0	0.907 ± 0.028	35.4 ± 1.9	0.898 ± 0.083	31.5 ± 4.6
PICS		0.912 ± 0.028	33.9 ± 0.4	0.814 ± 0.025	33.8 ± 3.7	0.856 ± 0.160	31.8 ± 10.0
RIM	94k	0.980 ± 0.008	39.0 ± 0.7	0.914 ± 0.027	36.0 ± 2.3	0.941 ± 0.063	36.0 ± 5.2
UNet	1.9 M	0.928 ± 0.022	28.1 ± 4.5	0.894 ± 0.033	34.2 ± 2.2	0.865 ± 0.087	30.3 ± 4.8
Zero-Filled		0.869 ± 0.056	20.1 ± 1.1	0.823 ± 0.017	22.8 ± 0.9	0.824 ± 0.084	21.0 ± 4.6

Table S.2.2: SSIM & PSNR scores of all methods evaluated on the T_1-Brain dataset (third and fourth column), the T_2-Knee dataset (fifth and sixth column), and the FLAIR-Brain dataset (seventh, eighth, ninth and tenth column). For all datasets performance is reported for six times acceleration using Gaussian 2D undersampling. The second column reports the total number of trainable parameters for each model. Best performing models are highlighted in bold. Methods are sorted in alphabetical order.

Method	Params	T_1-Brain Gaussian2D6x		T_2-Knee Gaussian2D6x		FLAIR-Brain Gaussian2D6x	
		SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑
CascadeNet	1.1 M	0.953 ± 0.022	32.5 ± 0.8	0.886 ± 0.035	34.0 ± 2.4	0.907 ± 0.079	32.1 ± 4.7
CIRIM	264k	0.975 ± 0.010	37.6 ± 0.3	0.901 ± 0.032	35.1 ± 2.2	0.932 ± 0.073	35.1 ± 5.3
E2EVN	19.6 M	0.963 ± 0.014	33.9 ± 1.0	0.901 ± 0.032	35.1 ± 2.2	0.891 ± 0.084	30.8 ± 5.4
IRIM	53k	0.974 ± 0.011	37.2 ± 0.6	0.894 ± 0.034	34.7 ± 2.4	0.921 ± 0.079	34.3 ± 5.0
KIKINet	1.9 M	0.948 ± 0.030	33.5 ± 0.4	0.868 ± 0.039	33.5 ± 2.0	0.856 ± 0.094	30.3 ± 4.6
LPDNet	118k	0.971 ± 0.010	36.3 ± 0.3	0.893 ± 0.031	34.6 ± 1.7	0.883 ± 0.093	30.9 ± 4.5
PICS		0.889 ± 0.029	32.4 ± 0.5	0.779 ± 0.031	32.1 ± 3.9	0.842 ± 0.163	31.0 ± 8.9
RIM	94k	0.974 ± 0.010	37.5 ± 0.7	0.896 ± 0.033	34.5 ± 2.3	0.926 ± 0.075	34.6 ± 5.2
UNet	1.9 M	0.910 ± 0.035	27.8 ± 3.6	0.872 ± 0.040	32.8 ± 2.4	0.829 ± 0.102	28.7 ± 4.7
Zero-Filled		0.821 ± 0.077	18.2 ± 2.1	0.746 ± 0.024	19.6 ± 1.2	0.739 ± 0.109	17.8 ± 4.5

Table S.2.3: SSIM & PSNR scores of all methods evaluated on the T_1-Brain dataset (third and fourth column), the T_2-Knee dataset (fifth and sixth column), and the FLAIR-Brain dataset (seventh, eighth, ninth and tenth column). For all datasets performance is reported for eight times acceleration using Gaussian 2D undersampling. The second column reports the total number of trainable parameters for each model. Best performing models are highlighted in bold. Methods are sorted in alphabetical order.

Method	Params	T_1-Brain Gaussian2D8x		T_2-Knee Gaussian2D8x		FLAIR-Brain Gaussian2D8x	
		SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑
CascadeNet	1.1 M	0.940 ± 0.031	31.8 ± 0.8	0.870 ± 0.040	33.0 ± 2.5	0.888 ± 0.090	31.1 ± 4.6
CIRIM	264k	0.970 ± 0.012	36.6 ± 0.5	0.888 ± 0.043	34.3 ± 2.3	0.922 ± 0.082	34.2 ± 5.1
E2EVN	19.6 M	0.952 ± 0.020	32.6 ± 1.3	0.887 ± 0.036	34.3 ± 2.5	0.870 ± 0.094	30.0 ± 4.8
IRIM	53k	0.968 ± 0.014	36.2 ± 0.4	0.881 ± 0.038	34.0 ± 2.2	0.908 ± 0.088	33.3 ± 4.9
KIKINet	1.9 M	0.936 ± 0.034	32.2 ± 0.9	0.853 ± 0.042	32.7 ± 1.8	0.833 ± 0.108	29.0 ± 4.7
LPDNet	118k	0.966 ± 0.013	35.4 ± 0.2	0.882 ± 0.035	33.9 ± 1.8	0.868 ± 0.102	30.2 ± 4.6
PICS		0.875 ± 0.030	31.5 ± 0.7	0.752 ± 0.036	30.7 ± 4.0	0.834 ± 0.164	30.5 ± 8.2
RIM	94k	0.969 ± 0.013	36.3 ± 0.5	0.883 ± 0.036	34.1 ± 2.2	0.914 ± 0.084	33.6 ± 5.0
UNet	1.9 M	0.891 ± 0.042	27.1 ± 3.3	0.857 ± 0.044	32.1 ± 2.1	0.800 ± 0.114	27.6 ± 4.5
Zero-Filled		0.790 ± 0.080	17.7 ± 2.0	0.702 ± 0.029	18.2 ± 1.2	0.688 ± 0.123	16.5 ± 4.6

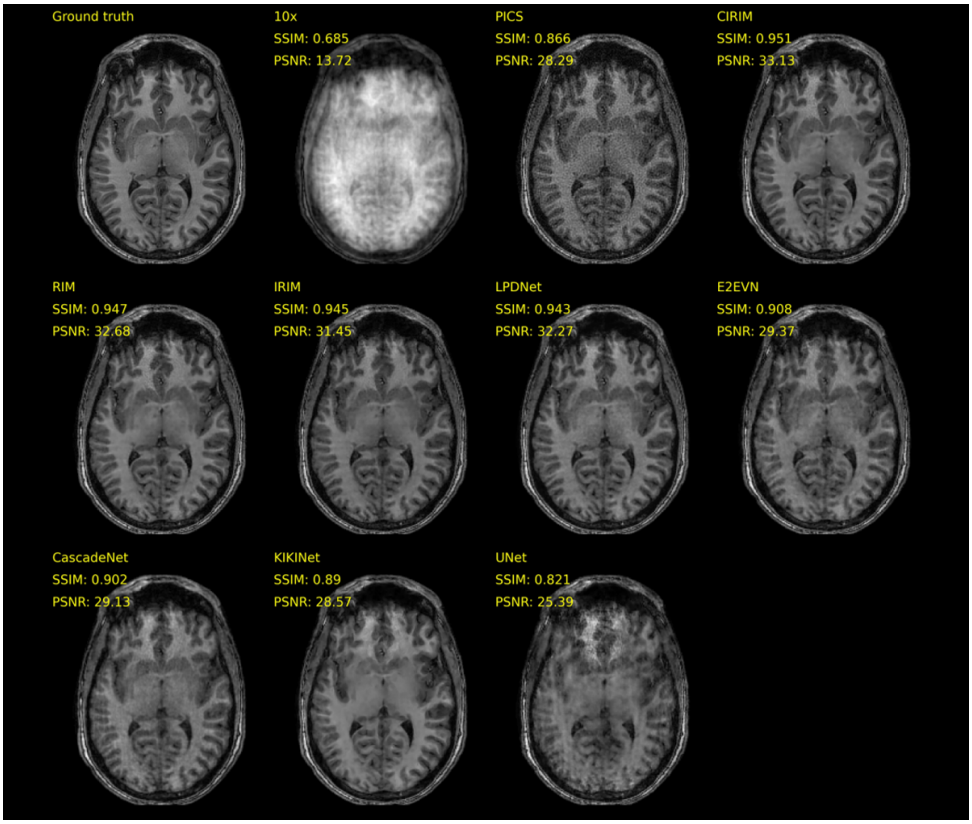


Figure S.2.3: Reconstructions of a ten times accelerated slice with a Gaussian 2D mask, from the validation set of the T_1-weighted brains dataset (first row-second). The ground truth is presented on the first row-first image. The CIRIM 7C (first row-fourth), the RIM (second row-first), and the IRIM (second row-second) enforced Data Consistency (DC) implicitly by gradient descent. The E2EVN 8C (second row-fourth), the CascadeNet (third row-first), and the KIKINet (third row-second) enforced DC explicitly by a formulated DC term.

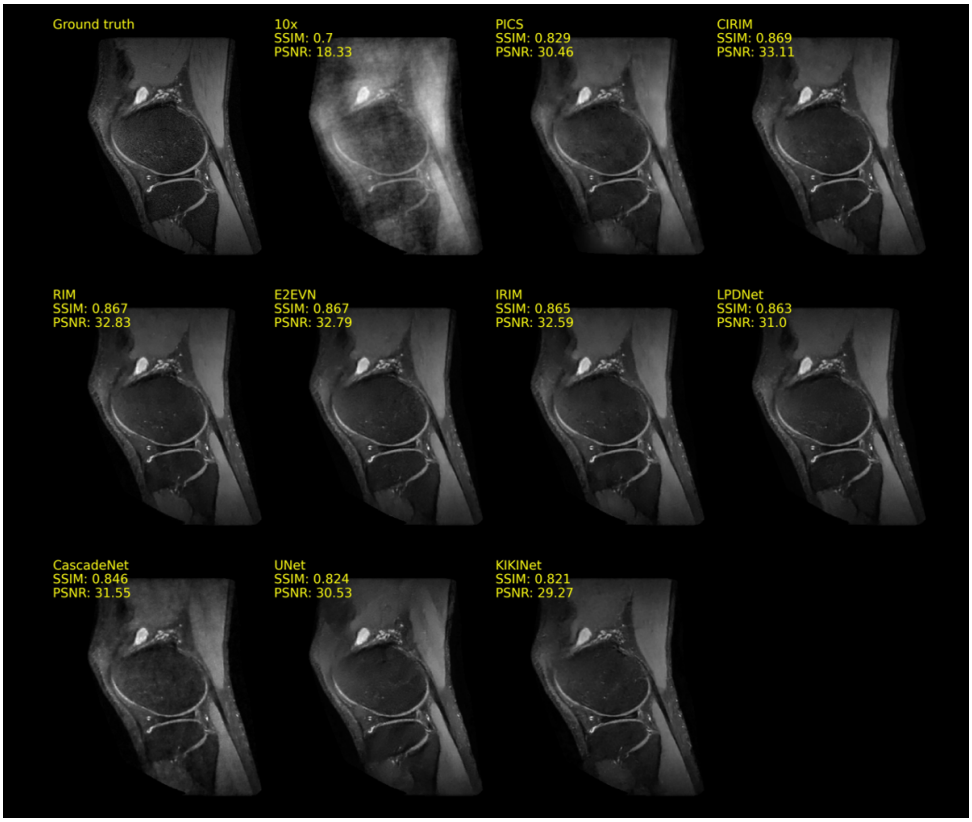


Figure S.2.4: Reconstructions of a ten times accelerated slice with a Gaussian 2D mask, from the validation set of the T_2-weighted knees dataset (first row-second). The ground truth is presented on the first row-first image. The CIRIM 5C (first row-fourth), the RIM (second row-first), and the IRIM (second row-third) enforced Data Consistency (DC) implicitly by gradient descent. The E2EVN 8C (second row-second), the CascadeNet (third row-first), and the KIKINet (third row-third) enforced DC explicitly by a formulated DC term.

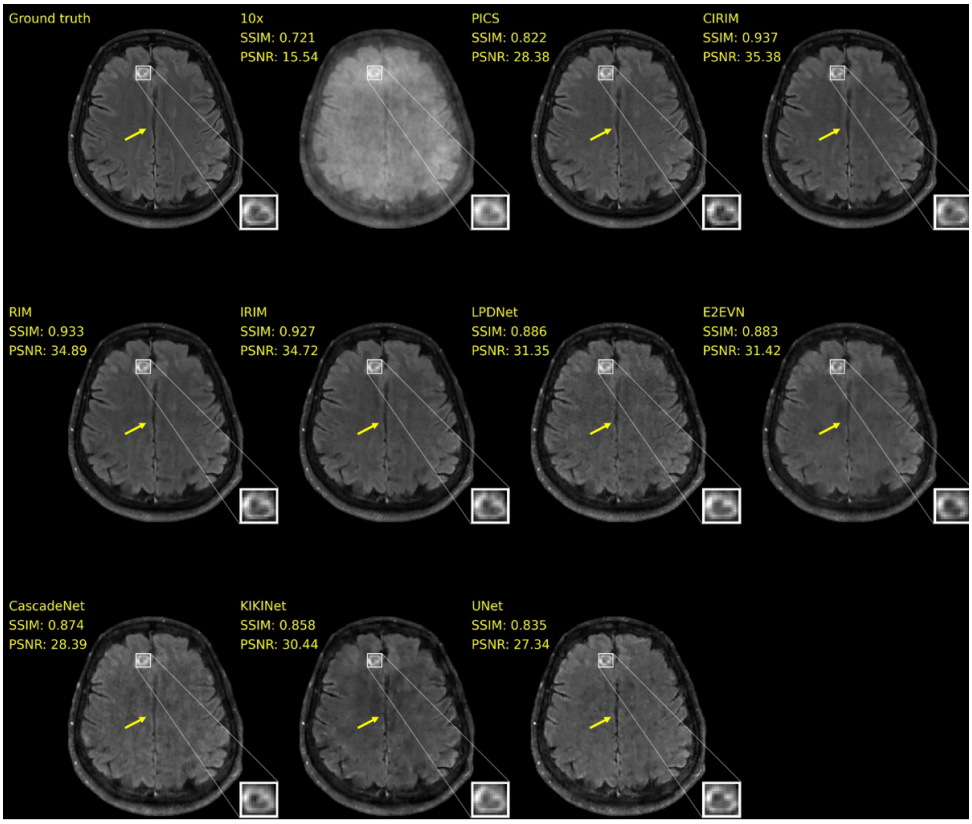


Figure S.2.5: Reconstructions of a ten times accelerated slice with a Gaussian 2D mask, from the validation set of the FLAIR brains dataset (first row-second). The ground truth is presented on the first row-first image. The inset focuses on a reconstructed White Matter lesion; obtained through the fastMRI+ annotations (Zhao et al., 2021). The arrow points out to a region of interested that some models failed to reconstruct. The CIRIM 5C (first row-fourth), the RIM (second row-first), and the IRIM (second row-second) enforced Data Consistency (DC) implicitly by gradient descent. The E2EVN 8C (second row-fourth), the CascadeNet (third row-first), and the KIKINet (third row-second) enforced DC explicitly by a formulated DC term.

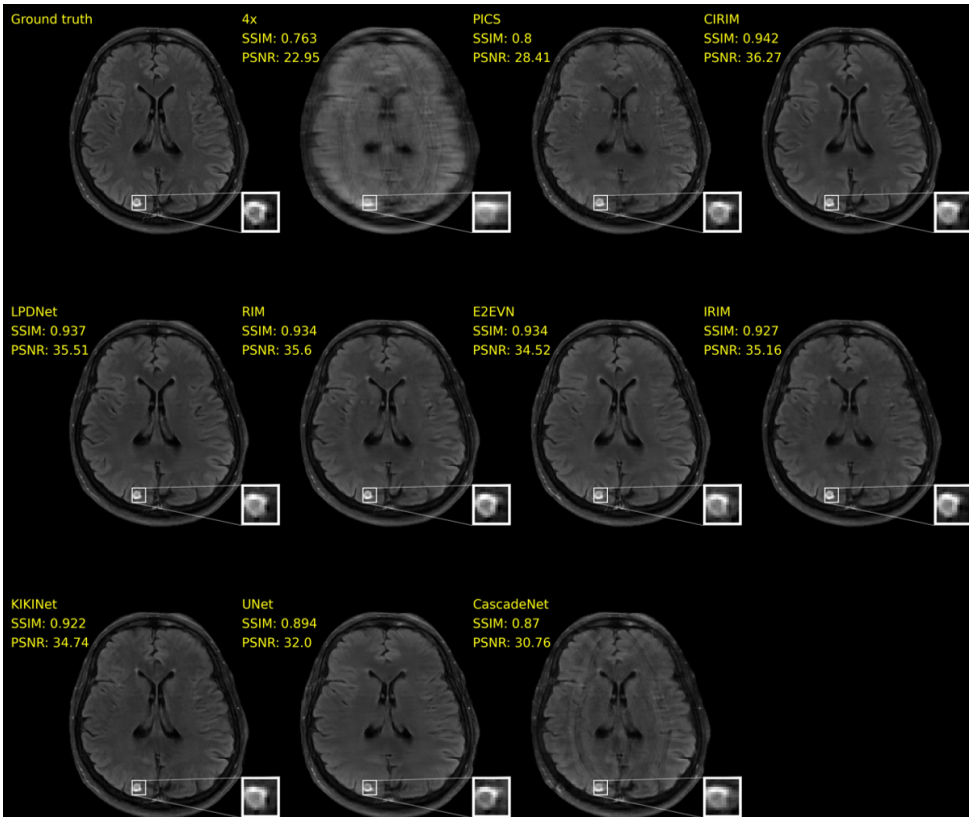


Figure S.2.6: Reconstructions of a four times accelerated slice with an equidistant 1D mask, from the validation set of the FLAIR brains dataset (first row-second). The ground truth is presented on the first row-first image. The inset focuses on a reconstructed White Matter lesion; obtained through the fastMRI+ annotations (Zhao et al., 2021). The CIRIM 5C (first row-fourth), the RIM (second row- second), and the IRIM (second row- fourth) enforced Data Consistency (DC) implicitly by gradient descent. The E2EVN 8C (second row-third), the KIKINet (third row-first), and the CascadeNet (third row-third) enforced DC explicitly by a formulated DC term.

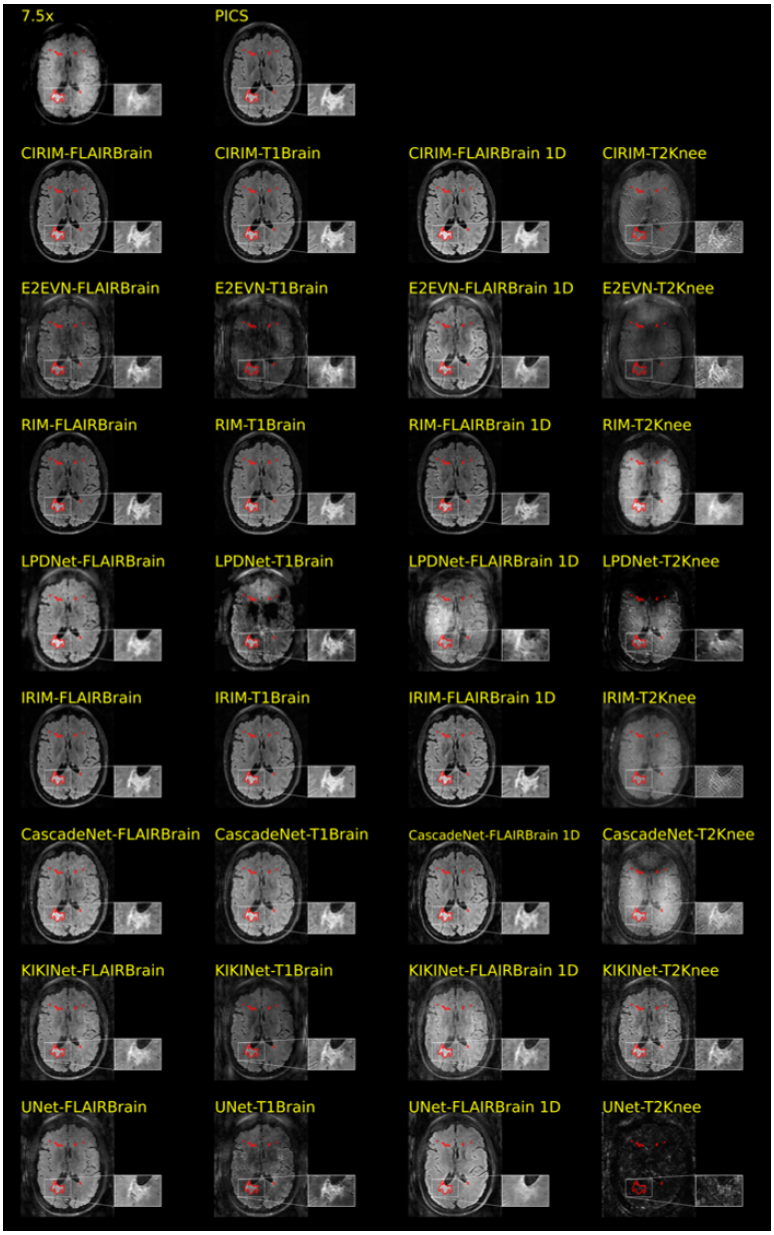


Figure S.2.7: Reconstructions of a representative axial slice of a 7.5x accelerated 3D FLAIR scan of a MS patient. Segmented MS lesions are depicted with red colored contours. Shown is the aliased linear reconstruction (first row-first image), PICS (first row-second image), and models' reconstructions on each trained scheme: the FLAIR-Brain dataset with Gaussian 2D undersampling (second-last row, first column), the T_1-Brain dataset with Gaussian 2D undersampling (second-last row, second column), the FLAIR-Brain dataset with equidistant 1D undersampling (second-last row, third column), and the T_2-Knee dataset with Gaussian 2D undersampling (second-last row, fourth column). The inset on the right bottom of each reconstruction focuses on a lesion region with high spatial detail.

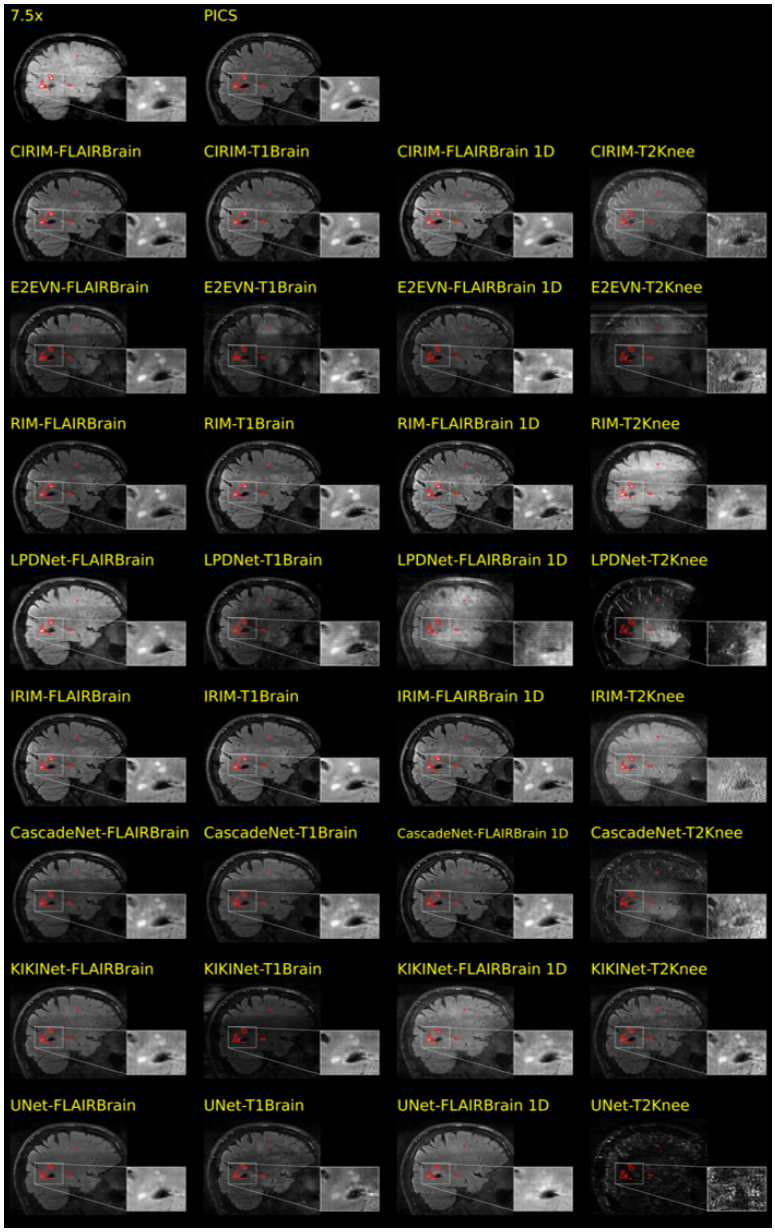


Figure S.2.8: Reconstructions of a representative sagittal slice of a 7.5x accelerated 3D FLAIR scan of a MS patient. Segmented MS lesions are depicted with red colored contours. Shown is the aliased linear reconstruction (first row-first image), PICS (first row-second image), and models' reconstructions on each trained scheme: the FLAIR-Brain dataset with Gaussian 2D undersampling (second-last row, first column), the T_1-Brain dataset with Gaussian 2D undersampling (second-last row, second column), the FLAIR-Brain dataset with equidistant 1D undersampling (second-last row, third column), and the T_2-Knee dataset with Gaussian 2D undersampling (second-last row, fourth column). The inset on the right bottom of each reconstruction focuses on a lesion region with high spatial detail.

3

Multi-Coil MRI Reconstruction Challenge—Assessing Brain MRI Reconstruction Models and Their Generalizability to Varying Coil Configurations

Beauferris, Y., Teuwen, J., **Karkalousos, D.**, Moriakov, N., Caan, M.W.A., Yiasemis, G., Rodrigues, L., Lopes, A., Pedrini, H., Rittner, L., Dannecker, M., Studenyak, V., Gröger, F., Vyas, D., Faghih-Roohi, S., Kumar Jethi, A., Chandra Raju, J., Sivaprakasam, M., Lasby, M., Nogovitsyn, N., … Souza, R. (2022).
Frontiers in neuroscience, 16, 919186.
DOI: 10.3389/fnins.2022.919186

Abstract

Deep-learning-based brain magnetic resonance imaging (MRI) reconstruction methods have the potential to accelerate the MRI acquisition process. Nevertheless, the scientific community lacks appropriate benchmarks to assess the MRI reconstruction quality of high-resolution brain images, and evaluate how these proposed algorithms will behave in the presence of small, but expected data distribution shifts. The multi-coil MRI (MC-MRI) reconstruction challenge provides a benchmark that aims at addressing these issues, using a large dataset of high-resolution, three-dimensional, T_1-weighted MRI scans. The challenge has two primary goals: (1) to compare different MRI reconstruction models on this dataset and (2) to assess the generalizability of these models to data acquired with a different number of receiver coils. In this paper, we describe the challenge experimental design and summarize the results of a set of baseline and state-of-the-art brain MRI reconstruction models. We provide relevant comparative information on the current MRI reconstruction state-of-the-art and highlight the challenges of obtaining generalizable models that are required prior to broader clinical adoption. The MC-MRI benchmark data, evaluation code, and current challenge leaderboard are publicly available. They provide an objective performance assessment for future developments in the field of brain MRI reconstruction.

3.1 Introduction

Brain magnetic resonance imaging (MRI) is a commonly used diagnostic imaging modality. It is a non-invasive technique that provides images with excellent soft-tissue contrast. Brain MRI produces a wealth of information, which often leads to a definitive diagnosis of a number of neurological conditions, such as cancer and stroke. Furthermore, it is broadly adopted in neuroscience and other research domains. MRI data acquisition occurs in the Fourier or spatial-frequency domain, more commonly referred to as k-space. IImage reconstruction consists of transforming the acquired k-space raw data into interpretable images. Traditionally, data is collected following the Nyquist sampling theorem [1], and for a single-coil acquisition, a simple inverse Fourier Transform operation is often sufficient to reconstruct an image. However, the fundamental physics, practical engineering aspects, and biological tissue response factors underlying the MRI data acquisition process make fully sampled acquisitions inherently slow. These limitations represent a crucial drawback when MRI is compared to other medical imaging modalities, impact both patient tolerance of the procedure and throughput, and more broadly neuroimaging research.

Parallel imaging (PI) [2–4] and compressed sensing (CS) [5, 6] are two proven approaches that are able to reconstruct high-fidelity images from sub-Nyquist sampled acquisitions. PI techniques leverage the spatial information available across multiple, spatially distinct, receiver coils to allow the reconstruction of undersampled kspace data. Techniques, such as generalized autocalibrating partially parallel acquisition (GRAPPA) [3], which operates in the k-space domain, and sensitivity encoding for fast MRI (SENSE) [2], which works in the image domain, are currently used clinically. CS methods leverage image sparsity properties to improve reconstruction quality from undersampled k-space data. Some CS techniques, such as compressed SENSE [6], have also seen clinical adoption. Those PI and CS methods that have been approved for routine clinical use are generally restricted to relatively conservative acceleration factors (*e.g.*, $R = 2\times$ to $3\times$ acceleration).

Currently employed comprehensive brain MRI scanning protocols, even those that use PI and CS, typically require between 30 and 45 min per patient procedure. Longer procedural times increase patient discomfort, thus lessening the likelihood of patient acceptance. It also increases susceptibility to both voluntary and involuntary motion artifacts.

In 2016, the first deep-learning-based MRI reconstruction models were presented [7, 8]. The excellent initial results obtained by these models caught the attention of the MR imaging community, and subsequently, dozens of deep-learning-based MRI reconstruction models were proposed, (*cf.*, [7–30]. Many of these studies demonstrated superior quantitative results from deep-learning-based methods compared to nondeep-learning-based MRI reconstruction algorithms [10, 16, 31]. These new methods are also capable of accelerating MRI examinations beyond traditional PI and CS methods. There is good evidence that deep-learning-based MRI reconstruction methods can accelerate MRI examinations by factors greater than 5 [32, 33].

A significant drawback, that hinders the progress of the brain MRI reconstruction field, is the lack of benchmark datasets. Importantly, the lack of benchmarks makes the comparison of different methods challenging. The fastMRI effort [32] is an important initiative that provides large volumes of raw MRI *k*-space data. The initial release of the fastMRI dataset provided two-dimensional (2D) MR acquisitions of the knee. A subsequent release added 2D brain MRI data with 5 mm slice thickness, which was used for the 2020 *fastMRI* challenge [34]. The *Calgary-Campinas* [35] initiative contains numerous sets of brain imaging data. For the purposes of this benchmark, we expanded the *Calgary-Campinas* initiative to include MRI raw data from three-dimensional (3D), high-resolution acquisitions. High-resolution images are crucial for many neuroimaging applications. Also importantly, 3D acquisitions allow for undersampling along two phase encoding dimensions, instead of one for 2D imaging. This potentially allows for further MRI acceleration. These *k*-space datasets correspond to either 12- or 32-channel data.

The goals of the Multi-Coil Magnetic Resonance Image (MC-MRI - https://www.ccdataset.com/mr-reconstruction-challenge) Reconstruction Challenge are to provide benchmarks that help improve the quality of brain MRI reconstruction, facilitate comparison of different reconstruction models, better understand the difficulties related to clinical adoption of these models, and investigate the upper limits of MR acceleration. The specific objectives of the challenge are as follows:

1. Compare the performance of different brain MRI reconstruction models on a large dataset, and

2. Assess the generalizability of these models to datasets acquired with different coils.

The results presented in this report correspond to benchmark submissions received up to 20 November, 2021. Four baseline solutions and three new benchmark solutions were presented and discussed during an online session at the Medical Imaging Deep Learning Conference held on 9 July, 2020.[1]. Two additional benchmark solutions were submitted after the online session. Collectively, these results provide a relevant performance summary of some state of the art MRI reconstruction approaches, including different model architectures, processing strategies, and emerging metrics for training and assessing reconstruction

[1]See video of session at https://www.ccdataset.com/mr-reconstruction-challenge/mc-mrrec-2020-midl-recording

models. The MC-MRI reconstruction challenge is ongoing and open to new benchmark submissions[2]. A public code repository with instructions on how to load the data, extract the benchmark metrics, and baseline reconstruction models are available at https://github.com/rmsouza01/MC-MRI-Rec.

3.2 Materials and Methods

3.2.1 Calgary-Campinas Raw MRI Dataset

The data used in this challenge were acquired as part of the Calgary Normative Study [36], which is a multi-year, longitudinal project that investigates normal human brain aging by acquiring quantitative MRI data using a protocol approved by our local research ethics board. Raw data from T_1-weighted volumetric imaging was acquired, anonymized, and incorporated into the *Calgary-Campinas* (*CC*) dataset [35]. The publicly accessible dataset currently provides k-space data from 167 3D, T_1-weighted, gradient-recalled echo, 1 mm3 isotropic sagittal acquisitions collected on a clinical 3T MRI scanner (Discovery MR750; General Electric Healthcare, Waukesha, WI). The brain scans are from presumed healthy subjects (mean ± standard deviation age: 44.5 ± 15.5 years; range: 20 years to 80 years; 71/167 (42.5%) male).

The datasets were acquired using either a 12-channel (117 scans, 70.0%) or 32-channel receiver coil (50 scans, 30.0%). Acquisition parameters were TR/TE/TI = 6.3 ms / 2.6 ms / 650 ms (93 scans, 55.7%) or TR/TE/TI = 7.4 ms / 3.1 ms / 400 ms (74 scans, 44.3%), with 170 to 180 contiguous 1.0 mm slices and a field of view of 256 mm × 218 mm. The acquisition matrix size $[N_x, N_y, N_z]$ for each channel was $[256, 218, 170-180]$, where x, y, and z denote readout, phase-encode, and slice-encode directions, respectively. In the slice-encode (k_z) direction, only 85% of the k-space data were collected; the remainder (15% of 170-180) was zero-filled. This partial acquisition technique is common practice in MRI. The average scan duration is 341 seconds. Because k-space undersampling only occurs in the phase-encode and slice-encode directions, the 1D inverse Fourier transform (iFT) along k_x was automatically performed by the scanner and hybrid (x, k_y, k_z) datasets were provided. This pre-processing effectively allows the MRI reconstruction problem to be treated as a 2D problem (in k_y and k_z). The partial Fourier reference data was reconstructed by taking the 2D iFT along the $k_y - k_z$ plane for each individual channel and combining these using the conventional square-root sum-of-squares algorithm [37].

3.2.2 MC-MRI Reconstruction Challenge Description

The MC-MRI Reconstruction Challenge was designed to be an ongoing investigation that will be disseminated through a combination of in-person sessions at meetings and virtual sessions, supplemented by periodic online submissions and updates. The benchmark is readily extensible and more data, metrics, and research questions are expected to be added in further updates. Individual research groups are permitted to make multiple submissions. The processing of submissions is semi-automated, and it takes on average 48 h to generate an update of the benchmark leaderboard.

Currently, the MC-MRI reconstruction challenge is split into two separate tracks. Teams can decide whether to submit a solution to just one track or to both tracks. Each track has

[2]See current leaders for the individual challenge tracks at https://www.ccdataset.com/

a separate leaderboard. The tracks are:

- **Track 01:** Teams had access to 12-channel data to train and validate their models. Models submitted are evaluated by only using the 12-channel test data.
- **Track 02:** Teams had access to 12-channel data to train and validate their models. Models submitted are evaluated for both the 12-channel and 32-channel test data.

In both tracks, the goal is to assess the brain MR image reconstruction quality and in particular note any loss of high-frequency details, especially at the higher acceleration rates. By having two separate tracks, we hoped to determine whether a generic reconstruction model trained on data from one coil would have decreased performance when applied to data from another coil.

Two MRI acceleration factors were tested: $R = 5$ and $R = 10$. These factors were chosen intentionally to exceed the acceleration factors typically used clinically with PI and CS methods. A Poisson disc distribution sampling scheme, where the center of k-space was fully sampled within a circle of radius of 16 pixels to preserve the low-frequency phase information, was used to achieve these acceleration factors. For brevity, we have only reported the results for $R = 5$, but the online challenge leaderboard contains the results for both acceleration factors.

The training, validation and test split of the challenge data is summarized in Table 3.1. The initial 50 and last 50 slices in each participant's image volume were removed because they have little anatomy present. The fully sampled k-space data of the training and validation sets were made public for teams to develop their models. Pre-undersampled k-space data corresponding to the test sets were provided for the teams for accelerations of $R = 5$ and $R = 10$.

Table 3.1: Summary of the raw MRI k-space datasets used in the first edition of the challenge. Reported are the number of slices in the test sets after removal of the initial 50 and last 50 slices (see text).

Coil	Category	# of datasets	# of slices
12-channel	Train	47	12,032
	Validation	20	5,120
	Test	50	7,800
32-channel	Test	50	7,800

3.2.3 Quantitative Metrics

In order to measure the quality of the image reconstructions, three commonly used, quantitative performance metrics were selected: peak signal-to-noise ratio (pSNR), structural similarity (SSIM) index [38], and visual information fidelity (VIF) [39]. The choice of performance metrics is challenging and it is recognized that objective measures such as pSNR, SSIM, and VIF may not correlate well with subjective human image quality assessments. Nonetheless, these metrics provide a broad basis to assess model performance in this challenge.

The pSNR is a metric commonly used for MRI reconstruction assessment and consists of the log ratio between the maximum value of the reference reconstruction and the root mean squared error (RMSE):

$$pSNR(y,\hat{y}) = 20\log_{10}\left(\frac{\max(y)}{\text{RMSE}}\right) = 20\log_{10}\left(\frac{\max(y)}{\sqrt{\frac{1}{M}\sum_{i=1}^{M}[y(i)-\hat{y}(i)]^2}}\right), \tag{3.1}$$

where y is the reference image, $\hat{y}$ is the reconstructed image, and M is the number of pixels in the image. Higher pSNR values represent higher-fidelity image reconstructions. However, pSNR does not take into consideration the factors involved in human vision. For this reason, increased pSNR can suggest that reconstructions are of higher quality, when in fact they may not be as well-perceived by the human visual system.

Unlike pSNR, SSIM and VIF are metrics that attempt to model aspects of the human visual system. SSIM considers biological factors, such as luminance, contrast, and structural information. SSIM is computed using:

$$SSIM(x,\hat{x}) = \frac{(2\mu_x\mu_{\hat{x}} + c_1)(2\sigma_{x\hat{x}} + c_2)}{(\mu^2{}_x + \mu^2{}_{\hat{x}} + c_1)(\sigma^2{}_x + \sigma^2{}_{\hat{x}} + c_2)} \tag{3.2}$$

where x and $\hat{x}$ represent corresponding image windows from the reference image and the reconstructed image, respectively; μ_x and σ_x represent the mean and standard deviation inside the image window, x; and $\mu_{\hat{x}}$ and $\sigma_{\hat{x}}$ represent the mean and standard deviation inside the reconstructed image window, $\hat{x}$. The constants c_1 and c_2 are used to avoid numerical instability. SSIM values for non-negative images are within $[0,1]$, where 1 represents two identical images.

The visual information fidelity metric is based on natural scene statistics [40, 41]. VIF models the natural scene statistics based on a Gaussian scale mixture model in the wavelet domain, and additive white Gaussian noise is used to model the human visual system. The natural scene of the reference image is modeled into wavelet components (C) and the human visual system is modeled by adding zero-mean white Gaussian noise in the wavelet domain (N), which results in the perceived reference image (E = C + N). In the same way, the reconstructed image, which is called the distorted image, is also modeled by a natural scene model (D) and the human visual system model (N'), leading to the perceived distorted image (F = D + N'). The VIF is given by the ratio of the mutual information of I(C, F) and I(C, E):

$$\text{VIF} = \frac{I(C,F)}{I(C,E)}, \tag{3.3}$$

where I represents the mutual information.

Mason *et al.* [42] investigated the VIF metric for assessing MRI reconstruction quality. Their results indicated that it has a stronger correlation with subjective radiologist opinion about MRI quality than other metrics such as pSNR and SSIM. The VIF Gaussian noise variance was set to 0.4 as recommended in [42]. All metrics were computed slice-by-slice in the test set. The reference and reconstructed images were normalized by dividing them by the maximum value of the reference image.

3.2.4 Visual Assessment

An expert observer (NN) with over 5 years of experience analyzing brain MR images and manually segmenting complex structures, such as the hippocampus and hypothalamus, visually inspected 25 randomly selected volumes for the 12-channel test set and other 25 volumes for the 32-channel test set for the best two submissions as determined from the quantitative metrics. The best two submissions were obtained by sorting the weighted average ranking. The weighted average ranking was generated by applying pre-determined weights to the ranking of the three individual quantitative metrics (0.4 for VIF, 0.4 for SSIM, and 0.2 for pSNR). We chose to give higher weights to VIF and SSIM because they have a better correlation with the human perception of image quality.

The visual assessment of the images was done by comparing the machine-learning-based reconstructions to the fully sampled reference images. This allowed the observer to distinguish between data acquisition related quality issues (e.g., motion) and problems associated with image reconstruction. The image quality assessment focused mostly on overall image quality and how well-defined was the contrast between white-matter, graymatter, and other relevant brain structures. The goal of the visual assessment was to compare the quality of the reconstructed MR images against the fully sampled reference images and not to compare the quality of the different submissions, because the benchmark is ongoing and we wanted to account for potential observer memory bias effects [43] in the qualitative metrics due to the difference between submission dates of the different solutions to the benchmark (*i.e.*, future submissions will be visually assessed at different dates compared to current submissions).

3.2.5 Models

Track 01 of the challenge included four baseline models, selected from the literature. These models are the zero-filled reconstruction, the U-Net model [44], the WW-net model [45], and the hybrid-cascade model [46]. To date, Track 01 has received six independent submissions from ResoNNance [47] (two different models), The Enchanted (two different models), TUMRI, and M-L UNICAMP teams.

The ResoNNance 1.0 model submission was a recurrent inference machine [48], ResoNNance 2.0 was a recurrent variational network [49]. The Enchanted 1.0 model was inspired by [50], where they used magnitude and phase networks, followed by a VS-net architecture [51]. The Enchanted 2.0 used an end-to-end variational network [52], and it was the only submission that used self-supervised learning [53] to initialize their model. The pretext task to initialize their models was the prediction of image rotations [54]. TUMRI used a similar model to the WW-net, but they implemented complex-valued operations [55]. They used a linear combination of VIF and MS-SSIM [56] as their loss function. M-L UNICAMP used a hybrid model with parallel network branches operating in k-space and image domains. Links to the source code for the different models are available in the benchmark repository. Some of the Track 01 models were designed to work with a specific number of coil channels, thus they were not submitted to Track 02 of the challenge.

Track 02 of the challenge included two baseline models (zerofilled reconstruction and the U-Net model). ResoNNance and The Enchanted teams submitted two models each to Track 02. The models submitted by ResoNNance and The Enchanted teams were the same models that were used for Track 01 of the challenge. Table 3.2 summarizes the processing

domains (image, *k*-space or dual/hybrid), the presence of elements, such as coil sensitivity estimation, data consistency, and the loss function used during training of the models. For more details about the models, we refer the reader to the source publications or the code repositories for the unpublished work.

3.3 Results

3.3.1 Track 01

The quantitative results for Track 01 are summarized in Table 3.3. There were in total 10 models (four baseline and six submitted) in Track 01. The zero-filled and U-Net reconstructions had the worst results. The M-L UNICAMP, Hybrid Cascade, WW-net, and TUMRI models were next with similar results in terms of SSIM and pSNR. Notably, the TUMRI submission achieved the second-highest VIF metric. ResoNNance and The Enchanted teams' submissions achieved the highest overall scores on the quantitative metrics. The ResoNNance 2.0 submission had the best SSIM and pSNR metrics and the fourth-best VIF metric. The Enchanted 1.0 submission obtained the best VIF metric. The Enchanted 2.0 submission achieved the second-best SSIM metric, and the third-best VIF and pSNR metrics. Representative reconstructions resulting from the different models for $R = 5$ are shown in Figure 3.1.

Table 3.2: Summary of the submissions including processing domain, presence of coil sensitivity estimation (SE), presence of data consistency (DC), and basis of the training loss functions. $*$ indicates a baseline model. Loss functions: Mean Absolute Error (MAE), Structural Similarity (SSIM), Mean Squared Error (MSE), Multi-Scale SSIM (MS-SSIM), and Visual Information Fidelity (VIF).

Model	Domain	SE	DC	Loss function
ResoNNance 2.0	Hybrid	Yes	Yes	MAE and SSIM
The Enchanted 2.0	Image	Yes	Yes	Cross entropy (pretext) and SSIM (main task)
ResoNNance 1.0	Image	Yes	Yes	MAE and SSIM
The-Enchanted 1.0	Image	Yes	Yes	MSE (first step) and SSIM (second step)
TUMRI	Hybrid	No	Yes	MS-SSIM and VIF
WW-Net*	Hybrid	No	Yes	MSE
Hybrid-cascade*	Hybrid	No	Yes	MSE
M-L UNICAMP	Hybrid	No	Yes	MSE
U-Net*	Image	No	No	MSE
Zero-filled*	N/A	No	N/A	N/A

Twenty five images in the test set were visually assessed by our expert observer for the two best submission (ResoNNance 2.0 and The Enchanted 2.0). Out of the 50 images assessed by the expert observer, only two (4.0%) were deemed to have minor deviations, such as shape, intensity, and contrast between the reconstructed images and the reference (cf., 3.2A). Twenty seven images (54.0%) were deemed to have similar quality to the fully sampled reference, and 21 (42.0%) were rated as having similar quality when compared to the reference, but exhibited filtering of the noise in the image background (cf.,3.2B).

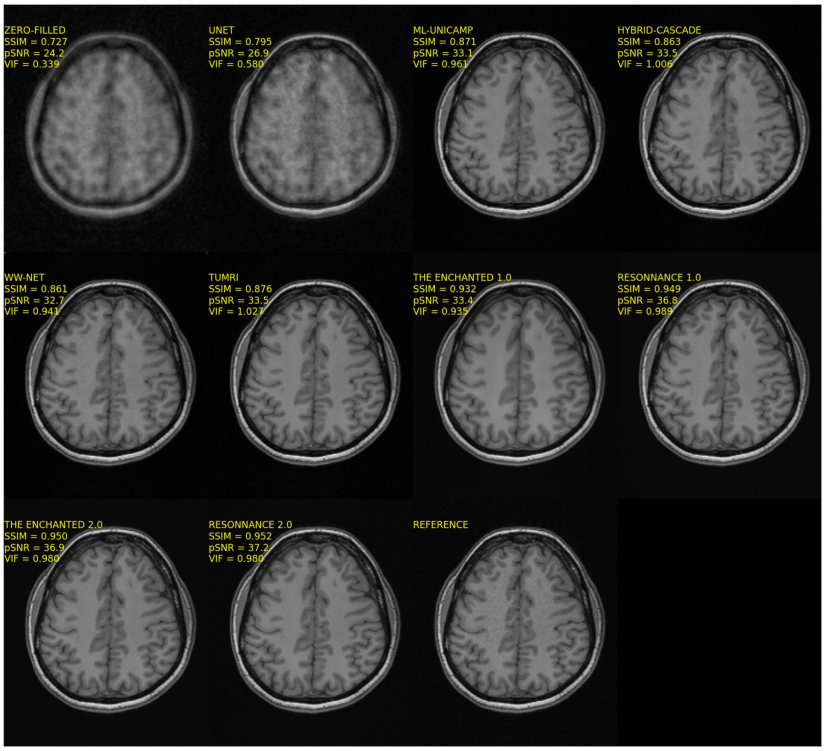


Figure 3.1: Representative reconstructions of the different models submitted to Track 01 (*i.e.*, 12-channel) of the challenge for $R = 5$. Note that the reconstructions from the top four methods, ResoNNance 1.0 and 2.0, and The Enchanted 1.0 and 2.0, try to match the noise pattern seen in the background of the reference image, while ML-UNICAMP, Hybrid-cascade, WW-net, and TUMRI seem to have partially filtered this background noise.

3.3.2 Track 02

Two teams, ResoNNance and The Enchanted, submitted a total of four models to Track 02 of the benchmark. Their results were compared to two baseline techniques. The models submitted to Track 02, except for the U-Net baseline, which has a higher input dimension (i.e., the input dimensions depends on the number of receiver coils), was the same as the models submitted for Track 01, so for the 12-channel test dataset, the results are the same as in Track 01 (see 3.3).

The results for Track 02 using the 32-channel test set are summarized in Table 3.4. For the 32-channel test dataset, The Enchanted 2.0 submission obtained the best VIF and pSNR metrics, and the second-best SSIM score. The ResoNNance 2.0 submission obtained the best SSIM metric, second-best pSNR, and third-best VIF metrics. The ResoNNance 1.0 submission obtained the third-best SSIM and pSNR metrics, and second-best VIF. The Enchanted 1.0 submission obtained the fourth-best SSIM and VIF, and fifth-best pSNR. The zero-filled and UNet reconstructions obtained the worse results. Representative reconstructions resulting from the different models are depicted in 3.3.

Twenty five images in the test set were visually assessed by our expert observer for

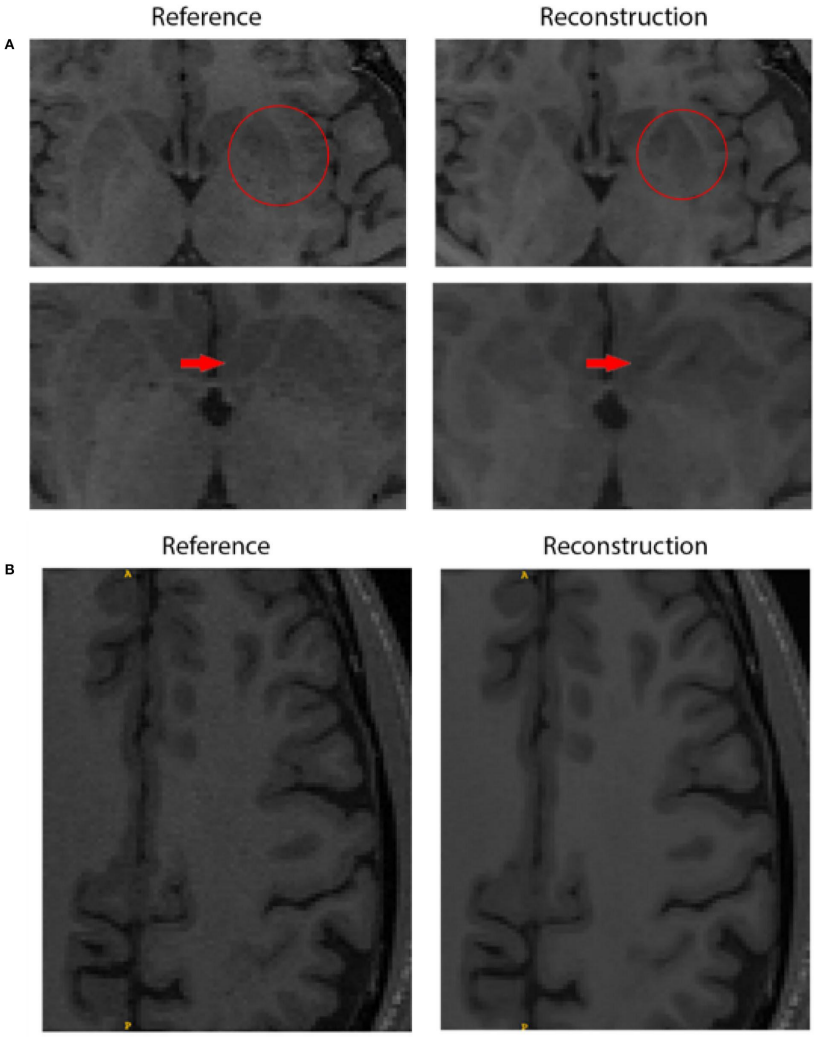


Figure 3.2: Quality assessment comparing the fully sampled reference and the reconstruction obtained by team ResoNNance 2.0. (A) The top row shows the border of the left putamen, where the reconstructed image has a discrepancy in shape compared to the reference image (highlighted with red circles). The bottom row shows that changes in the shape of the structure are also visible in the next slice of the same subject (highlighted with red arrows). It is important to emphasize that these | discrepancies are not restricted to the putamen, but a systematic evaluation of where these changes occur is out of scope for this work. (B) Illustration of a case where the expert observed rated that the deep-learning-based reconstruction improved image quality. In this figure, we can see smoothening of cortical white matter without loss of information as no changes appeared in the pattern of gyrification within cortical gray matter

the two best submissions (ResoNNance 2.0 and The Enchanted 2.0). Out of the 50 images assessed by the expert observer, 14 (28.0%) were deemed to have deviations from common anatomical borders. A total of 34 images (68.0%) were deemed to have a similar quality

Table 3.3: Summary of the Track 01 results for $R = 5$. The best value for each metric and acceleration is emboldened. Mean ± standard deviation are reported. ∗ indicates a baseline model.

Model	SSIM	pSNR (dB)	VIF
ResoNNance 2.0	**0.941 ± 0.029**	**35.7 ± 1.8**	0.957 ± 0.034
The Enchanted 2.0	0.937 ± 0.033	34.9 ± 2.4	0.973 ± 0.036
ResoNNance 1.0	0.936 ± 0.031	35.3 ± 1.8	0.960 ± 0.035
The-Enchanted 1.0	0.912 ± 0.034	30.3 ± 2.8	**0.993 ± 0.176**
TUMRI	0.868 ± 0.044	32.5 ± 1.7	0.989 ± 0.045
WW-Net∗	0.870 ± 0.043	32.5 ± 1.7	0.929 ± 0.049
Hybrid-cascade∗	0.860 ± 0.044	32.7 ± 1.6	0.954 ± 0.042
M-L UNICAMP	0.868 ± 0.044	32.4 ± 1.7	0.918 ± 0.053
U-Net∗	0.779 ± 0.039	26.8 ± 1.7	0.642 ± 0.068
Zero-filled∗	0.726 ± 0.045	25.2 ± 1.5	0.518 ± 0.066

Table 3.4: Summary of the Track 02 results for $R = 5$ using the 32-channel test set. The best value for each metric and acceleration is emboldened. Mean ± standard deviation are reported. ∗ indicates a baseline model.

Model	SSIM	pSNR (dB)	VIF
ResoNNance 2.0	**0.961 ± 0.027**	38.3 ± 2.2	0.955 ± 0.036
The Enchanted 2.0	0.960 ± 0.037	**38.34 ± 3.2**	**1.024 ± 0.034**
ResoNNance 1.0	0.947 ± 0.033	37.7 ± 2.9	0.992 ± 0.030
The Enchanted 1.0	0.907 ± 0.046	30.1 ± 2.7	0.834 ± 0.236
U-Net∗	0.832 ± 0.058	31.5 ± 2.6	0.804 ± 0.045
Zero-filled∗	0.780 ± 0.041	26.4 ± 1.5	0.472 ± 0.064

to the fully sampled reference, and only two images (4.0%) were rated as having similar quality when compared to the reference, but exhibited filtering of the noise in the image background.

3.4 Discussion

The first track of the challenge compared ten different reconstruction models (Table 3.3). As expected, the zero-filled reconstruction, which does not involve any training from the data, universally had the poorest results. The second worst technique was the U-Net model, which used as input the channel-wise zero-filled reconstruction and tried to recover the high-fidelity image. The employed U-Net [44] model did not include any data consistency steps. The remaining eight models all include a data consistency step, which seems to be an essential step for high-fidelity image reconstruction, as has been previously highlighted in [10, 14].

The M-L UNICAMP had the eighth-lowest pSNR and VIF metrics, and the seventh-lowest SSIM score. In contrast, the top ranked methods were either cascaded networks (Hybrid-cascade, WWnet, TUMRI, The Enchanted 1.0 and 2.0) or recurrent methods (ResoNNance 1.0 and 2.0).

The top four models in the benchmark were the ResoNNance 1.0 and 2.0 and The Enchanted 1.0 and 2.0 submissions. These four models estimated coil sensitivities and

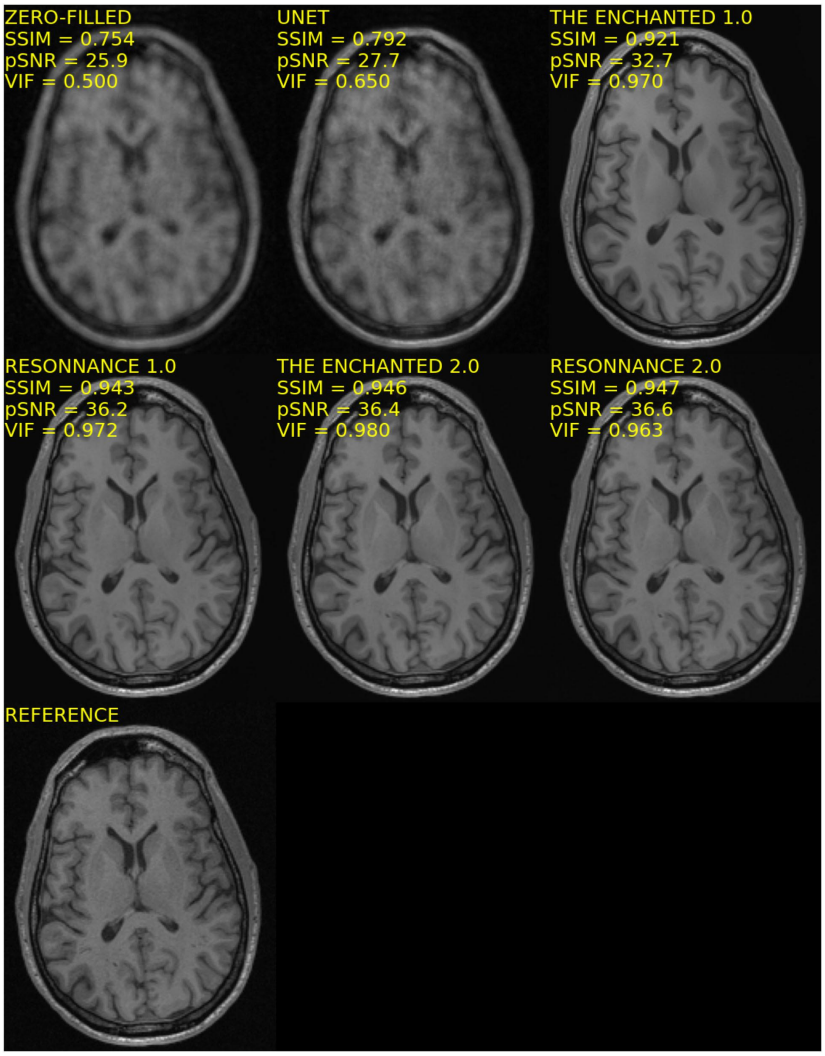


Figure 3.3: Representative reconstructions of the different models submitted to Track 02 of the challenge for $R = 5$ using the 32-channel coil.

combined the coil channels, which made these models flexible and capable of working with datasets acquired with an arbitrary number of receiver coils. The top two models ResoNNance 2.0 and Enchanted 2.0 are hybrid models. They are followed in rank by ResoNNance 1.0 and Enchanted 1.0, which are imagedomain methods. The other better performing models (M-L UNICAMP, Hybrid Cascade, WW-net, and TUMRI) used an approach that receives all coil channels as input, making these models tailored to a specific coil configuration (i.e., number of channels). Though the methods that combined the channels before reconstruction using coil sensitivity estimations similarly to [52], such

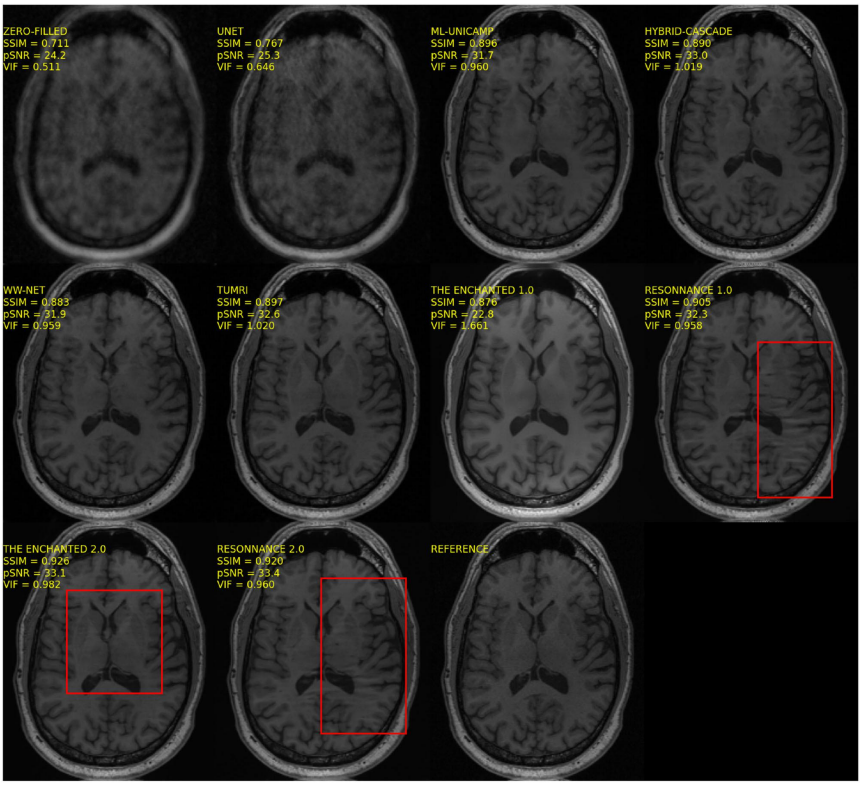


Figure 3.4: Sample reconstruction illustrating artifacts (highlighted in red boxes) that seem to be present on images reconstructed by models that used coil sensitivity estimation as part of their method.

as from ResoNNance and The Enchanted teams, demonstrated the best results so far, it is still unclear if this approach is superior to models that do not combine the channels before reconstruction. A recent work [29] indicated that the separate channel approach may be advantageous compared to models that combine the *k*-space channels before reconstruction.

All of the models submitted to the MC-MRI Reconstruction Challenge had a relatively narrow input convolutional layer (e.g., 64 filters), which may have resulted in the loss of relevant information. In [29], they used 15-channel data and the first layer had 384 filters. Another advantage of models that receive all channels as input is that they seem more robust to artifacts that can occur in the reconstructed images due to problems in coil sensitivity estimation. This finding was observed in our visual assessment mostly in methods that involved coil sensitivity estimation (ResoNNance and The Enchanted—Figure 3.4). Similar artifacts were not observed in images produced on models that do not require coil sensitivity estimation.

In our study, we also noted variability in the ranking across metrics (Table 3.3). For example, The Enchanted 1.0 submission had the best VIF score, but only the fourth-best SSIM and seventh-highest pSNR metrics. This variability reinforces the importance of

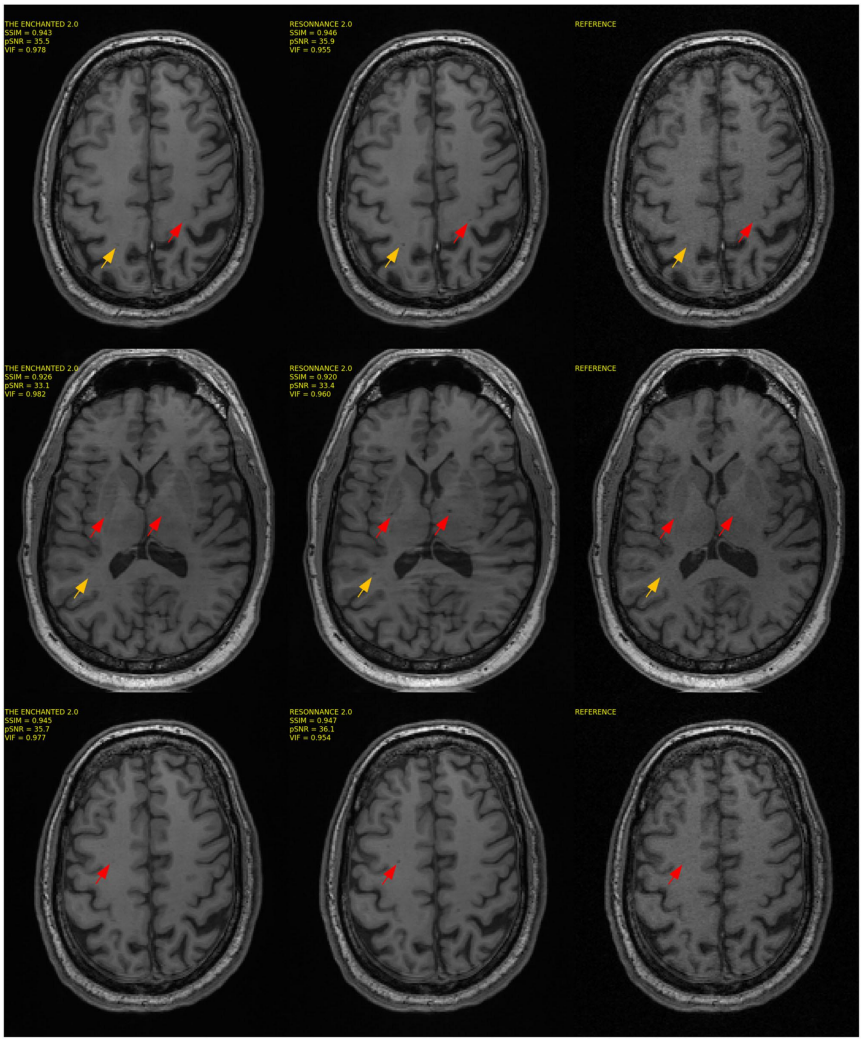


Figure 3.5: Three sample reconstructions, one per row, for the top two models. The Enchanted 2.0 and ResoNNance 2.0 and the reference are illustrated. The arrows in the figure indicate regions of interest that indicate deviations between the deep-learning-based reconstructions and the fully sampled reference.

including many benchmarks that can summarize the result of multiple submissions by using a consistent set of multiple metrics. Studies that use a single image quality metric, for example, are potentially problematic if the chosen measure masks specific classes of performance issues. While imperfect, the use of a composite score based on metric rankings attempts to reduce this inherent variability by examining multiple performance measures.

Visual inspection of the reconstructed MR images (cf., Figures 3.1, 3.3) indicates that with some models and for some samples in the test set, the reconstructed background noise is different from the background noise in the reference images. This observation, particularly

with the ResoNNance and The Enchanted teams' submissions, leads to questions on whether the evaluated quantitative metrics are best suited to determine the reconstruction quality. Given a noisy reference image, a noise-free reconstruction will potentially achieve lower pSNR, SSIM, and VIF than the same reconstruction with added noise. This finding is contrary to human visual perception, where noise impacts the image quality negatively and is, in general, undesired. During the expert visual assessment, 23 of 50 (46.0%) reconstructions were rated higher than the fully sampled reference due to the fact that the brain anatomical borders in these images were preserved, but the image noise was filtered out.

All trainable baseline models and the model submitted by M-L UNICAMP used mean squared error (MSE) as their cost function. The model submitted by TUMRI was trained using a combination of multi-scale SSIM (MS-SSIM) [56] and VIF as their cost function. The model The Enchanted 1.0 has two components in their cost function: (1) their model was trained using MSE as the cost function with the target being the coil-combined complex-valued fully sampled reference and then (2) their Down-Up network [57] received as input the absolute value of the reconstruction obtained in the previous stage, and the reference was the square-root sum-of-squares fully sampled reconstruction. The Down-Up network was trained using SSIM as the loss function. The model The Enchanted 2.0 is the only model that was pre-trained using a self-supervised learning pretext task of predicting rotations. The pretext task was trained using cross-entropy as the loss function. The main task (*i.e.*, reconstruction task) was trained using SSIM as the loss function.

The ResoNNance 1.0 and 2.0 models used a combination of SSIM and mean absolute error (MAE) as the training loss function, which is a combination that has been shown to be effective for image restoration [58]. Because the background in the images is quite substantial and SSIM is a bounded metric that is computed across image patches, this observation causes models trained using SSIM as part of their loss function to try to match the background noise in their reconstructions. This observation may offer a potential explanation for why the models submitted by The Enchanted and ResoNNance teams were able to preserve the noise pattern in their reconstructions. Metrics that are based on visual perception are important and evaluating the possibility of using these types of metrics as part of the loss functions is an interesting research avenue for the field of MRI reconstruction.

For $R = 5$, the top three models: ResoNNance 2.0, The Enchanted 2.0, and ResoNNance 1.0 produced the most visually pleasing reconstructions and also had the top performing metrics. It is important to emphasize that $R = 5$ in the challenge is relative to the 85% of k-space that was sampled in the slice-encode (k_z) direction. If we consider the equivalent full k-space, the acceleration factor would be $R = 5.9$. Based on the Track 01 results, we would say that an acceleration between 5 and 6 might be feasible to be incorporated into a clinical setting for a single-sequence MR image reconstruction model. Further analysis of the image reconstructions by a panel of radiologists is needed to better assess clinical value before achieving a definite conclusion.

The second track of the challenge compared six different reconstruction models (Tables 3.3, 3.4). The models, The Enchanted 2.0 and ResoNNance 2.0, achieved the best overall results. For the 12-channel test set (Figure 3.1), the results were the same as the results they obtained in Track 01 of the challenge since the models were the same. More interesting are the results for the 32-channel test set. Though the metrics for the 32-channel test set

are higher than the 12-channel test set, by visually inspecting the quality of the reconstructed images, it is clear that 32- channel image reconstructions are of poorer quality compared to 12-channel reconstructions (Figure 3.3). In total, 28% of the 32-channel images assessed by the expert observer were deemed to have poorer quality when compared the reference against 4% of the 12-channel images rated. This fact raises concerns about the generalizability of the reconstruction models across different coils. Potential approaches to mitigate this issue is to include representative data collected with different coils in the training and validation sets or employ domain adaptation techniques [59], such as data augmentation strategies, that simulate data acquired under different coil configurations, to make the models more generalizable.

Though the generalization of learned MR image reconstruction models and their potential for transfer learning has been previously assessed [60], the results from Track 02 of our challenge indicate that there is still room for improvements. Interestingly, the model The Enchanted 2.0 is the only model that employed self-supervised learning, which seems to have had a positive impact on the model generalizability for the 32-channel test data.

One important finding that we noticed during the visual assessment of the images is that some of the reconstructed images enhanced hypointensity regions within the brain white matter, while in others images, these hypointensities were blurred out of the image (*cf.*, Figure 3.5). In many cases, it was unclear from the fully sampled reference whether this hypointensity region corresponded to noise in the image or if it indicated the presence of relevant structures, such as lesions that appear as dark spots in T_1-weighted images. This finding is critical especially when targeting diseases that often present small lesions. Further investigation is necessary to determine its potential impact before the clinical adoption of these reconstruction models.

3.5 Summary

The MC-MRI reconstruction challenge provided an objective benchmark for assessing brain MRI reconstruction and the generalizability of models across datasets collected with different coils using a high-resolution, 3D dataset of T_1-weighted MR images. Track 01 compared ten reconstruction models and Track 02 compared six reconstruction models. The results indicated that although the quantitative metrics are higher for the test data not seen during training (*i.e.*, 32-channel data), visual inspection indicated that these reconstructed images had poorer quality. This conclusion that current models do not generalize well across datasets collected using different coils indicates a promising research field in the coming years that is very relevant for the potential clinical adoption of deep-learning-based MR image reconstruction models. The results also indicated the difficulty of reconstructing finer details in the images, such as lacunes. The MC-MRI reconstruction challenge continues and the organizers of the benchmark will periodically incorporate more data, which will potentially allow to train deeper models. As a long-term benefit of this challenge, we expect that the adoption of these deep-learning-based MRI reconstruction models in the clinical and research environments will be streamlined.

Data Availability Statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: www.ccdataset.com.

Ethics Statement

The studies involving human participants were reviewed and approved by Conjoint Health Research Ethics Board (CHREB), which reviews applications from researchers affiliated with the Faculties of Kinesiology, Medicine and Nursing at the University of Calgary. Approval number is REB 15-1285. The patients/participants provided their written informed consent to participate in this study.

Author Contributions

YB, WL, RF, and RS were responsible for data preparation, design of the challenge, implementation of the baseline models, and writing the bulk of the manuscript. JT, DK, NM, MC, and GY were responsible for submitting the ResoNNance reconstruction models. LRo, AL, HP, and LRi were responsible for the M-L UNICAMP submission. MD, VS, FG, DV, and SF-R submitted the TUMRI solution and helped with the VIF analysis. AK, JC, and MS submitted The Enchanted models. ML was responsible for creating the data repository and automating the extraction of metrics and challenge leader board creation. NN was the medical expert responsible for the visual assessment of the images. All authors reviewed the manuscript, provided relevant feedback across multiple rounds of reviews, and contributed to the article and approved the submitted version.

Funding

RF thanks the Canadian Institutes for Health Research (CIHR, FDN-143298) for supporting the Calgary Normative Study and acquiring the raw datasets. RF and RS thank the Natural Sciences and Engineering Research Council (NSERC - RGPIN/02858-2021 and RGPIN-2021-02867) for providing ongoing operating support for this project. We also acknowledge the infrastructure funding provided by the Canada Foundation of Innovation (CFI). The organizers of the challenge also acknowledge Nvidia for providing a Titan V Graphics Processing Unit and Amazon Web Services for providing computational infrastructure that was used by some of the teams to develop their models. DK and MC were supported by the STAIRS project under the Top Consortium for Knowledge and Innovation-Public, Private Partnership (TKI-PPP) program, co-funded by the PPP Allowance made available by Health Holland, Top Sector Life Sciences & Health. HP thanks the National Council for Scientific and Technological Development (CNPq #309330/2018-1) for the research support grant. LRi also thank the National Council for Scientific and Technological Development (CNPq #313598/2020-7) and São Paulo Research Foundation (FAPESP #2019/21964-4) for the support.

Acknowledgements

The authors would like to acknowledge that a preprint of this manuscript is available on arXiv (https://arxiv.org/abs/2011.07952).

Conflict of Interest

MC is a shareholder of Nico.lab International Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

3

Publisher's Note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

[1] Michael Lustig, David L. Donoho, Juan M. Santos, and John M. Pauly. Compressed Sensing MRI. *IEEE Signal Processing Magazine*, 25(2):72–82, 2008.

[2] Klaas Pruessmann, Markus Weiger, Markus Scheidegger, and Peter Boesiger. SENSE: sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–962, 1999.

[3] Mark A. Griswold, Peter M. Jakob, Robin M. Heidemann, Mathias Nittka, Vladimir Jellus, Jianmin Wang, Berthold Kiefer, and Axel Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, 47(6):1202–1210, 2002.

[4] Anagha Deshmane, Vikas Gulani, Mark A Griswold, and Nicole Seiberlich. Parallel MR imaging. *Journal of Magnetic Resonance Imaging*, 36(1):55–72, 2012.

[5] Michael Lustig, David Donoho, and John M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[6] Dong Liang, Bo Liu, Jiunjie Wang, and Leslie Ying. Accelerating SENSE using compressed sensing. *Magnetic Resonance in Medicine*, 62(6):1574–1584, 2009.

[7] Jian Sun, Huibin Li, and Zongben Xu. Deep ADMM-Net for compressive sensing MRI. In *Advances in Neural Information Processing Systems*, pages 10–18, 2016.

[8] Shanshan Wang, Zhenghang Su, Leslie Ying, Xi Peng, Shun Zhu, Feng Liang, Dagan Feng, and Dong Liang. Accelerating magnetic resonance imaging via deep learning. In *IEEE International Symposium on Biomedical Imaging*, pages 514–517, 2016.

[9] Kinam Kwon, Dongchan Kim, and HyunWook Park. A parallel MR imaging method using multilayer perceptron. *Medical Physics*, 44(12):6209–6224, 2017.

[10] Jo Schlemper, Jose Caballero, Joseph V Hajnal, Anthony N Price, and Daniel Rueckert. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Transactions on Medical Imaging*, 37(2):491–503, 2017.

[11] Jo Schlemper, Guang Yang, Pedro Ferreira, Andrew Scott, Laura-Ann McGill, Zohya Khalique, Margarita Gorodezky, Malte Roehl, Jennifer Keegan, and Dudley Pennell. Stochastic deep compressive sensing for the reconstruction of diffusion tensor cardiac MRI. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, pages 295–303, 2018.

[12] Jo Schlemper, Ilkay Oksuz, James Clough, Jinming Duan, Andrew King, Julia Schanbel, Joseph Hajnal, and Daniel Rueckert. dAUTOMAP: Decomposing AUTOMAP to achieve scalability and enhance performance. In *International Society for Magnetic Resonance in Medicine*, 2019.

[13] Muneer Ahmad Dedmari, Sailesh Conjeti, Santiago Estrada, Phillip Ehses, Tony Stöcker, and Martin Reuter. Complex fully convolutional neural networks for MR image reconstruction. In *International Workshop on Machine Learning for Medical Image Reconstruction*, pages 30–38, 2018.

[14] Taejoon Eo, Yohan Jun, Taeseong Kim, Jinseong Jang, Ho-Joon Lee, and Dosik Hwang. KIKI-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. *Magnetic Resonance in Medicine*, 80(5):2188–2201, 2018.

[15] Baran Gözcü, Rabeeh Karimi Mahabadi, Yen-Huan Li, Efe Ilıcak, Tolga Cukur, Jonathan Scarlett, and Volkan Cevher. Learning-based compressive MRI. *IEEE Transactions on Medical Imaging*, 37(6):1394–1406, 2018.

[16] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P. Recht, Daniel K. Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, 2018.

[17] Tran Minh Quan, Thanh Nguyen-Duc, and Won-Ki Jeong. Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. *IEEE Transactions on Medical Imaging*, 37(6):1488–1497, 2018.

[18] Maximilian Seitzer, Guang Yang, Jo Schlemper, Ozan Oktay, Tobias Würfl, Vincent Christlein, Tom Wong, Raad Mohiaddin, David Firmin, Jennifer Keegan, et al. Adversarial and perceptual refinement for compressed sensing MRI reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 232–240, 2018.

[19] Guang Yang, Simiao Yu, Hao Dong, Greg Slabaugh, Pier Luigi Dragotti, Xujiong Ye, Fangde Liu, Simon Arridge, Jennifer Keegan, and Yike Guo. DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1310–1321, 2018.

[20] Pengyue Zhang, Fusheng Wang, Wei Xu, and Yu Li. Multi-channel generative adversarial network for parallel magnetic resonance image reconstruction in k-space. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 180–188, 2018.

[21] Bo Zhu, Jeremiah Z Liu, Stephen F Cauley, Bruce R Rosen, and Matthew S Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487–492, 2018.

[22] Mehmet Akçakaya, Steen Moeller, Sebastian Weingärtner, and Kâmil Uğurbil. Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging. *Magnetic Resonance in Medicine*, 81(1):439–453, 2019.

[23] Morteza Mardani, Enhao Gong, Joseph Y Cheng, Shreyas S Vasanawala, Greg Zaharchuk, Lei Xing, and John M Pauly. Deep generative adversarial neural networks for compressive sensing MRI. *IEEE Transactions on Medical Imaging*, 38(1):167–179, 2019.

[24] Kamlesh Pawar, Zhaolin Chen, N Jon Shah, and Gary F Egan. A deep learning framework for transforming image reconstruction into pixel classification. *IEEE Access*, 7:177690–177702, 2019.

[25] Chen Qin, Jo Schlemper, Jose Caballero, Anthony N Price, Joseph V Hajnal, and Daniel Rueckert. Convolutional recurrent neural networks for dynamic MR image reconstruction. *IEEE Transactions on Medical Imaging*, 38(1):280–290, 2019.

[26] Roberto Souza and Richard Frayne. A hybrid frequency-domain/image-domain deep network for magnetic resonance image reconstruction. In *IEEE Conference on Graphics, Patterns and Images*, pages 257–264, 2019.

[27] K. Zeng, Y. Yang, G. Xiao, and Z. Chen. A very deep densely connected network for compressed sensing MRI. *IEEE Access*, 7:85430–85439, 2019.

[28] Seyed Amir Hossein Hosseini, Burhaneddin Yaman, Steen Moeller, Mingyi Hong, and Mehmet Akçakaya. Dense recurrent neural networks for accelerated mri: History-cognizant unrolling of optimization algorithms. *IEEE Journal of Selected Topics in Signal Processing*, 14(6):1280–1291, 2020.

[29] Anuroop Sriram, Jure Zbontar, Tullie Murrell, C Lawrence Zitnick, Aaron Defazio, and Daniel K Sodickson. GrappaNet: Combining parallel imaging with deep learning for multi-coil MRI reconstruction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14315–14322, 2020.

[30] Bo Zhou and S Kevin Zhou. DuDoRNet: Learning a dual-domain recurrent network for fast MRI reconstruction with deep T1 prior. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4273–4282, 2020.

[31] Florian Knoll, Tullie Murrell, Anuroop Sriram, Nafissa Yakubova, Jure Zbontar, Michael Rabbat, Aaron Defazio, Matthew J Muckley, Daniel K Sodickson, and C Lawrence Zitnick. Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge. *Magnetic Resonance in Medicine*, 2020.

[32] Jure Zbontar, Florian Knoll, Anuroop Sriram, Matthew J Muckley, Mary Bruno, Aaron Defazio, Marc Parente, Krzysztof J Geras, Joe Katsnelson, and Hersh Chandarana. fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv preprint arXiv:1811.08839*, 2018.

[33] Roberto Souza, Youssef Beauferris, Wallace Loos, Robert Marc Lebel, and Richard Frayne. Enhanced deep-learning-based magnetic resonance image reconstruction by leveraging prior subject-specific brain imaging: Proof-of-concept using a cohort of presumed normal subjects. *IEEE Journal of Selected Topics in Signal Processing*, 14(6):1126–1136, 2020.

[34] Matthew J. Muckley, Bruno Riemenschneider, Alireza Radmanesh, Sunwoo Kim, Geunu Jeong, Jingyu Ko, Yohan Jun, Hyungseob Shin, Dosik Hwang, Mahmoud Mostapha, Simon Arberet, Dominik Nickel, Zaccharie Ramzi, Philippe Ciuciu, Jean-Luc Starck, Jonas Teuwen, Dimitrios Karkalousos, Chaoping Zhang, Anuroop Sriram, Zhengnan Huang, Nafissa Yakubova, Yvonne W. Lui, and Florian Knoll. Results of the 2020 fastMRI Challenge for Machine Learning MR Image Reconstruction. *IEEE Transactions on Medical Imaging*, 40(9):2306–2317, 2021.

[35] Roberto Souza, Oeslle Lucena, Julia Garrafa, David Gobbi, Marina Saluzzi, Simone Appenzeller, Letícia Rittner, Richard Frayne, and Roberto Lotufo. An open, multi-vendor, multi-field-strength brain MR dataset and analysis of publicly available skull stripping methods agreement. *NeuroImage*, 170:482–494, 2018.

[36] Cheryl R McCreary, Marina Salluzzi, Linda B Andersen, David Gobbi, Louis Lauzon, Feryal Saad, Eric E Smith, and Richard Frayne. Calgary Normative Study: design of a prospective longitudinal study to characterise potential quantitative MR biomarkers of neurodegeneration over the adult lifespan. *BMJ open*, 10(8):e038120, 2020.

[37] Erik G Larsson, Deniz Erdogmus, Rui Yan, Jose C Principe, and Jeffrey R Fitzsimmons. SNR-optimality of sum-of-squares reconstruction for phased-array magnetic resonance imaging. *Journal of Magnetic Resonance*, 163(1):121–123, 2003.

[38] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004.

[39] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, Feb 2006.

[40] Eero P Simoncelli and Bruno A Olshausen. Natural Image Statistics and Neural Representation. *Annual Review of Neuroscience*, 24(1):1193–1216, 2001.

[41] Wilson S. Geisler. Visual Perception and the Statistical Properties of Natural Scenes. *Annual Review of Psychology*, 59(1):167–192, 2008.

[42] Allister Mason, James Rioux, Sharon E Clarke, Andreu Costa, Matthias Schmidt, Valerie Keough, Thien Huynh, and Steven Beyea. Comparison of objective image quality metrics to expert radiologists' scoring of diagnostic quality of MR images. *IEEE Transactions on Medical Imaging*, 39(4):1064–1072, 2019.

[43] Kristjan Kalm and Dennis Norris. Visual recency bias is explained by a mixture model of internal representations. *Journal of Vision*, 18(7):1–1, 2018.

[44] Kyong Jin, Michael McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017.

[45] Roberto Souza, Mariana Bento, Nikita Nogovitsyn, Kevin J Chung, Wallace Loos, R Marc Lebel, and Richard Frayne. Dual-domain cascade of U-Nets for multi-channel magnetic resonance image reconstruction. *Magnetic Resonance Imaging*, 71:140–153, 2020.

[46] Roberto Souza, R. Marc Lebel, and Richard Frayne. A hybrid, dual domain, cascade of convolutional neural networks for magnetic resonance image reconstruction. In *International Conference on Medical Imaging with Deep Learning*, volume 102, pages 437–446, 2019.

[47] George Yiasemis, Nikita Moriakov, Dimitrios Karkalousos, Matthan Caan, and Jonas Teuwen. DIRECT: Deep Image REConstruction Toolkit. https://github.com/directgroup/direct, 2021.

[48] Kai Lønning, Patrick Putzky, Jan-Jakob Sonke, Liesbeth Reneman, Matthan WA Caan, and Max Welling. Recurrent inference machines for reconstructing heterogeneous MRI data. *Medical Image Analysis*, 53:64–78, 2019.

[49] George Yiasemis, Clara I Sánchez, Jan-Jakob Sonke, and Jonas Teuwen. Recurrent variational network: A deep learning inverse problem Solver applied to the task of accelerated MRI reconstruction. *arXiv preprint arXiv:2111.09639*, 2021.

[50] Dongwook Lee, Jaejun Yoo, Sungho Tak, and Jong Chul Ye. Deep residual learning for accelerated MRI using magnitude and phase networks. *IEEE Transactions on Biomedical Engineering*, 65(9):1985–1995, 2018.

[51] Jinming Duan, Jo Schlemper, Chen Qin, Cheng Ouyang, Wenjia Bai, Carlo Biffi, Ghalib Bello, Ben Statton, Declan P O'Regan, and Daniel Rueckert. VS-Net: Variable splitting network for accelerated parallel MRI reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 713–722, 2019.

[52] Anuroop Sriram, Jure Zbontar, Tullie Murrell, Aaron Defazio, C Lawrence Zitnick, Nafissa Yakubova, Florian Knoll, and Patricia Johnson. End-to-end variational networks for accelerated MRI reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 64–73, 2020.

[53] Liang Chen, Paul Bentley, Kensaku Mori, Kazunari Misawa, Michitaka Fujiwara, and Daniel Rueckert. Self-supervised learning for medical image analysis using image context restoration. *Medical Image Analysis*, 58:101539, 2019.

[54] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised Representation Learning by Predicting Image Rotations. In *International Conference on Learning Representations*, 2018.

[55] Chiheb Trabelsi, Olexa Bilaniuk, Ying Zhang, Dmitriy Serdyuk, Sandeep Subramanian, Joao Felipe Santos, Soroush Mehri, Negar Rostamzadeh, Yoshua Bengio, and Christopher J Pal. Deep Complex Networks. In *International Conference on Learning Representations*, 2018.

[56] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, volume 2, pages 1398–1402 Vol.2, 2003.

[57] Songhyun Yu, Bumjun Park, and Jechang Jeong. Deep iterative down-up CNN for image denoising. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

[58] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging*, 3(1):47–57, 2016.

[59] Wouter M Kouw and Marco Loog. A review of domain adaptation without target labels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(3):766–785, 2019.

[60] Florian Knoll, Kerstin Hammernik, Erich Kobler, Thomas Pock, Michael P Recht, and Daniel K Sodickson. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magnetic Resonance in Medicine*, 81(1):116–128, 2019.

4

Deep learning for efficient reconstruction of highly accelerated 3D FLAIR MRI in neurological deficits

Liebrand, L.C.*, **Karkalousos, D.***, Poirion, É., Emmer, B.J., Roosendaal, S.D., Marquering, H.A., Majoie, C.B.L.M., Savatovsky, J., & Caan, M.W.A. (2024).
Magnetic Resonance Materials in Physics, Biology and Medicine.
DOI: 10.1007/s10334-024-01200-8
*Contributed equally.

Abstract

Object: *To compare compressed sensing (CS) and the Cascades of Independently Recurrent Inference Machines (CIRIM) with respect to image quality and reconstruction times when 12-fold accelerated scans of patients with neurological deficits are reconstructed.* ***Materials and Methods:*** *Twelve-fold accelerated 3D T_2-FLAIR images were obtained from a cohort of 62 patients with neurological deficits on 3 T MRI. Images were reconstructed offline via CS and the CIRIM. Image quality was assessed in a blinded and randomized manner by two experienced interventional neuroradiologists and one experienced pediatric neuroradiologist on imaging artifacts, perceived spatial resolution (sharpness), anatomic conspicuity, diagnostic confidence, and contrast. The methods were also compared in terms of self-referenced quality metrics, image resolution, patient groups and reconstruction time. In ten scans, the contrast ratio (CR) was determined between lesions and white matter. The effect of acceleration factor was assessed in a publicly available fully sampled dataset, since ground truth data are not available in prospectively accelerated clinical scans. Specifically, 451 FLAIR scans, including scans with white matter lesions, were adopted from the FastMRI database to evaluate structural similarity (SSIM) and the CR of lesions and white matter on ranging acceleration factors from four-fold up to 12-fold.* ***Results:*** *Interventional neuroradiologists significantly preferred the CIRIM for imaging artifacts, anatomic conspicuity, and contrast. One rater significantly preferred the CIRIM in terms of sharpness and diagnostic confidence. The pediatric neuroradiologist preferred CS for imaging artifacts and sharpness. Compared to CS, the CIRIM reconstructions significantly improved in terms of imaging artifacts and anatomic conspicuity ($p<0.01$) for higher resolution scans while yielding a 28% higher SNR ($p=0.001$) and a 5.8% lower CR ($p=0.04$). There were no differences between patient groups. Additionally, CIRIM was five times faster than CS was. An increasing acceleration factor did not lead to changes in CR ($p=0.92$), but led to lower SSIM ($p=0.002$).* ***Discussion:*** *Patients with neurological deficits can undergo MRI at a range of moderate to high acceleration. DL reconstruction outperforms CS in terms of image resolution, efficient denoising with a modest reduction in contrast and reduced reconstruction times.*

4.1 Introduction

Time is of utmost importance in diagnosing neurological deficits like Multiple Sclerosis (MS) [1, 2], and tumors [3], and initiating prompt treatment in patients with confirmed acute ischemic stroke (AIS). MS and tumor patients are commonly recalled to the clinic for repeated magnetic resonance imaging (MRI) to monitor disease progression. In the case of AIS, MRI plays a vital role in distinguishing between ischemic or hemorrhagic strokes or mimics [4, 5], visualizing any occlusions, and estimating the onset time [6, 7] and infarct core size [8]. The longer acquisition time required for an MRI examination than for computed tomography (CT) raises concerns about delaying immediate management and treatment decisions during emergencies in the case of a stroke. Given these considerations, exploring ways to accelerate MR sequences currently utilized in diagnosing neurological deficits is necessary.

In recent decades, there have been important advancements in accelerating clinical MRI. Initially, iterative reconstruction techniques such as SENSE [9] and GRAPPA [10] were proposed for parallel imaging. Compared with parallel imaging, compressed sensing (CS)

[11] pinvolves accelerating MRI through iterative reconstruction of irregularly undersampled data to a much greater degree. When imaging neurological deficits, most sequences can benefit from CS. One study used a combination of CS and SENSE to accelerate 3D T_1-echo-spoiled gradient echo and T_2-FLAIR sequences up to five times [3]. Others have also used CS and SENSE to accelerate the time-of-flight MR angiography (TOF-MRA) sequence by approximately ten times [12, 13].

Iterative image reconstruction techniques have two main limitations that hinder their use in a fast-paced clinical environment. First, iterative reconstruction may yield a prolonged reconstruction time. Especially for high-resolution images obtained after zero-filling, the reconstruction times can exceed the measurement times, which limit the clinical workflow. Second, the quality of the CS-reconstructed image may deteriorate when the acceleration factor increases [1]. Thus, novel reconstruction techniques applied to neurological deficits should ideally decrease scanning and reconstruction times while preserving image quality.

Deep learning (DL) can accelerate the imaging time by using graphical processing units (GPUs) for reconstruction while allowing for efficient denoising of the data. For example, one study used DL to reduce the scanning time by approximately 60% when reconstructing data [14]. Additionally, DL has exhibited promising results in reconstructing MRI data with pathologies [2, 15, 16], making it a valuable tool for clinical applications.

Physics-informed DL methods learn how to solve the inverse problem of accelerated MRI reconstruction from the data [17–19]. The objective is to map the undersampled k-space measurements to a denoised image. This approach benefits from generalizing well to modalities not seen during the network training. Recently, the Cascades of Independently Recurrent Inference Machines (CIRIM) were proposed, which balances efficiency and network complexity, and is fast with excellent denoising and generalization capabilities [20]. This could make this network a good candidate for use in neurological deficits. The CIRIM was shown to outperform CS reconstruction in terms of commonly computed metrics, i.e., structural similarity and the peak signal-to-noise ratio. In fast and potentially time-critical imaging settings, CS may thus render inferior image quality, increasing the need for improved image reconstruction under these conditions. Furthermore, an extensive clinical evaluation of this method is still lacking.

This work aims to achieve highly accelerated imaging and fast reconstruction in diagnosing patients with neurological deficits. We evaluate the image reconstruction performance of the previously proposed CIRIM in a representative clinical dataset. This dataset consists of highly accelerated (12X) 3D T_2-FLAIR images obtained as part of routine clinical practice and includes data from fifty-seven patients with neurological deficits, stroke, tumors, and multiple sclerosis (MS). Compressed sensing (CS) is the reference reconstruction method used in the clinic to which we compare our CIRIM method. Challenges lie in the dataset's inhomogeneity and preservation of pathologies unseen during training. Since the dataset was acquired as part of the clinical routine, no fully sampled scans are available; these scans take too long to acquire clinically. Therefore, the CIRIM is trained on a different dataset and compared to CS in terms of reconstruction times and image quality, which are scored both subjectively and objectively. Subdisciplines in radiology may have different requirements in terms of image quality and scanning time. We compare the perceived image quality rated by radiologists with diverse specializations, i.e., intervention neuroradiology and

pediatric neuroradiology.

4.2 Material and methods

4.2.1 Patients and ethics

The data for this retrospective study were routinely collected in our hospital (anonymized for review). All patients included in this study (n=62, 34 females) came to the hospital as part of the clinical routine, including patients with stroke (n=8), other vascular pathologies (n=6), multiple sclerosis (relapsing-remitting MS, n=10; progressive MS, n=3; undefined MS, n=4), tumors (n=8), and Meniere's disease (n=3). The mean age was 53±14 (range: 9 to 88) years. The sample size was chosen such that subjects over a broad age range with a spectrum of diseases were included. All the data were anonymized prior to analysis. Informed consent was not required according to the IRB.

4.2.2 Data acquisition

Patient data were consecutively acquired on a 3T Philips Ingenia Elition scanner equipped with a 32-channel head coil between 08/2021 and 02/2023. The scan parameters of the T_2-FLAIR sequence varied and were in the following ranges: field of view (FOV) from 249×249×180 to 251×251×180 mm, scanning matrix from 216×174×120 to 240×251×180, zero-filled reconstruction matrix from 336×336×240 to 528×528×360, acquisition resolution from 1.05×1.00×1.00 to 1.15×1.43×1.50 mm^3, and reconstruction resolution from 0.48×0.48×0.50 to 0.74×0.74×0.75 mm^3. The other parameters were TR=8000ms, TE=311ms, TI=2400ms, turbo factor=186, and scan time=1m52s to 3m00s. As per standard of care, the data were prospectively undersampled with a variable density mask with a radial shutter to a factor of 12. Sensitivity-reference scan data were obtained for coil sensitivity estimation. Raw data were retained in archive per clinical routine and exported in addition to on-scanner reconstructions in DICOM format.

4.2.3 Data (pre)processing

The raw data were preprocessed in a custom pipeline in MATLAB (version R2019b, MathWorks). Preprocessing for parallel-imaging CS (PICS) and CIRIM reconstruction was identical. The FLAIR k-space data were loaded, phase and offset-corrected, and sorted with MRecon (version 4.4.4, GyroTools). Oversampling was removed in the readout direction, and the matrix was zero-filled to match the original output resolution, leading to an eight-fold increase in matrix size. The sensitivity-reference scan was upsampled and brought into alignment with the FLAIR scan. Sensitivity maps were calculated with caldir (range 50) implemented in the BART toolbox [21]. Five subjects were discarded due to excessive motion artifacts, in which there was no exclusion bias toward a particular diagnostic label, resulting in a dataset of fifty-seven (n=57) subjects.

4.2.4 Parallel-Imaging Compressed Sensing (PICS)

Offline CS reconstructions were performed via the BART toolbox. We used the PICS algorithm with a $\ell 1$-wavelet sparsity transform. The regularization factor was heuristically set to 0.5 to balance artifacts and noise, for a maximum of 60 iterations.

4.2.5 Cascades of Independently Recurrent Inference Machines (CIRIM)

For DL reconstruction, we trained a CIRIM on fully sampled 3D T_1-weighted data of healthy volunteers, retrospectively undersampled twelve times from a 2D variable density Poisson distribution. Previous work has shown that a network trained on T_1-weighted data can generalize well to unseen FLAIR images [20]. Training data were acquired on a 3.0T Philips Ingenia scanner (Philips Healthcare, Best, The Netherlands), and comprised magnetization-prepared rapid gradient echo (MPRAGE) scans, no acceleration, an isotropic resolution of 1.0 mm^3 and a FOV of 256×240 mm^2. The training set consisted of ten subjects (approximately 2000 slices), and the validation set consisted of one subject (approximately 200 slices). No cross-validation was performed during training. Rather, this study serves as an independent study with an external validation test dataset. An overview of the network architecture is shown in Figure 4.1. The hyperparameters of the network were selected as follows. The number of channels was set to 128 for the recurrent and convolutional layers, the number of time steps was set to 8, and the number of cascades was set to 4. Additionally, we adopted and implemented a new stable backend using PyTorch Lighting 1.6.0 with floating-point 16 precision for fast reconstruction times. Model parameters were initialized randomly. The code is available online at https://github.com/wdika/atommic. +

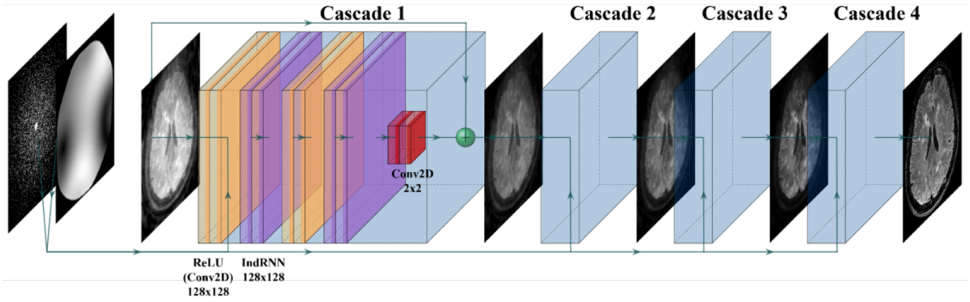


Figure 4.1: Schematic showing the architecture of the Cascades of Independently Recurrent Inference Machines (CIRIM) with four cascades. From left to right: raw k-space data and accompanying sensitivity maps are used to create an initial estimate entered into an IRIM block for calculating the gradient to update the image. An IRIM block consists of subsequent convolutional layers activated by a rectified linear unit (ReLU), recurrent layers (IndRNN), and a final convolutional layer. Four identical IRIM blocks are connected into cascades that share features but no parameters.

4.2.6 Reconstruction time

The reconstruction time was measured as the total time taken for reconstructing a 32-channel volume of size 432×432×278. Notably, when performing a reconstruction with the BART toolbox, there is a small overhead per slice of writing a temporary file and deleting it. For a fair comparison, we accumulated this overhead time, approximately three seconds, and subtracted it from the final reconstruction times. The measurements were repeated three times to ensure precision.

Reconstructions were performed offline on an Nvidia Tesla V100 GPU card with 32GB of memory.

4.2.7 Expert ratings

All the reconstructions were stored in DICOM format for subjective rating of image quality. Two experienced interventional neuroradiologists (J.S., B.E.) with 23 and 17 years of experience and one experienced pediatric neuroradiologist (S.R.) with 14 years of experience were asked to subjectively rate the CIRIM and CS images on multiple categories. The raters were blinded to the diagnosis of the cases whose images were processed for the study. The raters were only asked to review the FLAIR sequence. They scored images side-by-side on multiple categories on a 1 to 5 image quality scale. The scores were as follows: 1 for non-diagnostic quality, 2 for poor quality, 3 for acceptable quality, 4 for good quality, and 5 for excellent quality. The order of the reconstruction methods was randomized, and the raters remained unaware of the method used and the patients' clinical information. Inspired by previous work, five scoring categories were adopted. Imaging artifacts related to aliasing resulting from image acceleration, ranging from excessive artifacts that severely degrade images to no artifacts present. Perceived spatial resolution referred to image sharpness and the ability to discern small structures down to the voxel level sharpness, ranging from unacceptable, extreme blur levels to a high level of detail at the native level of the defined spatial resolution. Anatomic conspicuity ranged from being unable to discern (small) anatomical and pathological structures to perfect identification of structures. Diagnostic confidence summarized the certainty in the diagnosis of pathology, e.g., a lesion, on a scan, ranging from being unable and highly uncertain in diagnosis to perfect ability to diagnose a scan. Image contrast referred to the relative difference in the intensity of known tissue types and pathology ranging from no contrast visible to extremely good contrast [3, 22]. After the individual rating of the data, a review meeting was held with the readers, in which selected subjects with discrepancies in reading scores were re-evaluated.

4.2.8 Quantitative analyses

Since a fully sampled scan of the patient data is lacking, we calculated self-referenced quantitative measures of image quality using MRI Quality Control (MRIQC) [23]. Specifically, we selected the following set of metrics that we deemed relevant for the task of image reconstruction: coefficient of joint variation (CJV) [24], signal-to-noise ratio, a quality index (QI1) of the proportion of voxels corrupted by artifacts [25], the entropy focus criterion (EFC), being the Shannon entropy of voxel intensities as an indication of ghosting and blurring induced by head motion [26], the foreground to background energy ratio (FBER), being the mean energy of image values within the head relative to outside the head [27], and the full width at half maximum (FWHM) of the spatial distribution of the image intensity values in units of voxels [28]. To assess the dependency of FWHM on SNR, Gaussian noise was added post hoc to one randomly selected CIRIM reconstruction and FWHM was recalculated.

4.2.9 Statistics

Statistical analyses were performed using SciPy [29]. The statistical significance threshold was set at $p < 0.01$ for all tests. Bonferroni correction for multiple comparisons was used when necessary. A one-sample Wilcoxon signed-rank test was used to determine whether the expert scores significantly preferred one over the other reconstruction method. A paired t-test was used to determine whether the SNR differed significantly between methods.

Probabilistic ordinal linear regression was performed to evaluate the interaction effect of higher image resolution on improved rating scores, depending on the reconstruction method used. The voxel volume was used as image resolution metric. For each reconstruction method and per patient group, post hoc one-sample Wilcoxon signed-rank tests were used to determine significance at a Bonferroni-corrected threshold over multiple scoring categories.

4.3 Results

In total, the data of 57 subjects were complete and could be successfully reconstructed via the PICS and the CIRIM network. Figure 4.2 shows selected example images in which efficient denoising of the CIRIM compared with PICS can be seen. The CIRIM was also able to accurately reconstruct structures where PICS failed, such as the left internal capsule in the first example (Figure 4.2-top row). Figure 4.3 depicts example images where the raters did not clearly prefer one method. Ratings varied in terms of imaging artifacts and sharpness. In rare cases, PICS was preferred over the CIRIM in terms of image sharpness, as illustrated in Figure 4.4.

The raters significantly preferred the CIRIM over PICS in most cases, as shown in Table 4.1. For imaging artifacts, Rater 2 and Rater 3 significantly preferred the CIRIM (mean±SD = 3.8±0.6 vs. 3.4±0.5, p<0.01, and 2.9±0.4 vs. 2.2±0.6, p<0.01, respectively), whereas Rater 1 preferred PICS (4.6±0.7 vs. 4.1±0.8, p<0.01). With respect to sharpness, Rater 2 preferred the CIRIM (4.1±0.6 vs. 3.4±0.5, p<0.01), Rater 3 preferred PICS (2.8±0.5 vs. 2.3±0.5, p<0.01), and Rater 1 showed no significant difference. Rater 1 and Rater 3 preferred the CIRIM for anatomic conspicuity (4.5±0.7 vs. 3.3±0.8, p<0.01, and 3.0±0.5 vs. 2.6±0.6, p<0.01, respectively) and contrast (4.8±0.5 vs. 3.2±0.6, p<0.01, and 3.0±0.5 vs. 2.5±0.6, p<0.01, respectively). While Rater 1 significantly preferred the CIRIM in terms of diagnostic confidence (4.7±0.5 vs. 3.5±0.8, p<0.01), Rater 3 had no increased diagnostic confidence in the CIRIM after Bonferroni correction (2.9±0.6 vs. 2.6±0.7, p=0.029).

Table 4.1: Subjective ratings of the clinical cohort from two expert interventional neuroradiologists, Raters 1 and 2, and one expert pediatric neuroradiologist, Rater 3, of 57 side-by-side CIRIM and PICS reconstructions.

Category	Rater 1			Rater 2			Rater 3		
	Method (Mean ± SD)		p value	Method (Mean ± SD)		p value	Method (Mean ± SD)		p value
	PICS	CIRIM		PICS	CIRIM		PICS	CIRIM	
Imaging artifacts	4.6 ± 0.7	4.1 ± 0.8	**<0.001***	3.4 ± 0.5	3.8 ± 0.6	**0.005***	2.2 ± 0.6	2.9 ± 0.4	**<0.001***
Perceived spatial resolution (Sharpness)	4.5 ± 0.7	4.5 ± 0.7	0.639	3.4 ± 0.5	4.1 ± 0.6	**<0.001***	2.8 ± 0.5	2.3 ± 0.5	**<0.001***
Anatomic conspicuity	3.3 ± 0.8	4.5 ± 0.7	**<0.001***	3.8 ± 0.6	3.9 ± 0.6	0.463	2.6 ± 0.6	3.0 ± 0.5	**0.002***
Diagnostic confidence	3.5 ± 0.8	4.7 ± 0.5	**<0.001***	3.8 ± 0.6	4.1 ± 0.6	0.050	2.6 ± 0.7	2.9 ± 0.6	0.029**
Contrast	3.2 ± 0.6	4.8 ± 0.5	**<0.001***	3.9 ± 0.7	3.7 ± 0.7	0.088	2.5 ± 0.6	3.0 ± 0.5	**0.003***

Images were scored from 1 to 5. A score of 1 indicates nondiagnostic quality, 2 poor quality, 3 acceptable quality, 4 good quality, and 5 excellent quality. A one-sample Wilcoxon signed-rank test was used to determine significance at a Bonferroni-corrected threshold of p = 0.05/5 = 0.01, indicated in bold

SD standard deviation

*Significant difference

**Not significant after Bonferroni correction

Rater 1 reported two illustrative multiple sclerosis (MS) cases where lesions were better visible in CIRIM reconstructions. In another patient, Rater 1 noted that the internal capsule was not visible on the PICS reconstruction. In contrast, the CIRIM resulted in better image

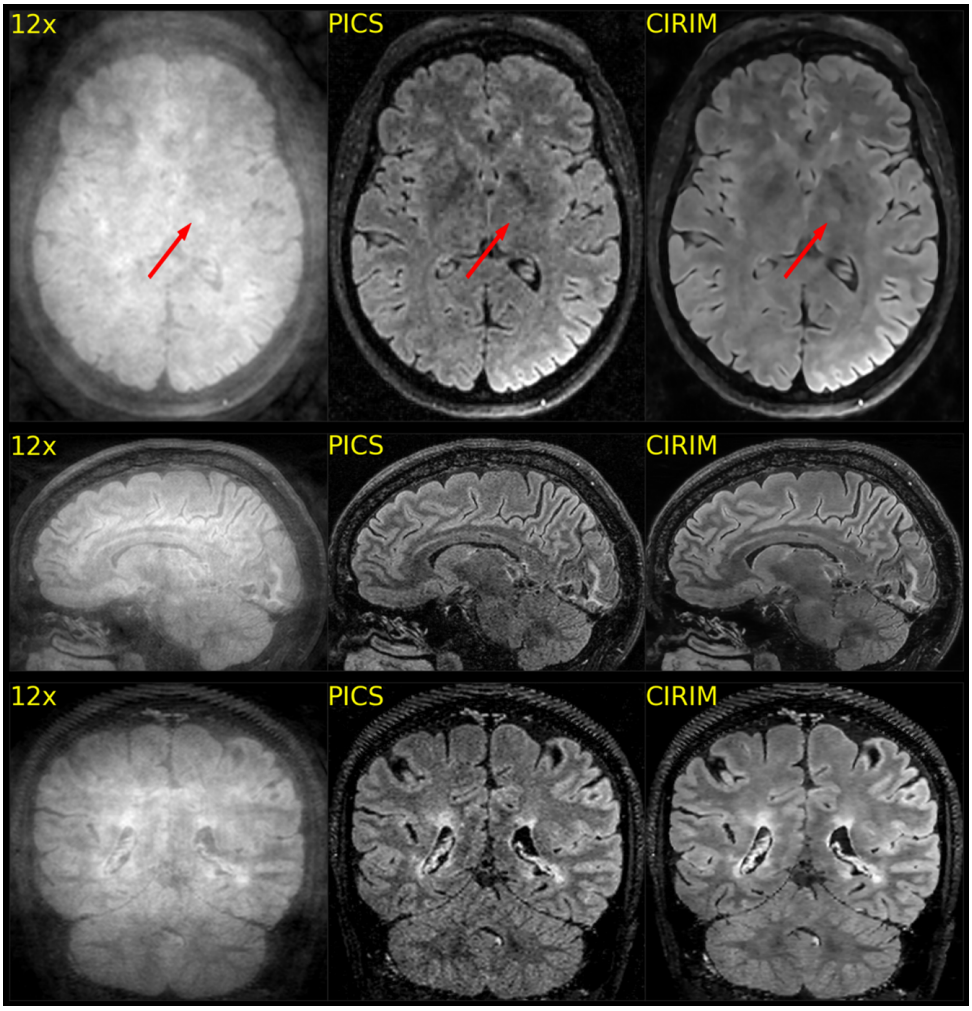


Figure 4.2: Reconstructions of 12 times accelerated FLAIR scans for three different subjects, where the CIRIM was able to generate better image quality. In the top row, PICS could not accurately depict the left internal capsule lesion (example indicated by the arrow), whereas the CIRIM preserved the contrast.

quality (Figure 4.2). Rater 3 reported that the CIRIM reconstruction was more blurred in 20 cases (scoring 2 on Sharpness): out of these cases PICS outscored the CIRIM by one point 14 times, and both scored equally on Sharpness six times. In one case, only the PICS reconstruction perceived by Rater 3 was more blurred (scoring 2 on sharpness). Rater 3 also stated that subtle MS lesions sometimes appeared slightly blurred, making it harder to discriminate them from artifacts. Raters 2 and 3 agreed that the CIRIM reconstructions were smoother than PICS reconstructions without apparent loss of detail. The interpretations of these two raters can be seen in Figure 4.3. The figure shows high-quality reconstructions of both PICS and the CIRIM, where PICS produces grainier images, whereas the CIRIM

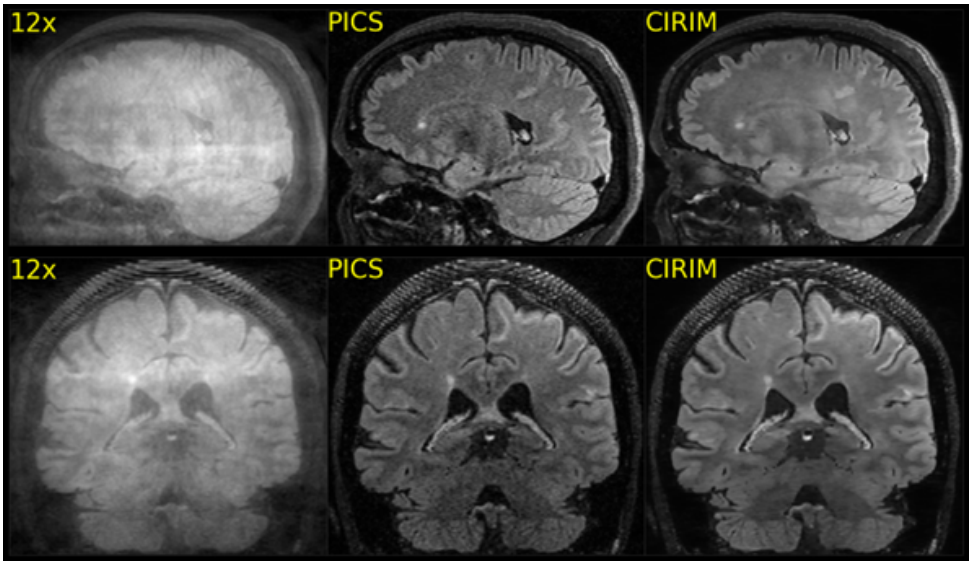


Figure 4.3: Reconstructions of 12 times accelerated FLAIR scans of two subjects, where the CIRIM and PICS provided high-quality reconstructions but were interpreted differently by two raters in a side-by-side comparison. Rater 2 interpreted the CIRIM reconstructions as having sharper edges, whereas Rater 3 interpreted the grainier PICS reconstructions as resulting in increased sharpness.

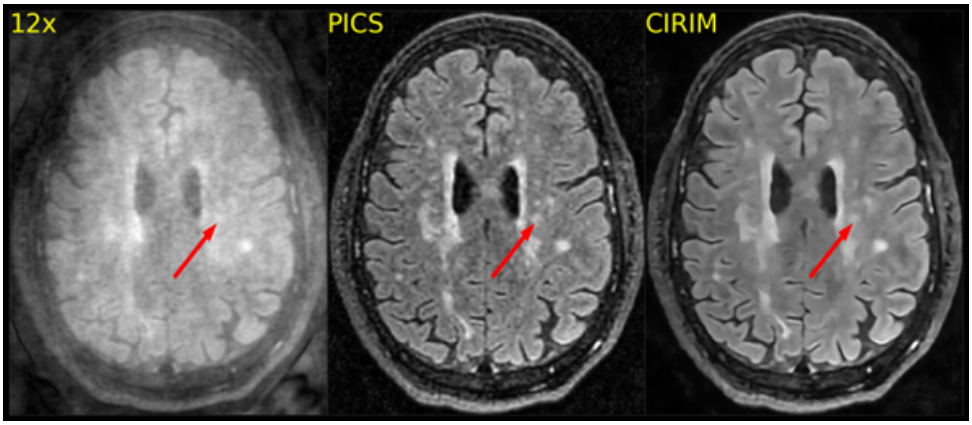


Figure 4.4: Reconstructions of a 12 times accelerated FLAIR scan, where the CIRIM yielded a more blurred reconstruction than PICS did for small T_2 high signal intensities among patients with small vessel disease (example indicated by the arrow).

results in smoother images. In a few selected cases, the grainy results of PICS resulted in higher sharpness scores than those of the CIRIM (Figure 4.4). Patient group analyses revealed no significant effect of disease on rating scores. Higher image resolution yielded significantly better imaging artifacts ($p<0.01$) and anatomic conspicuity ($p<0.01$) rating scores in CIRIM reconstructions (Table 4.2). For PICS, no improvements were observed.

The interaction effect between reconstruction method and resolution was non-significant (p>0.05).

Table 4.2: Probabilistic ordinal linear regression for evaluating the effect of image resolution depending on the reconstruction method used

Category	PICS		CIRIM	
	β [CI]	p	β [CI]	p
Imaging artifacts	0.098 [-0.287, 0.482]	0.619	0.528 [0.135, 0.922]	**<0.008***
Perceived spatial resolution (Sharpness)	0.222 [-0.166, 0.610]	0.262	0.503 [0.109, 0.896]	0.012**
Anatomic conspicuity	-0.008 [-0.391, 0.375]	0.967	0.562 [0.163, 0.961]	**<0.006***
Diagnostic confidence	0.159 [-0.224, 0.541]	0.417	0.494 [0.098, 0.891]	0.015**
Contrast	0.262 [-0.122, 0.645]	0.181	0.483 [0.082, 0.884]	0.018**

A one-sample Wilcoxon signed-rank test was used to determine significance at a Bonferroni-corrected threshold of p = 0.05/5 = 0.01, indicated in bold
*Significant difference
**Not significant after Bonferroni correction

Quantitative self-referenced MRI Quality Control (MRIQC) metrics are reported in Figure 4.5.A. CIRIM reconstructions had a significantly higher signal-to-noise ratio (SNR) and foreground to background energy ratio (FBER), yielding an improved outcome. The full width at half maximum (FWHM) was significantly higher in CIRIM reconstructions, reflecting a worse scoring. Adding Gaussian noise to a CIRIM reconstruction, resulting in a lowering in SNR from 11 to 8.5, yielded a lower estimated FWHM of 3.8 instead of 4.4.

In terms of reconstruction times is given in Figure 4.5.B,. CIRIM reconstructions were, on average, approximately five times faster (146±1.7s) than PICS reconstructions were (707±20s).

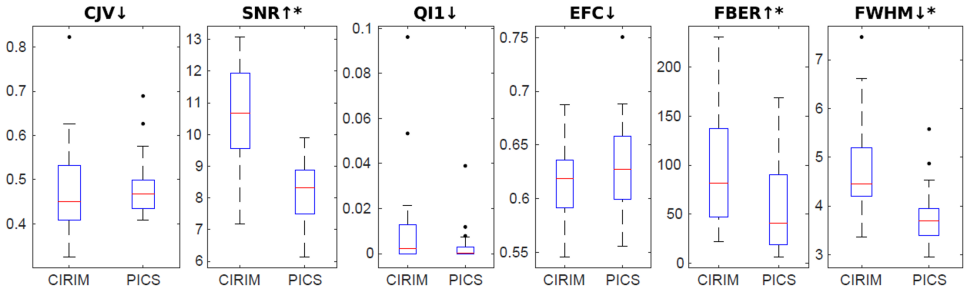


Figure 4.5: Quantitative self-referenced MRI Quality Control (MRIQC) metrics: coefficient of joint variation (CJV), signal-to-noise ratio (SNR), quality index 1 (QI1), entropy focus criterion (EFC), foreground to background ratio (FBER), full width at half maximum (FWHM). Arrows indicate better values. Significance is indicated with an asterisk (*) at p<0.05 (corrected).

Figure 4.6 shows contrast resolution (CR) as a function of the acceleration factor and highlights two example slices. No significant effect of the acceleration factor on the CR was observed (p=0.92). The SSIM decreased significantly with increasing acceleration factor (p=0.002), for which a plot is depicted in Figure 4.7.

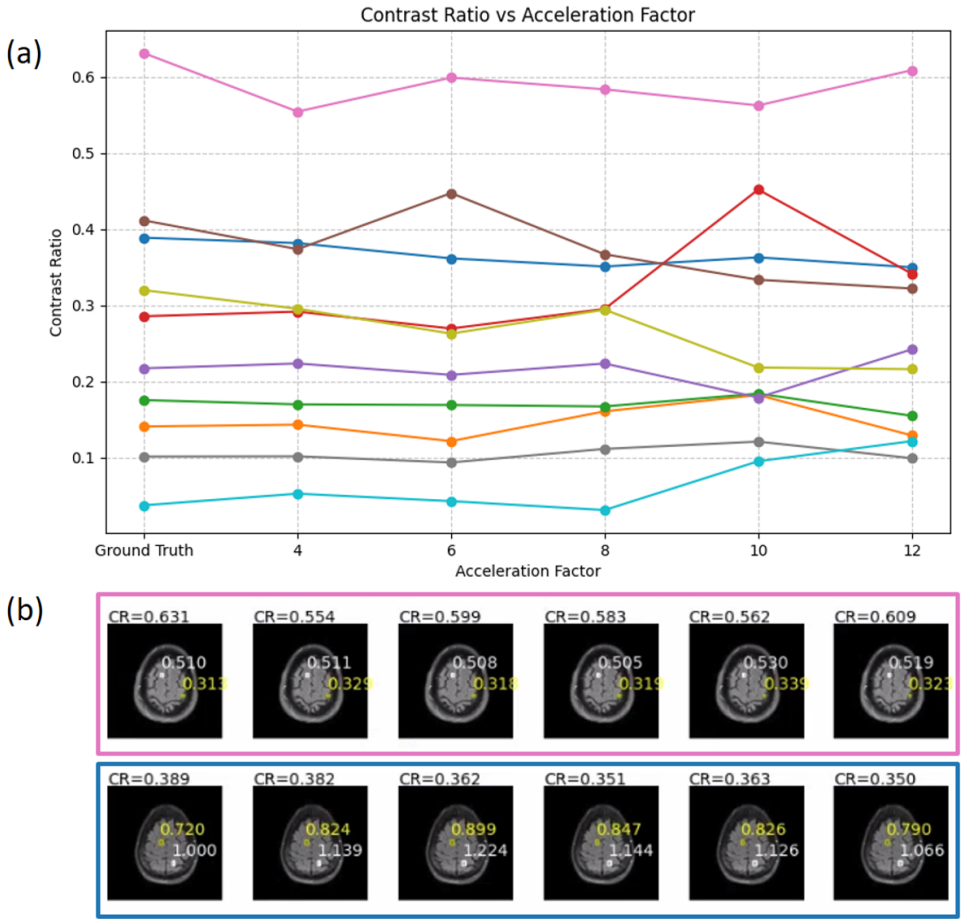


Figure 4.6: **a** Contrast ratio as a function of the acceleration factor for 10 FLAIR slices with an annotated white matter lesion in the FastMRI dataset. Different colors represent data from diferent subjects. **b** Selected reconstructions with bounding box colors matching the plotted lines in (**a**), and white and yellow bounding boxes positioned around the lesion and a selected white matter region.

When comparing lesion contrast in manually annotated lesions in our data, on average, the CR is 5.8% lower in CIRIM reconstructions than in PICS reconstructions, decreasing from 1.17 to 1.10. This difference was significant (p=0.04). Figure 4.8 illustrates that the CR is marginally lower but to a large extent preserved in CIRIM compared with PICS.

4.4 Discussion

We demonstrated the value of reconstructing highly accelerated clinical FLAIR data with DL. The CIRIM could generalize well to heterogeneous clinical data that had not been previously reported, as it was trained on another distribution with another contrast (i.e., 3D-T_1 scans of healthy volunteers). Rather than being explicitly trained on reconstructing a

specific contrast or tissue, the physics-informed network has learned to efficiently denoise FLAIR images. Notably, obtaining fully sampled FLAIR data in patients is infeasible as it leads to excessive scanning times of up to 30 minutes at the current isotropic resolution, with associated imaging artifacts. Despite not being optimized for this type of data, the CIRIM retained its efficient denoising capacity in a dataset with a high degree of clinically desirable zero-filling.

Compared with the default PICS method, the CIRIM network's denoising ability led to an almost 30% increase in SNR in brain tissue. In line with the improved calculated SNR, subjective metrics were primarily scored in favor of CIRIM. The increased SNR in CIRIM reconstructions led to significantly higher rating scores for imaging artifacts and anatomic conspicuity for higher-resolution images. While showing a smooth appearance, small features were still discernible in most images, which is clinically relevant [30]. The higher FWHM seen in CIRIM-reconstructions can be attributed to noise effects dominating the histogram distribution based on which this metric is computed. Image quality is thus maintained when reconstructing clinical data in a modality unseen during training, allowing for a reduction in scanning time or an increase in resolution while maintaining clinically acceptable image quality.

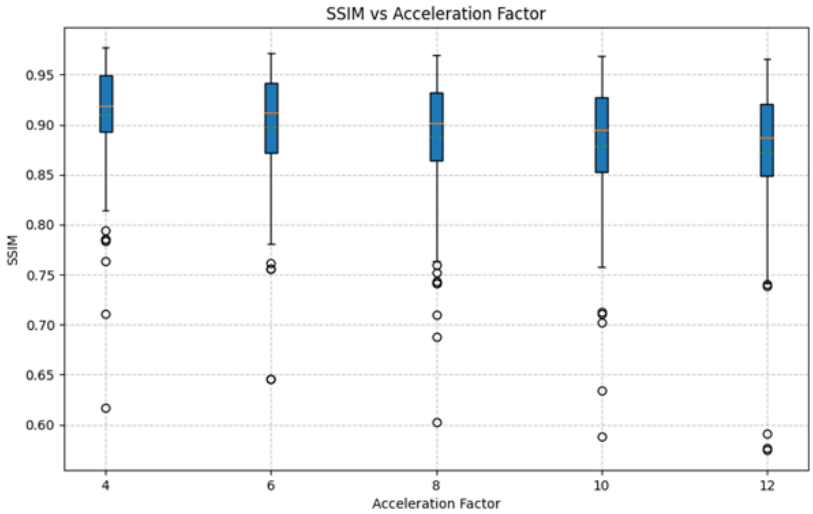


Figure 4.7: Structural similarity (SSIM) as a function of acceleration factor for the FastMRI validation set of 107 scans.

When embedding DL-enabled reconstruction methods in the clinic, it is interesting to note that the neuroradiologists, Raters 1 and 2, agreed on higher image quality in the CIRIM reconstructions than in the PICS reconstructions. On a more detailed level, a lower interrater agreement regarding sharpness and contrast in specific cases was observed. Specifically, Rater 1 significantly preferred CIRIM in terms of sharpness, suggesting that further denoised reconstructions yielded sharper edges than PICS did. Rater 2 perceived the denoising as a loss of spatial resolution or smoothness, resulting in a significant preference for PICS over CIRIM. Regarding image contrast, Rater 2 preferred CIRIM because of

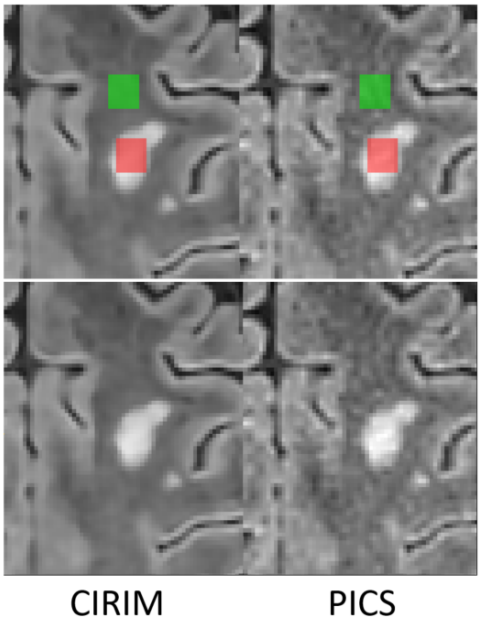


Figure 4.8: Manual annotation of a lesion (red) and proximal white matter (green) in CIRIM (left) and PICS (right) reconstructions.

improved dealiasing of adjacent regions, leading to enhanced visibility of lesions. Notably, not all the raters were accustomed to reading data with a high acceleration factor (12 times) in their daily routine. Rater 3 reported that in the field of pediatric neuroradiology, image quality at a slightly less aggressive acceleration is to preferable. Previous work also reported mixed interrater agreement [3]. It is evident that CIRIM reconstructions differ significantly from iterative reconstruction algorithms, highlighting the need for good interaction and communication when implementing DL reconstruction methods in the clinic.

Another essential advantage of the model used is its ability to increase reconstruction speed, which is highly relevant in neurological deficits where acquisition speed, reconstruction times, and image quality need to be balanced. At large FOVs, PICS might be too slow for clinical use despite being deployed on fast GPUs. Image quality and reconstruction time can be traded within the CIRIM by choosing a different number of cascades than those used here. Compared with previous work, we increased the number of channels of the network from 64 to 128, aiming to improve image quality further.

The main protocol used in this study was designed to have a high acceleration factor of 12x. Other related works chose a more conservative acceleration factor in the range of 2x to 4x [31]. This sequence was set up as multislice, allowing for acceleration along one phase-encoding dimension only. In contrast, here, a 3D FLAIR sequence is adopted with two phase encoding dimensions, allowing for much higher speed-up factors. Notably, this sequence is the standard of care in our clinic with compressed sensing reconstruction.

We intentionally designed the experiments to compare the performance of CIRIM and PICS on reconstruction alone and to do so offline (i.e., off-scanner). We wanted to exclude any on-scanner postprocessing to visually enhance the images, as this approach may be in place for some vendors. Postprocessing is typically performed with proprietary software. Thus, we disregarded this step to avoid an unbalanced comparison between methods. Moreover, leaving out post-processing makes the presented results more easily comparable with results from other vendors. However, in future research, it would be valuable to investigate the denoising capabilities of both CIRIM and on-scanner postprocessing filters for comparison purposes.

A few limitations need to be noted regarding the present study. In certain CIRIM reconstructions, a small amount of blur was introduced, possibly caused by the high acceleration factor (twelvefold) and the resulting low intrinsic signal in the data of some patients. Highly accelerated deep learning-based reconstructions need to be carefully evaluated in the clinic, since artifacts may appear differently than with conventional reconstruction methods. Metrics often used in image reconstruction in addition to SNR, such as the structural similarity index (SSIM), could not be computed since we did not have a fully sampled scan. Another limitation is that we could not compare with on-scanner reconstruction times since hardware and software differences hinder a comparison of the algorithm with the reconstructions performed on-scanner. Furthermore, a fully sampled reference could not be acquired, because of the risk of motion artifacts and image blurring in the 24 to 30 minute scanning time of a sequence without acceleration. For the prospectively acquired patient data, no ground truth data were available, and comparisons with clinically used CS reconstructions prohibited an assessment of changes in diagnosis. However, the analysis of FastMRI data demonstrated that CR is preserved over a broad range of acceleration factors.

We demonstrated the added value of deep learning in reconstructing 3D FLAIR scans in a clinically representative sample with neurological deficits. Reconstructions made with the physics-informed CIRIM model have increased SNRs and appear less noisy than PICS iterative reconstruction. The higher SNR in CIRIM reconstructions enables scanning at higher resolution. Moreover, the CIRIM achieves faster reconstruction times, which is crucial for the timely diagnosis of neurological deficits. Online inference on MRI scanners requires a graphical processing unit (GPU) to be installed on the reconstruction computer. The CIRIM, which balances the network size and is therefore memory efficient, does not place high demands on the specifications of GPU cards, requiring 1.6 GB of memory for 264k parameters in total. This work shows the promise of physics-informed neural networks in accelerated MRI reconstruction. Future work should evaluate whether the surplus in the SNR can be traded for further acceleration.

4.5 Funding information

This work was supported by the STAIRS project under the TKI-PPP program. The collaboration project is cofunded by the PPP Allowance made available by Health Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships.

4.6 Declaration of Interest

B.J. Emmer received funds from Healthcare Evaluation Netherlands, Health Holland. (unrelated to this project; all paid to institution). C.B.L.M. Majoie received funds from: CVON/-Dutch Heart Foundation, European Commission, Healthcare Evaluation Netherlands, TWIN Foundation and Stryker (unrelated to this project; all paid to institution. H.A. Marquering, C.B.L.M. Majoie and M.W.A. Caan are shareholders of Nicolab International Ltd. H.A. Marquering is a shareholder of TrianecT B.V. (unrelated to this project; all paid individually) J. Savatovsky received funds for travel costs and lecture fees: Koninklijke Philips N.V., Canon Inc., and Bayer AG. Julien Savatovsky is a shareholder of Chipiron, Guerbet, and GeodAIsics.

References

[1] S. Toledano-Massiah, A. Sayadi, R. De Boer, J. Gelderblom, R. Mahdjoub, S. Gerber, M. Zuber, M. Zins, and J. Hodel. Accuracy of the Compressed Sensing Accelerated 3D-FLAIR Sequence for the Detection of MS Plaques at 3T. *American Journal of Neuroradiology*, 39:454–458, 3 2018.

[2] T. Yamamoto, C. Lacheret, H. Fukutomi, R.A. Kamraoui, L. Denat, B. Zhang, V. Prevost, L. Zhang, A. Ruet, B. Triaire, V. Dousset, P. Coupé, and T. Tourdias. Validation of a Denoising Method Using Deep Learning–Based Reconstruction to Quantify Multiple Sclerosis Lesion Load on Fast FLAIR Imaging. *American Journal of Neuroradiology*, 43(8):1099–1106, 2022.

[3] J. E. Vranic, N. M. Cross, Y. Wang, D. S. Hippe, E. De Weerdt, and Mahmud M.D. Mossa-Basha. Compressed Sensing–Sensitivity Encoding (CS-SENSE) Accelerated Brain Imaging: Reduced Scan Time without Reduced Image Quality. *American Journal of Neuroradiology*, 40:92–98, 1 2019.

[4] Yuichi Okano, Kazuaki Ishimatsu, Yoichi Kato, Junichi Yamaga, Ken Kuwahara, Katsuki Okumoto, and Kuniyasu Wada. Clinical features of stroke mimics in the emergency department. *Acute Medicine & Surgery*, 5:241–248, 7 2018.

[5] Shannon Hextrum and José Biller. Clinical Distinction of Cerebral Ischemia and Triaging of Patients in the Emergency Department: Mimics, Wake-ups, Late Strokes, and Chameleons, 11 2018.

[6] M Ebinger, J B Fiebach, Götz Thomalla, Bastian Cheng, Martin Ebinger, Qing Hao, Thomas Tourdias, Ona Wu, Jong S Kim, Lorenz Breuer, Oliver C Singer, Steven Warach, Soren Christensen, Andras Treszl, Nils D Forkert, Ivana Galinovic, Michael Rosenkranz, Tobias Engelhorn, Martin Köhrmann, Matthias Endres, Dong-Wha Kang, Vincent Dousset, Gregory Sorensen, David S Liebeskind, Jochen B Fiebach, Jens Fiehler, and Christian Gerloff. DWI-FLAIR mismatch for the identifi cation of patients with acute ischaemic stroke within 4·5 h of symptom onset (PRE-FLAIR): a multicentre observational study. *The Lancet Neurology*, 10:978–986, 2011.

[7] Liang Jiang, Mingyang Peng, Wen Geng, Huiyou Chen, Haobo Su, Boxiang Zhao, Yu Chen Chen, and Xindao Yin. FLAIR hyperintensities-DWI mismatch in acute

stroke: associations with DWI volume and functional outcome. *Brain Imaging and Behavior*, 14:1230–1237, 8 2020.

[8] Álvaro García-Tornel, Daniel Campos, Marta Rubiera, Sandra Boned, Marta Olivé-Gadea, Manuel Requena, Ludovico Ciolli, Marian Muchada, Jorge Pagola, David Rodriguez-Luna, Matias Deck, Jesus Juega, Noelia Rodríguez-Villatoro, Estela Sanjuan, Alejandro Tomasello, Carlos Piñana, David Hernández, José Álvarez Sabin, Carlos A. Molina, and Marc Ribó. Ischemic Core Overestimation on Computed Tomography Perfusion. *Stroke*, 52(5):1751–1760, 2021.

[9] Klaas P. Pruessmann, Markus Weiger, Markus B. Scheidegger, and Peter Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–962, 1999.

[10] Mark A. Griswold, Peter M. Jakob, Robin M. Heidemann, Mathias Nittka, Vladimir Jellus, Jianmin Wang, Berthold Kiefer, and Axel Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, 47(6):1202–1210, 2002.

[11] Michael Lustig, David Donoho, and John M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[12] J. Ding, Y. Duan, Z. Zhuo, Y. Yuan, G. Zhang, Q. Song, B. Gao, B. Zhang, M. Wang, L. Yang, Y. Hou, J. Yuan, C. Feng, J. Wang, L. Lin, and Y. Liu. Acceleration of Brain TOF-MRA with Compressed Sensitivity Encoding: A Multicenter Clinical Study. *American Journal of Neuroradiology*, 42:1208–1215, 7 2021.

[13] S. S. Lu, M. Qi, X. Zhang, X. H. Mu, M. Schmidt, Y. Sun, C. Forman, P. Speier, and X. N. Hong. Clinical Evaluation of Highly Accelerated Compressed Sensing Time-of-Flight MR Angiography for Intracranial Arterial Stenosis. *American Journal of Neuroradiology*, 39:1833–1838, 10 2018.

[14] S. Bash, L. Wang, C. Airriess, G. Zaharchuk, E. Gong, A. Shankaranarayanan, and L.N. Tanenbaum. Deep Learning Enables 60% Accelerated Volumetric Brain MRI While Preserving Quantitative Performance: A Prospective, Multicenter, Multireader Trial. *American Journal of Neuroradiology*, 42:2130–2137, 12 2021.

[15] Roy Shaul, Itamar David, Ohad Shitrit, and Tammy Riklin Raviv. Subsampled brain MRI reconstruction by generative adversarial neural networks. *Medical Image Analysis*, 65:101747, 10 2020.

[16] Stephan Ursprung, Judith Herrmann, Natalie Joos, Elisabeth Weiland, Thomas Benkert, Haidara Almansour, Andreas Lingg, Saif Afat, and Sebastian Gassenmaier. Accelerated diffusion-weighted imaging of the prostate using deep learning image reconstruction: A retrospective comparison with standard diffusion-weighted imaging. *European Journal of Radiology*, 165, 8 2023.

[17] Hemant K. Aggarwal, Merry P. Mani, and Mathews Jacob. MoDL: Model-Based Deep Learning Architecture for Inverse Problems. *IEEE Transactions on Medical Imaging*, 38(2):394–405, 2019.

[18] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P. Recht, Daniel K. Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, 2018.

[19] Kai Lønning, Patrick Putzky, Jan-Jakob Sonke, Liesbeth Reneman, Matthan W.A. Caan, and Max Welling. Recurrent inference machines for reconstructing heterogeneous MRI data. *Medical Image Analysis*, 53:64–78, 2019.

[20] D. Karkalousos, S. Noteboom, H. E. Hulst, F. M. Vos, and M. W. A. Caan. Assessment of data consistency through cascades of independently recurrent inference machines for fast and robust accelerated MRI reconstruction. *Phys. Med. Biol.*, 67(12):124001, June 2022.

[21] Martin Uecker, Frank Ong, Jonathan I. Tamir, Dara Bahri, Patrick Virtue, Joseph Y Cheng, Tao Zhang, and Michael Lustig. Berkeley Advanced Reconstruction Toolbox. *Proc. Intl. Soc. Mag. Reson. Med.*, 23, 2015.

[22] Jeffrey D. Rudie, Tyler Gleason, Matthew J. Barkovich, David M. Wilson, Ajit Shankaranarayanan, Tao Zhang, Long Wang, Enhao Gong, Greg Zaharchuk, and Javier E. Villanueva-Meyer. Clinical Assessment of Deep Learning–based Super-Resolution for 3D Volumetric Brain MRI. *Radiology: Artificial Intelligence*, 4, 3 2022.

[23] Oscar Esteban, Daniel Birman, Marie Schaer, Oluwasanmi O. Koyejo, Russell A. Poldrack, and Krzysztof J. Gorgolewski. MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLOS ONE*, 12(9):1–21, 09 2017.

[24] Marco Ganzetti, Nicole Wenderoth, and Dante Mantini. Intensity Inhomogeneity Correction of Structural MR Images: A Data-Driven Approach to Define Input Algorithm Parameters. *Frontiers in Neuroinformatics*, 10, 2016.

[25] Bénédicte Mortamet, Matt A. Bernstein, Clifford R. Jack Jr., Jeffrey L. Gunter, Chadwick Ward, Paula J. Britson, Reto Meuli, Jean-Philippe Thiran, and Gunnar Krueger. Automatic quality assessment in structural brain magnetic resonance imaging. *Magnetic Resonance in Medicine*, 62(2):365–372, 2009.

[26] David Atkinson, Derek L.G. Hill, Peter N.R. Stoyle, Paul E. Summers, Stuart Clare, Richard Bowtell, and Stephen F. Keevil. Automatic compensation of motion artifacts in MRI. *Magnetic Resonance in Medicine*, 41(1):163–170, 1999.

[27] Zarrar Shehzad, Giavasis Steven, li Qingyang, Yassine Benhajali, Yan Chaogan, Zhen Yang, Milham Michael, Pierre Bellec, and Cameron Craddock. The Preprocessed Connectomes Project Quality Assessment Protocol - a resource for measuring the quality of MRI data. *Frontiers in Neuroscience*, 9, 01 2015.

[28] Steven D. Forman, Jonathan D. Cohen, Mark Fitzgerald, William F. Eddy, Mark A. Mintun, and Douglas C. Noll. Improved Assessment of Significant Activation in Functional Magnetic Resonance Imaging (fMRI): Use of a Cluster-Size Threshold. *Magnetic Resonance in Medicine*, 33(5):636–647, 1995.

[29] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R.J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, Aditya Vijaykumar, Alessandro Pietro Bardelli, Alex Rothberg, Andreas Hilboll, Andreas Kloeckner, Anthony Scopatz, Antony Lee, Ariel Rokem, C. Nathan Woods, Chad Fulton, Charles Masson, Christian Häggström, Clark Fitzgerald, David A. Nicholson, David R. Hagen, Dmitrii V. Pasechnik, Emanuele Olivetti, Eric Martin, Eric Wieser, Fabrice Silva, Felix Lenders, Florian Wilhelm, G. Young, Gavin A. Price, Gert Ludwig Ingold, Gregory E. Allen, Gregory R. Lee, Hervé Audren, Irvin Probst, Jörg P. Dietrich, Jacob Silterra, James T. Webber, Janko Slavič, Joel Nothman, Johannes Buchner, Johannes Kulick, Johannes L. Schönberger, José Vinícius de Miranda Cardoso, Joscha Reimer, Joseph Harrington, Juan Luis Cano Rodríguez, Juan Nunez-Iglesias, Justin Kuczynski, Kevin Tritz, Martin Thoma, Matthew Newville, Matthias Kümmerer, Maximilian Bolingbroke, Michael Tartre, Mikhail Pak, Nathaniel J. Smith, Nikolai Nowaczyk, Nikolay Shebanov, Oleksandr Pavlyk, Per A. Brodtkorb, Perry Lee, Robert T. McGibbon, Roman Feldbauer, Sam Lewis, Sam Tygier, Scott Sievert, Sebastiano Vigna, Stefan Peterson, Surhud More, Tadeusz Pudlik, Takuya Oshima, Thomas J. Pingel, Thomas P. Robitaille, Thomas Spura, Thouis R. Jones, Tim Cera, Tim Leslie, Tiziano Zito, Tom Krauss, Utkarsh Upadhyay, Yaroslav O. Halchenko, and Yoshiki Vázquez-Baeza. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17:261–272, 3 2020.

[30] Shigeru Kiryu, Hiroyuki Akai, Koichiro Yasaka, Taku Tajima, Akira Kunimatsu, Naoki Yoshioka, Masaaki Akahane, Osamu Abe, and Kuni Ohtomo. Clinical Impact of Deep Learning Reconstruction in MRI. *RadioGraphics*, 43(6):e220133, 2023. PMID: 37200221.

[31] Arne Estler, Till-Karsten Hauser, Annerose Mengel, Merle Brunnée, Leonie Zerweck, Vivien Richter, Mario Zuena, Martin Schuhholz, Ulrike Ernemann, and Georg Gohla. Deep Learning Accelerated Image Reconstruction of Fluid-Attenuated Inversion Recovery Sequence in Brain Imaging: Reduction of Acquisition Time and Improvement of Image Quality. *Academic Radiology*, 31(1):180–186, 2024.

Supplementary Material

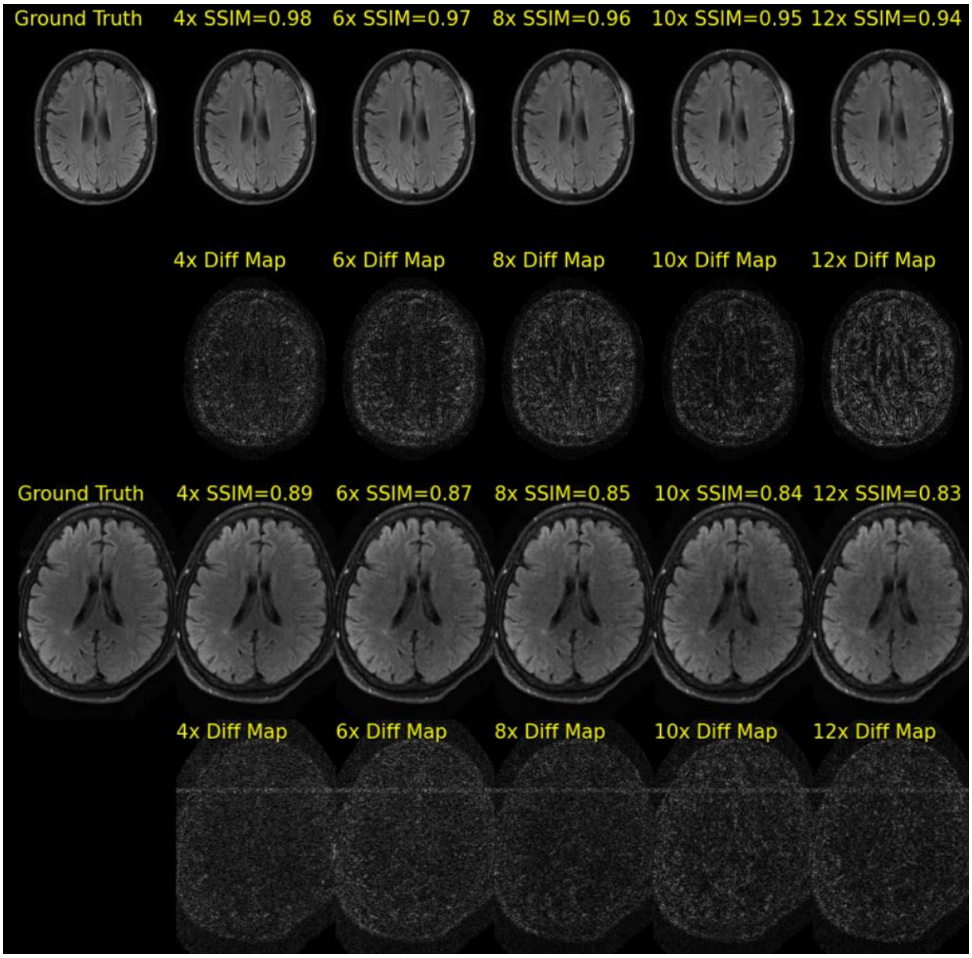


Figure S.4.1: Example reconstructions of FastMRI FLAIR scans with relatively high and low Structural Similary (SSIM) values, with absolute difference maps relative to the ground truth. Note that the line artefact in the lower scan is present in the ground truth and largely filtered out in the CIRIM reconstructions.

5

MultiTask Learning for accelerated-MRI Reconstruction and Segmentation of Brain Lesions in Multiple Sclerosis

Karkalousos, D., Išgum, I., Marquering, H.A., & Caan, M.W.A. (2024, January)
In Medical Imaging with Deep Learning (pp. 991-1005). PMLR.

Abstract

This work proposes MultiTask Learning for accelerated-MRI Reconstruction and Segmentation (MTLRS). Unlike the common single-task approaches, MultiTask Learning identifies relations between multiple tasks to improve the performance of all tasks. The proposed MTLRS consists of a unique cascading architecture, where a recurrent reconstruction network and a segmentation network inform each other through hidden states. The features of the two networks are shared and implicitly enforced as inductive bias. To evaluate the benefit of MTLRS, we compare performing the two tasks of accelerated-MRI reconstruction and MRI segmentation with pre-trained, sequential, end-to-end, and joint approaches. A synthetic multicoil dataset is used to train, validate, and test all approaches with five-fold cross-validation. The dataset consists of 3D FLAIR brain data of relapsing-remitting Multiple Sclerosis patients with known white matter lesions. The acquisition is prospectively undersampled by approximately 7.5 times compared to clinical standards. Reconstruction performance is evaluated by Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR). Segmentation performance is evaluated by Dice score for combined brain tissue and white matter lesion segmentation and by per lesion Dice score. Results show that MTLRS outperforms other evaluated approaches, providing high-quality reconstructions and accurate white matter lesion segmentation. A significant correlation was found between the performance of both tasks (SSIM and per lesion Dice score, $\rho = 0.92$, $p = 0.0005$). Our proposed MTLRS demonstrates that accelerated-MRI reconstruction and MRI segmentation can be effectively combined to improve performance on both tasks, potentially benefiting clinical settings.

5.1 Introduction

Acquisition, reconstruction, and analysis of Magnetic Resonance Imaging (MRI) are currently performed in a sequence of distinct tasks. Performing each task independently misses the opportunity to share valuable information between the tasks and jointly optimize their performance. MultiTask Learning (MTL) is a technique in which multiple domain-related tasks are trained in parallel using shared features, effectively acting as inductive bias. MTL can implicitly identify task-relatedness, yielding improved generalization [1]. By utilizing the information in multiple tasks, the performance of each task can be improved. Recently task-adapted reconstruction was proposed to combine reconstruction with related tasks [2] in different approaches.

In a pre-trained approach, a reconstruction network and a segmentation network are trained separately to perform the tasks individually. In a sequential approach, the segmentation network is fine-tuned using the predictions of the reconstruction network. In an end-to-end approach, the reconstruction and the segmentation networks are trained together at the same time. For performing end-to-end accelerated-MRI reconstruction and MRI segmentation, Huang et al. [3] proposed the SEgmentation Recurrent Attention Network (SERANET), starting from the subsampled k-space to result in a segmentation. In a joint approach, the reconstruction and segmentation networks are trained end-to-end, computing a joint reconstruction and segmentation loss with a weighting factor balancing the two tasks. For performing the two tasks jointly, Sun et al. [4] proposed the SegNet, consisting of cascades of U-Nets for reconstruction and a separate decoder for segmentation, using the output of all the reconstruction encoders. Similarly, the Image Deep Structured

Low-Rank (IDSLR) [5] and the RecSeg [6] methods perform joint reconstruction and segmentation. The IDSLR uses only the output of the final encoder for segmentation, while the RecSeg uses a second U-Net.

In this work, we formulate the inverse problem of accelerated-MRI reconstruction and the task of MRI segmentation as a multitask problem. In contrast to earlier methods, we show that performance on both tasks can be improved by informing each other through a recurrent approach. To this end, we leverage the Cascades of Independently Recurrent Inference Machines (CIRIM) [7] for accelerated-MRI reconstruction and we add a segmentation network to the cascades to inform MultiTask Learning for accelerated-MRI Reconstruction and Segmentation (MTLRS). Following [2], the aim is to find a forward operator that directly maps accelerated-MR images to MRI segmentation. In MTLRS, this direct operator is modeled by coupling the output of the hidden layers of the reconstruction network with the output of the segmentation network. We develop and evaluate the proposed MLTRS using five-fold cross-validation on a synthetic multicoil dataset of 3D FLAIR data of relapsing-remitting Multiple Sclerosis patients with known white matter lesions.

5.2 Methods

5.2.1 MultiTask Learning for accelerated-MRI Reconstruction and Segmentation

The inverse problem of accelerated-MRI reconstruction can be formalized through a forward model. Let $x \in \mathbb{C}^n$ with $n = n_x \times n_y$, be a true image and let $y \in \mathbb{C}^m$, with $m << n$, be the set of sparse k-space measurements. The forward model describes y as

$$y_i = A(x) + \sigma_i, i = 1, ..., c, \tag{5.1}$$

where i denotes the current receiver coil, for a total of c coils. $A : \mathbb{C}^n \mapsto \mathbb{C}^{n \times n_c}$ is the linear forward operator of accelerating MR acquisition, and $\sigma_i \in \mathbb{C}^n$ denotes the noise from the scanner for the $i-th$ coil. A is given by $A = U \odot \mathcal{F} \odot \epsilon$, where U denotes the subsampling operator and $\mathcal{F}$ the Fourier transform. $\epsilon : \mathbb{C}^n \times \mathbb{C}^{n \times n_c} \mapsto \mathbb{C}^{n \times n_c}$ is the expand operator, transforming x into x_c multicoil images, given by $\epsilon(x) = (S_0 \odot x, ..., S_x \odot x) = (x_0, ..., x_c)$ where S denote the coil sensitivity maps. Subsequently, the backward operator for projecting the sparse k-space to image space is given by $A^* = r \odot \mathcal{F}^{-1} \odot U^T$, where $\mathcal{F}^{-1}$ denotes the inverse Fourier transform. $r : \mathbb{C}^{n \times n_c} \times \mathbb{C}^{n \times n_c} \mapsto \mathbb{C}^n$ is the reduce operator computing a coil-combined image given by $r(x_0, ..., x_c) = \sum_{i=1}^{c} S_i^H \odot x_i$, where H denotes the Hermitian complex conjugate.

When solving the inverse problem of accelerated-MRI reconstruction, the $y \mapsto x$ mapping (Eq. 5.1) can be found through a Maximum A Posteriori (MAP) estimation. Formulating the MAP estimation into a non-convex optimization scheme [8] results in updates of the form

$$x_{\iota+1} = x_\iota + \theta_\phi \left(\nabla_{y|x_\iota}, x_\iota \right), \tag{5.2}$$

at iteration ι, for total number of iterations I. $\nabla_{y|x_\iota}$ is the gradient of the log-likelihood given by $\nabla_{y|x} := \frac{1}{\sigma^2} A^* (A(x) - y)$, assuming data are acquired under a Gaussian distribution. θ_ϕ explicitly models the update rule using a Recurrent Neural Network (RNN).

Here, we use a learned inverse problem solver, the Cascades of Independently Recurrent Inference Machines (CIRIM) [7]. The update equations of the network for the first cascade are given by

$$\begin{aligned} h_0^{k=1} &= 0, & \hat{x_0}^{k=1} &= A^*(y), \\ h_{t+1}^{k=1} &= \theta_\phi^* \left(\nabla_{y|\hat{x}_t}, \hat{x}_t, h_t \right), & \hat{x_{t+1}}^{k=1} &= \hat{x}_t + \theta_\phi \left(\nabla_{y|\hat{x}_t}, \hat{x}_t, h_{t+1} \right), \end{aligned} \tag{5.3}$$

where θ_ϕ^* is the updated model for the hidden state variable h and k denotes the current cascade, for total K cascades. For the rest $2 \leq k \leq K$ cascades, we extend the CIRIM by including a segmentation network and further informing it of the segmentation task described by

$$s = T(x), \tag{5.4}$$

where $T : x \mapsto s$ is the generic forward segmentation operator and can be replaced by any segmentation network. MultiTask Learning for accelerated-MRI Reconstruction and Segmentation (MTLRS) is then realized by coupling the output of the hidden states with s, resulting in updates of the form

$$\begin{aligned} h_0^{k\geq 2} &= \hat{x_I}^{k-1} * s^{k-1}, & \hat{x_0}^{k\geq 2} &= \hat{x_I}^{k-1}, \\ h_{t+1}^{k\geq 2} &= \theta_\phi^* \left(\nabla_{y|\hat{x_t}^k}, \hat{x_t}^k, \hat{x_I}^{k-1} * s^{k-1} \right), & \hat{x_{t+1}}^{k\geq 2} &= \hat{x_t}^k + \theta_\phi \left(\nabla_{y|\hat{x_t}^k}, \hat{x_t}^k, h_{t+1}^k \right). \end{aligned} \tag{5.5}$$

In this way, the reconstruction informs the segmentation network and vice versa. A schematic representation is shown in Fig. 5.1.

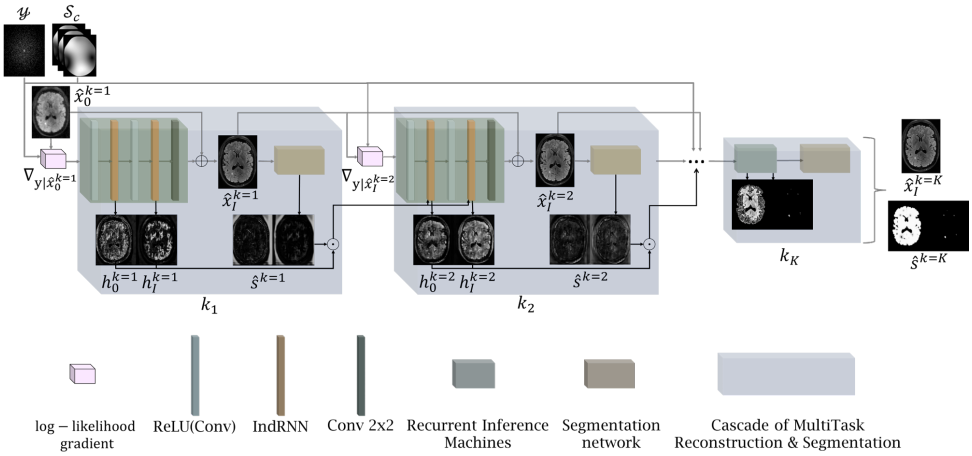


Figure 5.1: Schematic overview of the MultiTask Learning for accelerated-MRI Reconstruction and Segmentation (MTLRS) framework. MTLRS consists of K cascades of a reconstruction network (top-leftmost block on each cascade) and a segmentation network (top-rightmost block on each cascade). On each cascade, the network first performs a reconstruction ($\hat{x_I}^k$), next a segmentation ($\hat{s}^k$), and finally couples the segmented output with the output of the hidden layers (h_0^k and h_I^k), to initialize the hidden layers of the next cascade. After K cascades, the network outputs a final reconstruction ($\hat{x_I}^{k=K}$) and segmentation ($\hat{s}^{k=K}$) (top-rightmost).

5.2.2 Loss function

The loss function in MTLRS is described by a joint reconstruction $L_{recon}(\hat{x}, x)$ and segmentation $L_{seg}(\hat{s}, s)$ loss. The joint loss is given by

$$L^{joint} = \frac{1}{N}\sum_{n=1}^{N}(1-\alpha)L_{recon}(\hat{x}_n, x_n) + \alpha L_{seg}(\hat{s}_n, s_n), \tag{5.6}$$

where n is the current batch and N is the total number of training samples. x is the ground truth image, $\hat{x}$ the predicted reconstruction, s the ground truth segmentation label, and $\hat{s}$ the predicted segmentation. α, with $0 \leq \alpha \leq 1$, is a weighting factor, balancing the influence of each task to the final loss.

L^{recon} is usually computed on the magnitudes x and $\hat{x}$, where $\hat{x_0} = A(y)$ is the initially zero-filled reconstruction. In the case of the l_1-norm, the loss is given by

$$L^{l_1}(\hat{x}, x) = \frac{1}{N}\sum_{n=1}^{N}|\hat{x}_n - x_n|. \tag{5.7}$$

For MTLRS, L^{recon} is weighted over the number of recurrent iterations. Thus Eq. 5.7 is reformulated as

$$L^{l_1}(\hat{x}, x) = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{qI}\sum_{\tau=1}^{I} w_\tau |\hat{x_\tau}_n - x_n|\right), \tag{5.8}$$

where q is the total number of pixels and w_τ is a vector containing I weights, for a total number of iterations I, to emphasize the loss at later recurrent iterations. The weights are calculated as $w_\tau = 10^{-\frac{I-\tau}{I-1}}$.

For segmentation loss, we choose the commonly used binary cross-entropy loss and combine it with the Dice loss to ameliorate class imbalance given the very small size of white matter lesions compared to segmented brain tissue. Therefore, a combined weighted binary cross-entropy and Dice loss assures stable loss computation. L_{seg} is then given by

$$L_{seg}(\hat{s}, s) = \beta L^{CE}(\hat{s}, s) + (1-\beta)L^{Dice}(\hat{s}, s), \tag{5.9}$$

where $L^{CE}(\hat{s}, s) = -\frac{1}{N}\sum_{n=1}^{N} s_n \; log \; \hat{s_n} + (1-s_n) \; log \; (1-\hat{s_n})$ and $L^{Dice}(\hat{s}, s) = 1 - \frac{2\sum_{n=1}^{N}\hat{s_n}s_n}{\sum_{n=1}^{N}\hat{s_n}^2 + \sum_{n=1}^{N}s_n^2}$. Finally, β is a weighting factor balancing the contribution of each loss. In this work, we set $\beta = 0.5$.

5.2.3 Experiments

In our experiments, we evaluate the proposed MTLRS (Sec. 5.2.1) against other approaches which perform accelerated-MRI reconstruction and MRI segmentation without feature sharing. In a pre-trained approach, we train a reconstruction and a segmentation network separately and then use them independently at inference. In a sequential approach, we fine-tune the pre-trained segmentation network on the outputs of the reconstruction network. In an end-to-end approach, we train the two networks simultaneously but only compute a segmentation loss. In a joint approach, the two networks are trained with a joint reconstruction and segmentation loss (Eq. 5.6). The novelty of MTLRS lies in sharing

features between the reconstruction and the segmentation network. Through a sequence of cascades, the segmented output is concatenated with the output of the hidden layers to initialize the hidden layers of the subsequent cascade. In that way, MTLRS is informed by the outputs of both tasks, in addition to a joint loss. In a joint approach, the network is only informed by the joint loss.

In all these approaches, we choose the Cascades of Independently Recurrent Inference Machines (CIRIM) as the reconstruction network and the Attention-UNet [9] as the segmentation network, which empirically have been found to be well-performing models for each task. Additionally, we compare the performance of MTLRS with previously published methods. For this purpose, we implemented the end-to-end approach SEgmentation Recurrent Attention Network (SERANET) [3], and the joint approaches, RECSEGNET [6], Image Deep Structured Low-Rank (IDSLR) [5], and SEGNET [4]. All models were trained and tested on an Nvidia Tesla V100 GPU with 32GB memory. The hyperparameter settings for all methods can be found in the Appendix. The code is publicly available at https://github.com/wdika/mridc.

5.2.4 Dataset

A clinical dataset was used to train, validate, and test all methods using five-fold cross-validation. The dataset consisted of 3D FLAIR coil-combined magnitude brain images of 19 relapsing-remitting Multiple Sclerosis (MS) patients with white matter lesions. Data were acquired on a 3.0T scanner in our hospital. The local ethics review board approved this study, and the patients provided informed consent. Prospective undersampling was performed, accelerating imaging approximately 7.5 times under a Variable-Density Poisson disk distribution. Coil sensitivity maps were estimated using the caldir method of the BART toolbox (Uecker et al., 2015) on a fully-sampled reference.

The coil-combined magnitude images were used to synthesize multicoil complex data. To this end, we used a pre-trained CIRIM model trained only for reconstruction on 2D multislice FLAIR data [10], accelerated approximately eight times under a Variable-Density Poisson disk distribution. Minimal random gaussian noise was added to the synthetic data, with a relative weighting factor of $10^{-}5$. Data were then retrospectively accelerated by approximately 7.5 times under a Variable-Density Poisson disk distribution. Next, we used the reconstructed images to predict two segmentation classes, brain tissue (combined white and gray matter) and white matter lesions, as a reference standard for MRI segmentation. To obtain brain tissue segmentations, we used the statistical parametric mapping (SPM) toolbox [11]. To obtain white matter lesion segmentations, we used a pre-trained network for eye and tumor segmentation of retinoblastoma patients [12]. All segmentations were visually inspected and manually corrected when necessary to assure segmentation accuracy.

5.2.5 Evaluation

For evaluating reconstruction, we compute Structural Similarity Index Measure (SSIM) [13] and Peak Signal-to-Noise-Ratio (PSNR) on the normalized magnitude images between the synthesized ground truth x and the prediction $\hat{x}$. SSIM and PSNR are first computed per slice and per plane for each subject and then averaged to evaluate the reconstruction performance as a 3D volume. To evaluate segmentation, we calculate the Dice score as an overlap metric between the standard s and the prediction $\hat{s}$. Dice score is reported for the

combined (white and gray matter) tissue and white matter lesion segmentation and for only the white matter lesion segmentation. Dice scores are computed across all planes and slices for all subjects. To assess whether a correlation in performance between both tasks exists, we correlated SSIM and per lesion Dice scores using Spearman's rank test.

5.3 Results

Figure 5.2 shows an overall comparison, averaged over five-folds, of MTLRS to the Pre-Trained, Sequential, End-to-End, and Joint approaches. Note that the Sequential and the End-to-End approaches are optimized only for segmentation. MTLRS performed best on both reconstruction and segmentation. The Joint approach performed close to MTLRS but with a larger standard deviation. The Pre-Trained approach dropped in performance on both tasks, while it performed on par with the Sequential approach on segmentation, showing no apparent benefit when further optimizing the segmentation model on the reconstructed outputs. The End-to-End approach was the worst segmentation method, indicating the need for a joint loss rather than only segmentation loss.

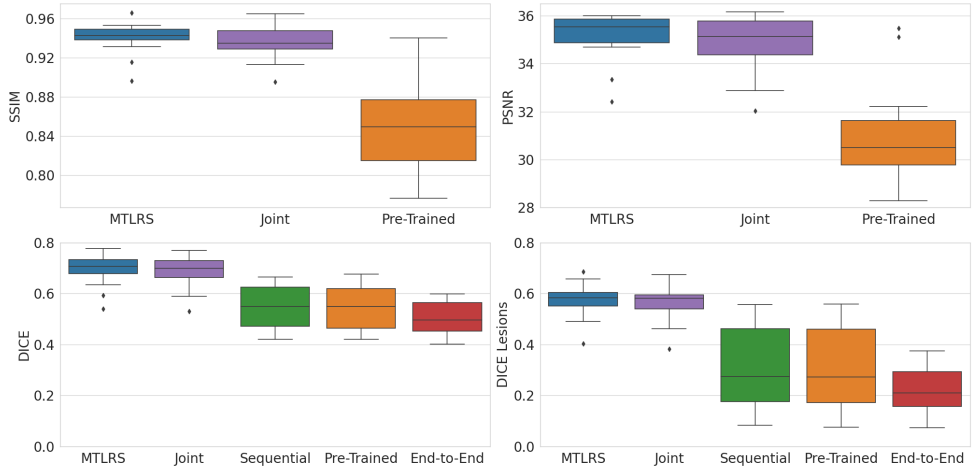


Figure 5.2: Quantitative evaluation averaged over five-folds when performing accelerated-MRI reconstruction and MRI segmentation with different approaches (x-axis). Data were retrospectively undersampled 7.5 times. SSIM and PSNR (top) evaluate reconstruction. DICE and DICE Lesions (bottom) evaluate segmentation.

Table 5.1 reports the performance of MTLRS and the evaluated previously published methods, averaged over five-folds. In both tasks, MTLRS outperformed the RECSEGNET, IDSLR, SEGNET, and SERANET, showing a clear advantage for the multitask approach. Overview tables reporting the performance of all approaches and previously published methods on each fold can be found in the Appendix.

Figure 5.3 shows an example of a reconstructed and segmented axial slice by MTLRS and the evaluated previously published methods. MTLRS provided the highest reconstruction quality (SSIM) and the most accurate lesion segmentation (Dice). The RECSEGNET performed comparably with MTLRS in reconstruction, while the IDSLR and SEGNET

reduced reconstruction performance further. The SERANET was the worst-performing method on reconstruction.

SSIM and lesions Dice scores were significantly correlated ($\rho = 0.92$, $p = 0.0005$). More examples of reconstructions and segmentations can be found in the Appendix.

Table 5.1: Overall comparison, averaged over five-folds, of MTLRS to previously published methods when performing accelerated-MRI reconstruction and MRI segmentation. SSIM and PSNR evaluate reconstruction. DICE and DICE Lesions evaluate segmentation. Metrics are computed on retrospectively undersampled data by 7.5 times. The arrow pointing upward indicates higher is better. Methods are sorted by DICE, while the best-performing method is shown in bold.

Method	SSIM ↑	PSNR ↑	DICE ↑	DICE Lesions ↑
MTLRS	**0.940 ± 0.017**	**35.26 ± 1.30**	**0.691 ± 0.065**	**0.574 ± 0.069**
RECSEGNET	0.787 ± 0.041	28.93 ± 0.99	0.512 ± 0.059	0.229 ± 0.086
SERANET			0.508 ± 0.063	0.221 ± 0.082
IDSLR	0.758 ± 0.034	27.31 ± 0.96	0.490 ± 0.054	0.186 ± 0.075
SEGNET	0.749 ± 0.039	27.07 ± 1.14	0.479 ± 0.056	0.178 ± 0.065

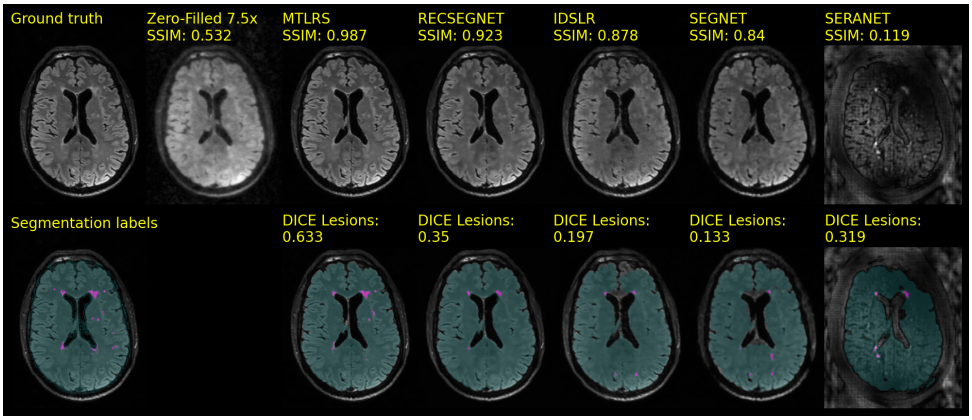


Figure 5.3: Reconstruction and segmentation of an axial slice with white matter lesions. An acceleration factor of approximately 7.5 was used to undersample the data retrospectively (top-second column). Methods are sorted by SSIM. SSIM is computed for evaluating reconstruction performance against the ground truth (top-first column). The per lesions Dice score is computed to evaluate segmentation performance against the reference labels (bottom-first column).

5.4 Discussion & Conclusion

We proposed MultiTask Learning for accelerated-MRI Reconstruction and Segmentation (MTLRS). MultiTask Learning was realized through a unique cascading network architecture consisting of a recurrent reconstruction network and segmentation network. The output of the hidden layers was combined with the segmented images to inform a sequence of

cascades, thus serving as an inductive bias. Performance was evaluated using five-fold cross-validation. MTLRS outperformed the Pre-Trained, Sequential, and End-to-End approaches and existing methods (RECSEGNET, SERANET, IDSLR, SEGNET) on reconstructing 7.5 times accelerated 3D FLAIR brain data of Multiple Sclerosis patients and on segmenting white matter lesions identified on this data. Additionally, it improved marginally upon the Joint approach. The reason could lie in the fact that the reconstruction network architecture used in MTLRS and the joint and pre-trained approaches was the same as the pre-trained network used in synthesizing the multicoil dataset. Therefore, future work will evaluate our method on a dataset where fully sampled reference data is available, e.g., knee data from the recently held KS-challenge [14]. Interestingly, a strong correlation was found between the quality metrics of both tasks. The results suggest that improved dealiasing during reconstruction leads to improved contrast and better-defined lesion boundaries, thereby supporting a more accurate segmentation. In future work, more tasks can be combined, such as classifying the underlying pathologies and improving performance by informing each other. Thus, MultiTask Learning is yet to be further explored, with potentially a high value if applied in the clinical setting, where aside from improving performance, the need for waiting time between multiple tasks would not be needed.

5.5 Acknowledgments

This publication is based on the STAIRS project under the TKI-PPP program. The collaboration project is co-funded by the PPP Allowance made available by Health Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships.

M.W.A. Caan is shareholder of Nico.Lab International Ltd.

We want to thank Hanneke E. Hulst for providing the dataset of the Multiple-Sclerosis patients, Samantha Noteboom for helping with the initial segmentation of the data, and Lysander de Jong for his contribution to assessing the reconstruction and segmentation as a joint problem.

References

[1] Rich Caruana. Multitask Learning. *Machine Learning*, 28:41–75, 1997.

[2] Jonas Adler, Sebastian Lunz, Olivier Verdier, Carola-Bibiane Schönlieb, and Ozan Öktem. Task adapted reconstruction for inverse problems. *Inverse Problems*, 38(7):075006, may 2022.

[3] Qiaoying Huang, Xiao Chen, Dimitris Metaxas, and Mariappan S. Nadar. Brain Segmentation from k-space with End-to-End Recurrent Attention Network. volume 11766 LNCS, pages 275–283. Springer Science and Business Media Deutschland GmbH, 2019.

[4] Liyan Sun, Zhiwen Fan, Xinghao Ding, Yue Huang, and John Paisley. Joint Cs-MRI Reconstruction and Segmentation with a Unified Deep Network. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11492 LNCS:492–504, 5 2019.

[5] Aniket Pramanik, Xiaodong Wu, and Mathews Jacob. Joint Calibrationless Reconstruction and Segmentation of Parallel MRI. 5 2021.

[6] Bin Sui, Jun Lv, Xiangrong Tong, Yan Li, and Chengyan Wang. Simultaneous image reconstruction and lesion segmentation in accelerated MRI using multitasking learning. *Medical Physics*, 48:7189–7198, 11 2021.

[7] Dimitrios Karkalousos, Samantha Noteboom, Hanneke E Hulst, Frans M Vos, and Matthan W A Caan. Assessment of data consistency through cascades of independently recurrent inference machines for fast and robust accelerated MRI reconstruction. *Physics in Medicine & Biology*, 5 2022.

[8] Marcin Andrychowicz, Misha Denil, Sergio Gómez Colmenarejo, Matthew W. Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando De Freitas. Learning to learn by gradient descent by gradient descent. *Advances in Neural Information Processing Systems*, pages 3988–3996, 2016.

[9] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention U-Net: Learning Where to Look for the Pancreas. 4 2018.

[10] Matthew J. Muckley, Bruno Riemenschneider, Alireza Radmanesh, Sunwoo Kim, Geunu Jeong, Jingyu Ko, Yohan Jun, Hyungseob Shin, Dosik Hwang, Mahmoud Mostapha, Simon Arberet, Dominik Nickel, Zaccharie Ramzi, Philippe Ciuciu, Jean-Luc Starck, Jonas Teuwen, Dimitrios Karkalousos, Chaoping Zhang, Anuroop Sriram, Zhengnan Huang, Nafissa Yakubova, Yvonne W. Lui, and Florian Knoll. Results of the 2020 fastmri challenge for machine learning mr image reconstruction. *IEEE Transactions on Medical Imaging*, 40(9):2306–2317, 2021.

[11] William Penny, Karl Friston, John Ashburner, Stefan Kiebel, and Thomas Nichols. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier, 2007.

[12] Victor I. J. Strijbis, Christiaan M. de Bloeme, Robin W. Jansen, Hamza Kebiri, Huu-Giao Nguyen, Marcus C. de Jong, Annette C. Moll, Merixtell Bach-Cuadra, Pim de Graaf, and Martijn D. Steenwijk. Multi-view convolutional neural networks for automated ocular structure and tumor segmentation in retinoblastoma. *Scientific Reports*, 11:14590, 12 2021.

[13] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13:600–612, 2004.

[14] Upasana Bharadwaj, Aniket Tolpadi, and Valentina Pedoia Pedoia.

Appendix

Hyperparameters

In our experiments, we set the hyperparameters of the related compared work according to what is reported by the authors in the original work.

For the CIRIM, we set the number of features to 64 for the convolutional and recurrent layers, cascades to 5, and recurrent iterations to 8. For the AttentionUNet, we set the number of features to 64, pooling layers to 2, and dropout to 0. For the SERANET, we chose the U-Net as the reconstruction network. We set the number of features to 32, pooling layers to 4, and dropout to 0 for the reconstruction, segmentation, and recurrent modules, and built three reconstruction blocks. For the RECSEGNET and the IDSLR, we set the number of features to 64, pooling layers to 2, and dropout to 0. For the IDSLR, the number of iterations was set to 5. α was set to 0.5 for the RECSEGNET and $10e-6$ for IDSLR. Finally, for the SEGNET, we set the number of features to 64, pooling layers to 2, dropout to 0, and cascades to 5.

For finding the optimal value for α in the joint loss for MTLRS (Eq. 5.6), we performed a hyperparameter search as presented in Figure 5.4. The tested values are $0.01, 0.1, 0.5, 0.9, 0.99$, going from favoring the reconstruction loss to balancing the loss to favoring the segmentation loss. The optimal α value was found to be 0.9.

We used ADAM as optimizer for all methods and set the learning rate to 10^{-4}.

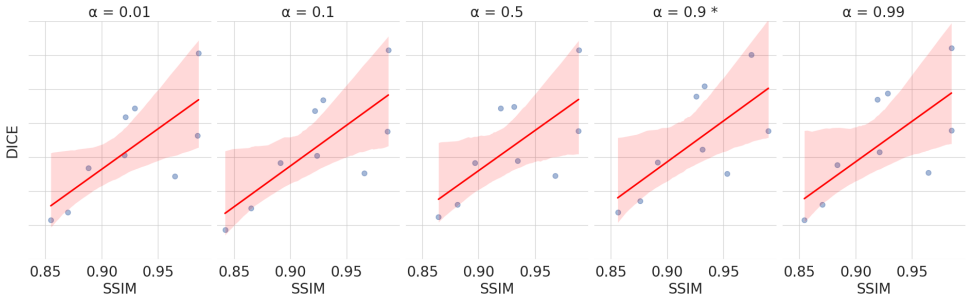


Figure 5.4: Hyperparameter search for finding the optimal α value in the joint loss (Eq. 5.6). Reconstruction and segmentation performance are realized on an SSIM (x-axis) over Dice (y-axis) plot. From left to right. The $*$ indicates the value resulting in the best SSIM & Dice scores.

Overview five-fold cross-validation

Tables 5.2 and 5.3 report performance on each of the five folds of the cross-validation, of MTLRS and all compared approaches and previously published methods when performing accelerated-MRI reconstruction and MRI segmentation. MTLRS was the best-performing method overall on all folds and all metrics. Only on the second fold and on PSNR the Joint approach scored higher than MTLRS.

Table 5.2: Overall performance of MTLRS and the Pre-Trained, Sequential, End-to-End, and Joint approaches for five-fold cross-validation when performing accelerated-MRI reconstruction and MRI segmentation. SSIM and PSNR evaluate reconstruction. DICE and DICE Lesions evaluate segmentation. Metrics are computed on retrospectively undersampled data by 7.5 times. The arrow pointing upward indicates higher is better. Methods are sorted by DICE, while the best-performing method is shown in bold.

Method	SSIM ↑	PSNR ↑	Dice ↑	Dice Lesions ↑
		Fold 1		
MTLRS	**0.936 ± 0.036**	**34.94 ± 3.50**	**0.656 ± 0.094**	**0.511 ± 0.098**
Joint	**0.936 ± 0.037**	34.79 ± 3.46	0.651 ± 0.096	0.505 ± 0.105
Pre-Trained	0.821 ± 0.090	29.84 ± 3.33	0.475 ± 0.111	0.197 ± 0.172
End-to-End			0.474 ± 0.075	0.178 ± 0.110
Sequential			0.473 ± 0.112	0.178 ± 0.110
		Fold 2		
MTLRS	**0.961 ± 0.025**	36.69 ± 2.75	**0.706 ± 0.074**	**0.588 ± 0.091**
Joint	0.959 ± 0.034	**37.13 ± 3.59**	0.704 ± 0.073	0.587 ± 0.091
Sequential			0.604 ± 0.150	0.407 ± 0.270
Pre-Trained	0.882 ± 0.075	31.43 ± 2.85	0.601 ± 0.151	0.400 ± 0.272
End-to-End			0.555 ± 0.138	0.319 ± 0.236
		Fold 3		
MTLRS	**0.944 ± 0.027**	**35.72 ± 3.20**	**0.677 ± 0.090**	**0.558 ± 0.107**
Joint	0.933 ± 0.033	34.96 ± 3.20	0.664 ± 0.095	0.528 ± 0.120
Pre-Trained	0.838 ± 0.067	30.65 ± 2.80	0.480 ± 0.147	0.211 ± 0.246
Sequential			0.489 ± 0.141	0.220 ± 0.238
End-to-End			0.487 ± 0.111	0.204 ± 0.180
		Fold 4		
MTLRS	**0.940 ± 0.032**	**35.26 ± 3.40**	**0.707 ± 0.051**	**0.572 ± 0.070**
Joint	0.937 ± 0.035	34.97 ± 3.67	0.697 ± 0.059	0.552 ± 0.080
Sequential			0.506 ± 0.093	0.222 ± 0.184
Pre-Trained	0.814 ± 0.080	29.63 ± 2.95	0.500 ± 0.094	0.219 ± 0.182
End-to-End			0.495 ± 0.083	0.201 ± 0.155
		Fold 5		
MTLRS	**0.923 ± 0.068**	**34.18 ± 4.57**	**0.654 ± 0.066**	**0.570 ± 0.085**
Pre-Trained	0.918 ± 0.064	33.82 ± 4.52	0.636 ± 0.087	0.540 ± 0.112
Joint	0.915 ± 0.061	33.64 ± 4.52	0.646 ± 0.070	0.562 ± 0.082
Sequential			0.634 ± 0.080	0.542 ± 0.114
End-to-End			0.490 ± 0.104	0.233 ± 0.153

Table 5.3: Overall performance of MTLRS and previously published methods for five-fold cross-validation when performing accelerated-MRI reconstruction and MRI segmentation. SSIM and PSNR evaluate reconstruction. DICE and DICE Lesions evaluate segmentation. Metrics are computed on retrospectively undersampled data by 7.5 times. The arrow pointing upward indicates higher is better. Methods are sorted by DICE, while the best-performing method is shown in bold

Method	SSIM ↑	PSNR ↑	Dice ↑	Dice Lesions ↑
		Fold 1		
MTLRS	**0.936 ± 0.036**	**34.94 ± 3.50**	**0.656 ± 0.094**	**0.511 ± 0.098**
RECSEGNET	0.781 ± 0.088	28.61 ± 2.51	0.481 ± 0.081	0.175 ± 0.120
SERANET			0.472 ± 0.063	0.160 ± 0.076
SEGNET	0.761 ± 0.072	27.47 ± 2.10	0.457 ± 0.066	0.129 ± 0.074
IDSLR	0.760 ± 0.075	27.48 ± 1.97	0.457 ± 0.067	0.129 ± 0.067
		Fold 2		
MTLRS	**0.961 ± 0.025**	**36.69± 2.75**	**0.706 ± 0.074**	**0.588 ± 0.091**
RECSEGNET	0.850 ± 0.079	30.29 ± 2.41	0.568 ± 0.128	0.322 ± 0.238
SERANET			0.548 ± 0.109	0.280 ± 0.200
IDSLR	0.808 ± 0.063	27.51 ± 2.24	0.526 ± 0.111	0.242 ± 0.204
SEGNET	0.791 ± 0.068	27.03 ± 2.33	0.505 ± 0.091	0.213 ± 0.150
		Fold 3		
MTLRS	**0.944 ± 0.027**	**35.72 ± 3.20**	**0.677 ± 0.090**	**0.558 ± 0.107**
SERANET			0.495 ± 0.096	0.202 ± 0.151
RECSEGNET	0.792 ± 0.072	29.40 ± 2.20	0.492 ± 0.110	0.199 ± 0.185
SEGNET	0.756 ± 0.067	27.86 ± 1.76	0.462 ± 0.096	0.157 ± 0.139
IDSLR	0.755 ± 0.068	27.83 ± 1.83	0.455 ± 0.093	0.137 ± 0.129
		Fold 4		
MTLRS	**0.940 ± 0.032**	**35.26 ± 3.40**	**0.707 ± 0.051**	**0.572 ± 0.070**
SERANET			0.526 ± 0.049	0.226 ± 0.095
RECSEGNET	0.773 ± 0.080	28.58 ± 2.33	0.493 ± 0.059	0.186 ± 0.118
SEGNET	0.747 ± 0.067	27.46 ± 2.01	0.477 ± 0.038	0.150 ± 0.071
IDSLR	0.744 ± 0.075	27.46 ± 2.04	0.480 ± 0.047	0.147 ± 0.096
		Fold 5		
MTLRS	**0.923 ± 0.068**	**34.18 ± 4.57**	**0.654 ± 0.066**	**0.570 ± 0.085**
RECSEGNET	0.749 ± 0.122	27.94 ± 3.21	0.507 ± 0.120	0.268 ± 0.205
IDSLR	0.719 ± 0.107	26.24 ± 2.32	0.492 ± 0.110	0.243 ± 0.188
SERANET			0.490 ± 0.104	0.233 ± 0.153
SEGNET	0.690 ± 0.103	25.48 ± 2.26	0.454 ± 0.111	0.198 ± 0.164

Overview examples

Figures 5.5 and 5.6 show examples of reconstructed and segmented slices of the coronal and sagittal view, with white matter lesions identified in all slices. An acceleration factor

of approximately 7.5 was used to undersample the data retrospectively. MTLRS provided the highest reconstruction quality (SSIM) and the most accurate lesions segmentation (Dice) in all cases. The RECSEGNET dropped significantly both in SSIM and Dice score by oversimplifying the reconstruction and slightly overestimating lesion volume. The same behavior is observed by the IDSLR and the SEGNET, reducing performance further. The SERANET performed poorly on reconstruction, while segmentation performance was comparable or better to the RECSEGNET, IDSLR, and SEGNET.

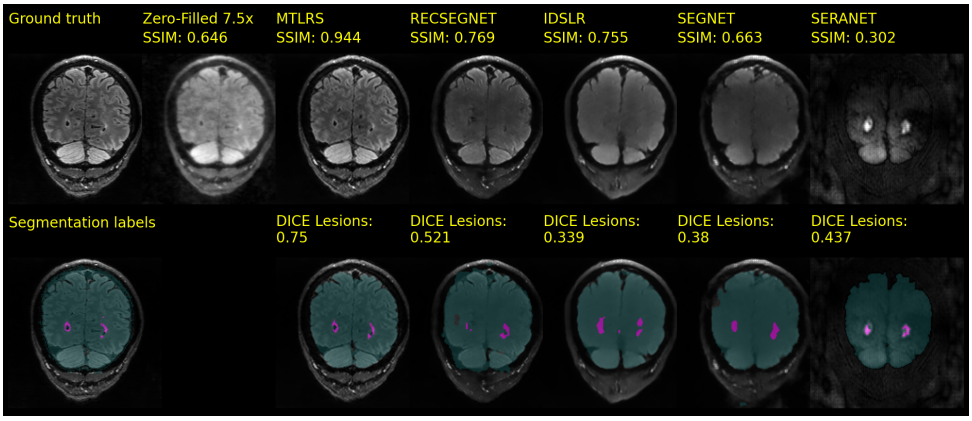


Figure 5.5: Reconstruction and segmentation of a coronal slice with white matter lesions. An acceleration factor of approximately 7.5 was used to undersample the data retrospectively (top-second column). Methods are sorted by SSIM. SSIM is computed for evaluating reconstruction performance against the ground truth (top-first column). The per lesions Dice score is computed to evaluate segmentation performance against the reference labels (bottom-first column).

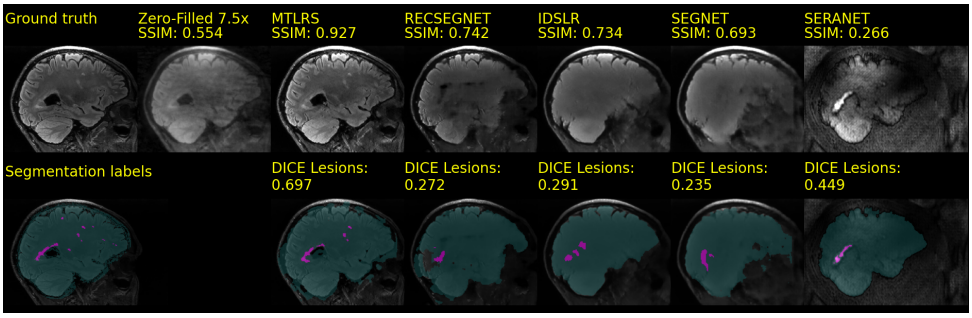


Figure 5.6: Reconstruction and segmentation of a sagittal slice with white matter lesions. An acceleration factor of approximately 7.5 was used to undersample the data retrospectively (top-second column). Methods are sorted by SSIM. SSIM is computed for evaluating reconstruction performance against the ground truth (top-first column). The per lesions Dice score is computed to evaluate segmentation performance against the reference labels (bottom-first column).

6

ATOMMIC: An Advanced Toolbox for Multitask Medical Imaging Consistency to facilitate Artificial Intelligence applications from acquisition to analysis in Magnetic Resonance Imaging

Karkalousos, D., Išgum, I., Marquering, H. A., & Caan, M.W.A. (2024).
Computer Methods and Programs in Biomedicine, 256, 108377.
DOI: 10.1016/j.cmpb.2024.108377

Abstract

Background and Objectives: *Artificial intelligence (AI) is revolutionizing Magnetic Resonance Imaging (MRI) along the acquisition and processing chain. Advanced AI frameworks have been applied in various successive tasks, such as image reconstruction, quantitative parameter map estimation, and image segmentation. However, existing frameworks are often designed to perform tasks independently of each other or are focused on specific models or single datasets, limiting generalization. This work introduces the Advanced Toolbox for Multitask Medical Imaging Consistency (ATOMMIC), a novel open-source toolbox that streamlines AI applications for accelerated MRI reconstruction and analysis. ATOMMIC implements several tasks using deep learning (DL) models and enables MultiTask Learning (MTL) to perform related tasks in an integrated manner, targeting generalization in the MRI domain.*

Methods: *We conducted a comprehensive literature review and analyzed 12,479 GitHub repositories to assess the current landscape of AI frameworks for MRI. Subsequently, we demonstrate how ATOMMIC standardizes workflows and improves data interoperability, enabling effective benchmarking of various DL models across MRI tasks and datasets. To showcase ATOMMIC's capabilities, we evaluated twenty-five DL models on eight publicly available datasets, focusing on accelerated MRI reconstruction, segmentation, quantitative parameter map estimation, and joint accelerated MRI reconstruction and segmentation using MTL.*

Results: *ATOMMIC's high-performance training and testing capabilities, utilizing multiple GPUs and mixed precision support, enable efficient benchmarking of multiple models across various tasks. The framework's modular architecture implements each task through a collection of data loaders, models, loss functions, evaluation metrics, and pre-processing transformations, facilitating seamless integration of new tasks, datasets, and models. Our findings demonstrate that ATOMMIC supports MTL for multiple MRI tasks with harmonized complex-valued and real-valued data support while maintaining active development and documentation. Task-specific evaluations demonstrate that physics-based models outperform other approaches in reconstructing highly accelerated acquisitions. These high-quality reconstruction models also show superior accuracy in estimating quantitative parameter maps. Furthermore, when combining high-performing reconstruction models with robust segmentation networks through MTL, performance is improved in both tasks.*

Conclusions: *ATOMMIC advances MRI reconstruction and analysis by leveraging MTL and ensuring consistency across tasks, models, and datasets. This comprehensive framework serves as a versatile platform for researchers to use existing AI methods and develop new approaches in medical imaging.*

6.1 Introduction

In recent years, Artificial Intelligence (AI) has led to significant advancements in medical imaging, spanning various tasks along the acquisition and processing chain. Deep Learning (DL) segmentation Convolutional Neural Networks (CNNs) enable fast and accurate segmentation of anatomy and pathology in Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) ([1–4]). DL models have also been developed to act as inverse problem solvers, improving the reconstruction quality in MRI and CT ([5]). In MRI specifically,

scanning time can be reduced considerably by accelerating the acquisition process, while DL reconstruction models offer high-quality images by training on the sparse signal representation ([6–8]). Similarly DL models can accurately estimate quantitative parameter maps from multiple accelerated MRI acquisitions while varying sequence parameters ([9]).

MultiTask Learning (MTL) can further improve the performance of models performing individual but related tasks by combining and jointly performing them ([10]). For instance, the segmentation efficacy depends on the reconstruction quality, as the former task invariably follows the latter. Therefore, since tasks are related, if they are performed simultaneously, performance can be improved on both while reducing the overhead time of performing the tasks separately. Although MTL has been successfully applied to combine reconstruction and segmentation ([11–14]), the challenge lies in maintaining consistency in performance while merging multiple single-task DL models and harmonizing data support for both complex-valued and real-valued domains. Dedicated medical imaging AI frameworks are usually employed to address the issue of regularization across tasks and data types. Primarily, such frameworks are task-specific, focusing on essential tasks like reconstruction ([15, 16]) and segmentation ([1, 17, 18]), or they are modality-specific ([19–21]), or focus on data pre-processing and data augmentations ([22]). The Medical Open Network for Artificial Intelligence (MONAI) ([23]) is a popular framework that supports multiple tasks, modalities, and data types. However, tasks can only be performed independently, and complex-valued data support is limited to the reconstruction task.

Nevertheless, integrating multiple data types support or multiple methods for MTL in existing frameworks can be complicated due to differences in data structures, formats, and programming languages, increasing the burden for researchers who need a more comprehensive range of options for medical image analysis. To address these inconsistencies, we present the Advanced Toolbox for Multitask Medical Imaging Consistency (ATOMMIC), an open-source toolbox supporting multiple independent MRI tasks, currently being reconstruction, segmentation, and quantitative parameter map estimation, and uniquely integrates MTL, aiming to streamline the application of DL in MRI tasks from reconstruction to analysis, offering a unified platform for advancing AI-driven medical imaging research.

ATOMMIC offers several key advantages as a comprehensive AI toolbox for MR Imaging. It provides standardized workflows for efficient training and benchmarking of a wide range of DL models, offering versatility across applications. The toolbox implements over 25 state-of-the-art DL models for various MRI tasks, enabling researchers to rapidly experiment multiple approaches with simple configuration steps. ATOMMIC uniquely harmonizes complex-valued and real-valued data support, ensuring consistency across tasks, DL models, datasets, and training and testing schemes. The toolbox's architecture is based on task-specific collections of data loaders, models, loss functions, evaluation metrics, and pre-processing transformations. This modular design facilitates easy integration of new datasets, models, and tasks. For example, the MTL collection builds upon independent tasks, extending them to combine related or subsequent processes, such as joint reconstruction and segmentation or reconstruction and quantitative parameter map estimation.

Beyond support of multiple MRI tasks, multi-task learning, availability of a wide range of DL models, support for different data types, and standardized workflows, ATOMMIC also provides various pre-processing transformations (Sec. 6.3.3) and undersampling schemes

(Sec 6.3.2) tailored to MR imaging research. The former include noise pre-whitening, coil compression, coil sensitivity map estimation, zero-filling, domain-specific cropping, various normalization options, and custom fast Fourier transformations. Additional features, such as motion simulation and signal-to-noise ratio estimation, are also available. Regarding the latter, the toolbox supports both prospectively and retrospectively undersampled data, with various 1D and 2D masking options using different noise distributions, such as equispaced, Gaussian, random, and Poisson, while partial Fourier sampling is also available to simulate realistic scanning scenarios. Furthermore, ATOMMIC offers a diverse range of loss functions for effective model training, including mean squared error (MSE), L1, structural similarity (SSIM), noise-aware, Wasserstein ([24]), cross-entropy, and DICE losses. It also implements popular unsupervised and self-supervised methods in MRI using DL, such as Noise-to-Recon ([25]) and self-supervised data undersampling ([26]). Essential evaluation metrics are provided for thorough model performance analysis across different tasks. Model training capabilities are further enhanced in ATOMMIC by including ten optimizers and fourteen learning rate schedulers. Advanced features such as exponential moving average, early stopping criterion, hyperparameter optimization, layer freezing, and export options to TensorBoard and Weights & Biases allow for tailored model training and easy monitoring (Sec 6.3.4). Importantly, ATOMMIC is built according to NVIDIA's NeMO ([27]), a computationally efficient conversational AI toolkit that allows for high-performance training and testing using multiple GPUs, multiple nodes, and mixed precision support.

To facilitate ease of use for researchers at various experience levels, ATOMMIC includes extensive documentation and several examples using publicly available datasets, including the Amsterdam ultra-high field adult lifespan dataset ([9]), the Calgary Campinas 359 dataset ([28]), the fastMRI brains multicoil and knees multicoil and singlecoil datasets ([29]), the Stanford fully sampled 3D Fast Spin Echo knee dataset ([30]), the brain tumor segmentation 2023 adult glioma challenge dataset ([31]), the ischemic stroke lesion segmentation 2022 challenge dataset ([32]), and the Stanford knee MRI with multi-task evaluation dataset ([33]).

In the following sections, we first explore the landscape of AI frameworks for medical imaging (Sec. 6.2) through a thorough literature search and, additionally, parsing GitHub repositories, showcasing the need for multitasking toolboxes that support multiple tasks and data types with detailed documentation and up-to-date maintenance. Next, we introduce ATOMMIC's main components (Sec. 6.3). To demonstrate ATOMMIC's capabilities, we conduct an extensive evaluation of twenty-five DL models across eight publicly available datasets, encompassing both brain and knee anatomies. This evaluation spans various MRI tasks, including accelerated reconstruction, quantitative parameter map estimation, segmentation, and MTL and assessing multiple undersampling schemes and acceleration factors. Concurrently, we evaluate the accuracy of segmentation for brain lesions, tumors, and knee pathologies in both standalone segmentation and MTL contexts, illustrating ATOMMIC's applicability across diverse scenarios (Sec. 6.4). Finally, in Sec. 6.5, we discuss how ATOMMIC aims to provide a multitask toolbox for the research community to use, develop, and share models and potentially datasets and pre-processing pipelines across various MRI tasks, targeting generalization in the MRI domain.

The datasets used in the experiments (Sec. 6.3.5) are publicly accessible, while pre-

processing pipelines, detailed API documentation, tutorials, and quick start guides are available on the open-source ATOMMIC repository[1], under Apache 2.0 license. Trained models' checkpoints are available on HuggingFace[2], allowing for full reproducibility of the results.

6.2 Related work

Several AI frameworks for MR Imaging have been developed over the past years. However, they are often limited to specific tasks ([1, 15–18, 23]) or are modality-specific ([19–22]). Most frameworks do not support MTL or unified complex/real-valued data handling across tasks. Additionally, comprehensive benchmarking of multiple models is not commonly supported. To demonstrate the landscape of available AI frameworks for MR Imaging, including their key advantages and limitations, we performed a thorough literature search.

Recognizing that the emphasis in research is often placed on the model implementation or dataset specifics rather than on frameworks, we extended our literature search of AI frameworks for MRI to include GitHub repositories, not limited to those published in scientific papers. Utilizing keywords such as 'medical-image-processing', 'medical-imaging', 'MRI', 'medical', 'MRI-reconstruction', 'MRI-segmentation', 'neuroimaging', 'nifti', 'dicom', 'compressed-sensing', 'image-reconstruction', 'brain', 'medical-image-analysis', and 'MRI-registration', we identified a total of 12,479 repositories. Removing duplicates and non-existent URLs resulted in 10,747 repositories. Next, we defined a minimum usage threshold based on the number of stars, where a star serves as a popularity and usage metric on GitHub. The minimum number of stars was 2, the maximum was 23,400, and the median was 13. Limiting our results to repositories with at least ten stars returned 3,623 repositories. We meticulously narrowed this list to 68 DL frameworks pertinent to MRI. In brief, we removed repositories irrelevant to MRI, not written in English, and containing data and file converters only. Furthermore, we discarded Graphic User Interfaces, specific model and paper implementations, theses, lab pages, and courses. A detailed list of the repositories, including URLs, is available on GitHub[3]. This comprehensive review highlights a gap in frameworks supporting MultiTask Learning for MRI, complex-valued data support, providing documentation, and up-to-date maintenance, as shown in Fig. 6.1.

Among the 68 AI frameworks for MRI identified, ATOMMIC and MONAI were notable for their up-to-date maintenance, detailed documentation, and support for multiple independent tasks. However, as shown in Fig. 6.1, ATOMMIC emerged as the only toolbox supporting MTL with harmonized complex-valued and real-valued data support, comprehensive documentation, and up-to-date maintenance.

[1] https://github.com/wdika/atommic
[2] https://huggingface.co/wdika
[3] https://github.com/wdika/atommic

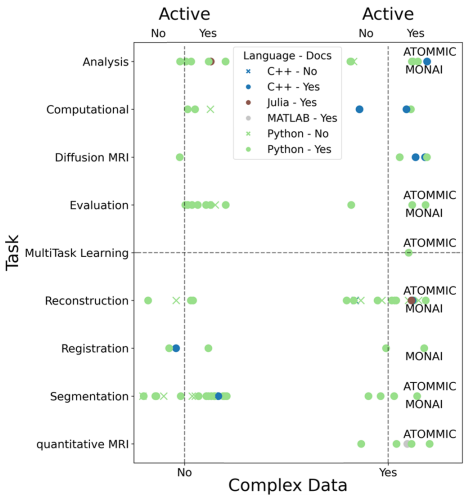


Figure 6.1: Overview of AI repositories for MRI tasks parsed from GitHub. The repositories are divided into two groups on the x-axis: those that do not support complex-valued data (left group) and those that do support complex-valued data (right group). The repositories are further categorized based on their activity level, on the top of the x-axis, split by vertical dashed lines: those that have committed updates within the last year (2023) are labeled as "Active Yes" (right side), and those that have not as "Active No" (left side). The supported tasks are depicted on the y-axis, while the horizontal dashed line showcases the multitasking toolboxes. Repositories are visualized as dots if documentation is available (Docs Yes) and cross marks if documentation is unavailable (Docs No). Each color signifies the language of the repository, with blue representing C++, brown representing Julia, gray representing MATLAB, and green representing Python.

6.3 Methods

This section presents an overview of the MRI tasks supported in ATOMMIC, including accelerated MRI reconstruction, segmentation, quantitative parameter map estimation, and MTL for joint reconstruction and segmentation. Furthermore, we describe the process of training and testing DL models in ATOMMIC and showcase the available pre-processing transformations. A schematic overview summarizing ATOMMIC's features and workflow is included for enhanced comprehension (Fig. 6.2). Finally, we present benchmarks and use cases to showcase the toolbox's advantages and capabilities.

6.3.1 MRI tasks

Starting from the acquisition process, the forward model of acquiring and accelerating MRI data can be expressed as follows:

$$y_i = P \cdot \mathcal{F}\left(S_i^H \odot x_i\right) + \sigma_i, i = 1, ..., N, \tag{6.1}$$

where x_i represents fully sampled multicoil complex-valued data for N total coils. S_i denotes the coil sensitivity maps, which homogenize the spatial intensities, H is the Hermitian complex conjugate, and $\odot$ is the Hadamard product. $\mathcal{F}$ is the Fourier transform, projecting the data onto the frequency domain, known as k-space in MRI. P is the undersampling scheme, which accelerates imaging by reducing the amount of data needed to acquire, and

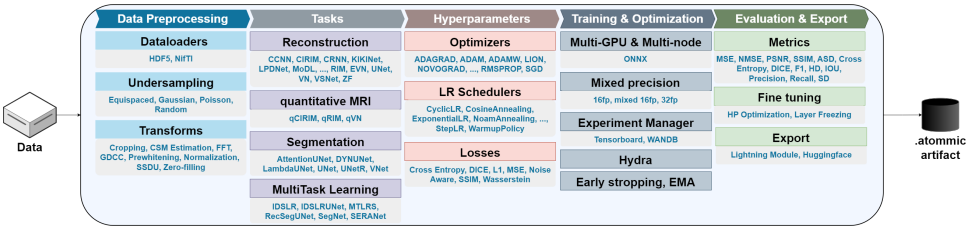


Figure 6.2: Schematic overview of ATOMMIC. Starting from the left to the right, MRI data are given as input. Next, configurations such as dataloaders, undersampling schemes, transforms, task(s) and models, optimizers, learning rate schedulers, losses, training and optimization settings, evaluation metrics, and exports are defined. The output is an atommic artifact (rightmost) containing the trained model's checkpoints and configurations, which can be directly used for inference on new datasets.

σ_i represents the noise inherent in the acquisition process. The resulting undersampled multicoil data are denoted by y_i.

While the forward process is well-defined, reconstructing high-quality images from undersampled data or finding the inverse mapping process $y \mapsto x$ remains challenging. The inverse problem of accelerated MRI reconstruction can be solved through a Bayesian estimation. The goal is to maximize the posterior distribution of y given x and the prior probability of x. This process is known as the Maximum A Posteriori (MAP) estimation and can be described as:

$$x_{\text{MAP}} = \arg\max_x \left(\log p(y|x) + \log p(x)\right). \tag{6.2}$$

Substituting Eq. 6.1 into Eq. 6.2 transforms the inverse problem into a minimization problem:

$$x = \arg\min_x \left\{ \sum_{i=1}^{N} \theta\left(y_i, P \cdot F\left(S_i^H \odot x_i\right) + \sigma_i\right) + \lambda R(x) \right\}, \tag{6.3}$$

where θ represents the discrepancy between the undersampled measurements and their predictions. The summation indicates that the multicoil data are transformed into a coil-combined image. λ denotes the weighting factor for the regularizer R. The regularizer can be modeled using a neural network.

Following accelerated MRI reconstruction by solving the inverse problem in Eq. 6.3, the task of estimating quantitative parameter maps can be expressed when data from multiple acquisitions with varying sequence parameters are available. The Multiple Echo Recombined Gradient Echo (ME-GRE) sequence varies the echo time (TE), such that the apparent transverse relaxation rate (R_2^*) may be computed from repeated acquisitions. The forward relaxation model describes the acquisition process for multiple TEs as

$$x_t = M \odot e^{-\text{TE}_t\left(R_2^* - B_0 i\right)}, \tag{6.4}$$

where M is the net magnetization, t denotes a single echo time, and B_0 is the off-resonance of the static magnetic field, and i notates complex-valued data. Inserting the forward relaxation model (Eq. 6.4) into the forward model of accelerated MRI acquisition (Eq. 6.1)

results in a unified quantitative MRI forward model

$$y_{t,i} = P \cdot \mathcal{F}\left(S_i^H \odot \left(M \odot e^{-\mathrm{TE}_t\left(R_2^* - B_0\right)}\right)\right) + \sigma_i. \tag{6.5}$$

The resulting parameter maps follow from minimizing Eq. 6.3.

In MultiTask Learning (MTL), multiple related tasks are combined and performed simultaneously instead of individually, aiming to identify relationships, leading to better generalization and enhancing the performance of each task ([34]). For example, reconstruction can be combined with subsequent tasks, such as quantitative parameter map estimation or segmentation, by modeling the regularizer R in Eq. 6.3 with a neural network and approximating x through an iterative training scheme. The predicted reconstruction, $\hat{x}$, is given as input to the subsequent task-specific network during each iteration. In the case of MTL for reconstruction and quantitative parameter map estimation, $\hat{x}$ is inserted into Eq. 6.5. When combining reconstruction with segmentation, $\hat{x}$ is mapped onto delineated anatomical structures. Features are shared, effectively acting as inductive bias for all tasks using a joint loss function ([11–14]).

6.3.2 Undersampling MRI

Data undersampling, as described in Eq. 6.1 by the undersampling mask P, is crucial to accelerate the acquisition process by partially sampling or sub-sampling the k-space. Prospective undersampling refers to accelerating imaging during the data acquisition phase. Retrospective undersampling refers to generating undersampling masks post-acquisition and applying them to fully sampled data, usually for research purposes. ATOMMIC supports both prospective and retrospective undersampling. Each undersampling scheme is implemented in a respective class as follows.

For equispaced 1D undersampling, the `Equispaced1DMaskFunc` class is utilized to generate a mask with evenly spaced lines in the fully sampled k-space ([35]). A number of fully sampled low frequencies in the center of k-space is defined as $N_{\text{low\_freqs}} = (N \cdot$ `center_fractions`$)$, where N is the size of the k-space and `center_fractions` is a parameter that can be adjusted. The chosen `accelerations` define the resulting undersampling rate, equal to $\frac{N}{\texttt{accelerations}}$. For 2D equispaced undersampling, the `Equispaced2DMaskFunc` class provides similar functionality. For a more randomized approach, the `Random1DMaskFunc` class allows for 1D undersampling with random spacing of the sampled k-space lines.

In the case of Gaussian density weighted undersampling, the `Gaussian1DMaskFunc` class generates a Gaussian 1D mask. Data points are sampled based on the probability density function of the Gaussian distribution. The half-axes of the ellipse are set to the `center_scale` percentage of the fully sampled region. The peripheral points are randomly sampled according to a Gaussian probability density function. Here, the `center_fractions` equivalent is the Full-Width at Half-Maximum (FWHM). Similarly, the `Gaussian2DMaskFunc` class allows Gaussian 2D undersampling, where data points near the center of the k-space are fully sampled within an ellipse. The half-axes of the ellipse are set to the `center_scale` percentage of the fully sampled region. The `Poisson2DMaskFunc` class allows for non-random sampling, generating a 2D mask following a Variable-Density Poisson-disc sampling pattern.

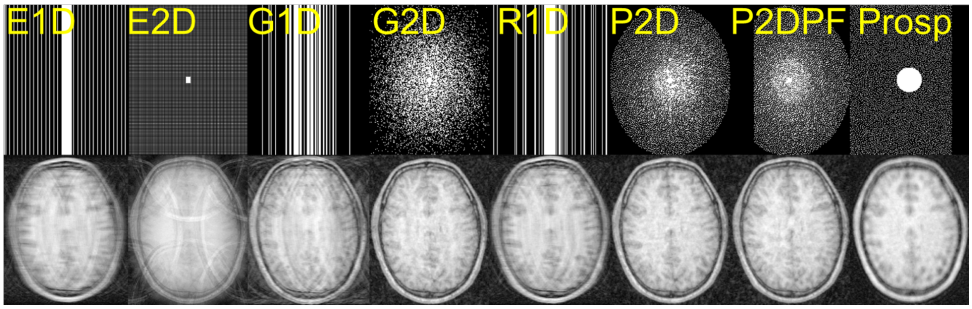


Figure 6.3: Overview of undersampling options in ATOMMIC. From left to right, columns one to seven present retrospective undersampling using Equispaced 1D (E1D), Equispaced 2D (E2D), Gaussian 1D (G1D), Gaussian 2D (G2D), Random 1D (R1D), Poisson 2D (P2D), and Poisson 2D with 20% Partial Fourier (P2DPF) masking, respectively. Note that Partial Fourier can be applied to any masking. The last column presents prospective undersampling (Prosp) using the Calgary-Campinas 359 dataset default 2D Poisson mask ([28]).

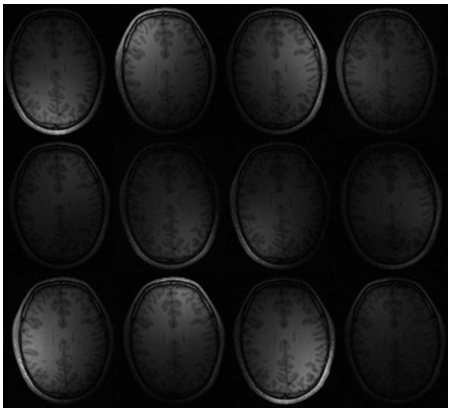
The `partial_fourier` parameter sets a percentage of outer k-space that is not sampled, resulting in a partially sampled k-space. An illustrative overview of the undersampling options is provided in Fig. 6.3.

6.3.3 MRI transforms

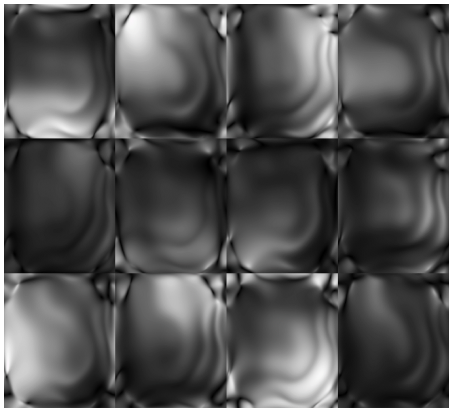
MRI transforms in ATOMMIC refer to pre-processing, i.e., data augmentations, intensity normalization, and multicoil-related transforms. The following transforms are implemented to handle both complex-valued and real-valued data for any task (Fig. 6.2).

The `NoisePreWhitening` class ensures that the inherent noise in the acquisition process, represented by σ in Eq. 6.1, will be independent and identically distributed by applying noise pre-whitening and decoupling or decorrelating coil signals [36]. While MRI acquisitions commonly include separate noise measurements, such information is only sometimes exported. When this information is unavailable, the physical properties are modeled, assuming that the periphery of k-space is dominated by noise, such that a patch can be defined to measure the noise level. Its size can be set manually with the `prewhitening_patch_start` and `prewhitening_patch_length` parameters or automatically by toggling the `find_patch_size` parameter. Alternatively, if the actual noise level is already measured and is available, it can be given as input instead of measuring the noise level with a patch. Also, the `scale_factor` parameter is used for setting an adequate noise bandwidth in outer k-space. A noise tensor is composed over all coil elements and multiplied by its conjugate transpose. Finally, Cholesky decomposition is performed, effectively minimizing noise correlation.

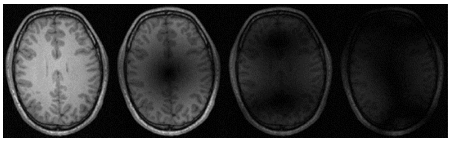
The `GeometricDecompositionCoilCompression` class can compress multicoil data using the geometric decomposition method ([37]). The `gcc_virtual_coilss` parameter defines the number of virtual coils to compress the multicoil data to. The `gcc_calib_lines` parameter is the number of calibration lines used for coil compression. The `gcc_align_data` parameter aligns the data before coil compression. An example of compressing 12-coil data to 4-virtual-coil data, with the `GeometricDecompositionCoilCompression` transformation is presented in Fig. 6.4c.



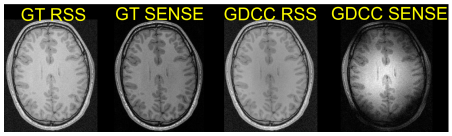
(a) Ground Truth (GT)



(b) Coil Sensitivity Maps (CSM)



(c) Geometric Decomposition Coil Compression (GDCC)



(d) Coil Combination Methods (CCM)

Figure 6.4: Multicoil-related transforms applied to example data from the CC359 dataset ([28]). Fig. 6.4a shows the fully sampled 12-coil Ground Truth (GT) data. Fig. 6.4b shows the estimated 12-coil Coil Sensitivity Maps (CSM) using the `EstimateCoilSensitivityMaps` class. In Fig. 6.4c, the 12-coil data are reduced to 4 coils after using the Geometric Decomposition Coil Compression (GDCC) method. Fig. 6.4d shows different Coil Combination Methods (CSM), such as the Root-Sum-of-Squares (RSS) and the SENsitivity Encoding (SENSE), applied to the Ground Truth (GT) (first and second, respectively), and the GDCC (third and fourth, respectively).

Cropping in both the image space and k-space may be performed using the `Cropper` class. Note that when cropping is applied in k-space, the Field-of-View (FOV) changes as a result. When applied in image space, the FOV remains the same while the spatial resolution changes. The `kspace_crop` parameter defines whether the cropping is performed in k-space or image space. The `crop_before_masking` parameter defines whether cropping will be applied before or after undersampling the k-space (Sec. 6.3.2). Note that cropping after undersampling alters the relative acceleration factor.

The `coil_combination_method` function allows to perform either Root-Sum-of-Squares (RSS) or SENSE ([38]) coil-combination (Fig. 6.4d). In the case of SENSE, coil sensitivity maps need to be available. The `EstimateCoilSensitivityMaps` class allows estimating coil sensitivity maps on the fly without needing to pre-compute and store them beforehand (Fig. 6.4b). The available options are adjusted according to the DIRECT toolkit ([16]) and include `ESPIRIT` ([39]), Root-Sum-of-Squares (RSS), and unitary methods. Also, the option of estimating coil sensitivity maps using a neural network (UNet) is available by toggling the `estimate_coil_sensitivity_map_with_nn` parameter. This option allows training a model end-to-end ([40, 41]) and can be combined with the `EstimateCoilSensitivityMaps` class for optimized coil sensitivity maps estimation.

Normalization can be applied through the `Normalizer` class. The `normalize_inputs` parameter determines whether the inputs will be normalized. The `normalization_type` parameter determines the normalization method. The `minmax` and `max` methods normalize the data as $\frac{\text{data}-\min(\text{data})}{\max(\text{data})-\min(\text{data})}$ and $\frac{\text{data}}{\max(\text{data})}$, respectively, in the range 0-1. The `mean_std` and the `mean_var` methods normalize the data as $\frac{\text{data}-\text{mean}(\text{data})}{\text{std}(\text{data})}$ and $\frac{\text{data}-\text{mean}(\text{data})}{\text{var}(\text{data})}$ respectively. The `grayscale` method first normalizes the data as in `minmax` and then multiplies by 255 to bring the data in the range 0-255. The options `fft` and `none` do not apply normalization. When handling complex-valued data, `fft` can be more intuitive. Finally, the `kspace_normalization` parameter determines whether the normalization is performed in k-space, when complex-valued data are available. By default normalization is performed in image space.

Finally, the `Composer` class allows composing a series of transforms into a single transform.

6.3.4 Training & Testing Deep Learning models for MRI tasks in ATOMMIC

Training and testing DL models in ATOMMIC requires a configuration (YAML) file and a single command to set its path, i.e., `atommic run -c path_to_configuration_file`. The configuration file allows setting various MRI transforms (as discussed in Sec. 6.3.3), undersampling options (as explained in Sec. 6.3.2), and hyperparameters (Fig. 6.2). The installation is simple through `pip install atommic`. Multi-GPU and multi-node training, mixed-precision (floating-point 16), early stopping, and Exponential Moving Average can also be configured. For exporting and logging models, tensorboard[4] and Weights & Biases[5] support is available.

6.3.5 Experiments

In our comprehensive evaluation, we demonstrate distinct applications of ATOMMIC in the tasks of accelerated MRI reconstruction, quantitative parameter map estimation, segmentation, and MTL for joint reconstruction and segmentation. Twenty-five DL models were benchmarked in various public datasets and with different hyperparameters, as presented in Table 6.1. Hyperparameters were randomly selected between Adam ([42]) and weighted Adam for the optimizer and cosine annealing and inverse square root annealing for the learning-rate scheduler, as they are among the most popular options when training DL models. For the loss functions, we presented different options depending on the task.

Datasets

For the task of accelerated MRI reconstruction, three datasets were used: the Calgary Campinas 359 dataset (CC359) ([28]), the fastMRI Brains multicoil dataset (fastMRIBrains) ([29]), and the Stanford Fully Sampled 3D FSE Knee dataset (StanfordKnee) ([30]).

The CC359 dataset comprises 117 3D T_1 weighted twelve-coil brain scans. The size of the acquisition matrix is 256×218, while it varies in the slice-encoding (kz) direction between 170 and 180 due to 15% zero-filling. In our experiments, every subject's first 50

[4] https://github.com/tensorflow/tensorboard
[5] https://github.com/wandb/wandb

and last 50 slices were excluded since these mainly resided outside the brain. The training set consisted of 47 subjects, the validation set of 20 subjects, and the test set of 50 subjects. A 2D Poisson disc distribution sampling pattern accelerated imaging by 5x and 10x times.

The fastMRIBrains dataset comprises of T_1-weighted, T_1-weighted with contrast agent (T_1POST), T_2-weighted, and Fluid-Attenuated Inversion Recovery (FLAIR) scans. In our experiments, we used the first batch of data out of 10, which contained 449 subjects in the training set and 457 subjects in the validation set. Nine subjects were removed due to containing not-a-number (NaN) values. The number of coils varied from four to twenty. The matrix size ranged from minimum 512 to maximum 768 × minimum 213 to maximum 396 and was cropped to 320 × minimum 213 to maximum 320. An Equispaced 1D sampling pattern accelerated imaging by 4x and 8x times.

The StanfordKnee dataset consists of Proton-Density (PD) 3D Fast-Spin Echo (FSE) eight-coil data with fat saturation. The dataset included 19 subjects, split into 13 subjects for the training set, 3 for the validation set, and 3 for the test set. The matrix size was 320 × 320 × 256. A Gaussian 2D sampling pattern was used to accelerate imaging by 12x times.

The Amsterdam Ultra-high field adult lifespan database (AHEAD) ([43]) dataset was used to estimate quantitative parameter maps, as it contains multi-echo data necessary for this task. It consists of thirty-two-coil T_1, T_{2^*} and Quantitative Susceptibility Mapping brain scans of four echo times MP2RAGE-ME 7 Tesla ([44]). Motion correction with Fat navigators (FatNavs) and defacing in the image domain was already applied to the dataset ([9]). The scanned image resolution is 0.7mm isotropic. The objective was to estimate the following quantitative maps: R_2^*, B_0, and the angle of the net magnetization M, denoted as $|M|$. We used the first ten subjects of the dataset, 001 to 010, of which the first six were used for training, the next two for validation, and the last two for testing. A Gaussian 2D sampling pattern was used to accelerate imaging by 12x times. Brain tissue segmentation masks were pre-computed and applied during training to avoid including NaN or infinity (Inf) values on the skull or the background. Brain tissue masks were computed by applying Otsu's thresholding, computing the largest connected component and the convex hull, and applying a series of binary erosions and dilations.

For the segmentation task, three datasets were used: the Brain Tumor Segmentation 2023 Adult Glioma challenge dataset (BraTS2023AdultGlioma) ([31]), the Ischemic Stroke Lesion Segmentation 2022 (ISLES2022SubAcuteStroke) challenge dataset ([32]), and the segmentation-only dataset of the Stanford Knee MRI with Multi-Task Evaluation (SKM-TEA) dataset ([33]).

The BraTS2023AdultGlioma dataset contains 1000 subjects on the training set and 251 on the validation set, while no ground truth test labels are available. The objective is to segment four classes: necrotic tumor core, peritumoral edematous/invaded tissue, gadolinium-enhancing tumor, and whole tumor. The ISLES2022SubAcuteStroke dataset includes Apparent Diffusion Coefficient (ADC) maps, FLAIR scans, and Diffusion Weighted Imaging (DWI) scans. The training set consisted of 172 subjects, the validation set 37, and the test set 38. The objective is to segment one class, specified as stroke lesions.

The SKM-TEA dataset contains complex-valued multicoil raw data, real-valued coil-combined data, and ground truth segmentation labels, allowing for both segmentation independently and MTL for combined reconstruction and segmentation. Data are of heterogeneous patient anatomy with potential distribution shifts being present as data

were acquired from multiple vendors ([33]). The SKM-TEA segmentation-only dataset provides data imaged in the sagittal plane, with four segmentation classes: lateral tibial cartilage, medial tibial cartilage, lateral meniscus, and medial meniscus. In contrast, for MTL, the complex-valued SKM-TEA dataset comprises data reconstructed in the axial plane with both phase-encoding dimensions. The data are stored as $x \times ky \times kz$, where x denotes the number of slices, and $ky \times kz$, the dimensions to apply the provided undersampling mask and coil sensitivity maps. The matrix size is 512 × 160 and is cropped to 416 × 80 to remove oversampling. Data are undersampled using a Poisson disc distribution 2D pattern with an acceleration factor 4x. Both for segmentation only and for MTL, we split the SKM-TEA dataset into 86 subjects in the training set, 33 in the validation set, and 36 in the test set.

Hyperparameters

For the tasks of reconstruction and quantitative parameter map estimation, models were trained for 20 epochs. For estimating quantitative parameter maps from an accelerated MRI acquisition, we first trained reconstruction models and then used them to initialize the quantitative parameter map estimation models. For the task of segmentation, models were trained for 20 epochs on the BraTS2023AdultGlioma and the SKM-TEA segmentation-only datasets and for 50 epochs on the ISLES2022SubAcuteStroke dataset since its size was significantly smaller than the other two. For MTL for joint reconstruction and segmentation, models were trained for 15 epochs.

The learning rate was set to 10^{-4}, and the floating point precision was set to mixed 16 for all models on all tasks. Normalization by the max value (Sec. 6.3.3) was applied to all models trained for reconstruction, segmentation, and MTL. To stabilize the training of quantitative parameter map estimation models on the AHEAD dataset, we heuristically scaled the input multi-echo k-space data by a factor of 10^4 and the input quantitative maps by a factor of 10^{-3} as in ([9]). The AHEAD data consist of four echo times, for which the values were 3 ms, 11.5 ms, 20 ms, and 28.5 ms, respectively.

Models were trained and tested on an Nvidia Tesla V100 GPU with 32GB memory. A detailed overview of the selected hyperparameters for each model is presented in the Appendix (Table 6.8). Trained models' checkpoints are available on HuggingFace[6].

Evaluation metrics

The performance in the reconstruction task (CC359, fastMRIBrains, and StanfordKnee datasets), in the task of reconstruction and quantitative parameter map estimation (AHEAD dataset), and in the task of reconstruction for MTL (SKM-TEA dataset) was evaluated by measuring the similarity of the predicted reconstructions and the ground truth images using the Structural Similarity Index Measurement (SSIM) ([52]) and assessing the perceived image quality using the Peak Signal-to-Noise Ratio (PSNR). For evaluating the performance of quantitative parameter map estimation models, the Normalized Mean Squared Error (NMSE) was also computed.

The accuracy of the segmentation models in the BraTS2023AdultGlioma and the SKM-TEA datasets was evaluated by quantifying the similarity between the predicted segmentations and the ground truth labels, measured by the DICE coefficient and the Intersection

[6] https://huggingface.co/wdika

Table 6.1: Comparative evaluation of DL models using ATOMMIC for different MRI tasks. The first column reports the task, specifically MultiTask Learning (MTL) (second row) for jointly performing accelerated MRI reconstruction (REC) (fourth row) and MRI segmentation (SEG) (fifth row), and quantitative MRI (qMRI) for estimating parameter maps (second row). The second column reports the publicly available datasets used for training and testing. The third column reports the coil sensitivity maps (CSM) estimation method and the coil combination method (CCM) (Sec. 6.3.3). When CSMs were not available, they were estimated with the `EstimateCoilSensitivityMaps` transformation of ATOMMIC or End-to-End during training with a UNet. The `GeometricDecompositionCoilCompression` (GDCC) transformation was applied to the fastMRI Brains Multicoil dataset, reducing various coils (four to twenty) into single coil data. The CCM was set either to Sensitivity Encoding (SENSE) or Root-Sum-of-Squares (RSS). The fourth column reports the used optimizer (Opt) and learning rate scheduler (LRS). For Optim, the Adam and Adam weighted (AdamW) were used, and for LR Sched, the Inverse Square Root Annealing (ISRA) and Cosine Annealing (CA) were used. The used loss function is reported in the fifth column, and the trained and tested DL models are reported in the sixth column.

Task	Dataset	CSM-CCM	Opt-LRS	Loss	Models
MTL	SKM-TEA [33]	Available -SENSE	Adam-ISRA	0.5*L1 + 0.5*DICE	Image domain Deep Structured Low-Rank Network (IDSLR) [13] Image domain Deep Structured Low-Rank UNet (IDSLRUNet) [13] Multi-Task Learning for MRI Reconstruction and Segmentation (MTLRS) [12] Segmentation Network MRI (SegNet) [14]
qMRI	AHEAD [43]	Available -SENSE	Adam-ISRA	SSIM	quantitative Cascades of Independently Recurrent Inference Machines (qCIRIM) quantitative End-to-End Variational Network (qVarNet) [9]
REC	CC359 [28] fastMRI Brains Multicoil [29] Stanford Knees [30]	End-to-End -RSS GDCC -SENSE ATOMMIC -SENSE	AdamW-CA Adam-ISRA AdamW-ISRA	0.9*SSIM + 0.1*L1 0.9*SSIM + 0.1*L1 Wasserstein [24]	Cascades of Independently Recurrent Inference Machines (CIRIM) [45] Convolutional Recurrent Neural Networks (CRNN) [46] Deep Cascade of Convolutional Neural Networks (CascadeNet) [47] End-to-End Variational Network (VarNet) [41] Joint Deep Model-Based MR Image and Coil Sensitivity Reconstruction Network (JointICNet) [40] KIKINet [48] Learned Primal-Dual Net [5] [5] Model-based Deep Learning Reconstruction (MoDL) [6] Recurrent Inference Machines (RIM) [8] Recurrent Variational Network (RVN) [49] UNet [4] Variable Splitting Network (VSNet) [50] XPDNet [51]
SEG	BraTS 2023 Adult Glioma [31] ISLES 2022 Sub Acute Stroke [32] SKM-TEA [33]		AdamW-ISRA Adam-CA AdamW-ISRA	DICE DICE DICE	Attention UNet [3] Dynamic UNet (DYNUNet) [1] UNet 2D [4] UNet 3D [4] VNet [2]

over Union (IOU). The accuracy of segmentation boundaries was assessed by computing the Hausdorff Distance 95% (HD95), which provides insights into the largest segmentation errors while minimizing the influence of outliers. Additionally, the significance of false positives and false negatives was measured by the F1 score.

Different metrics were used to evaluate the performance of segmentation models on the ISLES2022 SubAcuteStroke dataset, as specified in the challenge manuscript ([32]). Specifically, due to small lesions, such as punctiform infarcts, an increase in the DICE coefficient might result from detecting only a large lesion. Therefore, the Absolute Volume Difference (AVD) was used to compute voxel-wise differences in the infarct volume, while lesion-wise metrics such as the Absolute Lesion Difference (ALD) and the Lesion F1 (L-F1) score allowed measuring of the lesion detection and to count the lesion burden accurately.

6.4 Results

We demonstrate ATOMMIC's capabilities by evaluating twenty-five different DL models implemented and embedded in the toolbox for the tasks of accelerated MRI reconstruction, quantitative parameter map estimation, segmentation, and MTL for joint reconstruction and segmentation.

Table 6.2 presents the reconstruction task performance of models trained on the CC359 and fastMRIBrains datasets. The Variational Network (VarNet) achieved the highest SSIM and PSNR scores for 5x acceleration and the highest PSNR score for 10x acceleration on the CC359 dataset. The Joint Deep Model-Based MR Image and Coil Sensitivity Reconstruction Network (JointICNet) scored the highest SSIM for 10x acceleration on the CC359 dataset. In contrast, on the fastMRIBrains dataset, the Recurrent Variational Network (RVN) scored the highest SSIM and PSNR scores for 4x acceleration and the VarNet for 8x acceleration. The Cascades of Independently Recurrent Inference Machines (CIRIM) yielded the highest SSIM and PSNR scores on the StanfordKnee dataset for 12x acceleration, as presented in Table 6.3. Conversely, on the same dataset, the Convolutional Recurrent Neural Network (CRNN) was excluded from the analysis due to unstable gradient computation, although trained across a wide range of learning rates (10^{-4} to 10^{-9}).

Example reconstructions of brain data are shown in Fig. 6.5a and Fig. 6.5b, from the CC359 dataset, and Fig. 6.6a and Fig. 6.6b, from the fastMRIBrain dataset. Figure 6.7 shows example reconstructions of knee data from the StanfordKnee dataset.

Table 6.4 reports the performance of models trained on the AHEAD dataset for reconstruction and quantitative parameter map estimation. The CIRIM scored highest on reconstructing the AHEAD data for 12x acceleration, resulting in better initializations for the quantitative CIRIM (qCIRIM) model and, thus, more accurate quantitative parameter map estimation than the VarNet. The qCIRIM outperformed the quantitative VarNet (qVarNet) on accurately approximating the R_2^*, B_0, and $|M|$ quantitative maps. Example quantitative map estimations are shown in Fig. 6.8.

Table 6.5 presents the performance in the segmentation task of models trained on the BraTS2023 AdultGlioma and the SKM-TEA segmentation-only datasets. On the BraTS2023-AdultGlioma dataset, the UNet achieved the highest DICE score and the lowest HD95 score, while the UNet3D achieved the highest F1 score and the AttentionUNet the highest IOU. Although the high DICE scores, F1 and IOU scores were reported lower, which may be attributed to the heterogeneous tumors, leading to the inclusion of non-tumor regions in the

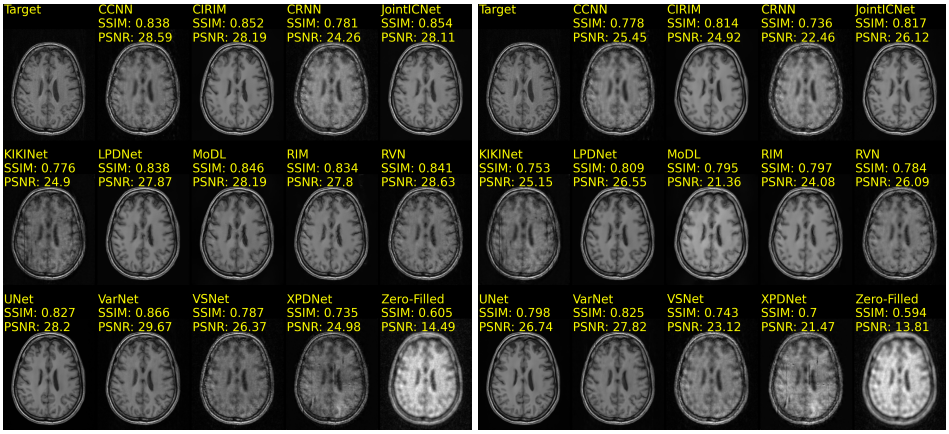
Table 6.2: Overview of performance on reconstructing accelerated brain data. In the first column, the name of the model is reported. The rest of the columns report SSIM and PSNR scores for each dataset, where up arrows indicate the highest the best. The second to fifth columns report the performance of each model on the CC359 dataset for 5x (second-third columns) & 10x (fourth and fifth column) Poisson 2D undersampling. The sixth to ninth columns report performance on the fastMRIBrain dataset, for 4x (sixth and seventh column) & 8x (eighth and ninth columns) Equispaced 1D acceleration. Best performing models are highlighted in bold. Methods are sorted in alphabetical order.

Model	CC359 - Poisson 2D				fastMRIBrains - Equispaced 1D			
	5x		10x		4x		8x	
	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑
CCNN	0.845 ± 0.064	28.36 ± 3.69	0.783 ± 0.089	25.95 ± 3.64	0.886 ± 0.192	33.47 ± 5.92	0.836 ± 0.202	29.40 ± 5.71
CIRIM	0.858 ± 0.074	28.79 ± 4.23	0.816 ± 0.094	26.92 ± 4.36	0.892 ± 0.184	33.83 ± 6.11	0.846 ± 0.202	30.23 ± 5.67
CRNN	0.774 ± 0.088	25.59 ± 4.19	0.722 ± 0.088	24.48 ± 3.39	0.868 ± 0.195	31.31 ± 5.46	0.806 ± 0.198	27.50 ± 5.57
JointICNet	0.872 ± 0.065	29.28 ± 3.99	**0.828 ± 0.086**	27.36 ± 4.10	0.832 ± 0.198	28.57 ± 5.50	0.772 ± 0.202	25.50 ± 5.38
KIKINet	0.788 ± 0.087	25.43 ± 4.16	0.742 ± 0.105	24.37 ± 3.88	0.856 ± 0.201	31.02 ± 5.68	0.805 ± 0.207	27.78 ± 5.82
LPDNet	0.849 ± 0.075	28.26 ± 4.22	0.810 ± 0.099	26.73 ± 4.23	0.882 ± 0.201	32.60 ± 6.78	0.840 ± 0.208	29.51 ± 5.93
MoDL	0.844 ± 0.068	27.97 ± 4.20	0.793 ± 0.088	25.89 ± 4.39	0.870 ± 0.188	31.44 ± 5.66	0.813 ± 0.192	27.81 ± 5.86
RIM	0.834 ± 0.077	27.45 ± 4.32	0.788 ± 0.091	25.56 ± 3.96	0.886 ± 0.188	33.12 ± 6.04	0.837 ± 0.199	29.49 ± 5.74
RVN	0.845 ± 0.067	28.14 ± 3.53	0.787 ± 0.093	26.03 ± 3.77	**0.894 ± 0.180**	**34.23 ± 5.97**	0.843 ± 0.195	30.08 ± 5.68
UNet	0.849 ± 0.070	28.85 ± 4.17	0.810 ± 0.091	27.20 ± 4.20	0.885 ± 0.182	33.09 ± 6.02	0.847 ± 0.197	29.87 ± 5.68
VarNet	**0.874 ± 0.061**	**29.49 ± 3.86**	0.827 ± 0.087	**27.51 ± 4.01**	0.892 ± 0.198	34.00 ± 6.30	**0.856 ± 0.216**	**30.73 ± 5.94**
VSNet	0.788 ± 0.079	25.51 ± 3.91	0.740 ± 0.089	24.19 ± 3.27	0.856 ± 0.196	30.37 ± 5.34	0.796 ± 0.197	26.88 ± 5.43
XPDNet	0.761 ± 0.100	24.27 ± 4.14	0.700 ± 0.112	22.65 ± 3.22	0.854 ± 0.212	31.03 ± 6.75	0.788 ± 0.218	26.96 ± 6.18
ZeroFilled	0.679 ± 0.103	19.89 ± 7.45	0.656 ± 0.092	19.24 ± 7.37	0.671 ± 0.194	24.12 ± 6.21	0.591 ± 0.213	21.03 ± 5.97

Table 6.3: Overview of performance on reconstructing accelerated knee data from the StanfordKnees dataset for 12x Gaussian 2D acceleration. In the first column, the name of the model is reported. The second and third columns report SSIM and PSNR scores, where up arrows indicate the highest the best. Best performing models are highlighted in bold. Methods are sorted in alphabetical order.

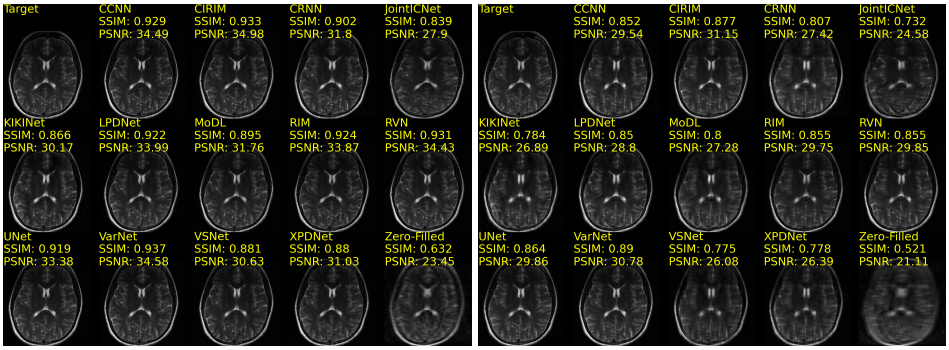
Model	StanfordKnees - Gaussian 2D 12x	
	SSIM ↑	PSNR ↑
CCNN	0.767 ± 0.299	31.64 ± 6.84
CIRIM	**0.795 ± 0.311**	**32.76 ± 7.20**
JointICNet	0.728 ± 0.291	29.59 ± 6.31
KIKINet	0.659 ± 0.241	27.35 ± 5.54
LPDNet	0.736 ± 0.297	29.75 ± 6.31
MoDL	0.566 ± 0.284	23.63 ± 4.66
RIM	0.769 ± 0.304	31.58 ± 6.74
RVN	0.778 ± 0.301	31.96 ± 7.00
UNet	0.771 ± 0.296	31.37 ± 6.54
VarNet	0.764 ± 0.302	31.48 ± 6.73
VSNet	0.708 ± 0.289	28.47 ± 5.82
XPDNet	0.654 ± 0.270	27.16 ± 5.81
ZeroFilled	0.548 ± 0.196	18.07 ± 6.20

predicted segmentation. Similar observations were made on the SKM-TEA segmentation-only dataset, where the UNet3D and VNet achieved the highest DICE score. While the UNet3D scored the highest IOU, the VNet scored the lowest HD95 and the highest F1. In general, the lower F1 scores may be attributed to the heterogeneity of the data since they were acquired from multiple vendors (Sec. 6.3.5). The variability of knee structures across different patients could cause low IOU scores. The impact on the variability between high DICE scores and low F1 and IOU scores can be seen in Fig. 6.9a and Fig. 6.9b, where the



(a) CC359 12-coil data - 5x acceleration (b) CC359 12-coil data - 10x acceleration

Figure 6.5: Reconstructions of 12-coil T_1-weighted data from the CC359 dataset, undersampled with a Poisson disc distribution 2D sampling pattern for 5x (Fig. 6.5a) and 10x (Fig. 6.5b) acceleration. The top row-first column shows the ground truth (Target) image. SSIM and PSNR scores are reported for each method and computed against the Target image. Methods are sorted alphabetically.



(a) fastMRIBrains 14-coil data - 4x acceleration (b) fastMRIBrains 14-coil data - 8x acceleration

Figure 6.6: Reconstructions of 14-coil T_2-weighted data from the fastMRI Brains dataset, undersampled with an Equispaced 1D sampling pattern for 4x (Fig. 6.6a) and 8x (Fig. 6.6b) acceleration. The top row-first column shows the ground truth (Target) image. SSIM and PSNR scores are reported for each method and computed against the Target image. Methods are sorted alphabetically.

DynUNet underestimated the segmented classes, resulting on the worst-performing model.

Table 6.6 reports the performance of segmentation models on the ISLES2022SubAcuteStroke dataset. The DynUNet achieved the lowest average lesion distance (ALD) and the highest DICE and Lesion-F1 scores. The UNet achieved the lowest average volume difference (AVD). The worst-performing model was the VNet, which overestimated the lesion segmentation, as shown in Fig. 6.9c.

Table 6.7 reports the performance in MTL for joint reconstruction and segmentation models trained on the SKM-TEA dataset. The IDSLRUNET achieved the highest SSIM

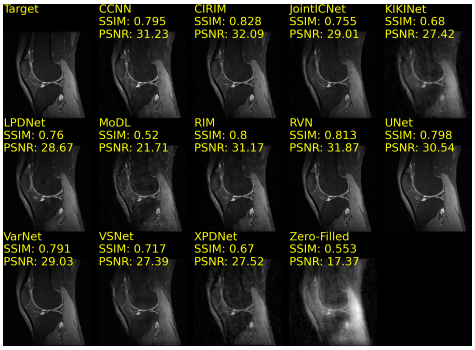


Figure 6.7: Reconstructions of 8-coil T_2-weighted Fast-Spin Echo data from the Stanford Knee dataset undersampled with a Gaussian 2D sampling pattern for 12x acceleration. The top row-first column shows the ground truth (Target) image. SSIM and PSNR scores are reported for each method and computed against the Target image. Methods are sorted alphabetically.

Table 6.4: Overview of performance on reconstructing and estimating quantitative parameter maps. The AHEAD dataset was used, while data were 12x accelerated with a Gaussian 2D undersampling pattern. In the first column, the name of the model is reported. Each model's SSIM, PSNR, and NMSE scores are reported in the second, third, and fourth columns. Up arrows indicate the highest, the best, and down arrows indicate the lowest, the best. The performance of the reconstruction models is reported in the fourth and fifth row, while the quantitative parameter map estimation models' performance is reported in the seventh and eighth row. Best performing models are highlighted in bold. Methods are sorted in alphabetical order.

Model	AHEAD - Gaussian 2D - 12x		
	SSIM ↑	PSNR ↑	NMSE ↓
	Reconstruction		
CIRIM	**0.910 ± 0.077**	**32.86 ± 8.51**	**0.043 ± 0.065**
VarNet	0.893 ± 0.055	32.37 ± 4.88	0.047 ± 0.054
	Quantitative parameter map estimation		
qCIRIM	**0.881 ± 0.178**	**28.36 ± 11.55**	**0.124 ± 0.363**
qVarNet	0.784 ± 0.206	24.35 ± 7.77	0.192 ± 0.334

and PSNR scores, while the SegNet achieved the highest DICE, F1, and IOU and lowest HD95 scores. The lower F1 and IOU scores, as also observed in the performance of the segmentation-only models, may be attributed to varying patient anatomy, data acquisition, and different knee structures. Example reconstructions and segmentations when performing MTL can be found in Fig. 6.10.

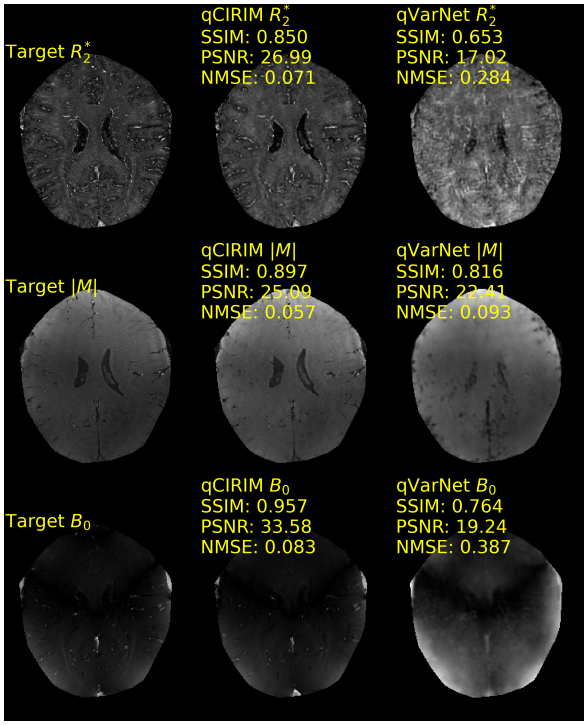


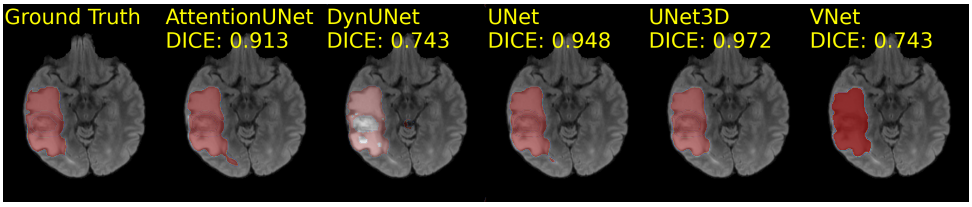
Figure 6.8: Quantitative parameter map estimation of 32-coil T_1-weighted data from the AHEAD dataset undersampled with a Gaussian 2D sampling pattern for 12x acceleration. The first column shows the ground truth (Target) quantitative parameter maps, R_2^*, $|M|$, and B_0 from top to bottom, respectively. The CIRIM and the VarNet were first used to reconstruct the undersampled data and give them as inputs to the qCIRIM and the qVarNet, respectively, to estimate the quantitative parameter maps, as shown in the second and third columns. SSIM, PSNR, and NMSE scores are reported for each method and computed against the Target quantitative parameter map.

Table 6.5: Overview of performance in the segmentation task of models trained on the BraTS2023AdultGlioma (third to seventh row) and the SKM-TEA segmentation-only (ninth to thirteenth row) datasets. Model name, DICE, F1, Hausdorff Distance 95% (HD95), and Intersection Over Union (IOU) scores are reported from left to right. Up and down arrows indicate higher and lower scores being better, respectively. Best performing models are highlighted in bold. Methods are sorted in alphabetical order.

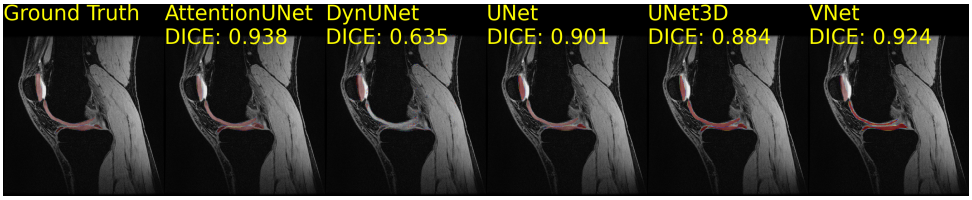
Model	DICE ↑	F1 ↑	HD95 ↓	IOU ↑
BraTS 2023 Adult Glioma				
AttentionUNet	0.930 ± 0.126	0.648 ± 0.763	3.836 ± 3.010	**0.537 ± 0.662**
DynUNet	0.806 ± 0.276	0.104 ± 0.580	5.119 ± 5.411	0.070 ± 0.419
UNet	**0.937 ± 0.118**	0.671 ± 0.787	**3.504 ± 2.089**	0.535 ± 0.663
UNet3D	0.936 ± 0.133	**0.674 ± 0.782**	3.550 ± 2.162	0.528 ± 0.652
VNet	0.733 ± 0.437	0.014 ± 0.234	6.010 ± 6.097	0.000 ± 0.004
SKM-TEA segmentation-only				
AttentionUNet	0.909 ± 0.088	0.637 ± 0.475	6.358 ± 2.209	0.529 ± 0.361
DynUNet	0.689 ± 0.136	0.059 ± 0.264	8.973 ± 4.507	0.015 ± 0.066
UNet	0.912 ± 0.058	0.651 ± 0.449	6.618 ± 1.793	0.516 ± 0.350
UNet3D	**0.918 ± 0.068**	0.789 ± 0.404	5.893 ± 2.995	**0.530 ± 0.347**
VNet	**0.918 ± 0.081**	**0.816 ± 0.426**	**5.540 ± 3.036**	0.507 ± 0.388

Table 6.6: Overview of performance in the segmentation task of models trained on the ISLES2022SubAcuteStroke dataset. Model name, Absolute Lesion Difference (AVD), Absolute Volume Difference (AVD), DICE, and Lesion F1 (L-F1) scores are reported from left to right. Up and down arrows indicate whether higher or lower values indicate better performance. Best performing models are highlighted in bold. Methods are sorted in alphabetical order.

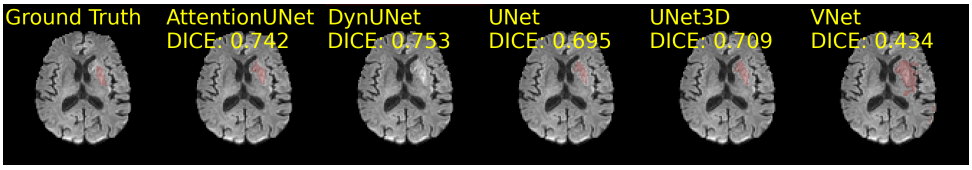
Model	ISLES 2022 Sub Acute Stroke			
	ALD ↓	AVD ↓	DICE ↑	L-F1 ↑
AttentionUNet	0.809 ± 2.407	0.548 ± 3.411	0.709 ± 0.552	0.799 ± 0.579
DynUNet	**0.752 ± 2.230**	0.586 ± 3.874	**0.729 ± 0.529**	**0.802 ± 0.564**
UNet	0.909 ± 3.953	**0.544 ± 3.921**	0.695 ± 0.559	0.786 ± 0.585
UNet3D	0.821 ± 2.167	0.691 ± 5.458	0.687 ± 0.547	0.798 ± 0.573
VNet	2.281 ± 10.72	3.257 ± 27.430	0.490 ± 0.694	0.600 ± 0.687



(a) Brain Tumor Segmentation 2023 Adult Glioma segmentations.



(b) Stanford Knee MRI with Multi-Task Evaluation segmentations.



(c) ISLES 2022 Sub Acute Stroke segmentations.

Figure 6.9: Segmentations on the Brain Tumor Segmentation 2023 Adult Glioma dataset (Fig. 6.9a), the Stanford Knee MRI with Multi-Task Evaluation (SKM-TEA) segmentation-only dataset (Fig. 6.9b), and the ISLES 2022 Sub Acute Stroke dataset (Fig. 6.9c). The first image on each figure shows the Ground Truth image with the segmentation labels. The rest of the images present the segmentations of different methods. Each segmentation method's DICE score is reported and computed against the Ground Truth labels. Methods are sorted alphabetically.

Table 6.7: Overview of performance in MTL for joint reconstruction and segmentation of models trained on the SKM-TEA dataset for Poisson 2D 4x undersampling. Model name, SSIM, PSNR, DICE, F1, Hausdorff Distance 95% (HD95), and Intersection Over Union (IOU) scores are reported from left to right. Up and down arrows indicate whether higher or lower scores indicate higher performance. Best performing models are highlighted in bold. Methods are sorted in alphabetical order.

Model	SKM-TEA - Poisson 2D 4x					
	SSIM ↑	PSNR ↑	DICE ↑	F1 ↑	HD95 ↓	IOU ↑
IDSLR	0.836 ± 0.106	30.38 ± 5.67	0.894 ± 0.127	0.256 ± 0.221	4.927 ± 2.812	0.298 ± 0.309
IDSLRUNET	**0.842 ± 0.106**	**30.53 ± 5.59**	0.870 ± 0.134	0.225 ± 0.194	8.724 ± 3.298	0.212 ± 0.199
MTLRS	0.832 ± 0.106	30.48 ± 5.30	0.889 ± 0.118	0.247 ± 0.203	7.594 ± 3.673	0.218 ± 0.194
SegNet	0.840 ± 0.107	29.95 ± 5.12	**0.915 ± 0.114**	**0.270 ± 0.284**	**3.002 ± 1.449**	**0.290 ± 0.349**

6.5 Discussion and Conclusions

We presented the Advanced Toolbox for Multitask Medical Imaging Consistency (ATOMMIC), a versatile toolbox designed to ensure consistency in the performance of various Deep Learning (DL) models applied in different MRI tasks such as reconstruction, quantitative parameter map estimation and segmentation (Sec. 6.3.1). Consistency is ensured by unifying the implementation of networks' components, hyperparameters, image transformations, and training configurations. Among existing AI frameworks for MRI analysis, ATOMMIC was found to be the only framework to uniquely harmonize complex-valued and real-valued data support, allowing the assessment of MultiTask Learning (MTL) by combining individual models designed for single tasks to perform joint tasks. To demonstrate ATOMMIC's capabilities we trained and tested twenty-five DL models on eight publicly available datasets, including brain and knee anatomies, for three different MRI tasks and presented applications of MTL for joint reconstruction and segmentation. Three undersampling schemes were evaluated, ranging from 4 times to 12 times acceleration, on the task of accelerated MRI reconstruction on three publicly available datasets, using different loss functions, optimizers, and learning rate schedulers (Fig. 6.2). We also assessed the effectiveness of end-to-end reconstruction and coil sensitivity maps estimation during training, removing the overhead of pre-computing coil sensitivity maps and increasing the storage space. Coil compression was evaluated on the reconstruction task, showing advantages in reducing training time while maintaining high performance (Sec. 6.3.3). Successful application of quantitative DL models was demonstrated in accurately estimating quantitative parameter maps of the brain, such as the R_2^* map, which allows for quantifying iron deposition related to aging and Parkinson's and Alzheimer's disease ([9]). Segmentation of brain lesions, tumors, and knee pathologies was also presented, while the tasks of segmentation and reconstruction were combined to assess MTL.

Physics-based DL models outperformed other DL models on the task of accelerated MRI reconstruction, showing an advantage in enforcing data consistency on the MRI domain either implicity (CIRIM) or explicitly (JointICNet, RVN, VarNet). The models trained and tested on the fastMRIBrains (Table 6.2) and StanfordKnee datasets (Table 6.3) exhibited higher standard deviations than the models trained on the CC359 dataset (Table 6.2), potentially due to the different undersampling patterns, multiple modalities, and varying numbers of coils for the fastMRIBrains dataset and the small number of coils (eight) for the StanfordKnee dataset (Sec. 6.3.5). These variations in data acquisition resulted in decreased

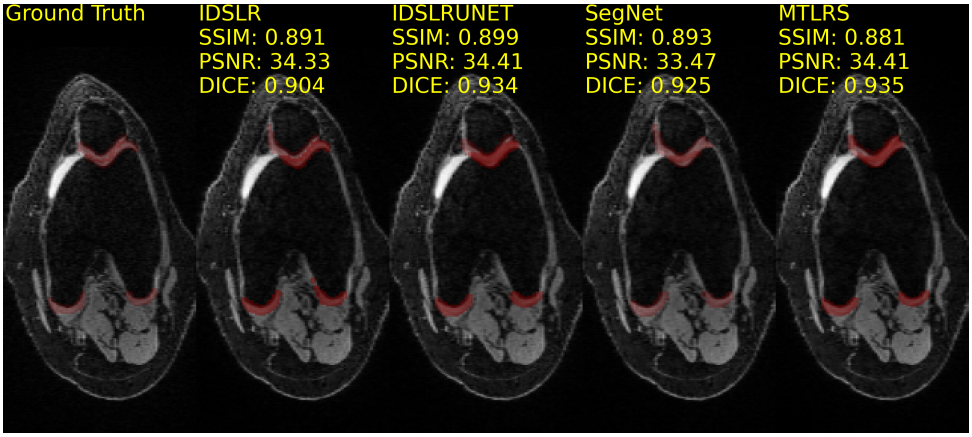


Figure 6.10: Reconstructions and segmentations on the Stanford Knee MRI with Multi-Task Evaluation (SKM-TEA) dataset, undersampled with a Poisson disc distribution 2D pattern for 4x acceleration. First image shows the Ground Truth image and segmentation labels. SSIM, PSNR, and DICE scores are reported for each method and computed against the Ground Truth image and segmentation labels. Methods are sorted alphabetically. Images are interpolated for visualisation purposes from 416×80 to 256×128.

SNR, adversely impacting the SSIM and PSNR scores. Physics-based models were also demonstrated to accurately approximate quantitative parameter maps (Table 6.4). This suggests that maintaining consistency, an essential feature in ATOMMIC, in the primary task of reconstruction can enhance performance in the subsequent quantitative parameter map estimation task, a promising advancement towards fast and robust quantification of neurological diseases. The baseline UNet performed the best on the BraTS2023AdultGlioma and SKM-TEA datasets on the segmentation task (Table 6.5), but F1 and IOU scores were low for all models, potentially due to the datasets' heterogeneity and distribution shifts. The UNet3D and the VNet showed strengths in handling heterogeneous data. However, the VNet struggled with small lesion segmentation in the ISLES2022SubAcuteStroke dataset (Table 6.6). The AttentionUNet achieved the highest IOU score on the BraTS2023AdultGlioma dataset, underscoring its ability to identify relevant tumor regions, although heterogeneous due to the integrated attention mechanisms. The dynamic nature of the DynUNet proved advantageous in the ISLES2022SubAcuteStroke dataset, where it outperformed others, scoring the highest ALD, DICE, and Lesion-F1 scores. ATOMMIC's advancement in utilizing MTL for combining tasks led to improved reconstruction quality and segmentation accuracy when performing the two tasks jointly (Table 6.7).

ATOMMIC offers a unique AI suite for MRI reconstruction and analysis with numerous embedded DL models, hyperparameters, training and testing schemes, and exporting options, which can significantly advance DL applications in MRI research. Evaluating multiple DL models on a single task using a robust framework provides a better understanding of the benefits of applying DL to medical imaging rather than just focusing on the performance of a single model. Utilizing MTL to combine tasks is a step towards end-to-end solutions that eliminate the overhead of splitting related tasks, leading to improved performance and faster processing speed. ATOMMIC enables the evaluation of many DL models on

multiple public datasets with standardized formats, thanks to the significant efforts made by various research groups. However, supporting private datasets can be challenging. In addition to privacy concerns, identifying a series of appropriate pre-processing steps to use the data is often necessary, which can be time-consuming and require expert knowledge. Raw MRI data, in particular, often come in vendor-locked proprietary formats. While the ISMRM-RD format ([53]) represents a step towards an open vendor-agnostic format for storing such data, integrating private datasets into open-source toolboxes remains limited. Such limitation is also identified in our work, and further limitations include the fact that ATOMMIC currently focuses solely on MRI, and essential tasks such as classification, registration, and motion correction remain to be implemented towards a robust end-to-end multitask framework.

Future work could focus on developing pre-processing pipelines for private medical imaging datasets, including additional tasks and imaging modalities, such as Computed Tomography. Nevertheless, open issues like data interoperability, model robustness, and assessing MTL still require attention. The availability of public datasets, open-source code, and comprehensive documentation is crucial to effectively adopting AI techniques and facilitating their further development by the scientific and broader research communities. With ATOMMIC, we aim to accelerate medical image analysis and provide a comprehensive framework for researchers to integrate and evaluate datasets, DL models, tasks, and imaging modalities.

Acknowledgments

This publication is based on the STAIRS project under the TKI-PPP program. The collaboration project is co-funded by the PPP Allowance made available by Health Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships.

H.A. Marquering and M.W.A. Caan are shareholders of Nicolab International Ltd. H.A. Marquering is a shareholder of TrianecT B.V. and inSteps B.V. (unrelated to this project; all paid individually).

References

[1] Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2):203–211, February 2021.

[2] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571, October 2016.

[3] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y. Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention U-Net: Learning Where to Look for the Pancreas, May 2018.

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger,

William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[5] Jonas Adler and Ozan Öktem. Learned primal-dual reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1322–1332, 2018.

[6] Hemant K. Aggarwal, Merry P. Mani, and Mathews Jacob. Modl: Model-based deep learning architecture for inverse problems. *IEEE Transactions on Medical Imaging*, 38(2):394–405, 2019.

[7] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P. Recht, Daniel K. Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated mri data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, 2018.

[8] Kai Lønning, Patrick Putzky, Jan-Jakob Sonke, Liesbeth Reneman, Matthan W. A. Caan, and Max Welling. Recurrent inference machines for reconstructing heterogeneous MRI data. *Medical Image Analysis*, 53:64–78, April 2019.

[9] Chaoping Zhang, Dimitrios Karkalousos, Pierre-Louis Bazin, Bram F. Coolen, Hugo Vrenken, Jan-Jakob Sonke, Birte U. Forstmann, Dirk H.J. Poot, and Matthan W.A. Caan. A unified model for reconstruction and r2* mapping of accelerated 7t data using the quantitative recurrent inference machine. *NeuroImage*, 264:119680, 2022.

[10] Rich Caruana. Multitask Learning. *Machine Learning*, 28:41–75, 1997.

[11] Qiaoying Huang, Xiao Chen, Dimitris Metaxas, and Mariappan S. Nadar. Brain Segmentation from k-Space with End-to-End Recurrent Attention Network. In Dinggang Shen, Tianming Liu, Terry M. Peters, Lawrence H. Staib, Caroline Essert, Sean Zhou, Pew-Thian Yap, and Ali Khan, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Lecture Notes in Computer Science, pages 275–283, Cham, 2019. Springer International Publishing.

[12] Dimitrios Karkalousos, Ivana Isgum, Henk Marquering, and Matthan W. A. Caan. MultiTask Learning for accelerated-MRI Reconstruction and Segmentation of Brain Lesions in Multiple Sclerosis. In *Medical Imaging with Deep Learning*, pages 991–1005. PMLR, January 2024.

[13] Aniket Pramanik and Mathews Jacob. RECONSTRUCTION AND SEGMENTATION OF PARALLEL MR DATA USING IMAGE DOMAIN DEEP-SLR. *Proceedings. IEEE International Symposium on Biomedical Imaging*, 2021:10.1109/isbi48211.2021.9434056, April 2021.

[14] Liyan Sun, Zhiwen Fan, Xinghao Ding, Yue Huang, and John Paisley. Joint CS-MRI Reconstruction and Segmentation with a Unified Deep Network. In Albert C. S. Chung, James C. Gee, Paul A. Yushkevich, and Siqi Bao, editors, *Information Processing in Medical Imaging*, Lecture Notes in Computer Science, pages 492–504, Cham, 2019. Springer International Publishing.

[15] Moritz Blumenthal, Guanxiong Luo, Martin Schilling, H. Christian M. Holme, and Martin Uecker. Deep, deep learning with BART. *Magnetic Resonance in Medicine*, 89(2):678–693, 2023.

[16] George Yiasemis, Nikita Moriakov, Dimitrios Karkalousos, Matthan Caan, and Jonas Teuwen. DIRECT: Deep Image REConstruction Toolkit. *Journal of Open Source Software*, 7(73):4278, May 2022.

[17] Eli Gibson, Wenqi Li, Carole Sudre, Lucas Fidon, Dzhoshkun I. Shakir, Guotai Wang, Zach Eaton-Rosen, Robert Gray, Tom Doel, Yipeng Hu, Tom Whyntie, Parashkev Nachev, Marc Modat, Dean C. Barratt, Sébastien Ourselin, M. Jorge Cardoso, and Tom Vercauteren. NiftyNet: A deep-learning platform for medical imaging. *Computer Methods and Programs in Biomedicine*, 158:113–122, May 2018.

[18] Guotai Wang, Xiangde Luo, Ran Gu, Shuojue Yang, Yijie Qu, Shuwei Zhai, Qianfei Zhao, Kang Li, and Shaoting Zhang. PyMIC: A deep learning toolkit for annotation-efficient medical image segmentation. *Computer Methods and Programs in Biomedicine*, 231:107398, April 2023.

[19] Elina Thibeau-Sutre, Mauricio Díaz, Ravi Hassanaly, Alexandre Routier, Didier Dormont, Olivier Colliot, and Ninon Burgos. ClinicaDL: An open-source deep learning software for reproducible neuroimaging processing. *Computer Methods and Programs in Biomedicine*, 220:106818, June 2022.

[20] J-Donald Tournier, Robert Smith, David Raffelt, Rami Tabbara, Thijs Dhollander, Maximilian Pietsch, Daan Christiaens, Ben Jeurissen, Chun-Hung Yeh, and Alan Connelly. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage*, 202:116137, November 2019.

[21] Nicholas J. Tustison, Philip A. Cook, Andrew J. Holbrook, Hans J. Johnson, John Muschelli, Gabriel A. Devenyi, Jeffrey T. Duda, Sandhitsu R. Das, Nicholas C. Cullen, Daniel L. Gillen, Michael A. Yassa, James R. Stone, James C. Gee, and Brian B. Avants. The ANTsX ecosystem for quantitative biological and medical imaging. *Scientific Reports*, 11(1):9068, April 2021.

[22] Fernando Pérez-García, Rachel Sparks, and Sébastien Ourselin. TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Computer Methods and Programs in Biomedicine*, 208:106236, September 2021.

[23] M. Jorge Cardoso, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang, Benjamin Murrey, Andriy Myronenko, Can Zhao, Dong Yang, Vishwesh Nath, Yufan He, Ziyue Xu, Ali Hatamizadeh, Andriy Myronenko, Wentao Zhu, Yun Liu, Mingxin Zheng, Yucheng Tang, Isaac Yang, Michael Zephyr, Behrooz Hashemian, Sachidanand Alle, Mohammad Zalbagi Darestani, Charlie Budd, Marc Modat, Tom Vercauteren, Guotai Wang, Yiwen Li, Yipeng Hu, Yunguan Fu, Benjamin Gorman, Hans Johnson, Brad Genereaux, Barbaros S. Erdal, Vikash Gupta, Andres Diaz-Pinto, Andre Dourson, Lena Maier-Hein, Paul F. Jaeger, Michael Baumgartner, Jayashree Kalpathy-Cramer,

Mona Flores, Justin Kirby, Lee A. D. Cooper, Holger R. Roth, Daguang Xu, David Bericat, Ralf Floca, S. Kevin Zhou, Haris Shuaib, Keyvan Farahani, Klaus H. Maier-Hein, Stephen Aylward, Prerna Dogra, Sebastien Ourselin, and Andrew Feng. MONAI: An open-source framework for deep learning in healthcare, November 2022.

[24] Marco Cuturi. Sinkhorn Distances: Lightspeed Computation of Optimal Transportation Distances, June 2013.

[25] Arjun D. Desai, Batu M. Ozturkler, Christopher M. Sandino, Robert Boutin, Marc Willis, Shreyas Vasanawala, Brian A. Hargreaves, Christopher Ré, John M. Pauly, and Akshay S. Chaudhari. Noise2Recon: Enabling SNR-robust MRI reconstruction with semi-supervised and self-supervised learning. *Magn Reson Med*, 90(5):2052–2070, November 2023. 3 citations (Crossref) [2024-03-13] 6 citations (Semantic Scholar/DOI) [2024-03-13].

[26] Burhaneddin Yaman, Seyed Amir Hossein Hosseini, Steen Moeller, Jutta Ellermann, Kâmil Uğurbil, and Mehmet Akçakaya. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magnetic Resonance in Medicine*, 84(6):3172–3191, 2020. 127 citations (Crossref) [2024-03-13] 180 citations (Semantic Scholar/DOI) [2024-03-13] _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.28378.

[27] Oleksii Kuchaiev, Jason Li, Huyen Nguyen, Oleksii Hrinchuk, Ryan Leary, Boris Ginsburg, Samuel Kriman, Stanislav Beliaev, Vitaly Lavrukhin, Jack Cook, Patrice Castonguay, Mariya Popova, Jocelyn Huang, and Jonathan M. Cohen. NeMo: A toolkit for building AI applications using Neural Modules, September 2019.

[28] Youssef Beauferris, Jonas Teuwen, Dimitrios Karkalousos, Nikita Moriakov, Matthan Caan, George Yiasemis, Lívia Rodrigues, Alexandre Lopes, Helio Pedrini, Letícia Rittner, Maik Dannecker, Viktor Studenyak, Fabian Gröger, Devendra Vyas, Shahrooz Faghih-Roohi, Amrit Kumar Jethi, Jaya Chandra Raju, Mohanasankar Sivaprakasam, Mike Lasby, Nikita Nogovitsyn, Wallace Loos, Richard Frayne, and Roberto Souza. Multi-Coil MRI Reconstruction Challenge-Assessing Brain MRI Reconstruction Models and Their Generalizability to Varying Coil Configurations. *Frontiers in Neuroscience*, 16:919186, 2022.

[29] Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J. Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, Hersh Chandarana, Zizhao Zhang, Michal Drozdzal, Adriana Romero, Michael Rabbat, Pascal Vincent, Nafissa Yakubova, James Pinkerton, Duo Wang, Erich Owens, C. Lawrence Zitnick, Michael P. Recht, Daniel K. Sodickson, and Yvonne W. Lui. fastMRI: An Open Dataset and Benchmarks for Accelerated MRI, December 2019.

[30] Kevin Epperson, Anne Marie Sawyer, Michael Lustig, Marcus Alley, Martin Uecker, Patrick Virtue, Peng Lai, and Shreyas Vasanawala. Creation of Fully Sampled MR Data Repository for Compressed Sensing of the Knee.

[31] Anahita Fathi Kazerooni, Nastaran Khalili, Xinyang Liu, Debanjan Haldar, Zhifan Jiang, Syed Muhammed Anwar, Jake Albrecht, Maruf Adewole, Udunna Anazodo, Hannah Anderson, Sina Bagheri, Ujjwal Baid, Timothy Bergquist, Austin J. Borja, Evan Calabrese, Verena Chung, Gian-Marco Conte, Farouk Dako, James Eddy, Ivan Ezhov, Ariana Familiar, Keyvan Farahani, Shuvanjan Haldar, Juan Eugenio Iglesias, Anastasia Janas, Elaine Johansen, Blaise V Jones, Florian Kofler, Dominic LaBella, Hollie Anne Lai, Koen Van Leemput, Hongwei Bran Li, Nazanin Maleki, Aaron S McAllister, Zeke Meier, Bjoern Menze, Ahmed W Moawad, Khanak K Nandolia, Julija Pavaine, Marie Piraud, Tina Poussaint, Sanjay P Prabhu, Zachary Reitman, Andres Rodriguez, Jeffrey D Rudie, Mariana Sanchez-Montano, Ibraheem Salman Shaikh, Lubdha M. Shah, Nakul Sheth, Russel Taki Shinohara, Wenxin Tu, Karthik Viswanathan, Chunhao Wang, Jeffrey B Ware, Benedikt Wiestler, Walter Wiggins, Anna Zapaishchykova, Mariam Aboian, Miriam Bornhorst, Peter de Blank, Michelle Deutsch, Maryam Fouladi, Lindsey Hoffman, Benjamin Kann, Margot Lazow, Leonie Mikael, Ali Nabavizadeh, Roger Packer, Adam Resnick, Brian Rood, Arastoo Vossough, Spyridon Bakas, and Marius George Linguraru. The Brain Tumor Segmentation (BraTS) Challenge 2023: Focus on Pediatrics (CBTN-CONNECT-DIPGR-ASNR-MICCAI BraTS-PEDs). *ArXiv*, page arXiv:2305.17033v5, February 2024.

[32] Moritz R. Hernandez Petzsche, Ezequiel de la Rosa, Uta Hanning, Roland Wiest, Waldo Valenzuela, Mauricio Reyes, Maria Meyer, Sook-Lei Liew, Florian Kofler, Ivan Ezhov, David Robben, Alexandre Hutton, Tassilo Friedrich, Teresa Zarth, Johannes Bürkle, The Anh Baran, Björn Menze, Gabriel Broocks, Lukas Meyer, Claus Zimmer, Tobias Boeckh-Behrens, Maria Berndt, Benno Ikenberg, Benedikt Wiestler, and Jan S. Kirschke. ISLES 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. *Scientific Data*, 9(1):762, December 2022.

[33] Arjun D. Desai, Andrew M. Schmidt, Elka B. Rubin, Christopher M. Sandino, Marianne S. Black, Valentina Mazzoli, Kathryn J. Stevens, Robert Boutin, Christopher Ré, Garry E. Gold, Brian A. Hargreaves, and Akshay S. Chaudhari. SKM-TEA: A Dataset for Accelerated MRI Reconstruction with Dense Image Labels for Quantitative Clinical Evaluation, March 2022.

[34] Jonas Adler, Sebastian Lunz, Olivier Verdier, Carola-Bibiane Schönlieb, and Ozan Öktem. Task adapted reconstruction for inverse problems. *Inverse Problems*, 38(7):075006, may 2022.

[35] Matthew J. Muckley, Bruno Riemenschneider, Alireza Radmanesh, Sunwoo Kim, Geunu Jeong, Jingyu Ko, Yohan Jun, Hyungseob Shin, Dosik Hwang, Mahmoud Mostapha, Simon Arberet, Dominik Nickel, Zaccharie Ramzi, Philippe Ciuciu, Jean-Luc Starck, Jonas Teuwen, Dimitrios Karkalousos, Chaoping Zhang, Anuroop Sriram, Zhengnan Huang, Nafissa Yakubova, Yvonne W. Lui, and Florian Knoll. Results of the 2020 fastmri challenge for machine learning mr image reconstruction. *IEEE Transactions on Medical Imaging*, 40(9):2306–2317, 2021.

[36] Michael Schacht Hansen and Thomas Sangild Sørensen. Gadgetron: An open source framework for medical image reconstruction. *Magnetic Resonance in Medicine*, 69(6):1768–1776, 2013.

[37] Tao Zhang, John M. Pauly, Shreyas S. Vasanawala, and Michael Lustig. Coil compression for accelerated imaging with Cartesian sampling. *Magnetic Resonance in Medicine*, 69(2):571–582, 2013.

[38] Klaas P. Pruessmann, Markus Weiger, Markus B. Scheidegger, and Peter Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–962, 1999.

[39] Martin Uecker, Peng Lai, Mark J. Murphy, Patrick Virtue, Michael Elad, John M. Pauly, Shreyas S. Vasanawala, and Michael Lustig. ESPIRiT — An Eigenvalue Approach to Autocalibrating Parallel MRI: Where SENSE meets GRAPPA. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 71(3):990–1001, March 2014.

[40] Yohan Jun, Hyungseob Shin, Taejoon Eo, and Dosik Hwang. Joint Deep Model-based MR Image and Coil Sensitivity Reconstruction Network (Joint-ICNet) for Fast MRI. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5266–5275, Nashville, TN, USA, June 2021. IEEE.

[41] Anuroop Sriram, Jure Zbontar, Tullie Murrell, Aaron Defazio, C. Lawrence Zitnick, Nafissa Yakubova, Florian Knoll, and Patricia Johnson. End-to-End Variational Networks for Accelerated MRI Reconstruction. In Anne L. Martel, Purang Abolmaesumi, Danail Stoyanov, Diana Mateus, Maria A. Zuluaga, S. Kevin Zhou, Daniel Racoceanu, and Leo Joskowicz, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, Lecture Notes in Computer Science, pages 64–73, Cham, 2020. Springer International Publishing.

[42] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[43] Anneke Alkemade, Martijn J Mulder, Josephine M Groot, Bethany R Isaacs, Nikita van Berendonk, Nicky Lute, Scott JS Isherwood, Pierre-Louis Bazin, and Birte U Forstmann. The Amsterdam Ultra-high field adult lifespan database (AHEAD): A freely available multimodal 7 Tesla submillimeter magnetic resonance imaging database. *NeuroImage*, 221:117200, November 2020.

[44] Matthan W. A. Caan, Pierre-Louis Bazin, José P. Marques, Gilles de Hollander, Serge O. Dumoulin, and Wietske van der Zwaag. MP2RAGEME: T1, T2*, and QSM mapping in one sequence at 7 tesla. *Human Brain Mapping*, 40(6):1786–1798, 2019.

[45] D. Karkalousos, S. Noteboom, H. E. Hulst, F. M. Vos, and M. W. A. Caan. Assessment of data consistency through cascades of independently recurrent inference machines for fast and robust accelerated MRI reconstruction. *Physics in Medicine & Biology*, 67(12):124001, June 2022.

[46] Chen Qin, Jo Schlemper, Jose Caballero, Anthony N. Price, Joseph V. Hajnal, and Daniel Rueckert. Convolutional Recurrent Neural Networks for Dynamic MR Image Reconstruction. *IEEE Transactions on Medical Imaging*, 38(1):280–290, January 2019.

[47] Jo Schlemper, Jose Caballero, Joseph V. Hajnal, Anthony N. Price, and Daniel Rueckert. A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction. *IEEE Transactions on Medical Imaging*, 37(2):491–503, February 2018.

[48] Taejoon Eo, Yohan Jun, Taeseong Kim, Jinseong Jang, Ho-Joon Lee, and Dosik Hwang. KIKI-net: Cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. *Magnetic Resonance in Medicine*, 80(5):2188–2201, 2018.

[49] George Yiasemis, Jan-Jakob Sonke, Clarisa Sanchez, and Jonas Teuwen. Recurrent Variational Network: A Deep Learning Inverse Problem Solver applied to the task of Accelerated MRI Reconstruction. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 722–731, New Orleans, LA, USA, June 2022. IEEE.

[50] Jinming Duan, Jo Schlemper, Chen Qin, Cheng Ouyang, Wenjia Bai, Carlo Biffi, Ghalib Bello, Ben Statton, Declan P. O'Regan, and Daniel Rueckert. VS-Net: Variable splitting network for accelerated parallel MRI reconstruction, July 2019.

[51] Zaccharie Ramzi, Philippe Ciuciu, and Jean-Luc Starck. Benchmarking MRI Reconstruction Neural Networks on Large Public Datasets. *Applied Sciences*, 10(5):1816, March 2020.

[52] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004.

[53] Souheil J. Inati, Joseph D. Naegele, Nicholas R. Zwart, Vinai Roopchansingh, Martin J. Lizak, David C. Hansen, Chia-Ying Liu, David Atkinson, Peter Kellman, Sebastian Kozerke, Hui Xue, Adrienne E. Campbell-Washburn, Thomas S. Sørensen, and Michael S. Hansen. ISMRM Raw data format: A proposed standard for MRI raw datasets. *Magnetic Resonance in Medicine*, 77(1):411–421, 2017.

Appendix

Table 6.8: Overview of selected hyperparameters for all trained models on all tasks.

Model	Hyperparameters
	Accelerated MRI Reconstruction
CCNN	num_cascades: 10, hidden_channels: 64, n_convs: 5
CIRIM	recurrent_layer: IndRNN, conv_filters: [128, 128, 2], conv_kernels: [5, 3, 3], conv_dilations: [1, 2, 1], conv_bias: [true, true, false], recurrent_filters: [128, 128, 0], recurrent_kernels: [1, 1, 0], recurrent_dilations: [1, 1, 0], recurrent_bias: [true, true, false], time_steps: 8, conv_dim: 2, num_cascades: 5
CRNN	num_iterations: 10, hidden_channels: 64, n_convs: 3
JointICNet	num_iter: 2, kspace_unet_num_filters: 16, kspace_unet_num_pool_layers: 2, imspace_unet_num_filters: 16, imspace_unet_num_pool_layers: 2, sens_unet_num_filters: 16, sens_unet_num_pool_layers: 2
KIKINet	num_iter: 2, kspace_unet_num_filters: 16, kspace_unet_num_pool_layers: 2, imspace_unet_num_filters: 16, imspace_unet_num_pool_layers: 2
LPDNet	num_primal: 5, num_dual: 5, num_iter: 5, primal_unet_num_filters: 16, primal_unet_num_pool_layers: 2, dual_unet_num_filters: 16
MoDL	unrolled_iterations: 5, residual_blocks: 5, channels: 64, regularization_factor: 0.1
RIM	recurrent_layer: GRU, conv_filters: [64, 64, 2], conv_kernels: [5, 3, 3], conv_dilations: [1, 2, 1], conv_bias: [true, true, false], recurrent_filters: [64, 64, 0], recurrent_kernels: [1, 1, 0], recurrent_dilations: [1, 1, 0], recurrent_bias: [true, true, false], time_steps: 8, conv_dim: 2
RVN	recurrent_hidden_channels: 64, recurrent_num_layers: 4, num_steps: 8, learned_initializer: true, initializer_initialization: "sense", initializer_channels: [32, 32, 64, 64], initializer_dilations: [1, 1, 2, 4]
UNet	channels: 64, pooling_layers: 4
VarNet	num_cascades: 8, channels: 18, pooling_layers: 4
VSNet	num_cascades: 10, imspace_model_architecture: CONV, imspace_conv_hidden_channels: 64, imspace_conv_n_convs: 4
XPDNet	num_primal: 5, num_dual: 1, num_iter: 10, use_primal_only: true, kspace_model_architecture: CONV, image_model_architecture: MWCNN, mwcnn_hidden_channels: 16, mwcnn_bias: true

quantitative MRI parameter map estimation

qCIRIM	quantitative_module_recurrent_layer: IndRNN, quantitative_module_conv_filters: [128, 128, 2], quantitative_module_conv_kernels: [5, 3, 3], quantitative_module_conv_dilations: [1, 2, 1], quantitative_module_conv_bias: [true, true, false], quantitative_module_recurrent_filters: [128, 128, 0], quantitative_module_recurrent_kernels: [1, 1, 0], quantitative_module_recurrent_dilations: [1, 1, 0], quantitative_module_recurrent_bias: [true, true, false], quantitative_module_time_steps: 8, quantitative_module_conv_dim: 2, quantitative_module_num_cascades: 5
qVarNet	quantitative_module_num_cascades: 8, quantitative_module_channels: 18, quantitative_module_pooling_layers: 4

MRI Segmentation

AttentionUNet	channels: 32, pooling_layers: 5
DynUNet	channels: 32, pooling_layers: 5, activation: leakyrelu, deep_supervision: true, deep_supervision_levels: 2
UNet	channels: 32, pooling_layers: 5
UNet3D	channels: 32, pooling_layers: 5, consecutive_slices: 3
VNet	channels: 16, pooling_layers: 5, activation: elu

MultiTask Learning for joint Accelerated MRI Reconstruction & MRI Segmentation

IDSLR	channels: 64, pooling_layers: 2, num_iters: 5
IDSLRUNET	channels: 64, pooling_layers: 2, num_iters: 5

MTLRS	num_cascades: 5, reconstruction_module_recurrent_layer: IndRNN, reconstruction_module_conv_filters: [64, 64, 2], reconstruction_module_conv_kernels: [5, 3, 3], reconstruction_module_conv_dilations: [1, 2, 1], reconstruction_module_conv_bias: [true, true, false], reconstruction_module_recurrent_filters: [64, 64, 0], reconstruction_module_recurrent_kernels: [1, 1, 0], reconstruction_module_recurrent_dilations: [1, 1, 0], reconstruction_module_recurrent_bias: [true, true, false], reconstruction_module_time_steps: 8, reconstruction_module_conv_dim: 2, reconstruction_module_num_cascades: 1, segmentation_module: AttentionUNet, segmentation_module_channels: 32, segmentation_module_pooling_layers: 5
SEGNET	channels: 64, pooling_layers: 2, num_cascades: 5, final_layer_conv_dim: 2, final_layer_kernel_size: 3, final_layer_dilation: 1, final_layer_bias: False, final_layer_nonlinear: relu

7

Discussion

This thesis introduced novel approaches for accelerating Magnetic Resonance Imaging (MRI) using deep learning (DL) and multitask learning (MTL). Key contributions included the development of the Cascades of Independently Recurrent Inference Machines (CIRIM), a physics-informed DL network that enabled fast and robust MRI reconstruction, as well as the extensive evaluation of DL-based reconstruction across a wide range of diverse datasets, encompassing large-scale challenges and real-world clinical settings. The boundaries of conventional reconstruction paradigms were further extended by establishing a novel MTL framework for joint reconstruction and segmentation, effectively leveraging the inherent interdependencies of these tasks. Finally, practical implementations of the concepts presented in this thesis were realized in the Advanced Toolbox for Multitask Medical Imaging Consistency (ATOMMIC). Through extensive ablation studies and theoretical analyses, these developments effectively conceptualize the theoretical and practical framework of deep multitask learning (DMTL) for accelerating MRI.

7.1 On the importance of data consistency and deep networks in MRI reconstruction

In **Chapter 2**, we introduced the CIRIM, a sophisticated DL-based reconstruction network that balances computational efficiency without compromising reconstruction quality. The architectural innovation of cascading Recurrent Inference Machines (RIM) [1] effectively implemented a form of deep supervision, where each cascade performed unrolled optimization through a fixed number of iterations, known as time-steps in Recurrent Neural Networks (RNNs). Backpropagation through reduced time-steps per cascade ensured robust gradient flow while enabling inter-cascade information propagation. In that way, the risk of gradients accumulation throughout lengthy sequences was mitigated, thereby decreasing the likelihood of vanishing and exploding gradients. The CIRIM was theoretically motivated by the need to balance depth in RNNs with computational stability, targeting high reconstruction quality and fast reconstruction times. This theoretical foundation proved essential for subsequent developments throughout the thesis, particularly in scaling

to clinical applications (**Chapter 4**) and integrating subsequent analysis tasks, such as segmentation, into the reconstruction task through the cascades (**Chapter 5**).

The assessment of data consistency in physics-informed DL reconstruction networks showed that implicit enforcement, through gradient descent, as implemented in the CIRIM, was advantageous to explicit enforcement, through a formulated term, as in other physics-informed reconstruction networks, such as the Variational Network (VN) [2]. The implicit enforcement of data consistency allowed the CIRIM to adapt more effectively to inherent variations in MRI data, such as noise levels, undersampling patterns, and pathologies absent from the training data. The observed robustness proved to be particularly crucial in the later clinical evaluation, presented in **Chapter 4**.

Furthermore, the CIRIM demonstrated superior performance in reconstructing data from three diverse datasets: 3D T_1-weighted brain images, 2D FLAIR brain images, and 3D T_2-weighted knee images, with multiple acceleration factors (ranging from four to ten), and two distinct undersampling patterns (equidistant and random Gaussian). Notably, the CIRIM outperformed seven other DL-based reconstruction networks, as well as the Compressed Sensing (CS) reconstruction method. Generalization capabilities were demonstrated in reconstructing out-of-training-distribution 7.5 times accelerated FLAIR data (under random Poisson sampling), from patients with multiple sclerosis (MS) lesions. Although the CIRIM was trained on data from healthy subjects with different contrast (T_1) and different undersampling pattern (Gaussian), it demonstrated improved dealiasing and denoising capabilities, represented by the successful preservation of the contrast of the lesion relative to surrounding white matter and the higher signal-to-noise ratio (SNR) compared to other networks and CS.

Computational efficiency was another critical advancement that was later shown to be crucial clinically (**Chapter 4**). Fast reconstruction times and high reconstruction quality were balanced in the CIRIM by substituting the Gated Recurrent Unit (GRU) [3], used in the original RIM, with the Independently RNN [4]. This replacement led to an approximate 56% decrease in total network parameters, which consequently allowed an increased number of RIM cascades and thus increased reconstruction quality while maintaining rapid reconstruction times. The relationship between acceleration and computational complexity emerged as a key theme throughout the thesis. This relationship exhibits non-linear characteristics, suggesting the existence of an optimal acceleration factor that minimizes the total time-to-solution across both acquisition and reconstruction phases.

The findings presented in **Chapter 2** contributed significantly to our understanding of the intricate interaction between network complexity, data consistency, and robustness in DL-based MRI reconstruction and laid the groundwork for subsequent large-scale and clinical evaluations. While data consistency was essential for maintaining fidelity to the acquired data, it simultaneously imposed inherent limitations on recovering high-frequency details absent in the undersampled k-space data. This understanding directly informed the development of the MTL for joint reconstruction and segmentation approach, presented in **Chapter 5**. Nevertheless, factors such as the SNR, the number of coils, the choice of the undersampling pattern, and the diversity of the training dataset were found to play a critical role in assessing robustness and generalizability of DL-based reconstruction. These findings were further validated in the following chapters.

7.2 Robustness and generalization assessment of deep learning based MRI reconstruction

The theoretical developments presented in **Chapter 2** were extensively evaluated and validated in several contexts, from controlled experiments to real-world clinical settings in **Chapters 3** and **4**. Large-scale initiatives, such as the multi-coil MRI (MC-MRI) reconstruction challenge, presented in **Chapter 3**, and the fastMRI challenges [5, 6], provided in-depth insights into robustness and generalization in diverse acquisition scenarios. While these standardized benchmarks have been instrumental in establishing baseline performance metrics and identifying promising architectural approaches, the ultimate validation of DL-based MRI reconstruction required rigorous assessment in clinical settings. The transition from controlled to clinical evaluation, presented in **Chapter 4**, confronted DL-based reconstruction with the full complexity of the clinical practice, including diverse pathologies and the nuanced requirements for diagnostic interpretation.

7.2.1 Large-scale challenges

The MC-MRI reconstruction challenge validated the findings of **Chapter 2** on the importance of enforcing data consistency for achieving robust high reconstruction performance. Networks that enforced data consistency, implicitly or explicitly, demonstrated superior performance to other networks, such as baseline U-Nets, with cascaded networks and RNNs emerging as particularly effective approaches[1].

Significant insights were obtained regarding the limitations of conventional metrics in assessing reconstruction quality. While models trained on 12-coil configurations exhibited strong performance in terms of Structural Similarity Index (SSIM) [7], Peak Signal-to-Noise Ratio (PSNR), and Visual Information Fidelity (VIF) [8], when evaluated on 32-coil data visual inspection revealed substantial artifacts that would impact clinical utility. At the same time, the MC-MRI reconstruction challenge illuminated a fundamental challenge in the field: the possible discrepancy between optimization criteria and clinically relevant image quality metrics. Although quantitative metrics serve as valuable tools for preliminary evaluation, they may not fully represent the level of image quality necessary for clinical assessment, particularly crucial in scenarios that involve out-of-training-distribution data. This observation directly informed our clinical evaluation approach in **Chapter 4**, where we incorporated quantitative metrics and an expert radiologists' assessment.

The fastMRI challenges further expanded our understanding of generalizability in DL-based reconstruction networks and evaluation criteria. The 2019 fastMRI challenge [5] provided an unprecedented large-scale dataset containing roughly 1,594 clinical knee scans, representing the largest publicly accessible MRI dataset containing raw k-space data at that time. Remarkable performance was demonstrated in single-coil [9][2] and multi-coil reconstruction [10], with high SSIM scores accompanied by high rankings from five radiologists, of the total of seven who participated in the evaluation. However, the image quality in single-coil reconstruction was inferior to that in multi-coil reconstruction. This

[1] The top performing method submitted by our team (ResoNNance) was a baseline RIM, which can also be thought as a CIRIM with one cascade.

[2] The top performing method in single-coil reconstruction was the i-RIM, a fully invertible RIM submitted by our team (AImsterdam).

divergence in performance between single-coil and multi-coil reconstruction was further explored in our work on coil handling within ATOMMIC, as detailed in **Chapter 6**.

Crucial insights into the failures of DL-based reconstruction were also recognized in the 2019 fastMRI challenge. Instances where the models performed poorly were manually identified. However, it was shown that the DL-based reconstruction quality was always better than that of the CS baseline. Notably, there were no cases where a model completely failed or exhibited considerable image degradation, even in images with artifacts from metallic implants.

The second fastMRI challenge [6], in 2020, provided an even larger dataset of 7,299 brain scans, with emphasis on pathological assessment over general image quality metrics. The challenge's transfer learning track revealed critical limitations in cross-vendor generalization. Models trained on Siemens data were unable to generalize in reconstructing General Electric data, where even the best-performing models produced false anatomical structures or altered existing ones[3].

Subsequent analysis by Johnson et al. [11] further evaluated robustness in handling discrepancies between training datasets and real-world clinical. To that end, several perturbations were applied to the test set of the second fastMRI challenge. The applied perturbations included structural changes, simulation of mismatched SNR, adjustments to the number of coils, and variations in the undersampling pattern. Overall, the best-performing models of the challenge, without undergoing retraining, while exhibited resilience to coil configuration changes, demonstrated sensitivity to structural modifications and SNR variations. Interestingly, altering the sampling pattern by sampling more lines in the center of k-space enhanced the visibility of pathologies, highlighting the importance of choosing suitable sampling patterns in clinical settings. The extensive set of evaluation schemes for assessing robustness in Johnson et al.'s study also motivated part of our experiments within the ATOMMIC toolbox (**Chapter 6**).

7.2.2 Real world clinical evaluation

The clinical evaluation presented in **Chapter 4** represented a crucial step toward validating DL-based reconstruction in real-world scenarios. The CIRIM was shown to successfully generalize from training on T_1-weighted data from healthy subjects, due to the lack of fully sampled clinical data with neurological conditions, in reconstructing highly accelerated (twelve-fold) pathological FLAIR data of patients with various neurological deficits, such as stroke, MS, tumors, and Meniere's disease. The observed 30% improvement in SNR over CS, coupled with a five-fold reduction in reconstruction times, suggested significant potential for enabling fast and high-quality reconstruction in the clinical setting. These improvements directly assessed the challenges of CS reconstruction, identified in the introduction (Section 1.1.1), and demonstrated the practical value of the CIRIM.

The assessment of reconstruction quality by three neuroradiologists revealed additional insights into the subjective nature of image quality assessment, which were in line with the findings of the challenges (Section 7.2.1). The superior performance of the CIRIM was corroborated by the consensus among the two (interventional) neuroradiologists, particu-

[3]The ResoNNance submission from our team secured third place in the transfer track, utilizing a RIM. Nonetheless, it was rated lowest by radiologists due to poor performance in reconstructing General Electric data and the occurrence of false or modified anatomical structures.

larly in reducing artifacts (dealiasing) and clear visualization of pathologies. Conversely, the third neuroradiologist perceived the sharp image characteristics as blurring. This subjective view was ascribed to the specialization in pediatric neuroradiology, with this reader reporting that image quality at lower acceleration factors is preferable in the field of pediatrics.

These findings highlighted that the enhanced denoising capabilities of the CIRIM, also realized in the reconstruction of data with MS lesions in **Chapter 2**, can improve diagnosis by allowing the detection of subtle pathological features that might otherwise be obscured by noise. Despite the encouraging findings, the discrepancies in radiologists' perceptions of image quality highlighted the need for specialized training in interpreting DL-based reconstructed images while raised considerations about the impact of the reconstruction quality on subsequent analysis tasks. These considerations were further assessed in the development of MTL for joint reconstruction and segmentation (**Chapter 5**) and the development of ATOMMIC (**Chapter 6**).

7.3 Multitask Learning

Chapter 5 represented a significant theoretical advancement in our approach for accelerating MR imaging by incorporating the segmentation task into the reconstruction task, utilizing MTL [12]. Unlike traditional methods that handle tasks independently, sequentially, or merely simultaneously through a joint loss function [13], our proposed MTL for accelerated-MRI reconstruction and segmentation (MTLRS) framework treated the two tasks as intrinsically coupled computational problems that share underlying representational spaces.

MTLRS outperformed traditional methodologies, such as single-task, end-to-end, and joint, across multiple evaluation metrics in reconstructing and segmenting 7.5 times accelerated 3D FLAIR MRI data from MS patients with white matter lesions. The observed enhancements in SSIM and PSNR for reconstruction, coupled with improved Dice coefficients for tissue and lesion segmentation, suggested interconnectedness in the learned representations between the two tasks. Statistical analysis of the connection between reconstruction quality and segmentation accuracy validated the strong and significant association between these two tasks, revealed by Spearman's rank correlation. Visual inspection of the results identified more clearly defined lesion boundaries and enhanced contrast in MTRLS, further underscoring the synergistic nature of the learning processes within the framework. The insights gained from this study greatly influenced the development of ATOMMIC (**Chapter 6**), informing its integrated approach to multiple imaging tasks.

Additionally, the theoretical implications of our findings in **Chapter 5** provide insights into representation learning mechanisms within complex medical imaging tasks. The success of MTLRS in leveraging shared information challenges the traditional paradigm of treating reconstruction and analysis tasks as independent processes. Our results suggest the existence of rich, interconnected latent spaces that simultaneously encode structural and semantic information in medical imaging data. The emergent regularization patterns arising from task interdependence serve a dual purpose: they enhance computational efficiency by eliminating the redundancy of separate task execution while contributing to robustness through intrinsic constraints imposed by shared representational structures.

The MTLRS capability to enforce consistency across both data and task domains established a theoretical foundation for deep multitask learning (DMTL). By connecting multiple single-task DL networks through MTL, DMTL provides empirical evidence for the hypothesis that shared representation learning can simultaneously optimize multiple objectives in medical imaging.

7.4 The Advanced Toolbox for Medical Imaging Consistency

Building directly upon the theoretical foundations established in the previous chapters, in **Chapter 6** we presented the Advanced Toolbox for Medical Imaging Consistency (ATOMMIC). ATOMMIC aimed to systematically address crucial challenges in developing and deploying DMTL methodologies for accelerating MR imaging, from reconstruction to analysis. The efficacy of ATOMMIC was assessed through a rigorous evaluation of twenty-five DL models in multiple datasets and imaging tasks.

The superior performance of physics-informed networks in the reconstruction task, particularly of the CIRIM and the VN, validated the fundamental role of data consistency in MRI reconstruction, established throughout the thesis. The segmentation analysis across diverse anatomical regions and pathologies, including intracranial tumors, knee structures, gliomas, and ischemic lesions, not only demonstrated the versatility of ATOMMIC, but also confirmed the discrepancy observed in segmentation performance metrics, similar to the discrepancy between optimization criteria and clinically relevant image quality metrics in the reconstruction task (**Chapter 3**).

ATOMMIC's capabilities were further evaluated through its application to quantitative parameter map estimation from multi-echo MRI data. Expanding on the innovations of Zhang et al. [14], the quantitative RIM demonstrated high precision in estimating quantitative parameter maps, not only confirming the importance of incorporating data consistency also in the quantitative MRI (qMRI) task but additionally showcasing ATOMMIC's potential to substantially reduce qMR imaging times. This advancement was directly related to the clinical needs identified in **Chapter 4**, particularly regarding faster treatment decisions and improved patient outcomes in neurological applications. For example, the successful reconstruction of R_2^* maps can provide crucial information on (local) iron deposition in the brain, a biomarker associated with aging and neurodegenerative conditions, including MS, Parkinson's, and Alzheimer's disease [15].

Central to ATOMMIC's design was its multifaceted approach to consistency, synthesizing insights from our previous research. Consistency was encompassed in three ways: data consistency, ensuring that the reconstructed images accurately reflect the acquired data, based on the principles established in **Chapter 2**; task consistency, allowing for a comprehensive evaluation of individual tasks and DMTL, extending the findings of **Chapter 5**; and workflow consistency, standardizing model training and evaluation processes, incorporating insights from **Chapter 4**. In this way, ATOMMIC aimed to address one of the most significant challenges in the field: generalizability in diverse heterogeneous datasets and patient populations.

7.5 Limitations & Future directions of Deep Multitask Learning for accelerating MRI

Moving from the findings presented in this thesis, future DMTL directions for accelerating MRI encompass several interconnected challenges. At their core, successful DMTL implementations demand a balance between shared and task-specific approaches while addressing fundamental challenges in data availability and computational efficiency.

The theoretical landscape of DMTL presents particularly intriguing challenges concerning the behavior and stability of shared latent representations in high-dimensional feature spaces. The interaction between multiple tasks introduces complex optimization landscapes that necessitate sophisticated theoretical frameworks for analysis. Understanding these dynamics will be crucial when examining the behavior of a DMTL system under various conditions, particularly in scenarios characterized by data sparsity or extreme noise conditions, such as the scenarios presented in **Chapters 2** and **6**. Future work should focus on developing formal metrics to quantify task-relatedness and establish robustness.

Robustness can be enforced through data consistency. The integration of more complex constraints within DMTL represents a natural extension of the principles established in **Chapter 2**. While current approaches incorporate basic physical principles, future frameworks could benefit from more sophisticated regularization schemes capable of capturing complex MRI phenomena. For instance, modeling B_0 inhomogeneity effects and gradient nonlinearities could lead to enhanced robustness and more interpretable models.

At the same time, a significant barrier lies in data scarcity, a limitation identified throughout our work. Our studies in the clinical assessment of DL-based reconstruction, in **Chapter 4**, and MTL, in **Chapter 5**, highlighted the need for large annotated datasets. For example, to establish MTLRS as a proof of concept and due to the lack of fully sampled reference MRI data accompanied by expert annotations, we were compelled to synthesize a multi-coil dataset from a relatively small group of MS patients. Although this synthetic dataset allowed us to introduce the MTLRS framework, it emphasized the need for more extensive annotated datasets.

Semi-supervised [16] and self-supervised [17] learning approaches have shown potential in learning robust representations from limited labeled data; yet, their theoretical foundations in the context of DMTL require further investigation. Another approach to address data scarcity could lie in leveraging ATOMMIC's flexibility in handling different data types. ATOMMIC could be employed to develop robust privacy-preserving techniques that allow model training across multiple institutions without sharing raw patient data.

However, the integration of data from multiple sources poses additional challenges. Private datasets, especially when comprised of raw MRI data, are commonly stored in proprietary formats, due to privacy concerns, and their usage often requires expert knowledge, due to their complex pre-processing pipelines. To that end, efforts to develop standardized pipelines and data formats, such as the ISMRM raw data format (ISMRMRD) [18], are necessary but merit further research.

From a computational perspective, the scalability of DMTL also presents significant challenges that would need to be addressed through efficient training strategies and computational resource management. Exploring efficient optimization algorithms for multi-objective learning, coupled with adaptive architectural solutions that dynamically allocate

resources based on task complexity and data characteristics, could optimize DMTL. Nevertheless, such advancements extend beyond immediate clinical applications to fundamental advances in understanding multi-objective learning and optimization in complex, high-dimensional spaces.

Careful consideration should also be given to potential sources of bias when developing DMTL approaches. Although incorporating longitudinal information presents promising opportunities, the temporal relationships between different data modalities should be carefully evaluated to avoid systematic biases. Such biases, especially if models are trained on datasets that do not adequately capture diverse groups, raise serious concerns that transcend technical aspects, touching on the core issues of equity and inclusion within healthcare. Future research must emphasize diversity in data acquisition and model training, encompassing not only a broad spectrum of medical cases and diseases, but also heterogeneous demographic populations. Additionally, the development of novel approaches that can learn effectively from decentralized data sources would be necessary for preserving data privacy.

Currently, the integration of DL into clinical practice has reached a notable milestone, with major MRI manufacturers already incorporating DL-based reconstruction methods directly into their systems. However, the findings presented in this thesis suggest that the path to advancement lies not exclusively in complete automation, but rather in strategic improvements to accuracy, efficiency, and workflow optimization [19]. Furthermore, this thesis emphasized the need for a collaborative approach, where computer scientists actively incorporate radiologists' expertise during system development, while radiologists develop proficiency in DL principles to optimize clinical practice. This synergistic collaboration has the potential to transform personalized medicine, as coordinated efforts of computer scientists, medical physicists, and clinicians can translate these theoretical advances into tangible improvements in patient care, fundamentally advancing modern medicine.

References

[1] Kai Lønning, Patrick Putzky, Jan-Jakob Sonke, Liesbeth Reneman, Matthan WA Caan, and Max Welling. Recurrent inference machines for reconstructing heterogeneous MRI data. *Medical image analysis*, 53:64–78, 2019.

[2] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P. Recht, Daniel K. Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, 2018.

[3] Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *CoRR*, abs/1406.1078, 2014.

[4] Shuai Li, Wanqing Li, Chris Cook, Ce Zhu, and Yanbo Gao. Independently recurrent neural network (indrnn): Building a longer and deeper rnn. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5457–5466, 2018.

[5] Florian Knoll, Tullie Murrell, Anuroop Sriram, Nafissa Yakubova, Jure Zbontar, Michael Rabbat, Aaron Defazio, Matthew J. Muckley, Daniel K. Sodickson, C. Lawrence

Zitnick, and Michael P. Recht. Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge. *Magnetic Resonance in Med*, 84(6):3054–3070, December 2020.

[6] Matthew J. Muckley, Bruno Riemenschneider, Alireza Radmanesh, Sunwoo Kim, Geunu Jeong, Jingyu Ko, Yohan Jun, Hyungseob Shin, Dosik Hwang, Mahmoud Mostapha, Simon Arberet, Dominik Nickel, Zaccharie Ramzi, Philippe Ciuciu, Jean-Luc Starck, Jonas Teuwen, Dimitrios Karkalousos, Chaoping Zhang, Anuroop Sriram, Zhengnan Huang, Nafissa Yakubova, Yvonne W. Lui, and Florian Knoll. Results of the 2020 fastMRI Challenge for Machine Learning MR Image Reconstruction. *IEEE Transactions on Medical Imaging*, 40(9):2306–2317, 2021.

[7] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13:600–612, 2004.

[8] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, Feb 2006.

[9] Patrick Putzky, Dimitrios Karkalousos, Jonas Teuwen, Nikita Miriakov, Bart Bakker, Matthan Caan, and Max Welling. i-RIM applied to the fastMRI challenge, October 2019.

[10] Nicola Pezzotti, Sahar Yousefi, Mohamed S. Elmahdy, Jeroen Hendrikus Fransiscus Van Gemert, Christophe Schuelke, Mariya Doneva, Tim Nielsen, Sergey Kastryulin, Boudewijn P. F. Lelieveldt, Matthias J. P. Van Osch, Elwin De Weerdt, and Marius Staring. An Adaptive Intelligence Algorithm for Undersampled Knee MRI Reconstruction. *IEEE Access*, 8:204825–204838, 2020.

[11] Patricia M. Johnson, Geunu Jeong, Kerstin Hammernik, Jo Schlemper, Chen Qin, Jinming Duan, Daniel Rueckert, Jingu Lee, Nicola Pezzotti, Elwin De Weerdt, Sahar Yousefi, Mohamed S. Elmahdy, Jeroen Hendrikus Franciscus Van Gemert, Christophe Schülke, Mariya Doneva, Tim Nielsen, Sergey Kastryulin, Boudewijn P. F. Lelieveldt, Matthias J. P. Van Osch, Marius Staring, Eric Z. Chen, Puyang Wang, Xiao Chen, Terrence Chen, Vishal M. Patel, Shanhui Sun, Hyungseob Shin, Yohan Jun, Taejoon Eo, Sewon Kim, Taeseong Kim, Dosik Hwang, Patrick Putzky, Dimitrios Karkalousos, Jonas Teuwen, Nikita Miriakov, Bart Bakker, Matthan Caan, Max Welling, Matthew J. Muckley, and Florian Knoll. Evaluation of the Robustness of Learned MR Image Reconstruction to Systematic Deviations Between Training and Test Data for the Models from the fastMRI Challenge. In Nandinee Haq, Patricia Johnson, Andreas Maier, Tobias Würfl, and Jaejun Yoo, editors, *Machine Learning for Medical Image Reconstruction*, volume 12964, pages 25–34. Springer International Publishing, Cham, 2021. Series Title: Lecture Notes in Computer Science.

[12] Rich Caruana. Multitask Learning. *Machine Learning*, 28:41–75, 1997.

[13] Jonas Adler, Sebastian Lunz, Olivier Verdier, Carola-Bibiane Schönlieb, and Ozan Öktem. Task adapted reconstruction for inverse problems. *Inverse Problems*, 38(7):075006, may 2022.

[14] Chaoping Zhang, Dimitrios Karkalousos, Pierre-Louis Bazin, Bram F. Coolen, Hugo Vrenken, Jan-Jakob Sonke, Birte U. Forstmann, Dirk H.J. Poot, and Matthan W.A. Caan. A unified model for reconstruction and R2* mapping of accelerated 7T data using the quantitative recurrent inference machine. *NeuroImage*, 264:119680, 2022.

[15] Steven D Brass, Nan-kuei Chen, Robert V Mulkern, and Rohit Bakshi. Magnetic resonance imaging of iron deposition in neurological disorders. *Topics in Magnetic Resonance Imaging*, 17(1):31–40, 2006.

[16] Arjun D. Desai, Batu M. Ozturkler, Christopher M. Sandino, Robert Boutin, Marc Willis, Shreyas Vasanawala, Brian A. Hargreaves, Christopher Ré, John M. Pauly, and Akshay S. Chaudhari. Noise2Recon: Enabling SNR-robust MRI reconstruction with semi-supervised and self-supervised learning. *Magn Reson Med*, 90(5):2052–2070, November 2023.

[17] Burhaneddin Yaman, Seyed Amir Hossein Hosseini, Steen Moeller, Jutta Ellermann, Kâmil Uğurbil, and Mehmet Akçakaya. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magnetic Resonance in Medicine*, 84(6):3172–3191, 2020.

[18] Souheil J. Inati, Joseph D. Naegele, Nicholas R. Zwart, Vinai Roopchansingh, Martin J. Lizak, David C. Hansen, Chia-Ying Liu, David Atkinson, Peter Kellman, Sebastian Kozerke, Hui Xue, Adrienne E. Campbell-Washburn, Thomas S. Sørensen, and Michael S. Hansen. ISMRM Raw data format: A proposed standard for MRI raw datasets. *Magnetic Resonance in Medicine*, 77(1):411–421, 2017.

[19] Eric J. Topol. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*, 25(1):44–56, January 2019.

Summary

The intricate balance between acquisition speed and image quality in Magnetic Resonance Imaging (MRI) presents a fundamental challenge. Imaging times can be significantly reduced by accelerating the acquisition process. However, higher acceleration factors generally lead to degradation of image quality. Low image quality increases the risk of inaccurate analysis and, subsequently, misdiagnosis. This thesis presented a comprehensive assessment of the integration of deep learning (DL) with multitask learning (MTL) and introduced a unified deep multitask learning (DMTL) framework to accelerate the acquisition and analysis of MRI.

Chapter 2 introduced the Cascades of Independently Recurrent Inference Machines (CIRIM), a novel DL reconstruction network that adeptly balances network complexity and robustness. The cascading architecture of the CIRIM allowed for enhanced dealiasing and denoising capabilities while maintaining stable gradient calculations, despite being a deep Recurrent Neural Network (RNN). Through an extensive evaluation, the CIRIM demonstrated superior performance against seven other state-of-the-art DL approaches and the clinical standard reconstruction method, Compressed Sensing (CS). The evaluation included diverse MRI datasets, encompassing T_1-weighted, T_2-weighted, and FLAIR brain and knee scans, different undersampling schemes, and varying acceleration factors. In addition, robustness and generalization of DL-based reconstruction was tested on reconstructing data unseen during training, specifically from patients with multiple sclerosis (MS) lesions. The CIRIM successfully preserved the contrast of the lesion relative to surrounding white matter and achieved a higher signal-to-noise ratio (SNR) compared to other methods and CS, highlighting the promising potential for clinical integration.

Furthermore, in **Chapter 2** data consistency enforcement in physics-informed DL reconstruction networks was assessed. It was shown that the CIRIM's implicit data consistency enforcement, through gradient descent, was advantageous to explicit enforcement, through a designed term, common in other physics-informed DL-based reconstruction networks, such as the Variational Network. This finding provided valuable insights into designing optimal MRI reconstruction strategies in the following chapters.

Chapter 3 detailed our participation in the Multi-Coil MRI Reconstruction Challenge (MC-MRI), a worldwide collaborative initiative that extensively evaluated the generalization capabilities and robustness of DL-based reconstruction networks. The experimentation with high-resolution 3D T_1-weighted brain MRI scans across 12-coil and 32-coil configurations highlighted the superiority of physics-informed networks featuring cascaded and recurrent architectures. At the same time, crucial insights regarding cross-coil generalization were obtained. Specifically, while models trained on 12-coil and tested on 32-coil configurations exhibited strong performance in quantitative metrics, such as Structural Similarity Index, Peak Signal-to-Noise Ratio, and Visual Information Fidelity, visual inspection revealed substantial artifacts that would impact clinical utility. This observation highlighted the

fundamental challenge of the possible discrepancy between optimization criteria and clinically relevant image quality metrics.

Chapter 4 evaluated the application of the CIRIM in a real-world clinical setting. The CIRIM was compared with CS in reconstructing highly accelerated (twelve-fold) 3D FLAIR MRI scans of patients with neurological deficits, such as stroke, MS, tumors, and Meniere's disease. Reconstruction quality was quantitatively assessed and subjectively scored by three expert neuroradiologists in terms of artifacts, sharpness, anatomical clarity, diagnostic confidence, and contrast. Despite little variation in expert opinions, the CIRIM exhibited statistically significant improvements, including enhanced artifact reduction, superior anatomical conspicuity, and a substantial increase in SNR, while only marginally reduced contrast ratio. Notably, the CIRIM's five-fold increase in reconstruction speed over CS underscored its potential for significantly accelerating imaging times.

Chapter 5 introduced a sophisticated MTL framework for simultaneous reconstruction and segmentation (MTLRS) of accelerated MRI data, with a specific focus on brain lesions in MS patients. MTLRS employed a unique cascading architecture, in which the CIRIM and a segmentation network shared information through hidden states. This approach allowed for leveraging shared features as inductive bias to enhance overall performance. Through comparative testing against conventional pre-trained, sequential, end-to-end, and joint methodologies, MTLRS was shown to outperform all other methods in both the reconstruction and the segmentation task. Furthermore, a strong and significant correlation was found between the performance of the two tasks, indicating the advantages of addressing interrelated tasks in a unified MTL manner.

Chapter 6 presented the Advanced Toolbox for Multitask Medical Imaging Consistency (ATOMMIC), an open-source toolbox designed to streamline AI applications for accelerated MRI reconstruction and analysis. Within ATOMMIC, twenty-five DL models were rigorously evaluated in eight diverse publicly available datasets, for the tasks of reconstruction, segmentation, quantitative parameter map estimation, and MTL for joint reconstruction and segmentation. ATOMMIC demonstrated versatility in the integration of multiple tasks while maintaining consistency across data, tasks, and workflows.

In conclusion, the research presented in this thesis contributed to the field of accelerated MRI reconstruction and analysis. The development of the CIRIM demonstrated significant improvements in reconstruction times, while allowing for high diagnostic image quality. The subsequent introduction of the MTLRS framework and its validation in simultaneous reconstruction and segmentation revealed the synergistic benefits of MTL in medical imaging. The creation and comprehensive validation of the ATOMMIC toolbox represented a step toward standardizing AI applications in medical imaging. These contributions collectively conceptualized the DMTL framework and its potential for accelerating MR imaging.

Samenvatting

De complexe balans tussen acquisitiesnelheid en beeldkwaliteit in *Magnetic Resonance Imaging* (MRI) vormt een fundamentele uitdaging. Scantijden kunnen aanzienlijk worden verminderd door het acquisitieproces te versnellen. Hogere versnellingsfactoren leiden echter in het algemeen tot verslechtering van de beeldkwaliteit. Een lage beeldkwaliteit maakt beeldanalyse vatbaarder voor onnauwkeurigheden en verhoogt daarmee het risico op misdiagnoses. Dit proefschrift bevat een uitgebreide evaluatie van de integratie van *deep learning* (DL) met *multitask learning* (MTL) en introduceert een geïntegreerd *deep multitask learning* (DMTL) raamwerk voor het versnellen van MRI-acquisitie en -analyse.

Hoofdstuk 2 introduceerde de *Cascades of Independently Recurrent Inference Machines* (CIRIM), een nieuw DL-reconstructienetwerk dat de balans vindt tussen netwerkcompexiteit en robuustheid. Door middel van cascades zorgt de CIRIM voor verbeterde mogelijkheden voor het verwijderen van *aliasing* en ruis, terwijl stabiele gradiëntberekeningen behouden blijven in het diepe *Recurrent Neural Network* (RNN). Uitgebreide validatie toonde de superioriteit van CIRIM aan ten opzichte van zeven andere *state-of-the-art* DL-methoden en de klinische reconstructiemethode, *Compressed Sensing* (CS). Deze validatie omvatte verschillende MRI-datasets, waaronder T_1-gewogen, T_2-gewogen, en FLAIR hersen- en kniescans met verschillende manieren van onderbemonsteren bij een variërend bereik aan versnellingsfactoren. De consistente resultaten van CIRIM bij het reconstrueren van data van patiënten met multiple sclerose (MS) laesies, die niet in de trainingsdataset waren opgenomen, onderstreept de potentie voor klinische implementatie.

Bovendien werd in **Hoofdstuk 2** de handhaving van dataconsistentie in natuurkundig-geïnformeerde DL-reconstructienetwerken beoordeeld. Er werd aangetoond dat de impliciete handhaving van dataconsistentie door CIRIM, via een iteratief netwerk, voordeliger was dan expliciete handhaving via een ontworpen term, wat gebruikelijk is in andere natuurkundig-geïnformeerde DL-reconstructienetwerken, zoals het *Variational Network*. Deze bevinding leverde waardevolle inzichten op voor het ontwerpen van optimale MRI-reconstructiestrategieën in de volgende hoofdstukken.

Hoofdstuk 3 beschreef onze deelname aan de *Multi-Coil MRI Reconstruction Challenge* (MC-MRI), een wereldwijd samenwerkingsinitiatief, waarin uitgebreid werd geëvalueerd hoe generaliseerbaar en robuust DL-gebaseerde reconstructienetwerken zijn. De experimenten met hoge-resolutie 3D T_1-gewogen hersen-MRI-scans met 12-kanaals- en 32-kanaalsspoelen benadrukten de superioriteit van natuurkundig-geïnformeerde netwerken met cascades en iteratieve architecturen. Ook werd duidelijk hoe netwerken wel en niet toepasbaar waren in data met meer of minder spoelelementen. Modellen getraind op 12-kanaals en getest op 32-kanaals configuraties lieten goede prestaties zien in kwantitatieve metrieken, zoals de *Structural Similarity Index*, *Peak Signal-to-Noise Ratio*, en *Visual Information Fidelity*. Toch onthulde visuele inspectie substantiële artefacten die de klinische toepasbaarheid beïnvloeden. Deze observatie benadrukte de fundamentele

uitdaging van mogelijke discrepantie tussen optimalisatiecriteria en klinisch relevante beeldkwaliteitsmetrieken.

Hoofdstuk 4 evalueerde de toepassing van de CIRIM in een realistische klinische omgeving. Het model werd vergeleken met CS bij het reconstrueren van sterk versnelde (twaalfvoudige) 3D FLAIR MRI-scans van patiënten met neurologische aandoeningen, zoals beroerte, MS, tumoren en de ziekte van Menière. De reconstructiekwaliteit werd kwantitatief beoordeeld en subjectief gescoord door drie expert-neuroradiologen op het gebied van artefacten, scherpte, anatomische duidelijkheid, diagnostische zekerheid en contrast. Ondanks kleine variaties in expertbeoordelingen, liet de CIRIM statistisch significante verbeteringen zien, waaronder verbeterde artefactreductie, superieure anatomische zichtbaarheid, en een substantiële toename in SNR, met slechts een beperkte reductie in contrastratio. Met name de vijfvoudige toename in reconstructiesnelheid van CIRIM ten opzichte van CS onderstreepte de potentie voor het significant versnellen van scantijden.

Hoofdstuk 5 introduceerde een geavanceerd MTL-raamwerk voor gelijktijdige reconstructie en segmentatie (MTLRS) van versnelde MRI-data, met een specifieke focus op hersenlaesies bij MS-patiënten. MTLRS maakte gebruik van een unieke cascade-architectuur, waarin het CIRIM en een segmentatienetwerk informatie deelden via *hidden states*. Deze aanpak maakte het mogelijk om gedeelde kenmerken te benutten als inductieve bias om de algehele prestaties te verbeteren. Door middel van vergelijkende tests met conventionele vooraf getrainde, sequentiële, *end-to-end* en gezamenlijke methodologieën, werd aangetoond dat MTLRS beter presteerde dan alle andere methoden in zowel de reconstructie- als de segmentatietaak. Bovendien werd een significante en sterke correlatie gevonden tussen de prestaties van de twee taken, wat wijst op de voordelen van het gezamenlijk aanpakken van onderling gerelateerde taken.

Hoofdstuk 6 presenteerde de *Advanced Toolbox for Multitask Medical Imaging Consistency* (ATOMMIC), een *open-source* toolbox ontworpen om AI-toepassingen in versnelde MRI-reconstructie en -analyse te stroomlijnen. Binnen ATOMMIC werden vijfentwintig DL-modellen geëvalueerd voor de taken van reconstructie, segmentatie, kwantitatieve parameterschatting en MTL voor gezamenlijke reconstructie en segmentatie, over acht diverse openbaar beschikbare datasets. ATOMMIC toonde veelzijdigheid in de integratie van meerdere taken met behoud van robuuste consistentie over data, taken en workflows.

Concluderend heeft het onderzoek gepresenteerd in dit proefschrift bijgedragen aan het veld van versnelde MRI-reconstructie en -analyse. De ontwikkeling van het CIRIM toonde significante verbeteringen in reconstructietijd met behoud van hoge diagnostische beeldkwaliteit. De daaropvolgende introductie van het MTLRS-raamwerk en de validatie hiervan in gelijktijdige reconstructie- en segmentatietaken onthulde de synergetische voordelen van MTL in medische beeldvorming. De ontwikkeling en uitgebreide validatie van de ATOMMIC-toolbox vertegenwoordigde een stap richting het standaardiseren van AI-toepassingen in medische beeldvorming. Deze bijdragen conceptualiseerden gezamenlijk het DMTL-raamwerk en de potentie hiervan voor het versnellen van MR-beeldvorming.

List of Publications

In this thesis

Karkalousos, D., Noteboom, S., Hulst, H. E., Vos, F. M., & Caan, M. W. A. (2022). Assessment of data consistency through cascades of independently recurrent inference machines for fast and robust accelerated MRI reconstruction. Physics in Medicine and Biology, 67(12), 10.1088/1361-6560/ac6cc2.

Beauferris, Y., Teuwen, J., **Karkalousos, D.**, Moriakov, N., Caan, M., Yiasemis, G., Rodrigues, L., Lopes, A., Pedrini, H., Rittner, L., Dannecker, M., Studenyak, V., Gröger, F., Vyas, D., Faghih-Roohi, S., Kumar Jethi, A., Chandra Raju, J., Sivaprakasam, M., Lasby, M., Nogovitsyn, N., … Souza, R. (2022). Multi-Coil MRI Reconstruction Challenge-Assessing Brain MRI Reconstruction Models and Their Generalizability to Varying Coil Configurations. Frontiers in neuroscience, 16, 919186. 10.3389/fnins.2022.919186.

Liebrand, L.C.*, **Karkalousos, D.***, Poirion, É., Emmer, B. J., Roosendaal, S. D., Marquering, H. A., Majoie, C. B. L. M., Savatovsky, J., & Caan, M. W. A. (2024). Deep learning for efficient reconstruction of highly accelerated 3D FLAIR MRI in neurological deficits. Magma (New York, N.Y.), 10.1007/s10334-024-01200-8. Advance online publication. *Contributed equally

Karkalousos, D., Išgum, I., Marquering, H. & Caan, M.W. (2024, January). MultiTask Learning for accelerated-MRI Reconstruction and Segmentation of Brain Lesions in Multiple Sclerosis. In Medical Imaging with Deep Learning (pp. 991-1005). PMLR.

Karkalousos, D., Išgum, I., Marquering, H. A., & Caan, M. W. A. (2024). ATOMMIC: An Advanced Toolbox for Multitask Medical Imaging Consistency to facilitate Artificial Intelligence applications from acquisition to analysis in Magnetic Resonance Imaging. Computer Methods and Programs in Biomedicine, 256, 108377. 10.1016/j.cmpb.2024.108377.

Other publications

Muckley, M. J., Riemenschneider, B., Radmanesh, A., Kim, S., Jeong, G., Ko, J., Jun, Y., Shin, H., Hwang, D., Mostapha, M., Arberet, S., Nickel, D., Ramzi, Z., Ciuciu, P., Starck, J. L., Teuwen, J., **Karkalousos, D.**, Zhang, C., Sriram, A., Huang, Z., … Knoll, F. (2021). Results of the 2020 fastMRI Challenge for Machine Learning MR Image Reconstruction. IEEE transactions on medical imaging, 40(9), 2306–2317. 10.1109/TMI.2021.3075856.

Johnson, P. M., Jeong, G., Hammernik, K., Schlemper, J., Qin, C., Duan, J., Rueckert, D., Lee, J., Pezzotti, N., de Weerdt, E., Yousefi, S., Elmahdy, M. S., van Gemert, J. H. F., Schülke, C., Doneva, M., Nielsen, T., Kastryulin, S., Lelieveldt, B. P. F., van Osch, M. J. P., …,

Putzky, P., **Karkalousos, D.**, Teuwen, J., Miriakov, N., Bart Bakker, Caan, M.W.A., Welling, M., Muckley, M.J., & Knoll, F. (2021). Evaluation of the robustness of learned MR image reconstruction to systematic deviations between training and test data for the models from the fastMRI challenge. In Machine Learning for Medical Image Reconstruction: 4th International Workshop, MLMIR 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings 4 (pp. 25-34). Springer International Publishing. 10.1007/978-3-030-88552-6_3.

Yiasemis, G., Moriakov, N., **Karkalousos, D.**, Caan, M., & Teuwen, J. (2022). DIRECT: Deep Image REConstruction Toolkit. Journal of Open Source Software, 7(73), 4278. 10.21105/joss.04278.

Zhang, C., **Karkalousos, D.**, Bazin, P. L., Coolen, B. F., Vrenken, H., Sonke, J. J., Forstmann, B. U., Poot, D. H. J., & Caan, M. W. A. (2022). A unified model for reconstruction and $R2^*$ mapping of accelerated 7T data using the quantitative recurrent inference machine. NeuroImage, 264, 119680. 10.1016/j.neuroimage.2022.119680.

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.

Karkalousos, D., Lønning, K., Hulst, H. E., Dumoulin, S. O., Sonke, J. J., Vos, F. M., & Caan, M. W. (2020). Reconstructing unseen modalities and pathology with an efficient Recurrent Inference Machine. arXiv preprint arXiv:2012.07819.

Portfolio

Name PhD student: Dimitrios Karkalousos PhD period: December 2019 to December 2023 Names of PhD supervisors & co-supervisors: dr. ir. M. W. A. Caan, prof. dr. I. Išgum, prof. dr. H. A. Marquering		
	Year	ECTS
1. PhD training		
General courses		
AMC World of Science	2020	0.5
Writing a Scientific Paper	2021	1.5
VERSA: Video Games for PhD Soft Skills Training	2021	0.5
Project Management	2021	0.7
Specific courses		
Pulse Programming	2021	1.5
MRI - basic understanding for (bio)medical researchers	2021	1.0
Seminars, workshops and master classes		
MR Physics research group meetings, Radiology (weekly)	2019-2023	4.0
Cardiovascular Engineering research group meetings, BMEP (weekly)	2019-2023	4.0
Ischemic stroke research group meetings, BMEP (monthly)	2019-2023	2.0
Machine Learning reading group meetings, BMEP (bi-weekly)	2019-2023	1.0
Bernstein MRI reading group meetings, BMEP & Radiology (monthly)	2019-2021	0.7
AI in MRI reconstruction reading group meetings, BMEP (weekly)	2020-2022	2.0
Moore lectures, BMEP (monthly)	2021-2023	1.0
Presentations		
Oral		
Efficient and Robust Reconstruction using the Recurrent Inference Machine, European Society for Magnetic Resonance in Medicine and Biology, Rotterdam, Netherlands	2019	0.3

	Year	ECTS
Efficient and Robust accelerated MRI Reconstruction using the Recurrent Inference Machine, Amsterdam Neuroscience annual meeting, Amsterdam, Netherlands	2019	0.3
Recurrent Inference Machine applied to the Multi-channel MR Image Reconstruction Challenge (Calgary-Campinas), Medical Imaging with Deep Learning, Online	2020	0.3
Recurrent Variational Inference for fast and robust reconstruction of accelerated FLAIR MRI in Multiple Sclerosis, International Society for Magnetic Resonance in Medicine Benelux, Arnhem, Netherlands	2022	0.3
Poster		
Recurrent Variational Inference for fast and robust reconstruction of accelerated FLAIR MRI in Multiple Sclerosis, International Society for Magnetic Resonance in Medicine, London, United Kingdom	2022	0.3
MultiTask Learning for accelerated-MRI Reconstruction and Segmentation of Brain Lesions in Multiple Sclerosis, Medical Imaging with Deep Learning, Nashville, United States of America	2023	0.3
The Advanced Toolbox for Multitask Medical Imaging Consistency (ATOMMIC): A framework to facilitate Deep Learning in Magnetic Resonance Imaging, Medical Imaging with Deep Learning, Paris, France	2024	0.3
(Inter)national conferences		
NeurIPS, Vancouver, Canada	2019	1.3
ISMRM Benelux Chapter meeting, Arnhem, Netherlands	2020	0.4
Medical Imaging with Deep Learning, Online	2020	1.3
Amsterdam Neuroscience Annual Meeting, Online	2020	0.4
ESO-WSO Conference, Online	2020	
NeurIPS, Online	2020	1.3
CONTRAST Consortium meeting, 09/12, Online	2020	0.4
ISMRM Benelux Chapter meeting, Online	2021	0.4
GIDRM, The promises and dark sides of Artificial Intelligence in NMR, MRI and Neuroscience, Online	2021	0.4
Medical Imaging with Deep Learning, Online	2021	1.3
NeurIPS, 06/12 – 14/12, Online	2021	1.3
ISMRM Benelux, Maastricht, Netherlands	2022	0.4

	Year	ECTS
ISMRM, London, United Kingdom	2022	1.3
RNG Research Day, Amsterdam, Netherlands	2022	0.4
Amsterdam Neuroscience Annual Meeting, Amsterdam, Netherlands	2022	0.4
Medical Imaging with Deep Learning, Nashville, United States of America	2023	1.3
Medical Imaging with Deep Learning, Paris, France	2024	1.3
Other		
Contributed on the open-source project "Deep Image Reconstruction Toolkit (DIRECT)", https://github.com/directgroup/direct	2020	1.0
Developed a Deep Learning reconstruction tool, "MRI Data Consistency", https://github.com/wdika/mridc	2021	1.0
Developed the "Advanced Toolbox for Multitask Medical Imaging Consistency (ATOMMIC)", https://github.com/wdika/atommic	2023	1.0
2. Teaching		
Tutoring, Mentoring		
Advanced medical image processing (4604MM109Y)	10/2020	0.5
Informatie in Medische Beelden	11/2020	0.5
Medical AI	10/2022	0.5
Supervising		
Kevin Pancras, *Evaluating the Robustness of the Recurrent Inference Machine Based on MRI Lesion Data from Multiple Sclerosis Patients*, Research Thesis for Master in Mechanical Engineering, TU Delft, Netherlands	2020–2021	1.0
Samuel van Gurp, *Artifact suppression of machine learning based reconstructions of accelerated MR images*, Research end project for Bachelor in Physics and Astronomy, UvA, Netherlands	2021–2021	1.0
Lysander de Jong, *Joint Reconstruction & Segmentation for accelerated MR imaging using Deep Learning*, Research Thesis for Master in AI, UvA, Netherlands	2021-2022	1.0
Jerke van den Berg, *Accelerated MR Angiography Reconstruction using Deep Learning (3D)*, Research Thesis for Master in AI, UvA, Netherlands	2021-2022	1.0

	Year	ECTS
Iason Skylitsis, *Evaluation of Supervision in Deep Learning networks for accelerated-MRI reconstruction*, Research end project for Bachelor in Informatics and Computer Engineering, University of West Attica, Greece	2022-2023	1.0
Maurice Kingma, *Accelerated high-resolution and motion-robust 3D volume reconstruction for neonatal MRI*, Research Thesis for Master in Medical Informatics, UvA, Netherlands	2022-2023	1.0
3. Parameters of Esteem		
Awards and Prizes		
Won the fastMRI 2019 Challenge on accelerated MRI Reconstruction of single-coil data.	2019	0.2
Won the Multi-channel MR Image Reconstruction 2020 Challenge (Calgary-Campinas).	2020	0.2
Third place on the fastMRI 2020 Challenge on transfer learning for accelerated MRI Reconstruction.	2020	0.2

Curriculum Vitæ

Dimitrios Karkalousos

7^{th} of April 1990. Born in Athens, Greece

Dec. 2023 - present:	Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands Postdoctoral Researcher at the department of Biomedical Engineering & Physics and at the Institute of Informatics. Project: "*Artificial Intelligence for early detection of non-communicable disease risk in people with breast cancer*".
Dec. 2019 - Dec. 2023:	Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands PhD candidate at the department of Biomedical Engineering & Physics. Project: "*Stroke treatment decision support by Artificial Intelligence for analysis of MR images*".
Aug. 2019 - Dec. 2019:	Amsterdam Machine Learning Lab, University of Amsterdam, Amsterdam, Netherlands Research officer. Project: "*Invertible Recurrent Inference Machine applied to the fastMRI reconstruction challenge*".
Apr. 2018 - Oct. 2018:	Spinoza Centre of Neuroimaging, Amsterdam, Netherlands Research internship for M.Sc. Thesis "*Optimizing a Recurrent Inference Machine for Accelerated MRI Reconstruction using Deep Learning*".
Oct. 2016 - Oct. 2018:	University of West Attica, Athens, Greece, & University of Limoges, Limoges, France M.Sc. in Informatics, Image Synthesis and Graphics Design
Sep. 2008 - May 2014:	University of Ioannina, Arta, Greece B.Sc. in Information and Telecommunication Engineering

Acknowledgments

My journey toward becoming a Doctor of Philosophy began with the overarching aim of contributing meaningfully to the fields of Artificial Intelligence and Medical Imaging. This path has been shaped by remarkable individuals who have guided, supported, and inspired me along the way.

First and foremost, I would like to express my deepest gratitude to my supervisor, dr. Matthan Caan. Our collaboration began when Wietske at the Spinoza Centre for Neuroimaging connected us for what would become my master's thesis internship. This internship with you and Kai opened doors I never imagined possible. From working with an incredible team – Jonas, Nikita, Patrick, Nicola, and Max – on the fastMRI challenge (which we won and led to my first experience at a top AI conference, NeurIPS in Canada), to offering me the opportunity to pursue a PhD. Matthan, you have been more than a supervisor – you have been a mentor, friend, and inspiration. Your brilliance is only matched by your kindness, and I am profoundly grateful for your guidance throughout this journey.

I am equally grateful to my promoters, prof. dr. Henk Marquering and prof. dr. Ivana Išgum. Henk, your ability to provide simple yet transformative feedback – whether through a single question mark or reminding me that "I don't have to bring world peace when writing a paper" – has been invaluable. Your successful transition from "CT for the earth" to medical imaging for the brain continues to inspire me. Ivana, our paths crossed when about the same time as I was beginning my PhD you moved with your group from Utrecht UMC to Amsterdam UMC and eventually became one of my supervisors. You have always been an example of excellence in science for me. Your deep involvement in multiple projects while maintaining the highest standards has taught me invaluable lessons about scientific rigor and dedication. Thank you for also giving me the opportunity to work further with you and your group after finishing my PhD.

I extend my sincere appreciation to my committee, prof. dr. ir. M. Staring, prof. dr. L. Reneman, prof. dr. ir. J.J. Sonke, prof. dr. ir. G.J. Strijkers, dr. E. Gavves, dr. H.v. Hoof, and dr. ir. D.H.J. Poot, for agreeing to be members of my committee and for reading and providing thoughtful feedback on my thesis.

My research journey was enriched by many other collaborations and wonderful colleagues. I am grateful to Kai, for guiding me in my first steps in Deep Learning and Medical Imaging, and to the people I met at the Spinoza Centre for Neuroimaging - Bob, Icaro, Martin, Serge, Tomas, Wietske, and many others - for their warm welcome when I first came to the Netherlands.

The PhD project (STAIRS) also introduced me to Luka, who began as a colleague working at the same project and became a friend. Our shared passions for science, athletics, video games, and food have made this journey particularly memorable, especially during our conference travels. Thank you for being a great colleague and, of course, for agreeing to be my paranymph.

I also would like to thank our collaborators on the STAIRS project, Emilie and Julien, for welcoming us in Paris and showing me how real clinical practice is performed in stroke MRI. It was the first time I came into contact with stroke patients, real people whose images I have been looking for so long, and it has definitely been an eye-opening experience.

I also would like to thank a few other people with whom I had the pleasure of working: Chaoping, for teaching me about quantitative MRI and sharing insights about China; Kevin and Lysander for their valuable contributions to my research; Maurice for our collaboration on neonatal imaging and making work so enjoyable; Iasonas for his contributions and humble approach to collaboration; and Jerke for his research contributions and our engaging lunch discussions about video games.

I also want to acknowledge my colleagues at the BMEP department, the Radiology department, and the qurAI group. The PhDs and Post-Docs: Ajey, Anas, Annemiek, Arnolt-jan, Bhavesh, Bobby, Carolina, Caroline, Coen, Daisy, Daphne, Fatima, George, Giannis, Gino, Henk, Ilse, Jackie, Jin, Kelly, Kevin, Koen, Laura, Laura, Linda, Lisa, Lucas, Luis, Maarten, Mahsa, Mancy, Manon, Maria, Matej, Meagan, Michel, Mohammad, Natalia, Nerea, Nils, Pierandrea, Pranneta, Renske, Riaan, Ricardo, Roel, Sanyog, Sisi, Soufiane, Susi, Thijs, Tom, Vikram, Wiktor, and Xiaojin, and the faculty members: Aart, Bart, Bram, Carla, Charles, Clarisa, Gustav, Hoel, Liesbeth, Navchetan, Oliver, Paul, Richard, Simone, Stefan, and Ton. Thank you all for creating such a collaborative, friendly, and supportive environment.

I am also grateful to the people who enriched my PhD journey in one way or another: Cigdem, Frans, Gerasimos, George, Jinpeng, Nikos, Samuel, and Tim. I probably forgot to mention some other people as well and I deeply apologize, but I thank you all.

Certainly, I would like to thank my friends in Greece. My childhood friends from Piraeus, Nikos and Takis, and my classmates and friends from Arta, Antonis, Giorgos, Marios, Spyros, Spyros, and Thodoris. Takis, it is an honor to have you as a lifetime friend, I sincerely wish you all the best in your life. Spyros and Thodoris, thank you for your friendship and support for half a life now, I wish you all the best for you and your families. Spyros, thank you for checking out every single day, which always brings me back to reality.

Foremost, I want to express my deepest gratitude to my partner, Eleni, who walked beside me throughout this journey, sharing countless days and nights pursuing our goals together. Our scientific and philosophical discussions have enriched my knowledge and

have also helped me trace so many AI concepts back to their foundations in psychology. Your deep intuition, dedication, and hard work have always inspired me, and I am certain you will achieve great things on your own journey. Thank you for all our adventures and simple moments of joy, for your unwavering support, love, and for sharing the journey of life together.

Τέλος, και πάνω από όλα, θα ήθελα να ευχαριστήσω την οικογένειά μου, τους γονείς μου, Γεωργία και Χρήστο, και τον αδερφό μου, Νίκο. Σας ευχαριστώ για την αγάπη, για την αμέριστη υποστήριξη, για όλες τις θυσίες που κάνατε όλα αυτά τα χρόνια για να μου προσφέρετε την καλύτερη δυνατή εκπαίδευση και για την δυνατότητα που μου δώσατε να πραγματοποιήσω όλες μου τις σπουδές και αυτό το διδακτορικό. Χωρίς εσάς τίποτα από αυτά δεν θα ήταν εφικτό.