



DI Kerstin Hammernik, BSc

# Variational Networks for Medical Image Reconstruction

## DOCTORAL THESIS

to achieve the university degree of  
Doktorin der technischen Wissenschaften

submitted to

**Graz University of Technology**

Supervisor

Prof. Dr. Thomas Pock

Institute of Computer Graphics and Vision,  
Graz University of Technology, Austria

Prof. Dr. Daniel Rueckert

Department of Computing,  
Imperial College London, United Kingdom

Graz, Austria, 2019



Quiet the mind, and the soul will speak.

---

*Ma Jaya Sati Bhagavati*



## Abstract

During the past years, deep learning has shown a great potential for segmentation and classification tasks not only in computer vision but also in medical imaging. However, many more opportunities for applying deep learning techniques exist in medical imaging, ranging from data acquisition and reconstruction to computer aided diagnosis. In this thesis, we explore the potential of deep learning for medical image reconstruction. First, this thesis provides a broad overview of how deep learning techniques can be used to improve image reconstruction. Second, we propose Variational Networks to solve various inverse problems. Variational Networks are rooted in variational methods and deep learning. Their aim is to keep the structure of variational models and use deep learning techniques to improve on both the regularization of these models and the reconstruction algorithm itself. In this thesis, we explore different ways to further improve the results of the Variational Networks. We study different types of regularization used in the Variational Networks ranging from Fields of Experts regularizers, that concentrate on low-level features, to deep regularizers, that are able to incorporate mid- and high-level features. Furthermore, we investigate the impact of different pixel-based and patch-based loss functions. The main focus of this thesis is the application of Variational Networks to various inverse problems in medical image reconstruction. First, we setup a Variational Network for image enhancement to correct structured artifacts in limited-angle Computed Tomography. The second application for the proposed Variational Networks is accelerated Magnetic Resonance image reconstruction of static multi-coil 2D data. Here, we investigate the impact and generalization potential of Variational Networks on clinical knee data, including a reader study on image quality. In the third application, we explore the potential of Variational Networks for dynamic Magnetic Resonance image reconstruction using cardiac multi-coil data and spatio-temporal regularization. All studied applications have in common that typical structured artifacts arise in the reconstructed images due to the acquisition process, which cannot be removed efficiently with current state-of-the-art techniques. Using the proposed Variational Networks, we achieve promising results in terms of improved image quality and reduced artifacts. Once the Variational Networks are learned, new data can be reconstructed efficiently and

no further parameter tuning is required, which are major limitations of current state-of-the-art reconstruction approaches. Consequently, this new approach to image reconstruction offers the potential to be directly integrated into the clinical workflow and improve patient comfort, patient safety and healthcare costs.

**Keywords.** Variational Networks, Deep Learning, Variational Methods, Inverse Problems, Magnetic Resonance Imaging, Limited-Angle Computed Tomography

## Kurzfassung

Forschungsergebnisse der letzten Jahre zeigen ein großes Potential von Maschinellern für diverse Anwendungen in der Bildverarbeitung, wie Segmentierung und Klassifizierung. Jedoch gibt es viele weitere Möglichkeiten, Maschinelles Lernen im Bereich der medizinischen Bildverarbeitung anzuwenden, zum Beispiel bei der Datenakquisition, Bildrekonstruktion oder in der computerassistierten Diagnose. Diese Arbeit gibt einen Überblick über laufende Entwicklungen in der medizinischen Bildrekonstruktion basierend auf Maschinellern. Hierzu wird das Konzept von Variationsnetzwerken vorgestellt, in denen die Strukturiertheit von Variationsmethoden mit Ideen aus dem Maschinellern vereint werden, um nicht nur verbesserte Bildqualität zu erreichen, sondern auch den Rekonstruktionsalgorithmus ansich zu lernen. In dieser Arbeit werden auch Erweiterungen von Variationsnetzwerken vorgestellt, um die Bildqualität weiter zu verbessern. Dazu werden unterschiedliche Arten von Regularisierung, die in den Variationsnetzwerken verwendet wird, vorgestellt. Diese beinhalten Fields of Experts Regularisierung, die sich auf low-level Features von Bildern beschränkt, sowie tiefe Regularisierung, die auch mid-level und high-level Features von Bildern in den Lernprozess miteinbeziehen kann. Des Weiteren werden diverse Pixel-basierte und Patch-basierte Ähnlichkeitsmaße, die für die Qualität des Lernens verantwortlich sind, vorgestellt. Der Fokus dieser Arbeit liegt auf der Anwendung von Variationsnetzwerken für diverse Inverse Probleme aus dem Bereich der medizinischen Bildrekonstruktion. Zuerst werden Variationsnetzwerke zur Nachbearbeitung und Artefaktkorrektur in der Computertomographie verwendet. Im Bereich der Magnetresonanztomographie wird zuerst auf die Rekonstruktion von statischen Knie-Scans, die im Rahmen von klinischen Untersuchungen akquiriert wurden, eingegangen. Dies beinhaltet eine umfassende Analyse um den Einfluss und das Generalisierungspotential von Variationsnetzwerken zu zeigen, und inkludiert auch eine radiologische Studie, um die Bildqualität von Knie-Rekonstruktionen mittels Variationsnetzwerken zu erfassen. Zuletzt wird die Anwendung von Variationsnetzwerken auf dynamischen Daten des Herzens gezeigt, die sowohl einer räumlichen als auch einer zeitlichen Regularisierung bedürfen. Obwohl es sich um unter-

schiedlichste Applikationen handelt, haben diese etwas gemeinsam: Die Art der Artefakte, die bei der Datenakquisition entstehen sind so strukturiert, dass sie mit gewöhnlichen Algorithmen nicht zufriedenstellend rekonstruiert werden können. Mit den vorgestellten Variationsnetzwerken können diese Artefakte erheblich reduziert und gleichzeitig die Bildqualität verbessert werden. Des Weiteren können neu akquirierte Daten effektiv rekonstruiert werden und bedürfen keiner Feineinstellung von Parametern, wie dies bei gewöhnlichen Rekonstruktionsalgorithmen oft der Fall ist. Infolgedessen eröffnen die vorgestellten Variationsnetzwerke das Potential um direkt in den klinischen Arbeitsablauf integriert zu werden, um so Patientenkomfort sowie -sicherheit zu verbessern, als auch die Kosten im Gesundheitssystem zu reduzieren.

**Schlagwörter.** Variationsnetzwerke, Maschinelles Lernen, Variationsmethoden, Inverse Probleme, Magnetresonanztomographie, Computertomographie



**Affidavit**

*I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used.*

*The text document uploaded to TUGRAZonline is identical to the present doctoral thesis.*

---

Date

---

Signature



## Acknowledgments

Writing this thesis establishes the end of a unique and formative stage in my life comprising major setbacks, enlightening steps forward and new perspectives. Here, I have now the opportunity to thank all people that have been part of these exceptional experiences, which would not have been possible without their support, friendship, love, inspiration, mentoring, and unforgettable adventures inside and outside the office, at conferences and during my internship.

First and foremost, I would like to thank Thomas Pock, for encouraging, inspiring, challenging and motivating me, for giving me the freedom to conduct research in the area of medical imaging, for his patience, enthusiasm and guidance, for supporting my conference visits and my internship, and last but not least for supervising this thesis. I would also like to thank Daniel Rueckert for agreeing to be my second supervisor.

Next, I would like to express my gratitude to our cooperation partners from the Department of Radiology at New York University Langone Health, namely Florian Knoll, Daniel Sodickson and Michael Recht, and I am grateful for the invaluable opportunity to host me as an intern. Without their support, the work in deep learning for accelerated MR image reconstruction of musculoskeletal data would not have been possible. I would also like to thank Florian Knoll for sharing his enthusiasm and comprehensive expertise in MR image reconstruction.

During the course of my PhD studies, I am greatly thankful for the experience I have made with my colleagues in the exceptional and inspiring environment at the Institute of Computer Graphics and Vision, Graz University of Technology. Here, I want to specifically mention Erich Kobler and my former colleague and friend Teresa Klatzer from the Vision, Learning and Optimization group. Thank you for the fruitful, enlightening and crazy whiteboard discussions, for the valuable feedback, for the lively duck debugging sessions, for those moments when the tragedy on the day of paper submission turned into a comedy in front of the whiteboard, and for those moments when paper correction turned into abstract artwork.

I am also greatly thankful for the support that I received from the Institute of Medical Engineering, Graz University of Technology, namely Rudolf Stollberger, Matthias Schlögl and

Christoph Aigner. Thank you for sharing your MRI expertise and introducing me to the exciting world of cardiac MR image acquisition and reconstruction.

Furthermore, I want to express my special thanks to Alexander Effland, Christina Graf, Joana Grah, Erich Kobler and Matthias Schlögl for proof-reading my thesis and for their valuable feedback.

I am deeply grateful to my colleagues and friends from various local and international institutions, namely Christoph Aigner, Christina Graf, Joana Grah, Teresa Klatzer, Florian Knoll, Erich Kobler, Alexander Lassnig, Andreas Lesch, Rebecca Ramb, Matthias Schlögl, Martin Urschler, Jelle Veraart, and many, many others for sharing numerous hours of fruitful, inspiring and philosophical discussions, skiing, hiking and biking, and for sharing great moments at the MR.Graz Stammtisch, Glühwein sessions, conferences and various other intra- and extra-office events.

Finally, I would like to thank all my friends from the Grazy SBKZ latin dance community for keeping up my work-dance-balance and for all those wonderful moments that have filled and brightened my life with joy, love, truth and passion.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	The Potential of Artificial Intelligence in Medical Imaging . . . . .	2
1.2	Contributions and Outline . . . . .	3
<b>2</b>	<b>Machine Learning for Image Reconstruction</b>	<b>5</b>
2.1	Inverse Problems in Imaging . . . . .	6
2.2	Unsupervised Learning in Image Reconstruction . . . . .	9
2.3	Supervised Learning in Image Reconstruction . . . . .	11
2.3.1	Learning an Improved Regularization Function . . . . .	12
2.3.2	Learning an Iterative Reconstruction Model . . . . .	16
2.3.3	Deep Learning for Image and Data Enhancement . . . . .	22
2.3.4	Learning a Direct Mapping . . . . .	23
2.3.5	Other Approaches for Image Reconstruction . . . . .	25
2.4	Training data . . . . .	25
2.5	Image Quality . . . . .	26
2.6	Reproducible Research . . . . .	29
<b>3</b>	<b>Variational Networks</b>	<b>31</b>
3.1	From Linear Reconstruction to a Variational Network . . . . .	32
3.2	Variational Network Parameters . . . . .	35
3.3	Variational Network Training . . . . .	36
3.4	Loss Functions and Quantitative Evaluation . . . . .	37
3.5	Deep Regularization for Variational Networks . . . . .	39
3.6	Implementation Details . . . . .	40

---

<b>4</b>	<b>Variational Networks for Image Enhancement in Limited-Angle Computed Tomography (CT)</b>	<b>41</b>
4.1	Fundamental Principles of CT	42
4.2	Artifact Correction in Limited-Angle CT	44
4.3	Methods	47
4.4	Results	49
4.5	Discussion	49
<b>5</b>	<b>Variational Networks for 2D Cartesian MR Image Reconstruction</b>	<b>53</b>
5.1	Fundamental Principles of Magnetic Resonance (MR) Image Reconstruction	54
5.1.1	From Signals to Images	55
5.1.2	The Discrete World of $k$ -Space	57
5.1.3	Acquisition Strategies in Clinical Practice	60
5.1.4	Accelerated MR Imaging	61
5.1.5	Definition of the Static MR Forward Model	66
5.2	Data Acquisition	69
5.3	Insights into Learning a Variational Network (VN) for Accelerated Magnetic Resonance Imaging (MRI) Data	71
5.3.1	Methods	72
5.3.2	Results	76
5.3.3	Discussion	92
5.4	Exploring the Generalization Potential of VNs for Accelerated MR Imaging	96
5.4.1	Methods	96
5.4.2	Results	97
5.4.3	Discussion	98
5.5	Influence of Loss Function Design for Accelerated MR Image Reconstruction	98
5.5.1	Methods	99
5.5.2	Results	100
5.5.3	Discussion	100
5.6	Improved Regularization for Accelerated MR Image Reconstruction	101
5.6.1	Methods	101
5.6.2	Experimental Setup	102
5.6.3	Results	103
5.6.4	Discussion	103
5.7	Intra-Vendor Reproducibility	104
5.8	Conclusion and Outlook	104
<b>6</b>	<b>Variational Networks for Dynamic MR Image Reconstruction</b>	<b>111</b>
6.1	Fundamentals of Accelerated Cardiac MR Image Reconstruction	112
6.2	Methods	113
6.2.1	A Variational Network for Dynamic Applications	113

---

6.2.2	Data Acquisition . . . . .	114
6.2.3	Experimental Setup and Evaluation . . . . .	115
6.3	Results . . . . .	116
6.4	Discussion . . . . .	117
<b>7</b>	<b>Conclusion and Outlook</b>	<b>125</b>
<b>A</b>	<b>List of Acronyms</b>	<b>129</b>
<b>B</b>	<b>List of Publications</b>	<b>133</b>
<b>C</b>	<b>Algorithms</b>	<b>137</b>
C.1	Inertial Incremental Proximal Gradient (IIPG) . . . . .	138
C.2	Proximal ADAM with Block-Preconditioning . . . . .	138
<b>D</b>	<b>Gradient Calculations</b>	<b>141</b>
D.1	Variational Network Parameters . . . . .	141
D.2	Loss functions . . . . .	143
	<b>Bibliography</b>	<b>149</b>





## List of Figures

2.1	Example measurement spaces and image spaces. . . . .	7
2.2	Schematic illustration of the training process. . . . .	12
2.3	Overview over learning-based approaches for image reconstruction. . . . .	13
2.4	Generative Adversarial Networks (GANs) for image reconstruction. . . . .	28
3.1	Basic VN structure. . . . .	34
3.2	Examples for regularizers. . . . .	34
3.3	Gradient visualization of different regularizers. . . . .	34
3.4	Approximation of functions using Gaussian radial basis functions (RBFs). . . . .	35
4.1	Examples for different beam geometries in CT. . . . .	42
4.2	CT projection principle for a point source. . . . .	45
4.3	CT projection principle for a circle phantom. . . . .	45
4.4	Examples for Filtered Back-Projection (FBP) reconstructions. . . . .	46
4.5	Deep learning architecture for limited-angle CT reconstruction. . . . .	48
4.6	Qualitative comparison of non-linear filtering methods for limited-angle CT. . . . .	50
5.1	Comparison of a reconstruction VN and denoising VN. . . . .	55
5.2	Cartesian sampling: Example pulse sequence of a Spin Echo (SE) and sampling trajectory. . . . .	57
5.3	Radial sampling: Example pulse sequence of a SE and sampling trajectory. . . . .	58
5.4	Relationship between $k$ -space and image space. . . . .	61
5.5	Effects of a small FoV, phase oversampling and phase resolution. . . . .	62
5.6	Acquired $k$ -space data from a 15-channel knee coil. . . . .	63
5.7	Reconstructed coil images for fully sampled data acquired with a 15-channel knee coil. . . . .	64
5.8	Coil sensitivity maps for data acquired with a 15-channel knee coil. . . . .	65

5.9	Examples for aliasing of Cartesian and radially undersampled data. . . . .	67
5.10	Influence of sampling mask design for reconstructions based on Sensitivity Encoding (SENSE). . . . .	70
5.11	Proposed image reconstruction pipeline for static MR image reconstruction. . .	73
5.12	VN structure for accelerated static MR image reconstruction. . . . .	74
5.13	Different Cartesian sampling patterns along with their corresponding Point Spread Function (PSF) . . . . .	75
5.14	Coronal PDw scan with acceleration $R = 3$ of a 32-year-old male. . . . .	80
5.15	Difference images to reference image for the reconstructed coronal PDw scans with acceleration $R = 3$ presented in Figure 5.14. . . . .	81
5.16	Coronal PDw-FS scan with acceleration $R = 3$ of a 57-year-old female. . . . .	82
5.17	Difference images to reference image for the reconstructed coronal PDw-FS scans with acceleration $R = 3$ presented in Figure 5.16. . . . .	83
5.18	Coronal PDw scan with acceleration $R = 4$ of a 32-year-old male. . . . .	84
5.19	Difference images to reference image for the reconstructed coronal PDw scans with acceleration $R = 4$ presented in Figure 5.18. . . . .	85
5.20	Coronal PDw-FS scan with acceleration $R = 4$ of a 57-year-old female. . . . .	86
5.21	Difference images to reference image for the reconstructed coronal PDw-FS scans with acceleration $R = 4$ presented in Figure 5.20. . . . .	87
5.22	Reconstruction results for sagittal $T_2w$ -FS, sagittal PDw and axial $T_2w$ -FS sequences of a complete knee protocol for regular undersampling of $R = 4$ . . .	88
5.23	Difference images for sagittal $T_2w$ -FS, sagittal PDw and axial $T_2w$ -FS sequences of a complete knee protocol. . . . .	89
5.24	Reconstruction results of prospectively undersampled data for $R = 4$ . . . . .	90
5.25	Examples of learned parameters of the VN. . . . .	91
5.26	Results for the generalization potential of the proposed VNs. . . . .	106
5.27	Reconstruction errors depending on the number of Conjugate Gradient (CG) SENSE iterations. . . . .	107
5.28	CG SENSE reconstructions for PDw-FS data at $R = 3$ . . . . .	107
5.29	VN reconstructions trained with different loss functions. . . . .	108
5.30	VN reconstruction results using different types of regularization. . . . .	109
5.31	Intra-vendor reproducibility: Retrospective and prospective VN reconstructions. . . . .	110
6.1	VN for 2D+t multi-coil MR image reconstruction. . . . .	114
6.2	$k$ -t sampling patterns of size $192 \times 17$ for dynamic MR image reconstruction at different acceleration rates. . . . .	116
6.3	bSSFP: 2CH view for Variable Density Incoherent Spatio-Temporal Acquisition (VISTA) sampling at $R = 8$ . . . . .	119
6.4	bSSFP: Short Axis (SA) view for VISTA sampling at $R = 12$ . . . . .	120
6.5	bSSFP: SA view for Variable Density Random Sampling (VRS) sampling at $R = 12$ . . . . .	121

---

6.6	FLASH: SA view for VISTA sampling at $R = 12$ . . . . .	122
6.7	FLASH: SA view for VRS sampling at $R = 12$ . . . . .	123
6.8	FLASH: 4CH view for VRS sampling at $R = 16$ . . . . .	124
6.9	Learned VN parameters for dynamic CINE MRI. . . . .	124
7.1	The iceberg theory of deep learning . . . . .	127



## List of Tables

4.1	Quantitative comparison of non-linear filtering methods for limited-angle CT. . . . .	49
5.1	Acquisition times for a clinical knee protocol with different acceleration factors. . . . .	69
5.2	Overview of sequence parameters for the used clinical knee protocol along with details about the scanned patient population. . . . .	71
5.3	Quantitative evaluation results for a clinical knee protocol and acceleration factor $R = 3$ for regular sampling and VRS. . . . .	78
5.4	Quantitative evaluation results for a clinical knee protocol and acceleration factor $R = 4$ for regular sampling and VRS. . . . .	78
5.5	Image quality reader scores for a clinical knee protocol and acceleration factor $R = 4$ for regular sampling. . . . .	79
5.6	Quantitative results for generalization experiments using VNs. . . . .	97
5.7	Quantitative results for VNs trained with different loss functions. . . . .	100
5.8	Quantitative results for VNs trained with different regularizers. . . . .	103
6.1	Quantitative results for the bSSFP dataset for different sampling patterns and acceleration rates. . . . .	117
6.2	Quantitative results for the FLASH dataset for different sampling patterns and acceleration rates. . . . .	118



## Introduction

Radiologists who use AI will replace radiologists who don't.

---

*Curtis Langlotz, RSNA 2017*

### Contents

---

1.1	The Potential of Artificial Intelligence in Medical Imaging . . .	2
1.2	Contributions and Outline . . . . .	3

---

During the past years, deep learning [73, 148] has caused a paradigm shift in computer vision from using handcrafted features and classifiers to data-driven and learning-based approaches which are able to learn the feature representations in combination with statistical classifiers from suitable training data. Impressive improvements in image quality and accuracy have been achieved for various applications in computer vision such as image classification [142], semantic segmentation [42], optical flow [61] and image restoration [271]. In medical imaging, the deep learning techniques have mostly focused on image classification [159] and segmentations tasks [121]. In these applications, deep learning was mainly used as a tool for image processing and interpretation, however, the concept of learning can also be used at earlier stages of image formation such as data acquisition or image reconstruction, which is of particular interest in medical imaging. Although first results in using Artificial Neural Networks (ANNs) for reconstruction in Magnetic Resonance Imaging (MRI) [124], Computed Tomography (CT) [166] and Single Photon Emission Computed Tomography (SPECT) [67, 127, 139, 184] have already been shown in the 1990s, a breakthrough of deep learning for medical image reconstruction started in 2016 [171, 244, 245]. The major reason for today's success of deep learning methods for medical imaging is that the training data and computational power have increased tremendously over the past years. Furthermore, continuous developments of dedicated algorithms in combination with freely available toolboxes allow us to implement new architectures with a

few lines of code. Although impressive results are being achieved, these "black box" results are hard to interpret and incomprehensible, which does not only give rise to hope but also fear concerning the application of Artificial Intelligence (AI) in healthcare.

## 1.1 The Potential of Artificial Intelligence in Medical Imaging

In the last years, deep learning algorithms have shown astonishing performance on medical tasks that could only be performed by trained Medical Doctors (MDs). One example was shown in 2017 where researchers at Google applied deep learning to detect metastatic breast cancer on microscopy images [160]. Their deep learning algorithm achieved comparable or greater accuracy on detecting cancer than human MDs which underwent extensive training to detect tiny signs of cancer on these microscopic images. However, even if these first results are promising, the employed algorithms have to be assessed thoroughly on large scale studies if they have the potential to improve diagnostic accuracy and how these algorithms are able *assist* MDs. Another application of AI is precision medicine: A vast amount of data is available nowadays such as medical records, treatment plans, the patients' history and background information or even data from genome sequences, medical sensors and wearables. Processing and analyzing these data might allow MDs to set up or improve new treatment plans or prevention concepts on a single-patient level to improve individual diagnosis and patient monitoring. These developments might generate new view points for MDs by moving away from general, narrow 'textbook' solutions to a more open, holistic view.

Indeed, there are a lot of opportunities how AI could change the world of radiology. In the area of medical imaging, these opportunities range from computer aided diagnosis over semantic image segmentation, image quantification to image acquisition and reconstruction. This thesis is focused on deep learning for medical image reconstruction. Medical imaging modalities such as CT and MRI are extremely powerful and have become indispensable in clinical workflow, however, these modalities suffer from specific drawbacks that need to be improved. Magnetic Resonance (MR) imaging offers excellent soft-tissue contrast and protocol flexibility, which makes it a very powerful imaging modality for neurological, musculoskeletal, and oncological diseases. However, the long acquisition time in MRI, which can easily exceed 30 minutes for a typical imaging protocol, leads to low patient throughput, problems with patient comfort and compliance, artifacts from patient motion, and high exam costs. The acquisition time can be decreased by combinations of improved (1) hardware concepts, such as stronger gradients, phased array receive coils, higher field strengths, (2) signal generation and acquisition design and (3) reconstruction concepts such as Parallel Imaging (PI) [77, 190, 227] and Compressed Sensing (CS) [31, 60]. While MR is not hazardous, CT relies on ionizing radiation, hence, a major goal is to decrease radiation dose in CT examinations. This can be achieved by different concepts such as lowering the tube current, interrupted beam acquisitions or acquiring less projections. However, there are some applications where it is not possible to acquire data over the full angular range, known as limited-angle CT.

For all medical imaging technologies, the employment of more sophisticated image re-



construction techniques requires to compute solutions *iteratively* and every single exam and corresponding reconstruction task is treated as a new optimization problem. Hence, these optimization problems do not use additional information explicitly, such as the expected appearance of the anatomy or the known structure of artifacts, which appear due to reduced amount of acquired data. This stands in stark contrast to how human radiologists read images. Radiologists are trained extensively throughout their careers to look for certain reproducible patterns. They obtain remarkable skills to “read through” known image artifacts [105], even though local noise provides a measure of uncertainty. Translating this conceptual idea of human learning to deep learning allows us to shift the key effort of optimization from the online reconstruction stage to an up-front offline training task, which enables both accelerated data acquisition and accelerated image reconstruction.

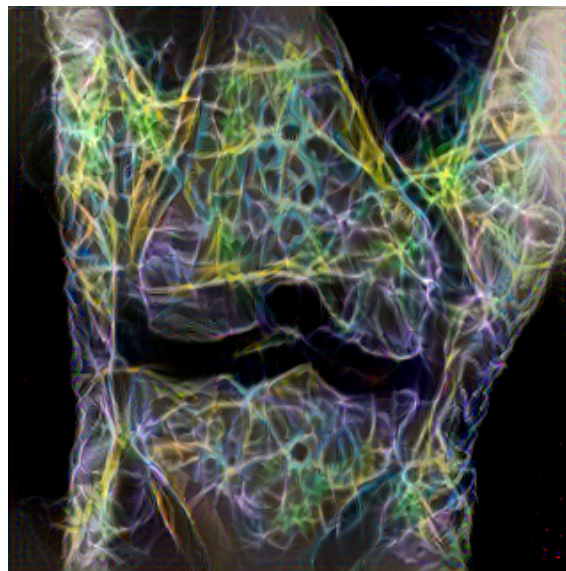
AI leverages the potential to change the complete imaging workflow in radiology. Most of the big tech-companies such as Google, Facebook, Microsoft, IBM and Amazon as well as many smaller start-ups have started to develop deep learning algorithms for medical applications. However, many of currently existing algorithms were developed regardless of practical relevance in medical imaging. Hence, one of the biggest challenges will be the interdisciplinary work of AI specialists and medical professionals, who have the profound medical training and understanding of the imaging workflow, to develop solutions with high practical relevance.

## 1.2 Contributions and Outline

The field of learning-based medical image reconstruction evolved tremendously during the course of my PhD studies. Due to the enormous development of new algorithms and application of deep learning algorithms to medical image reconstruction, this topic became almost unmanageable within a short amount of time. The first contribution is to provide an overview and deeper understanding how deep learning can improve image reconstruction. Inspired by variational models and deep learning, the second contribution is to leverage the potential of Variational Networks (VNs), which were initially proposed as a reaction-diffusion model [43, 46], for medical image reconstruction and radiology. First, we formulate image reconstruction as an image enhancement problem and use VNs for artifact-correction in limited-angle CT. This study was conducted in cooperation with the Friedrich-Alexander-University Erlangen-Nürnberg, Germany. Second, we formulate image reconstruction as an iterative reconstruction problem with a limited number of iterations and explore the impact of VNs on accelerated MRI reconstruction of multi-coil patient data, which was the primary focus of this thesis. This study was conducted in close collaboration with medical and MRI experts at the Department of Radiology at New York University, NY, USA, who provided clinical patient data of knee exams and experts’ knowledge, which served as a basis to develop an algorithm with respect to a real clinical setting. This formulation was further extended to dynamic MR image reconstruction, which was conducted in collaboration with the Institute of Medical Engineering at Graz University of Technology, Austria.

This thesis is organized as follows: Chapter 2 provides an extensive overview how ma-

chine learning can improve medical image reconstruction. The theory of VNs is introduced in Chapter 3, along with improvements on the network architecture and the loss function, which measures the similarity of the reconstructed image to the reference during training. The following chapters illustrate different applications of VNs for medical image reconstruction. In Chapter 4, a VN is formulated for image enhancement to correct artifacts of limited-angle CT data. Chapter 5 presents the impact of VNs for the reconstruction of accelerated MR data in a static 2D setting. Chapter 5 involves also practical details about the acquisition strategies in clinical practice in Section 5.1 and data acquisition for the used clinical knee protocol in Section 5.2. In Section 5.3, we gain insights into how VNs impact accelerated MR image reconstruction for different sequences used in clinical knee imaging and Cartesian sampling patterns, including a reader study on image quality. The generalization potential for a large range of datasets and the impact of different loss functions on the image reconstructions are explored in Section 5.4 and Section 5.5. Chapter 6 presents insights in using VNs with spatio-temporal regularization for dynamic MRI applications. Each application chapter involves a review of fundamental principles of the underlying applications. However, these introductory sections have their own notation which should not be confused with the general notation of the related work presented in Chapter 2 or the theory of VNs presented in Chapter 3. Chapter 7 summarizes the contributions of this thesis and presents an outlook on future research.



*AI - New perspectives on the future of radiology<sup>1</sup>*

---

<sup>1</sup>This image was generated by style-transfer [69] of an image containing colored wired structures on an MR image of the knee.

## Machine Learning for Image Reconstruction

Study the past if you want to  
define the future.

---

*Confucius*

This chapter is based on the book chapter, which is currently under review:

K. Hammernik and F. Knoll. Machine Learning for Image Reconstruction. In D. Rueckert, G. Fichtinger, and S. K. Zhou, editors, *Handbook of Medical Image Computing and Computer Assisted Intervention*. Elsevier, 2018

### Contents

---

<b>2.1</b>	<b>Inverse Problems in Imaging</b>	<b>6</b>
<b>2.2</b>	<b>Unsupervised Learning in Image Reconstruction</b>	<b>9</b>
<b>2.3</b>	<b>Supervised Learning in Image Reconstruction</b>	<b>11</b>
<b>2.4</b>	<b>Training data</b>	<b>25</b>
<b>2.5</b>	<b>Image Quality</b>	<b>26</b>
<b>2.6</b>	<b>Reproducible Research</b>	<b>29</b>

---

This chapter introduces how deep learning is able to improve upon the state-of-the-art in medical image reconstruction of high-quality images from incomplete data. We present a general overview over existing techniques in computer vision and medical imaging in order to give an intuition of how deep learning can be applied to image reconstruction. Furthermore, this chapter provides a deeper mathematical understanding of how deep learning techniques can be employed for image reconstruction tasks. Therefore, we draw connections to traditional approaches of solving inverse problems, rather than presenting details about the used neural

network architectures. Approaches for both unsupervised, such as dictionary learning, and supervised learning are covered in this chapter. The reviewed approaches for supervised learning range from learning of hyper-parameters and more general regularizers in a variational model to replace steps in an iterative reconstruction algorithm with Convolutional Neural Networks (CNNs), to learning image enhancement and learning a full mapping between measured data and reconstructed images. With the major focus on supervised learning, this chapter also includes key challenges such as suitable training data, transfer learning, selection of the loss function for training and evaluation of image quality.

## 2.1 Inverse Problems in Imaging

Inverse problems arise in various applications ranging from medicine over geophysics to economics. In the area of computer vision, this includes, but is not limited to image segmentation, motion correction, image registration, object detection and image reconstruction. Considering medical image reconstruction, the goal is to reconstruct an image  $x \in \mathbb{K}^{N_x}$  from measurement data  $y \in \mathbb{K}^{N_y}$  given the following system of equations

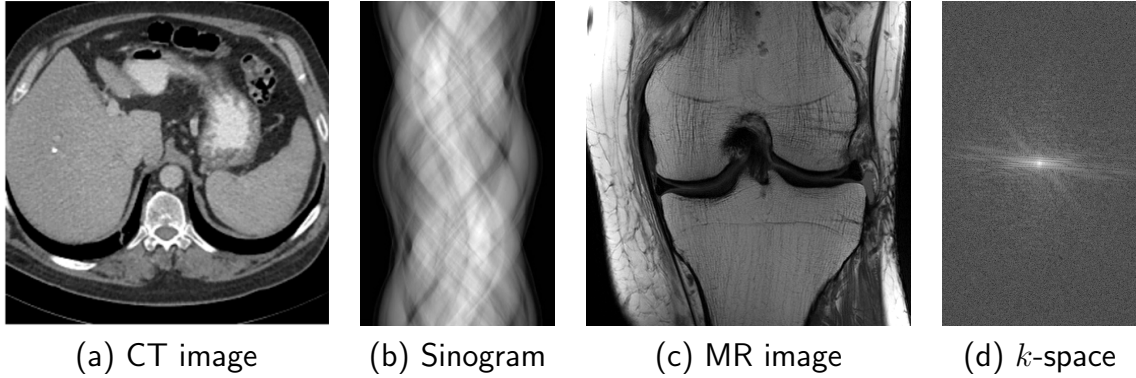
$$y = A(x) + \nu, \quad (2.1)$$

where  $\nu \in \mathbb{K}^{N_y}$  is an additive random noise variable, the field  $\mathbb{K} = \mathbb{R}, \mathbb{C}$  is the set of either real numbers  $\mathbb{R}$  or complex numbers  $\mathbb{C}$ . The dimensions of the vector space of the image  $x$  and data  $y$  is denoted by  $N_x$  and  $N_y$ . The forward operator  $A : \mathbb{K}^{N_x} \rightarrow \mathbb{K}^{N_y}$  defines a mapping between the normed vector spaces of the image data and the measurement data. In this work, we mainly consider the  $\ell_p$ -norm for  $p \geq 1$  over a vector  $z \in \mathbb{K}^N$  which is defined for an  $N$ -dimensional field  $\mathbb{K}^N$  as

$$\|z\|_p = \left( \sum_{i=1}^N |z_i|^p \right)^{\frac{1}{p}}.$$

The forward operator  $A$  describes the acquisition process to measure  $y$  given the physical and technical conditions and limitations, which often involves approximations of the real physics. In Magnetic Resonance Imaging (MRI), the forward operator includes the Fourier Transform (FT) to map an image to the measurement space ( $k$ -space), while the fundamental relationship to map images to sinograms in Computed Tomography (CT) is mainly described by the Radon transform. Figure 2.1 illustrates examples for image and measurement data for MRI and CT. Typical examples for medical image reconstruction problems are the reconstruction from Cartesian or non-Cartesian undersampled  $k$ -space data in MRI from single or multiple receiver channels, where the latter is commonly termed Parallel Imaging (PI). In CT typical applications are low-dose, sparse-view and limited-angle CT, which are based on parallel-beam, cone-beam, fan-beam or spiral acquisitions. In Positron Emission Tomography (PET), typical applications are dose reduction, improved resolution, attenuation correction and motion correction. In

Photoacoustic Tomography (PAT), one aims at recovering the initial pressure distribution which is related to the optical absorption of biological tissue. Common techniques here are sparse sampling or limited-angle acquisitions. Generally speaking, one often seeks to acquire as few measurements as possible. This reduces acquisition time, ionizing radiation dose for CT or PET and decreases motion artifacts, which greatly increases patient comfort, reduces risks for patients and lowers overall healthcare costs.



**Figure 2.1:** Example measurement spaces and image spaces for CT and MRI. A CT image (a) is reconstructed from a sinogram (b) based on the Radon transform. For Magnetic Resonance (MR) images (c), the data are acquired in Fourier domain, termed  $k$ -space (d).

Recovering  $x$  from the measurement data  $y$  is often an ill-posed problem: A solution might not exist, the solution might not be unique or the solution might be unstable with respect to small variations in the data [81]. Obtaining a solution for  $x$  in Equation (2.1) is ill-posed in most practical cases, because of uncertainties due to measurement errors, low Signal-to-Noise Ratio (SNR), incomplete data and hardware limitations. Hence, no explicit solution can be obtained for  $x$ . A natural approach is to minimize the data-misfit using

$$x^* \in \arg \min_{x \in \mathbb{K}^{N_x}} \mathcal{D}[A(x), y], \quad (2.2)$$

where  $\mathcal{D} : \mathbb{K}^{N_y} \times \mathbb{K}^{N_y} \rightarrow \mathbb{R}$  is a mapping representing the statistical properties of the data. For normally distributed noise, an assumption that is true for MRI  $k$ -space data, a common approach here is to estimate a least-squares solution

$$x^* \in \arg \min_{x \in \mathbb{K}^{N_x}} \frac{1}{2} \|A(x) - y\|_2^2. \quad (2.3)$$

For different noise statistics, e.g., Poisson distributed samples encountered in PET and to some degree in CT, the Kullback-Leibler divergence is commonly used

$$x^* \in \arg \min_{x \in \mathbb{K}^{N_x}} \sum_{i=1}^{N_y} (A(x)_i - y_i \log(A(x)_i)),$$

where  $i$  denotes discrete sampling locations. Calculating the minimizer of these problems often leads to over-fitting the noisy measurement data. There exists several methods to restrict the solution space and search for an approximate solution in order to avoid solutions that are dominated by noise. One possibility is to perform early stopping [97], which acts as regularization. Another possibility is to impose additional constraints on  $x$  and add an regularization term to Equation (2.2) leading to the variational model

$$x^* \in \arg \min_{x \in \mathbb{K}^{N_x}} \lambda \mathcal{D} [A(x), y] + \mathcal{R} [x], \quad (2.4)$$

where  $\lambda > 0$  is a weight parameter that controls the influence of the regularization term  $\mathcal{R}$  and the data consistency term  $\mathcal{D}$ . In traditional optimization, simple generic and robust image priors are used. In the context of imaging, a common choice for the regularization term is a discrete approximation of the Total Variation (TV) semi-norm [207], which reads as

$$\mathcal{R} [x] = \|Dx\|_{2,1} = \sum_{i=1}^{N_x} \sqrt{\sum_d \left( |Dx|_i^{(d)} \right)^2}, \quad (2.5)$$

where the operator  $D : \mathbb{K}^{N_x} \rightarrow \mathbb{K}^{N_x \cdot d}$  approximates the image gradient by finite differences [35], the index  $i$  indicates the discrete pixel locations in image domain and  $d$  denotes the dimension of the image space. Due to the convexity of the TV semi-norm and, hence, the entire variational model in Equation (2.4), it allows for efficient global optimization. One particular feature of the TV semi-norm is that it approximates sparsity in the image edges, hence, it allows for sharp discontinuities in the reconstruction. However, it favors piece-wise constant solutions which are often not a suitable regularization to describe the content of natural and medical images. Especially when it comes to the acceleration of the image acquisition process, the choice of the regularization part and thus the weight parameter has a strong influence on how the final solution appears.

In the context of medical image reconstruction, Compressed Sensing (CS) [19, 31, 60, 164] is a widely used technique to reconstruct images from only a few measurements, sampled below the Nyquist rate [177, 220]. CS requires three conditions to be fulfilled. The first condition is the incoherence of artifacts arising due to the encoding of the undersampled measurements  $y$  [31, 60]. In MRI this can be achieved by non-Cartesian [19] or pseudo-random [164] sampling trajectories. The second condition states that the image has to be represented sparsely in a certain transform domain, which can be achieved by Wavelets [53, 164] or the TV semi-norm including extensions to higher-order Total Generalized Variation (TGV) [19, 21, 132, 207] as presented in Equation (2.5). To enforce sparsity and additionally allow for efficient global optimization, the  $\ell_1$  norm is used to approximate sparsity of the transformed image in the convex setting. Finally, both conditions, i.e., the consistency to the measured data and the sparsity in the transform domain, are combined in a non-linear reconstruction given in Equation (2.4), defining the third CS condition.

CS approaches show highly promising results for various imaging applications, however,

the translation to clinical practice is not trivial. In the case of MRI, Cartesian sampling schemes, which violate the incoherence assumption of artifacts, form the majority of clinical examinations [105]. Another observation is that the sparsifying transforms used in CS are too simple to capture the complex image content and structure of artifacts of medical images. Images reconstructed with the simple, handcrafted priors are often criticized by radiologists because the reconstructions can appear unnatural or blocky [105]. Another drawback, not only for CS-based approaches but also for many other iterative approaches, is the computational complexity and long reconstruction times for many of the algorithms used to solve the non-linear optimization problems. Finally, the sensitivity of hyper-parameters to the final reconstruction results makes it challenging to translate CS to clinical examinations, where imaging situations might vary. If the hyper-parameters are selected poorly, the reconstruction might be either under-regularized and still show artifacts or it might be over-regularized and thus biased. The latter case often results in a loss of details, which could be pathologies in the worst case, and an unnatural appearance of the images.

If we take a closer look at iterative reconstruction approaches, we observe that every new task is treated as a new optimization problem and no prior knowledge of the known structure of artifacts and the image content is taken into account. However, this contradicts how human radiologists read images, as they are trained throughout their careers to recognize certain patterns and read through diagnostic non-relevant patterns as residual artifacts [105]. When translating this observation to machine learning, this means that the optimization task can be transferred to an off-line training task where the key parameters are learned from undersampled data and clean images instead of solving an on-line optimization task for each new dataset based on handcrafted features. In the following sections, we will see how deep learning strategies in unsupervised and supervised learning improve upon the regularizer and overcome other challenges of CS-based approaches.

## 2.2 Unsupervised Learning in Image Reconstruction

We briefly review unsupervised learning strategies such as dictionary and sparsifying transform learning along with a K-sparse Autoencoder (KSAE). The dictionary or the sparsifying transform can either be pre-trained from a set of distorted training data or learned simultaneously with the reconstruction, which is termed *blind compressed sensing*.

Dictionary learning works at the level of image patches  $x_p \in \mathbb{K}^{N_p}$  of size  $p \times p$ . The dimension of the patch is defined by  $N_p = p^2$ . An arbitrary patch  $x_{p,i} \in \mathbb{K}^{N_p}$  can be extracted from an image  $x$  for the  $i^{\text{th}}$  patch location using a patch extraction matrix  $P_i : \mathbb{K}^{N_p \times N_x}$ . This patch extraction matrix  $P_i$  is defined for all  $N_s$  possible patch locations  $i$  in the image and considers the correct boundary conditions, e.g., symmetric or zero boundary conditions. The aim of synthesis dictionary learning is that an image patch can be approximated by a sparse linear combination of a dictionary  $D \in \mathbb{K}^{N_p \times N_\alpha}$  with a sparse  $N_\alpha$ -dimensional vector  $\alpha_i \in \mathbb{K}^{N_\alpha}$ . The dictionary  $D = [d_1, \dots, d_{N_\alpha}]$  consists of a normalized set of basis vectors  $d_j, j = 1, \dots, N_\alpha$ , called atoms. The according optimization problem for blind compressed

sensing is formulated as [195]

$$\min_{\substack{x \in \mathbb{K}^{N_x} \\ D \in \mathbb{K}^{N_p \times N_\alpha}}} \sum_{i=1}^{N_s} \frac{1}{2} \|P_i x - D \alpha_i\|_2^2 + \frac{\lambda}{2} \|Ax - y\|_2^2 \quad \text{s.t.} \quad \|\alpha_i\|_0 \leq K_0 \forall i, \|d_j\|_2 \leq 1 \forall j$$

where  $\lambda > 0$  and the parameter  $K_0$  defines the sparsity level. However, the above problem is NP-hard, even if the  $\ell_0$  quasi-norm is replaced by the convex  $\ell_1$  norm. Ravishankar et al. [195] proposed to solve this problem using an alternating minimization scheme. In the first step, the dictionary is learned using a K-Singular Value Decomposition (SVD) [8] algorithm. Due to the high computational requirement, only a fraction of all patches is used for this step. The sparse codes  $\alpha_i$  for all image patches are then estimated using the Orthogonal Matching Pursuit (OMP) algorithm [28]. In the second step, the reconstruction is updated, while the dictionary  $D$  and sparse codes  $\alpha_i$  are fixed. The reconstruction quality highly depends on the sparsity level  $K_0$  and the size of the dictionary. Promising results using dictionary learning have been shown for MRI [26, 195] and CT [258].

Instead of learning a synthesis dictionary, Ravishankar et al. [196] also proposed to learn the sparsifying transform, which can be seen as a generalization of the analysis model. Applied to blind compressed sensing, the corresponding model reads [197, 198]

$$\min_{\substack{x \in \mathbb{K}^{N_x} \\ W \in \mathbb{K}^{N_p \times N_p} \\ \alpha \in \mathbb{K}^{N_\alpha \times N_s}}} \sum_{i=1}^{N_s} \left( \frac{1}{2} \|W P_i x - \alpha_i\|_2^2 + \beta \|\alpha_i\|_0 \right) + \frac{\lambda}{2} \|Ax - y\|_2^2 \quad \text{s.t.} \quad W^* W = I,$$

where  $\beta > 0$ ,  $\lambda > 0$  and  $\alpha = [\alpha_1, \dots, \alpha_{N_s}]$  is a matrix of all sparse vectors  $\alpha_i$ . The sparsifying transform is represented by the matrix  $W \in \mathbb{K}^{N_p \times N_p}$ . Similar to dictionary learning, this problem can be solved by alternating minimization of the sparsifying transform  $W$ , the sparse codes  $\alpha_i \forall i$  and the reconstruction  $x$  [197, 198]. In general, learning of the sparsifying transform is computationally inexpensive compared to dictionary learning, because a closed-form solution based on thresholding exists for the sparse codes. In medical imaging, transform-based learning has led to promising results for both MRI and CT [197, 198, 200, 275]. Recently, the dictionary, transform and thresholding operators for MRI reconstruction were learned based on a supervised learning scheme [199] (see Section 2.3.2 for more details).

In both dictionary learning and sparsifying transform learning, patches are approximated by a sparse combination of dictionary atoms. To incorporate the whole image, an alternative approach is to learn a convolutional sparse representation [48, 250], where an image is approximated by convolving a number of  $K$  dictionary filters  $\zeta_k \in \mathbb{K}^{N_\zeta}$  with sparse coefficient maps  $c_k \in \mathbb{K}^{N_x}$ . The convolutional dictionary problem in synthesis form is formulated as

$$\min_{\substack{\zeta_k \in \mathbb{K}^{N_\zeta} \\ c_k \in \mathbb{K}^{N_x}}} \frac{1}{2} \sum_{s=1}^S \left( \left\| x_s - \sum_{k=1}^K \zeta_k * c_{k,s} \right\|_2^2 + \lambda \sum_{k=1}^K \|c_{k,s}\|_1 \right) \quad \text{s.t.} \quad \|\zeta_k\|_2 \leq 1 \forall k,$$



where  $\lambda > 0$  and  $x_s$  is the  $s^{\text{th}}$  of  $S$  training samples. To reconstruct a new image, the filters  $\zeta_k$  are fixed and the sparse coefficient maps  $c_k$  are learned. The final image is reconstructed by computing  $x^* = \sum_{k=1}^K \zeta_k * c_k$ .

A very recent approach for unsupervised learning learns a non-linear sparse prior based on KSAE [252] with application to iterative low-dose CT reconstruction. The goal of the KSAE is [167] to train an encoder  $f_{\text{enc}}$  and decoder  $f_{\text{dec}}$

$$(\theta_{\text{enc}}^*, \theta_{\text{dec}}^*) \in \arg \min_{\theta_{\text{enc}}, \theta_{\text{dec}}} \sum_{i=1}^{N_s} \frac{1}{2} \|x_i - f_{\text{dec}}(f_{\text{enc}}(x_i))\|_2^2 \quad \text{s.t.} \quad \|f_{\text{enc}}(x_i)\|_0 \leq K_0 \forall i,$$

where  $K_0$  defines the sparsity level and  $\theta_{\text{enc}}^*$  and  $\theta_{\text{dec}}^*$  denote the optimal parameters for the encoder  $f_{\text{enc}}$  and decoder  $f_{\text{dec}}$ , respectively. Here,  $x_i$  denotes the  $i^{\text{th}}$  of  $N_s$  training patches of clean reference images, i.e., normal-dose CT images. The encoder and decoder are realized using fully-connected networks and Rectified Linear Unit (ReLU) activations. Once the encoder and decoder networks are learned, a new image can be reconstructed according to the following unconstrained optimization problem

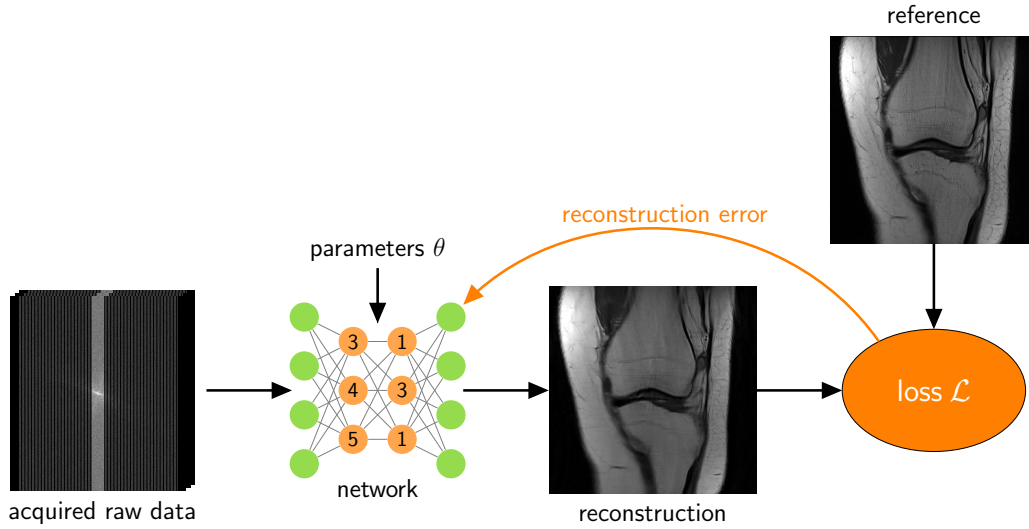
$$\min_{\substack{x \in \mathbb{K}^{N_x} \\ \hat{x} \in \mathbb{K}^{N_x \times N_s}}} \frac{1}{2} \sum_{i=1}^{N_s} \|P_i x - f_{\text{dec}}(f_{\text{enc}}(\hat{x}_i))\|_2^2 + \frac{\lambda}{2} \|Ax - y\|_2^2,$$

where  $\lambda > 0$ ,  $P_i$  is the patch-extraction matrix as defined for dictionary and transform learning, and  $\hat{x} = [\hat{x}_1, \dots, \hat{x}_{N_s}]$  are the patches projected on the set trained by the autoencoder. Wu et al. [252] propose this optimization problem in an alternating manner using the Separable Quadratic Surrogate (SQS) algorithm.

## 2.3 Supervised Learning in Image Reconstruction

Supervised learning approaches require three major ingredients for successful learning: A network architecture  $f_\theta$ , suitable training data and an appropriate loss function  $\mathcal{L}$ . In an off-line training procedure as depicted in Figure 2.2, the corrupted data are fed to a network parametrized by the parameters  $\theta$  that are optimized during training. The output of the network is compared to a reference image using a loss function  $\mathcal{L}$  that measures the similarity between the two images. This determines the reconstruction error that is back-propagated [147] through the network to obtain a new set of updated parameters. This process is repeated until the training has converged. Once all the parameters are learned, new images can be reconstructed efficiently as a simple application of  $f_\theta$ .

Figure 2.3 illustrates different strategies of how neural networks can be employed in image reconstruction. In the first method, a data-driven regularizer is learned from pairs of corrupted and clean images. This includes learning optimal weight parameters using bi-level optimization or learning an improved denoiser, which replaces fixed proximal steps in an iterative reconstruction scheme involving operations that ensure consistency to the given measurement data. The

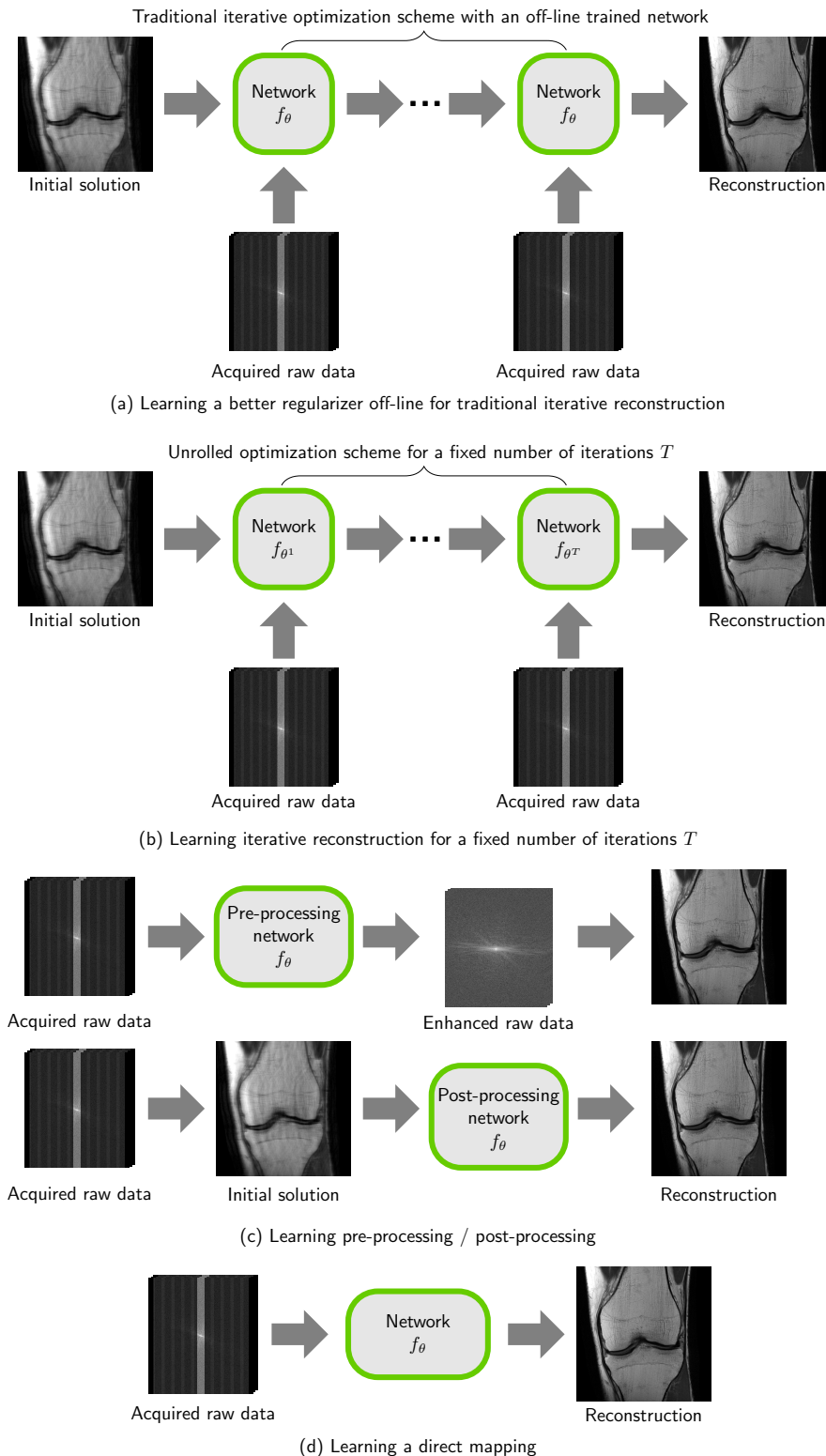


**Figure 2.2:** Schematic illustration of the training process. The corrupted data are propagated through a network with parameters  $\theta$  that should be learned to obtain an artifact-free reconstruction. The quality corresponds to the selected loss function  $\mathcal{L}$  that measures the similarity between the reconstructed image and the reference image. The reconstruction error is propagated back through the network to achieve a new set of parameters. Once the training has converged, new images can be reconstructed efficiently by simply forward-propagating the new data through the network using the fixed learned parameters.

second approach focuses on learning an unrolled iterative optimization scheme, involving data consistency term operations in every single unrolled step. These two approaches ensure data consistency to the measured data, which is not ensured by the following two approaches. The third approach defines image reconstruction as an image enhancement problem, which tries to improve upon an initial corrupted reconstruction from the measured data without including the measured data in the learning process. This approach also includes improvements upon the measured data directly in the measurement domain and performs, e.g., artifact correction before the image is finally reconstructed. The fourth approach focuses on mapping the measured data directly to the image domain, where the whole transform or parts of the image formation process can be learned. The four different approaches will be reviewed in detail in the following subsections.

### 2.3.1 Learning an Improved Regularization Function

Given the variational model in Equation (2.4), a natural question is how to select not only the correct weight parameter  $\lambda$ , but also the regularization function  $\mathcal{R}$  itself. While *bi-level optimization* [29, 55–57, 185, 209, 240] addresses the first question, the underlying handcrafted regularizer might still be too weak for more complex reconstruction tasks. Thus, developments have focused on the use of non-convex regularizers, embedded in a bi-level optimization approach. An overview over modern regularization techniques for inverse problems can be



**Figure 2.3:** Overview of how image reconstruction can be defined as deep learning-based problem. (a) Data-driven regularization in a traditional optimization problem with an off-line trained denoising network. (b) Learning an unrolled optimization scheme for a fixed number of iterations  $T$  (c) Image enhancement in data or image domain (d) Direct Mapping.

found in [15]. In recent publications, CNNs are trained to act as additional regularization in a variational model or to replace a proximal mapping in a Proximal Gradient (PG) scheme.

**Non-Convex Regularization.** Simple convex regularizers such as the TV semi-norm as defined in Equation (2.5) enable efficient global optimization where a unique solution exists. The used  $\ell_1$  norm is a convex approximation to the sparsity-inducing  $l_0$  norm. However, the  $\ell_1$  norm is not a suitable model to describe the prior distribution of natural and medical images. The quasi-convex  $\ell_q$  norm with  $q \in (0, 1)$  approximates the prior distribution better and already yields superior results [175]. However, when leaving convexity assumptions behind and moving to non-convex optimization, finding a good solution is even more challenging and highly sophisticated optimization algorithms are needed to overcome spurious local minima while allowing for efficient computation.

An example for non-convex regularization is the Fields of Experts (FoE) model [206], which can be seen as an extension of the TV semi-norm

$$\mathcal{R}[x] = \sum_{i=1}^{N_k} \langle \rho_i(K_i x), \mathbf{1} \rangle.$$

Here, the single term of the TV semi-norm is extended to  $N_k$  terms consisting of non-linear potential functions  $\rho_i(z) = (\rho_{i,1}(z_1), \dots, \rho_{i,N}(z_N)) \in \mathbb{K}^{N_x}$  and linear convolution operators  $K_i : \mathbb{K}^{N_x} \rightarrow \mathbb{K}^{N_x}$ . The symbol  $\mathbf{1} \in \mathbb{K}^{N_x}$  indicates a vector of ones. The free parameters of the FoE model are the convolution operators and the parametrization of the non-linear potential functions, which are both learned from data [206, 209]. The selected potential functions are associated with the statistics of natural images [119]. An example are the log-student-t functions  $\rho_{i,j}(z_j) = \alpha_{i,j} \log(\beta_{i,j} z_j^2 + 1)$ , applied in a point-wise manner, where  $j$  represents the discrete pixel locations in image domain. Learning strategies of the FoE model such as bi-level optimization [44, 45, 209] and learned iterative reconstructions, termed Variational Networks (VNs) [46, 84, 140], will be presented in this chapter and in Chapter 3.

**Bi-level Optimization.** One natural question that arises when solving Equation (2.4) is how to select the weight parameter  $\lambda$ . Different approaches exist to tackle this problem such as reinforcement learning [221] and bi-level optimization. Bi-level optimization problems consist of two problems: (i) A higher-level problem which defines the goal of the optimization, e.g., minimizing the Mean Squared Error (MSE). The outcome of this problem depends on (ii) the solution of the lower-level problem, which could be the solution to Equation (2.4). Mathematically speaking, bi-level optimization reads as

$$\min_{\theta \geq 0} \mathcal{L}(x^*(\theta), x_{\text{ref}}) \quad \text{s.t.} \quad x^*(\theta) \in \arg \min_{x \in \mathbb{K}^{N_x}} E(x, \theta),$$

where  $x^*$  is a solution to the lower-level problem  $\arg \min_{x \in \mathbb{K}^{N_x}} E(x, \theta)$ , depending on trainable parameters  $\theta$ , and minimizing  $\mathcal{L}$  defines the higher-level problem that compares  $x^*$  with a

reference  $x_{\text{ref}}$  using a certain mapping  $\mathcal{L}$ . To solve this bi-level optimization problem, the lower-level problem needs to be solved with high accuracy and is in general required to be twice differentiable in order to compute the gradient of the higher-level problem

$$\frac{\partial \mathcal{L}}{\partial \theta} = -\frac{\partial^2 E}{\partial x^* \partial \theta} \left( \frac{\partial^2 E}{(\partial x^*)^2} \right)^{-1} \frac{\partial \mathcal{L}}{\partial x^*},$$

where  $\beta > 0$  and  $\lambda > 0$ . In [178], a way to consider non-smooth functions as lower-level problems in bi-level optimization is proposed. Bi-level optimization helps to tune the reconstruction results, as the right weight parameters are learned as shown for analysis prior learning [185] and various image restoration problems [29, 55–57, 240]. However, the underlying regularizers are often still too simple to capture the characteristics of natural and medical images. Hence, the focus moved from learning single parameters to learning function parametrizations and filters from data [44, 45, 144, 209] which fit the statistics of natural images better [119]. However, the major drawbacks of bi-level optimization problems are the huge computational effort and the requirement to solve the lower-level problem exactly, which is especially critical in the context of medical image reconstruction due to, e.g., large-scale problems or the expensive computations of the forward operator  $A$  and the adjoint operator  $A^*$ .

**Convolutional Neural Networks as Regularization.** There are different ways to include CNNs in existing models. The earliest example in the context of MR image reconstruction trains a CNN that maps artifact-corrupted images  $A^*(y)$  to artifact-free images  $x_{\text{ref}}$  [246]. The trained CNN is then embedded in a CS formulation, where the CNN output acts as a reference image for the new optimization problem

$$x^* \in \arg \min_{x \in \mathbb{K}^{N_x}} \frac{\beta}{2} \|f_\theta(A^*(y)) - x\|_2^2 + \frac{\lambda}{2} \|A(x) - y\|_2^2 + \mathcal{R}[x].$$

The parameters  $\beta > 0$  and  $\lambda > 0$  are weight parameters that have to be selected individually for this optimization problem and  $\mathcal{R}[x]$  is an arbitrary fixed regularization function. The operators  $A$  and  $A^*$  define the forward and adjoint operator, respectively.

Another possibility how CNNs can improve the reconstruction quality is to replace existing proximal mappings in iterative reconstruction procedures to estimate a solution for the variational model in Equation (2.4). In many iterative reconstruction algorithms such as the PG method, the proximal mapping  $\text{prox}_{\tau \mathcal{R}}$  is employed to fulfill the regularization constraint while a gradient step is performed wrt. the data consistency term  $\mathcal{D}$ . This leads to the following iterative scheme

$$x^{t+1} = \text{prox}_{\tau \mathcal{R}} \left( x^t - \tau \nabla_x \mathcal{D} [A(x^t), y] \right),$$

where the proximal operator is defined as

$$\text{prox}_{\tau\mathcal{R}}(\hat{x}) = \arg \min_{x \in \mathbb{K}^{N_x}} \frac{1}{2\tau} \|x - \hat{x}\|_2^2 + \mathcal{R}[x]. \quad (2.6)$$

Heide et al. [102] and Venkatakrishnan et al. [241] observed that the proximal mapping in Equation (2.6) coincides with the Maximum-A-Posteriori (MAP) estimation of a Gaussian denoiser. It was suggested to replace the proximal operator by more sophisticated denoising algorithms such as BM3D [50] and Non-Local Means (NLM) [25]. In fact, this does not have to be necessarily done for the projected gradient method, but can also be achieved with any first-order method such as the Alternating Direction Method of Multipliers (ADMM) [20], used in the plug-and-play priors framework by [241], or the Primal-Dual (PD) algorithm [34], as shown in [102]. Replacing the denoising steps by BM3D and NLM denoising steps in medical image reconstruction has been successfully applied to MRI [4, 5, 64] and CT [47, 192].

Inspired by [102, 241], Meinhardt et al. [174] replaced the proximal operator by a learned denoising CNN

$$x^{t+1} = f_{\theta}(x^t - \tau \nabla_x \mathcal{D}[A(x^t), y]).$$

As before, this is not only restricted to a single PG method, but can also be used for other methods like ADMM or PD. The training of the used CNN is based on pairs of corrupted and clean reference images. The CNN can be trained more effectively by additionally feeding pairs of intermediate reconstruction results and reference images to the CNN [80, 126]. Promising results towards learning a proximal operator were shown for CT [37, 80, 126] and PET [71]. Obviously, learning a proximal operator has the advantage that it can be trained offline, meaning that the content of natural or medical images is reflected in the CNN. Once the denoiser is learned, it offers flexible usage for any new optimization problem with arbitrary data consistency terms. The learned denoiser, i.e., the proximal operator, remains the same in every single step of the optimization procedure.

### 2.3.2 Learning an Iterative Reconstruction Model

Instead of solving a new optimization problem for each task, the whole iterative reconstruction procedure itself can be learned. This can be motivated by the Landweber method [146]. Given an iteration-dependent step size  $\tau^t$  and an initial solution  $x^0$ , the Landweber method performs a gradient descent on the least-squares problem Equation (2.3), leading to the following iterative algorithm

$$x^{t+1} = x^t - \tau^t \nabla_x \mathcal{D}[A(x^t), y].$$

In order to prevent over-fitting to the measurement data  $y$ , it is beneficial to stop the Landweber method after a finite number of iterations  $T$  [97], which is known as early stopping. When including the regularization term, this idea can be extended to unroll a (proximal) gradient

**Algorithm 1** Learned Proximal Gradient, gradient in  $\mathcal{R}$  [58, 213]**Input:**  $x_0 \in \mathbb{K}^{N_x}$ **for**  $t = 0, \dots, T - 1$  **do**

$$x^{t+\frac{1}{2}} = x^t - f_{\theta^t}$$

$$x^{t+1} = \text{prox}_{\lambda^t \mathcal{D}[A(\cdot), y]} \left( x^{t+\frac{1}{2}} \right) = \arg \min_{x \in \mathbb{K}^{N_x}} \frac{1}{2} \left\| x - x^{t+\frac{1}{2}} \right\|_2^2 + \lambda^t \mathcal{D}[A(x), y]$$

**end for**

algorithm for a finite number of iterations  $T$ , where the data consistency term and regularization term are applied in an alternating manner

$$x^{t+\frac{1}{2}} = f_1(x^t), \quad (2.7)$$

$$x^{t+1} = f_2(x^{t+\frac{1}{2}}). \quad (2.8)$$

Here, the functions  $f_1$  and  $f_2$  consider the prior term and the data term, or vice versa, depending on the algorithm. The goal is now to learn this optimization scheme for a fixed number of  $T$  iterations, including step sizes, the prior model and weight parameters. We also see that this formulation enforces data consistency steps in every iteration. In 2005, Gregor and LeCun made first attempts towards learning an unrolled scheme by proposing the Learned Iterative Shrinkage and Thresholding Algorithm (LISTA) algorithm [76], which learns an encoder network to approximate the sparse codes in Iterative Shrinkage and Thresholding Algorithm (ISTA) [33, 52]. As stated in the previous section, this scheme is not only restricted to (proximal) gradient methods [1, 46, 84, 213], but can be also extended to a broader class of optimization algorithms [2, 58, 243, 247, 263].

The first possibility to solve Equation (2.7) and Equation (2.8) is with a PG method where  $f_1$  computes the gradient wrt.  $\mathcal{R}$  and  $f_2$  models the proximal operator on  $\mathcal{D}$ . This results in

$$x^{t+\frac{1}{2}} = x^t - \tau^t \nabla_x \mathcal{R} [x^t], \quad (2.9)$$

$$x^{t+1} = \text{prox}_{\tau^t \lambda \mathcal{D}[A(\cdot), y]} \left( x^{t+\frac{1}{2}} \right) = \arg \min_{x \in \mathbb{K}^{N_x}} \frac{1}{2\tau^t} \left\| x - x^{t+\frac{1}{2}} \right\|_2^2 + \lambda \mathcal{D}[A(x), y]. \quad (2.10)$$

Diamond et al. [58] and Schlemper et al. [213] suggest to replace the gradient update in Equation (2.9) by a trainable function  $f_{\theta^t}$  representing a CNN, which is allowed to change in every iteration  $t$ , or the weights  $\theta$  can be shared across the different iterations [6, 7, 193]. Then, Equation (2.9) naturally turns into a residual neural network [101]. The step size in Equation (2.10) can be neglected as it is implicitly contained in the weight parameter  $\lambda$  and the CNN, which are both allowed to change in every iteration. This version of the learned PG with the gradient in  $\mathcal{R}$  is depicted in Algorithm 1. To train Algorithm 1 end-to-end, a closed-form solution of the proximal operator has to exist.

**Algorithm 2** Learned Proximal Gradient, gradient in  $\mathcal{D}$  [100]

---

**Input:**  $x_0 \in \mathbb{K}^{N_x}$   
**for**  $t = 0, \dots, T - 1$  **do**  
 $x^{t+1} = f_{\theta^t}(x^t, \nabla_x \mathcal{D}[A(x^t), y])$   
**end for**

---

**Example: Single-coil MRI reconstruction [213].** An example in medical imaging for which the proximal operator has a closed-form solution is single-receive-coil MRI reconstruction. The data consistency term reads as  $\mathcal{D}[A(x), y] = \frac{1}{2} \|Ax - y\|_2^2$  with  $A = P\mathcal{F}$ , where  $\mathcal{F}$  is the Fourier transform and  $P$  models the encoding matrix, filling the missing  $k$ -space lines with zeros. The proximal mapping to this data consistency term is computed as

$$x^{t+1} = \mathcal{F}^* \Lambda^t \mathcal{F} x^{t+\frac{1}{2}} + \frac{\lambda^t}{1 + \lambda^t} \mathcal{F}^* P^* y,$$

$$\Lambda_{ii}^t = \begin{cases} 1 & \text{if } i \notin \Omega \\ \frac{1}{1 + \lambda^t} & \text{if } i \in \Omega \end{cases}$$

where  $\Lambda$  is a diagonal matrix and  $\Omega$  defines the index set of acquired  $k$ -space samples. As this proximal mapping enforces data consistency, this step is also termed *data consistency layer* in [193, 213].

Various applications in MRI exist to solve Equation (2.4) with algorithms like Algorithm 1. Schlemper et al. [213] proposed a deep cascaded CNN with the focus on dynamic MRI in a single-coil setting. This approach was improved by Qin et al. [193] who introduced a recurrent neural network architecture that shares weights over layers and considers additional memory in both spatial and time domain. Aggarwal et al. [6, 7] proposed an approach for both single-coil and multi-coil MRI for static and dynamic imaging. A similar idea was used by [170], which consists of a single iteration of CNN regularization and data consistency step. An improved version using a recurrent architecture was presented in [169].

A second possibility to solve Equation (2.7) and Equation (2.8) is a variant of the PG method where  $f_1$  now computes the gradient wrt.  $\mathcal{D}$  and  $f_2$  models the proximal operator on  $\mathcal{R}$ , resulting in

$$x^{t+\frac{1}{2}} = x^t - \tau^t \nabla_x \mathcal{D}[A(x^t), y],$$

$$x^{t+1} = \text{prox}_{\tau^t \mathcal{R}}(x^{t+\frac{1}{2}}).$$

In principle, the proximal operator can be replaced by a trainable function  $f_{\theta^t}$ , however, it is challenging to define the correct step size  $\tau^t$ . Therefore, Adler et al. [1, 2] and Hauptmann et al. [100] proposed to let the CNN learn how to combine  $x^t$  with the gradient of the data consistency term  $\nabla_x \mathcal{D}[A(x^t), y]$ . Hence,  $f_{\theta^t}$  takes  $x^t$  stacked with  $\nabla_x \mathcal{D}[A(x^t), y]$  as an input, such that the update  $x^{t+1}$  can be reformulated as shown in Algorithm 2. Motivated



**Algorithm 3** Learned Proximal Gradient with memory, gradient in  $\mathcal{D}$  [1]

---

**Input:**  $x_0 \in \mathbb{K}^{N_x}$ ,  $x_{\text{mem}}^0 \in \mathbb{K}^{N_{x,\text{mem}}}$   
**for**  $t = 0, \dots, T - 1$  **do**  
 $(x^{t+1}, x_{\text{mem}}^{t+1}) = f_{\theta^t}(x^t, \nabla_x \mathcal{D}[A(x^t), y], x_{\text{mem}}^t)$   
**end for**

---

by accelerated gradient methods, information of earlier iterates can be used by extending the formulation with an additional memory variable  $x_{\text{mem}} \in \mathbb{K}^{N_{x,\text{mem}}}$ , leading to Algorithm 3, where  $N_{x,\text{mem}}$  defines the dimension of the memory.

To introduce another class of optimizers, let us first go back to traditional optimization. For this purpose, we formulate the variational model in Equation (2.4) as a general model

$$\min_{x \in \mathbb{K}^{N_x}} F(H(x)) + G(x), \quad (2.11)$$

where  $F : \mathbb{K}^{N_p} \rightarrow \mathbb{R}$  is a proper convex lower-semicontinuous (l.s.c.) and possibly non-smooth function, the function  $G : \mathbb{K}^{N_x} \rightarrow \mathbb{R}$  is smooth and proper convex l.s.c., and  $H : \mathbb{K}^{N_x} \rightarrow \mathbb{K}^{N_p}$  is a differentiable, possibly non-linear, operator. In some cases, the proximal mapping on  $F$  does not have a closed-form solution, so PD methods [34, 35], which were extended to non-linear operators [239], offer a possibility to tackle this problem. By introducing a dual variable  $p \in \mathbb{K}^{N_p}$ , the primal problem in Equation (2.11) can be cast to a PD saddle-point problem with following structure

$$\min_{x \in \mathbb{K}^{N_x}} \max_{p \in \mathbb{K}^{N_p}} \langle H(x), p \rangle + G(x) - F^*(p),$$

where  $F^*$  is the Fenchel conjugate of  $F$ . A solution can be approximated by using the (non-linear) PD hybrid gradient method depicted in Algorithm 4, where  $\tau, \sigma$  are the primal and dual step sizes, and the expression  $[\partial H(x^t)]^* : \mathbb{K}^{N_p} \rightarrow \mathbb{K}^{N_x}$  denotes the adjoint of the derivative of  $H$ , evaluated in  $x^t$ . When inspecting Algorithm 4, we observe two proximal operators in the primal and dual space, as well as an over-relaxation step. This is the motivation of Adler et al. [2] to learn the proximal operators along with the update steps, resulting in the learned PD Algorithm 5, where the parameters that are learned during training are  $\theta = [\theta_{\text{primal}}, \theta_{\text{dual}}]$ .

The learned PG methods have been applied to different medical inverse problems. Adler et al. [1, 2] showcase Algorithm 3 and Algorithm 5 on a 2D CT problem, including a discussion on the performance using linear and non-linear operators. Hauptmann et al. [100] used Algorithm 2 for PAT reconstruction. An extension to a multi-scale U-net architecture was presented in [99].

Another possibility to tackle the problem defined in Equation (2.4) is to take gradient steps in the direction of the prior term and the data consistency term, either in an alternated way or in a single combined step. This concept was initially used in learning an optimized reaction-diffusion process for image restoration [46], and further led to the formulation of Variational Networks [84, 140], which have a strong connection to both variational models and

**Algorithm 4** Non-Linear Primal-Dual Hybrid Gradient [34, 239]

---

**Input:**  $x^0 \in \mathbb{K}^{N_x}$ ,  $p^0 \in \mathbb{K}^{N_p}$   
**Choose:**  $\sigma, \tau > 0$  s.t.  $\sigma\tau \|H\|_2^2 \leq 1$ ,  $\vartheta \in [0, 1]$   
**for**  $t = 0, \dots, T - 1$  **do**  
 $p^{t+1} = \text{prox}_{\sigma F^*} (p^t + \sigma H(\tilde{x}^t))$   
 $x^{t+1} = \text{prox}_{\tau G} (x^t - \tau [\partial H(x^t)]^* (p^{t+1}))$   
 $\tilde{x}^{t+1} = x^{t+1} + \vartheta (x^{t+1} - x^t)$   
**end for**

---

**Algorithm 5** Learned Primal-Dual [2]

---

**Input:**  $x_0 \in \mathbb{K}^{N_x}$ ,  $p_0 \in \mathbb{K}^{N_p}$ ,  $x_{\text{mem}}^0 \in \mathbb{K}^{N_{x,\text{mem}}}$ ,  $p_{\text{mem}}^0 \in \mathbb{K}^{N_{p,\text{mem}}}$   
**for**  $t = 0, \dots, T - 1$  **do**  
 $(p^{t+1}, p_{\text{mem}}^{t+1}) = f_{\theta_{\text{dual}}^t} (p^t, p_{\text{mem}}^t, H(x_{\text{mem}}^{(1),t}), y)$   
 $(x^{t+1}, x_{\text{mem}}^{t+1}) = f_{\theta_{\text{primal}}^t} (x^t, x_{\text{mem}}^t, [\partial H(x^t)]^* (p^{t+1}))$   
**end for**

---

deep learning. As VNs are the main focus of this thesis, their theory is described in detail in Chapter 3. The basic concept of VNs relies on unrolling a gradient descent step for a finite number of iterations  $T$  in direction of the gradient of a smooth variational model defined in Equation (2.4)

$$x^{t+1} = x^t - \tau^t (\nabla_x \mathcal{R} [x^t] + \nabla_x \mathcal{D} [A(x^t), y]).$$

The step size  $\tau^t$  can be neglected during learning as it is implicitly contained in  $\mathcal{R}$  and  $\mathcal{D}$ . VNs are characterized by the special structure of the regularization term and data consistency term which is related to variational models as depicted in Algorithm 6. For the regularization part, this relation is typically fulfilled by the FoE model. In case the exact data consistency term is not known, it can be replaced by a trainable data consistency term, e.g.,  $\mathcal{D} [A(x^t), y] = \rho_D^t (Ax^t - y)$  [130, 243] where  $\rho_D$  are potential functions defined in the same way as  $\rho$ . Due to the strong connection to variational models, VNs allow us to gain some insights into what is learned by the model, as filters and activation / potential functions can be visualized. The first works on restricting the potential functions to convex functions and studies on the relation to energy minimization were shown in [140]. Furthermore, VNs are a versatile architecture that have shown state-of-the-art results not only in image restoration tasks, including image denoising [46], JPEG deblocking [46], demosaicing [130] and image inpainting [267], but also in MRI reconstruction [84, 136] and low-dose CT reconstruction [141]. In Real-Time Hand-Held Sound-Speed Imaging [243], an unrolled gradient descent scheme with momentum is learned along with an additional parametrization of the data consistency term. Inspired by the idea of VNs, Chen et al. [39] replaced the gradient of the regularization function by a CNN part for sparse-view CT reconstruction. In a similar approach, a CNN with trainable activation

**Algorithm 6** VNs [46, 84, 140]**Input:**  $x^0 \in \mathbb{K}^{N_x}$ **for**  $t = 0, \dots, T - 1$  **do**

$$x^{t+1} = x^t - f_{\theta_{\mathcal{R}}}^t(x^t) - f_{\theta_{\mathcal{D}}}^t(x^t, y) \text{ s.t. } f_{\theta_{\mathcal{R}}}^t = \nabla_x \mathcal{R}_{\theta_{\mathcal{R}}}^t, f_{\theta_{\mathcal{D}}}^t = \nabla_x \mathcal{D}_{\theta_{\mathcal{D}}}^t$$

**end for****Algorithm 7** Recurrent Inference Machines [191]**Input:**  $x^0 \in \mathbb{K}^{N_x}, x_{\text{mem}}^0 = 0$ **for**  $t = 0, \dots, T - 1$  **do**

$$\begin{aligned} x_{\text{mem}}^{t+1} &= \hat{f}_{\theta}(\nabla_x \log p(y|x)(x^t), x^t, x_{\text{mem}}^{t+1}) \\ x^{t+1} &= x^t + f_{\theta}(\nabla_x \log p(y|x)(x^t), x^t, x_{\text{mem}}^{t+1}) \end{aligned}$$

**end for**

functions was used for joint denoising and segmentation of tumor cells [63].

A different approach for learned iterative MR reconstruction using Recurrent Inference Machines [191] was shown in [161] and is stated in Algorithm 7. An aliased image along with the gradient of its log-likelihood  $\nabla_x \log p(y|x)(x^t)$  is fed to a recurrent neural network  $f_{\theta}$ , where weights are shared across iterations. This produces an incremental update of the input  $x$  and the latent memory variable  $x_{\text{mem}}$ , where  $\hat{f}_{\theta}$  in Algorithm 7 is the part of  $f_{\theta}$  generating the update for  $x_{\text{mem}}$ .

There are also other approaches that fall into the class of learning an iterative reconstruction. Dictionary-transform learning has been presented in Section 2.2, where alternating updates wrt. to data consistency term and dictionary / transform update steps are performed. Recently, Ravishanker et al. [199] proposed to unroll this scheme and learn the regularizer layer-by-layer in a supervised way, involving the dictionary, transforms and thresholds using a neural network approach. For MRI reconstruction, Eo et al. [65] proposed the KIKI-net that performs iterative updates in the  $k$ -space and image domain which are both trained with CNNs, with intermediate data consistency layers.

The presented unrolled model-based iterative approaches were defined in a very generic way which makes them convenient to adapt to a new model when the forward and adjoint operators  $A$  and  $A^*$  are known. Thus, data consistency to the measured data  $y$  is ensured in every single iteration. However, the complexity of some operators in medical image reconstruction such as PAT [100] makes it infeasible to train the whole algorithm end-to-end. In this case, the algorithm has to be trained in a greedy manner iteration-by-iteration, although end-to-end training would result in a performance increase [46]. Hence, a trade-off between tractable computation time and quality of the output has to be made. If the parameters in every iteration vary, the model is more flexible to adapt to training data [140]. Special recurrent architectures share the weights across iterations and additionally transport information through the iterative procedure via a memory state variable, which greatly reduces the number of parameters while still being flexible.

### 2.3.3 Deep Learning for Image and Data Enhancement

In contrast to iterative reconstruction methods, where the measured raw data are accounted in every reconstruction step, enhancement methods do not include the physical model. Enhancement can be performed in the data domain as pre-processing  $y^* = f_\theta(y)$  or in the image domain as post-processing  $x^* = f_\theta(x^0)$ . In the case of pre-processing, the reconstruction is performed on the enhanced measurement data  $y^*$ , while in the case of post-processing an initial solution  $x^0$  is first reconstructed and then enhanced to obtain  $x^*$ . An obvious advantage is that the training is highly efficient as the physical model, which is often heavily time consuming to evaluate numerically, is not considered. The drawback is that the used network architecture can only work on information that is provided in the initial guess and the information that is extracted from the training data. Consequently, consistency to measured raw data is not maintained for post-processing methods. As the algorithms solely work in the image domain, various types of post-processing methods such as image denoising, artifact removal and image super-resolution fall into this class of problems. These algorithms are either applied after reconstruction of  $x_0$  or on top of other correction methods.

In MR imaging, the input to the used algorithms are mainly coil-combined or sum-of-squares-combined images. In [111, 261], MR images were reconstructed from aliased single-coil images using a U-net architecture. Extensions to 3D [118] and dynamic MRI [210] were made using similar approaches. Jin et al. [118] additionally suggest to provide input and reference images that underwent a Wavelet transform to suppress image modality-specific noise, similar to [123]. Instead of learning the mapping from the aliased input image to the artifact-free reference image, residual learning [101] provides a way to learn the aliasing artifacts instead, which are then subtracted from the aliased input image as presented in [96, 151]. Using a similar approach, Lee et al. [150] included both image magnitude and phase information and showed that this acts as iterative  $k$ -space interpolation. While the aforementioned algorithms mainly work on single-coil images, Kwon et al. [145] introduced a Multi Layer Perceptron (MLP) that unfolds the single coil images of Cartesian undersampled data line-by-line, similar to classical Sensitivity Encoding (SENSE) [190]. However, this architecture has to be adapted and retrained for different matrix sizes and sampling patterns. Besides undersampled MR image reconstruction, a residual CNN approach was used for single-shot  $T_2$  mapping using overlapping-echo detachment planar imaging [27]. Another research area for post-processing methods is image superresolution. Deep learning approaches offer a way to directly learn the mapping from low-resolution to high-resolution images, where learning-based results generally outperform conventional interpolation approaches such as cubic interpolation. This topic was studied in the context of dynamic MRI [179], static MRI [36, 186, 223], fetal MRI [172] and Diffusion Tensor Imaging (DTI) [228].

Image denoising of low-dose CT data is a well studied example for the application of CNNs in medical image reconstruction [40, 41, 154, 219, 251, 262]. Low-dose CT has been also studied within the framework of deep-convolutional framelets [122, 123, 266], which draws a mathematical connection between CNNs and Framelet representations where the underly-

ing encoder-decoder structure originates from the Hankel matrix decomposition. Within this framework, the input to the used U-net underwent a Wavelet transform. A similar Wavelet residual learning approach was proposed for limited-angle CT [78]. VNs [46, 140] provide another alternative for both low-dose CT [141] and limited-angle CT [92]. To account for the structured streaking artifacts in sparse-view CT, opposed to an increase in unstructured noise that arise from low-dose CT, algorithms such as residual learning using a U-net architecture [93, 117, 256], deep convolutional framelets [94], or a combination of DenseNet and deconvolution [273] were proposed. Image super-resolution was also studied in the context of CT to map thick slices to thin slices [183]. Another application of image enhancement methods is CNN-based Metal Artifact Correction (MAR) in CT [272]. Here, a CNN is trained to generate an image with reduced artifacts from uncorrected and pre-corrected images. The forward projections of the predicted image are then used to correct the projections which are affected by metals such that the Filtered Back-Projection (FBP) yields an artifact-reduced image.

Image enhancement in PET imaging has been studied in [260]. Here, an enhanced PET image was generated from multiple MAP solutions with different regularization weights using an MLP. Xu et al. [257] proposed to map a set of adjacent low-dose PET slices to a standard-dose PET image using a U-net architecture. Xiang et al. [255] fed both a low-dose PET and a  $T_1$ -weighted MR image in a CNN to obtain a standard-dose PET image.

In the area of PAT, Antholzer et al. [12] presented a CNN approach to improve the quality of sparsely sampled reconstructions. Schwab et al. [217] first reconstructed limited-view PAT data and applied a dynamic aperture length correction algorithm, before a U-net is applied to improve the final reconstruction quality.

Instead of performing post-processing to enhance reconstructions in the image domain, another approach is to perform pre-processing to enhance the measurement data itself. Examples here are artifact detection and removal, such as MAR in CT. While Claus et al. [49] used a MLP to inpaint the missing data in the sinogram on patches, Park et al. [182] learned a U-net to correct the areas affected by metals on the whole sinogram. In PAT, reflection artifacts are created by waves due to echogenic structures, which might appear as a true signal. Allman et al. [11] locate and classify wave sources and artifacts and correct the measurement data before reconstruction.

#### 2.3.4 Learning a Direct Mapping

Up to now, the presented methods train a neural network in the image domain. Image reconstruction can be also seen from a different perspective where the complete or partial mapping from the measured data to the image is learned. Learning the complete mapping means that no information of the underlying physics is incorporated, i.e.,

$$x^* = f_{\theta}(y).$$

To be able to realize this mapping, a high dimensional parameter space as well as diverse training data that account for all kinds of different data are required. However, this is typically infeasible in medical imaging due to the limited amount of training data. Besides the high memory requirements of full direct mappings, there is also no guarantee that the final reconstruction is consistent to the measured data. Furthermore, it is a challenging question if the valid and well known operations such as FT or Radon transform should be replaced by a neural network, although efficient implementations exist.

An example for a full direct mapping from the measured data to the image is Automated Transform by Manifold Approximation (AUTOMAP) [276] which was proposed for a broad range of applications. The key of the network architecture is that fully connected layers are used as first layers, followed by a CNN, similar to ReconNet [143] which acts on sparsely sampled image patches. In the context of MRI, the inverse FT is learned. Although results for small image sizes in [276] demonstrate this to be a promising research direction, a major practical challenge of AUTOMAP is that the algorithm does not scale to large problems due to the huge memory requirements of fully connected layers.

Another example that falls into this class of problems is learning a transform from a small number of Digitally Reconstructed Radiograph (DDR) projections to a CT reconstruction using a CNN [229, 230, 265]. In [229, 230], images are generated from 1D sum projections DDRs which are repeated in the direction of the acquisition angle. The images of the different acquisition angles are stacked and form the input to a U-net architecture, whose output is the final CT image. Similarly, sinograms can be directly mapped to image space using a CNN architecture. This direct reconstruction approach using an encoder-decoder network was proposed by Häggström et al. [82] for PET reconstruction.

Instead of learning the entire model, only parts of the reconstruction process can be replaced by deep neural networks, while known mathematical transforms contained in  $A^*$  are fixed and fully connected layers become needless

$$x^* = A^*(f_\theta(y)).$$

Hence, parts such as the ramp filter for Backprojection (BP) could be learned [67]. A similar approach was proposed for limited-angle CT, where projection-domain weights are learned to account for missing data for different geometries [253, 254]. The backprojection layer is mapped to a neural network with fixed weights corresponding to the true mathematical transform, enabling full end-to-end training of the architecture.

Motivated by the  $k$ -space interpolation by the Annihilating Filter-Based Low-Rank Hankel Structured Matrix Completion Approach (ALOHA), Ye et al. [266] drew connections between the Hankel matrix decomposition and deep neural networks by convolutional framelets, which have been presented for image enhancement problems in Section 2.3.3. Similar U-net architectures and residual learning strategies were used to inpaint the missing  $k$ -space data while the loss function itself is defined in image domain, with application to accelerated MRI [95], MR angiography [32] and Echo Planar Imaging (EPI) ghost correction [152].

### 2.3.5 Other Approaches for Image Reconstruction

Besides the presented approaches so far, there are also other applications where learning can be helpful to improve image reconstruction. One example is MRI where the selection of the sampling pattern has a strong influence on the appearance of undersampling artifacts. Gözcü et al. [75] proposed an learning-based approach to optimize for this sampling patterns, justified by statistical learning theory.

In PET/MR, the attenuation correction of PET images is challenging, because the necessary information cannot be extracted from MR images directly. Gong et al. [72] proposed to use Dixon MR images together with Zero Echo Time (ZTE) images to generate pseudo-CT images using a U-net architecture. These pseudo-CT images can then be used for attenuation correction.

## 2.4 Training data

Training data is one of the key ingredients why learning-based solutions became successful in the last years. In the computer vision community, the ImageNet database [116] became one of the most important databases for deep learning, consisting of about 14 million labeled images from 1000 different classes. This freely available database allows the researchers to develop new methods with a huge amount of data and serves as a standardized test benchmark that makes it possible to compare the newly developed methods objectively to other existing approaches in literature. However, obtaining training data for medical solutions comprises various challenges. While crowdsourcing made it possible that ImageNet grew within a short time to this huge amount of labeled images in an inexpensive way, collecting training data for medical solutions is expensive and requires highly trained Medical Doctors (MDs) to perform tedious labeling tasks. Furthermore, data has to be anonymized before it is shared. Hence, big data becomes available only slowly in medical imaging. The UK Biobank<sup>1</sup> aims to provide medical data from 500,000 volunteers to improve the diagnoses, treatment and prevention of a broad range of life-threatening and serious illnesses. Further examples for ongoing projects that provide data to the public are the human connectome project<sup>2</sup>, the Montreal Neurological Institute's Brain Images of Tumors for Evaluation database<sup>3</sup>, the National Alzheimer's Coordinating Center (NACC)<sup>4</sup>, the Alzheimer's disease Neuroimaging Initiative (ADNI)<sup>5</sup> or DeepLesion<sup>6</sup> [259].

While these projects offer a great potential for a wide range of applications, including classification and segmentation, they cannot be used for image reconstruction, because only the reconstructed Digital Imaging and Communications in Medicine (DICOM) images and not the raw measurement data is included. Most of currently available data repositories

<sup>1</sup><http://www.ukbiobank.ac.uk/>

<sup>2</sup><http://www.humanconnectomeproject.org/>

<sup>3</sup><http://www.bic.mni.mcgill.ca/laurence/data/>

<sup>4</sup>[https://www.alz.washington.edu/WEB/researcher\\_home.html](https://www.alz.washington.edu/WEB/researcher_home.html)

<sup>5</sup><http://adni.loni.usc.edu/>

<sup>6</sup><https://nihcc.app.box.com/v/DeepLesion>

consist only of small collections of data. This data were either associated with dedicated research challenges, like the Low Dose CT challenge organized by Mayo Clinic<sup>7</sup> or provided by individual groups in connection with specific research papers, e.g., in our case we provided 100 datasets of a full clinical knee protocol with our paper on accelerated MRI reconstruction [84]. However, these fragmented datasets were obtained with the same scanner hardware and same imaging protocol and do not reflect the wide range of scanner hardware from multiple vendors, scanner protocols and imaging artifacts associated with the acquisition. Even sharing raw data across multiple vendors is challenging, as different data formats are used. A first multi-center endeavor was recently started<sup>8</sup> with the goal to provide raw  $k$ -space data from multiple vendors in the standardized ISMRMD raw data format [112]. However, due to the limited availability of raw data for image reconstruction, every research group uses its own dataset for learning. Additionally, numerous assumptions are made in many cases and algorithms are only deployed for simulated datasets and not for real clinical settings. In summary, the current lack of a big raw data archive, including data from multiple scanner hardware, makes it challenging to both train models from large data sets, and validate the performances of a new approaches. Recently, a first step towards a generalized data archive was made by the cooperation between Facebook Artificial Intelligence Research (FAIR) and New York University by releasing an MRI-dataset consisting of more than 1,500 volumes of raw  $k$ -space data of clinical knee examinations collected with different Siemens scanners [268]. In addition to this raw data, DICOM data from 10,000 clinical knee examinations were released, which could be used for, e.g., transfer learning.

**Transfer Learning.** Current network architectures consist of a large number of free parameters and require extensive computations, training time and large datasets. Especially in cases where it is challenging or impossible to acquire large-scale datasets, a common scenario in medical imaging, the concept of transfer learning [59] provides a way to still be able to obtain reasonable results for learning-based approaches. The idea of transfer learning [59] is that networks are pre-trained on large available datasets such as ImageNet [116] or simulated data. Afterwards, the network is fine-tuned on a limited amount of application-dependent data such that the networks adapt on the current application. Transfer learning was recently investigated for MR image reconstruction of neurological data [51] with a cascaded CNN architecture proposed in [213], and for musculoskeletal data [136] using a VN [92] where the architectures were pre-trained on natural images. Learning destreaking on CT images and fine-tune on radially undersampled MR data was presented in [96].

## 2.5 Image Quality

The performance of learning-based approaches does not only depend on the network architecture and training data, but also on the loss functions used during training. Quantitative image

<sup>7</sup><https://www.aapm.org/GrandChallenge/LowDoseCT/>

<sup>8</sup><http://mridata.org/>



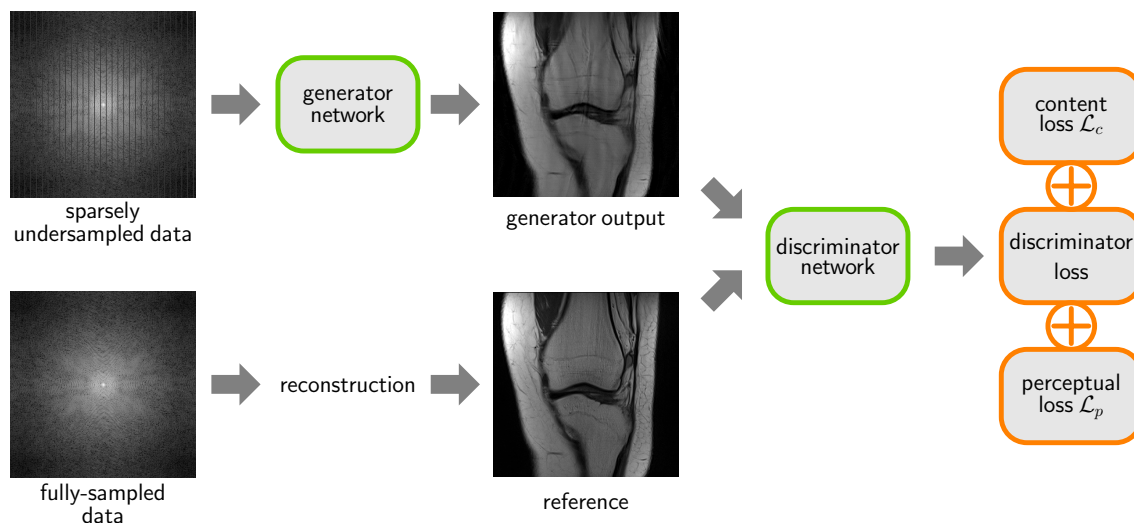
quality measures, such as the pixel-wise squared  $\ell_2$  loss or the patch-based Structural Similarity Index (SSIM) [248], serve as typical loss functions. However, the reconstructed images trained with these measures can appear oversmoothed due to the averaging nature of these approaches and their sensitivity to noise [274]. This effect is particularly prominent if the training data are noisy. More details on loss functions can be found in Section 3.4.

Generative Adversarial Networks (GANs) [74] have shown a great potential in increasing the perceptual image quality by training a loss function from data. Figure 2.4 shows the basic architecture of GANs for image reconstruction. GANs consist of two networks, a generator network and a discriminator network, that try to compete with each other: The generator generates images from a source distribution, whereas the discriminator tries to distinguish the generated images from a clean reference distribution. However, GANs are inherently difficult to train. The initial formulation involves a sigmoid cross-entropy loss which often leads to mode collapse, vanishing gradients and thus unstable training. Many other strategies have been proposed to stabilize the training such as Wasserstein Generative Adversarial Networks (wGANs) [13], which were further stabilized with gradient penalty [79], or least-squares GANs [168].

While GANs are capable to generate images that appear with a similar texture to the target distribution, this does not necessarily conclude that the resulting images are anatomically correct. Especially, if only few data are available, GANs might introduce new artificial structures, an effect commonly known as hallucination. While this behavior can be desirable in the context of natural images, when the goal is only to create images that appear sharp and natural, it must be avoided in medical imaging and diagnosis. Therefore, pixel-based content losses such as  $\ell_2$  or  $\ell_1$  norms are added to the GAN loss to guide the reconstruction and thus provide general information about the image content [114, 149]. This idea was further extended by introducing a perceptual loss [14, 149], which is based on feature responses of the VGG network [225].

Wolterink et al. [251] used a GAN together with an  $\ell_2$  loss for noise reduction in low-dose CT. In another approach for low-dose CT image denoising, a wGAN architecture with gradient penalty was combined with MSE as content loss and a perceptual VGG loss [262]. Thaler et al. [229] included an  $\ell_1$  content loss in a wGAN architecture with gradient penalty for CT reconstruction from DDR projections. They also provide first insights in the importance of the content loss to guide the reconstruction process and the limits for data acquisition to obtain reasonable results.

A classical GAN approach combined with stabilizing content losses for enhancing MRI reconstructions was proposed in [128, 224, 261]. Least-squares GANs were studied in [169, 170] in combination with an  $\ell_2$  loss for accelerated MR. Quan et al. [194] adopted a GAN with cyclic loss [277] that ensures that the generated images are similar to the input distribution using a second generator. In another approach for MRI, Seitzer et al. [218] refined reconstructions trained with MSE by learning a residual image using a GAN architecture with additional feature matching loss [208] and VGG loss.



**Figure 2.4:** For image reconstruction, GANs are combined with a content loss  $\mathcal{L}_c$  such as MSE to stabilize the training. Various authors showed that using a perceptual loss  $\mathcal{L}_p$  based on features from the VGG network [225] results in even more improved reconstructions.

**Evaluation.** The loss function used for training has an impact on the quality of the reconstruction. These measures serve also for quantitative, hence, objective comparison for the newly developed approaches. Although this kind of evaluation is essential, these measures give only limited insight into the real diagnostic value of the reconstructed images. While the presence or absence of tiny and subtle structures in the reconstructed images is indispensable in a diagnostic setting, these structures are usually not captured well by quantitative criteria. Therefore, an important open question in the context of medical imaging is if the currently developed learning-based approaches are robust and general enough so that they can be translated successfully to clinical practice. This requires large scale prospective clinical evaluation studies, where inter-subject anatomical and pathological variations encountered in a typical patient population are assessed. A radiologists reader study of the visual image quality of the reconstructions regarding criteria like image sharpness, residual aliasing artifacts and apparent SNR was performed in our initial publication on VNs for 20 knee exams [92], see Section 5.3. In [38], the performance of this architecture was evaluated with similar criteria in 157 abdominal patients. In both studies, learned image reconstruction was found to outperform conventional PI and CS with respect to these criteria. While these studies are examples of evaluations that go beyond objective quantitative metrics, the diagnostic value of these studies is still limited because the diagnostic information are not considered in these reader studies. Current learning-based solutions might result in a degraded image quality and remaining artifacts, which might not influence the radiologists to make the correct diagnosis. However, if the degraded images and remaining artifacts lead to an incorrect diagnosis and, hence, an incorrect patient management decision this case can be considered as a truly failed imaging exam. A small-scale prospective study of the diagnostic content of learned image

reconstruction was performed in our work [134] for 25 knee MRI patients. The reconstructions were assessed for the presence or absence of typical pathologies like meniscal tears or ligament tears and the diagnostic outcome was compared to a clinical gold standard data acquisition and for some cases was followed up with a scope exam to obtain a true ground truth. Excellent agreement between the two acquisitions was found in this study, demonstrating that an accelerated acquisition with a learned reconstruction had the same diagnostic content and, hence, led to the same clinical decision as the existing gold standard. However, this study is still limited by its comparably small sample size, the restriction to a very specific type of exam and the analysis of only one contrast out of a usually more diverse MR acquisition.

## 2.6 Reproducible Research

Using deep learning approaches, impressive results are presented more than ever and it is often challenging to reproduce the claims and results authors make in their publications, especially because validations are usually only performed on their own individual datasets. Reproducible research promotes that data and source code should be made available along with the publication of the paper. In the special issue for learning-based image reconstruction of IEEE Transactions on Medical Imaging [245], sharing code was mandatory with publication of the paper. Most authors shared their code on GitHub<sup>9</sup>. Available software repositories for the presented approaches in this chapter can be found alongside the publication entry in the references.

During the course of my PhD, I contributed to several open-source projects which were used for various experiments throughout this thesis:

- ImageUtilities<sup>10</sup>: Bridging the gap between Central Processing Unit (CPU) and Graphics Processing Unit (GPU) to overcome the tedious memory management in CUDA code, with python and Matlab interfaces.
- Primal-Dual Toolbox<sup>11</sup>: Implementation of various TV and second-order TGV problems using the primal-dual algorithm [34], including python and Matlab wrappers. This toolbox includes algorithms for TGV-based MRI reconstruction for Cartesian and radial sampled data as well as real-valued and complex-valued denoising.
- Tensorflow-ICG<sup>12</sup>: Fork of the original tensorflow repository<sup>13</sup>. Includes custom operators which are necessary for the VN implementations such as trainable activation functions, Inertial Proximal Alternating Linearized Minimization (IPALM) optimizer [188], core framework for VNs as well as complex convolutions and Fourier operators, including centered 2D (inverse) FT and (inverse) fftshift.

---

<sup>9</sup><https://github.com>

<sup>10</sup><https://github.com/VLOGroup/imageutilities>

<sup>11</sup><https://github.com/VLOGroup/primal-dual-toolbox>

<sup>12</sup><https://github.com/VLOGroup/tensorflow-icg>

<sup>13</sup><https://github.com/tensorflow/tensorflow>

- Denoising VN<sup>14</sup>: Tensorflow implementation of the VN for image denoising.
- MRI VN<sup>15</sup>: Tensorflow implementation of the VN for MRI reconstruction of 2D Cartesian sampled data.

---

<sup>14</sup><https://github.com/VLOGroup/denoising-variationalnetwork>

<sup>15</sup><https://github.com/VLOGroup/mri-variationalnetwork>

## Variational Networks

There is nothing more practical  
than a good theory.

---

*Kurt Lewin*

This chapter is based on the publication:

K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, 2018, <https://github.com/VLOGroup/mri-variationalnetwork>

### Contents

---

<b>3.1</b>	<b>From Linear Reconstruction to a Variational Network . . . . .</b>	<b>32</b>
<b>3.2</b>	<b>Variational Network Parameters . . . . .</b>	<b>35</b>
<b>3.3</b>	<b>Variational Network Training . . . . .</b>	<b>36</b>
<b>3.4</b>	<b>Loss Functions and Quantitative Evaluation . . . . .</b>	<b>37</b>
<b>3.5</b>	<b>Deep Regularization for Variational Networks . . . . .</b>	<b>39</b>
<b>3.6</b>	<b>Implementation Details . . . . .</b>	<b>40</b>

---

In this chapter, we introduce the theory and a general notation of Variational Networks (VNs), holding for a variety of applications. For the sake of clarity, some basic concepts of inverse problems presented in Chapter 2 are revised. Application-specific details, e.g., how complex values are considered in Magnetic Resonance Imaging (MRI), are described in the subsequent chapters.

### 3.1 From Linear Reconstruction to a Variational Network

We consider the ill-posed linear inverse problem of finding a reconstructed image  $x \in \mathbb{K}^{N_x}$  from measurement data  $\hat{y} \in \mathbb{K}^{N_y}$  described by a linear forward model  $A : \mathbb{K}^{N_x} \rightarrow \mathbb{K}^{N_y}$ , satisfying the following system of equations

$$Ax = \hat{y}. \quad (3.1)$$

Since Equation (3.1) is ill-posed, we cannot solve for  $x$  explicitly. Therefore, a natural idea is to compute  $x$  by minimizing the least squares error

$$\min_x \frac{1}{2} \|Ax - \hat{y}\|_2^2. \quad (3.2)$$

In practice we do not have access to the true  $\hat{y}$  but only to a noisy variant  $y$  satisfying

$$\|\hat{y} - y\|_2 \leq \delta,$$

where  $\delta$  is the noise level. A closed-form solution  $x^*$  to Equation (3.2) is given by the pseudo-inverse

$$x^* = (A^*A)^{-1} A^*y.$$

However, in some applications computing the pseudo-inverse is computationally intractable, hence, iterative methods such as the Conjugate Gradient (CG) algorithm are required. Another idea is to simply perform a gradient descent on the least squares problem Equation (3.2) that leads to an iterative algorithm, which is known as the Landweber method [146]. It is given by choosing some initial  $x^0$  and performing the iterations with step sizes  $\tau^t$

$$x^t = x^{t-1} - \tau^t A^*(Ax^{t-1} - y), \quad t \geq 1, \quad (3.3)$$

where  $A^*$  is the adjoint linear sampling operator. To prevent over-fitting to the noisy data  $y$ , it is beneficial to stop the Landweber iterative algorithm early [97], i.e., after a finite number of iterations  $T$ . Instead of early stopping, we can also extend the least squares problem by an additional regularization term  $\mathcal{R}[x]$  to prevent over-fitting. The associated (variational) minimization problem is given by

$$\min_x \left\{ \mathcal{R}[x] + \frac{\lambda}{2} \|Ax - y\|_2^2 \right\}.$$

The minimizer of the regularized problem depends on the trade-off between the regularization term and the least squares data fidelity term controlled by  $\lambda > 0$ . One of the most influential regularization terms in the context of images is the Total Variation (TV) semi-norm [207] defined in Equation (2.5). The main advantage of the TV semi-norm is that it allows for

sharp discontinuities (edges) in the solution while being a convex functional enabling efficient global optimization. From a sparsity point of view, the TV semi-norm induces sparsity in the image edges and hence, favors piece-wise constant solutions. However, it is also clear that the piece-wise constant approximation is not a suitable criterion to describe the complex structure of medical images and a more general regularizer is needed. A generalization of the TV semi-norm is the Fields of Experts (FoE) model [206]

$$\mathcal{R}[x] = \sum_{i=1}^{N_k} \langle \rho_i(K_i x), \mathbf{1} \rangle. \quad (3.4)$$

Here, the regularization term is extended to  $N_k$  terms and  $\mathbf{1}$  denotes a vector of ones. The linear operator  $K : \mathbb{K}^{N_x} \rightarrow \mathbb{K}^{N_x/N_f}$  models convolutions with filter kernels  $k$  of size  $s$  and depth  $N_f$ , which is expressed as

$$Kx \Leftrightarrow x * k.$$

The filter depth  $N_f$  equals the number of features in  $x$ . The non-linear potential functions  $\rho(z) = (\rho(z_1), \dots, \rho(z_N))^\top : \mathbb{K}^{N_x/N_f} \rightarrow \mathbb{K}^{N_x/N_f}$  are composed by scalar functions  $\rho$ . In the FoE model [206], both the convolution kernels and the parametrization of the non-linear potential functions, such as student-t functions, are learned from data. Plugging the Fields of Experts model Equation (3.4) into the Landweber iterative algorithm Equation (3.3) yields

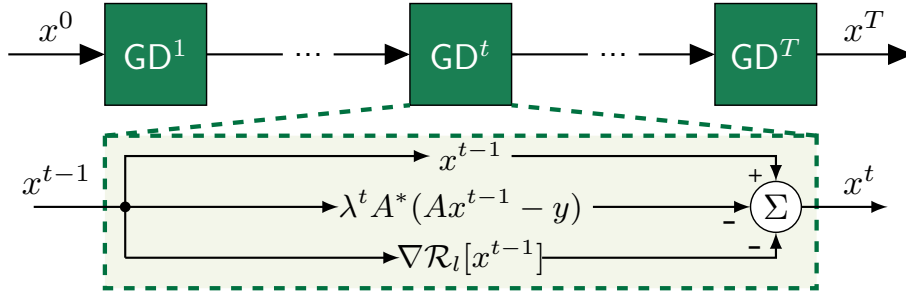
$$x^t = x^{t-1} - \tau^t \left( \sum_{i=1}^{N_k} (K_i)^\top \phi_i(K_i x^{t-1}) + \lambda A^*(Ax^{t-1} - y) \right), \quad (3.5)$$

where  $\phi_i(z) = \text{diag}(\phi_i(z_1), \dots, \phi_i(z_N))$  are the activation functions defined by the first derivative of potential functions  $\phi_i$ . Observe that the application of the transpose operation  $(K_i)^\top$  can be implemented as a convolution with filter kernels  $k_i$  rotated by  $180^\circ$ . Chen et al. [43, 46] introduced a trainable reaction-diffusion approach that performs early stopping on the gradient scheme Equation (3.5) and allows the parameters, i.e., filters, activation functions and data term weights, to vary in every gradient descent step  $t$ . All parameters of the approach are learned from data. We rewrite the trainable gradient descent scheme with time-varying parameters  $K_i^t$ ,  $\phi_i^t$  and  $\lambda^t$  as

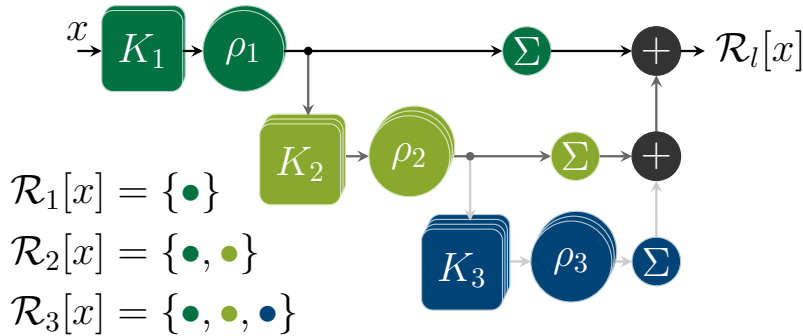
$$x^t = x^{t-1} - \sum_{i=1}^{N_k} (K_i^t)^\top \phi_i^t(K_i^t x^{t-1}) - \lambda^t A^*(Ax^{t-1} - y), \quad 1 \leq t \leq T. \quad (3.6)$$

Additionally, we omit the step size  $\tau^t$  in Equation (3.5) because it is implicitly contained in the activation functions and data term weights. By unfolding the single iterations of Equation (3.6), we obtain the Variational Network structure depicted in Figure 3.1. Essentially, one iteration of an iterative reconstruction can be related to one step in the network. As different applications can be modeled via the forward operator  $A$  and its adjoint operator  $A^*$ , we can directly use

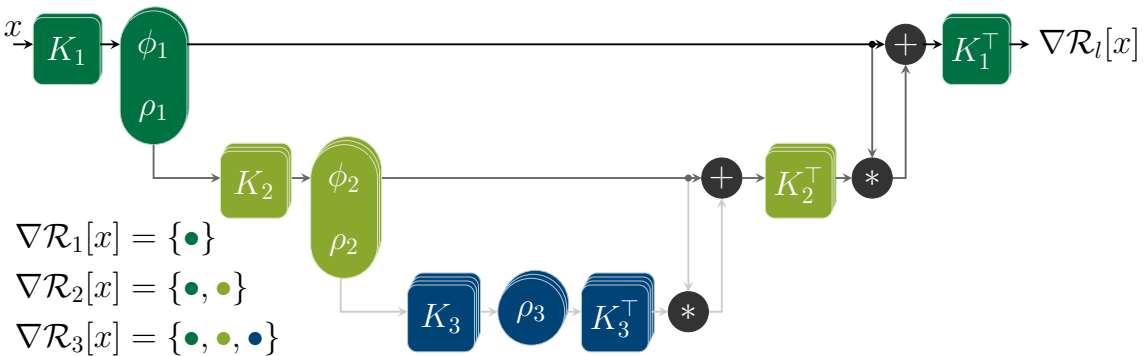
the raw measurement data as input. In the case of image enhancement, the operator  $A$  simplifies to the identity matrix. For the regularization part, the FoE model is commonly used. Theoretically, any regularizer that can be formulated as energy can be used. An example for deep regularization is presented in Section 3.5. Different regularizers are depicted in Figure 3.2, along with their gradients in Figure 3.3.



**Figure 3.1:** Basic VN structure. The VN is defined as an unrolled Gradient Descent (GD) scheme with  $T$  steps. Each of these steps can be seen as a residual block [140]. The gradient of these GD steps correspond to the gradient of a variational model which consists in our case of a square data term and a regularization term. Examples for different regularizers are depicted in Figure 3.2.

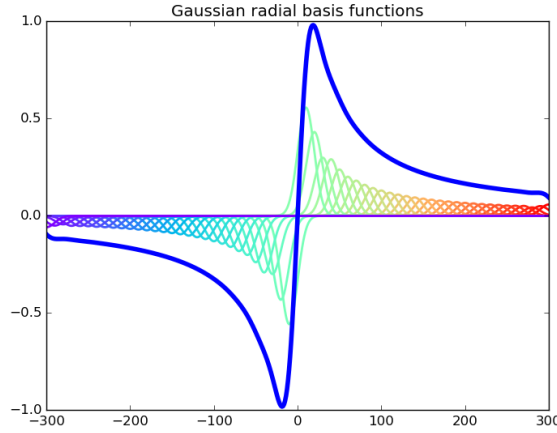


**Figure 3.2:** Examples for different regularizers. The corresponding gradients are depicted in Figure 3.3.



**Figure 3.3:** Gradient visualization of the regularizers depicted in Figure 3.2.





**Figure 3.4:** RBFs provide a way to smoothly approximate any function. The single Gaussian bases with fixed standard deviation appear at equidistant nodes within a pre-defined interval.

## 3.2 Variational Network Parameters

The VN defined by Equation (3.6) contains a number of parameters: Filter kernels  $k_i^t$ , activation functions  $\phi_i^t$ , and data term weights  $\lambda^t$ . First, we constrain the filters to be zero-mean, i.e.,  $\sum_j k_{i,j}^t = 0 \forall i$ . Additionally, the whole kernel is constrained to lie in the unit-sphere, i.e.,  $\|k_i^t\|_2 \leq 1 \forall i$ , to avoid a scaling problem of the activation functions. To learn the activation functions, we require a suitable function parametrization between filters and corresponding activation functions. A standard choice to smoothly approximate any function are Gaussian radial basis functions (RBFs) as illustrated in Figure 3.4. We define the scalar activation functions  $\phi_i^t$  as a weighted combination of  $N_w$  RBFs with equidistant nodes  $\mu$  and standard deviation  $\sigma = \frac{2I_{\max}}{N_w - 1}$ ,

$$\phi_i^t(z) = \sum_{j=1}^{N_w} w_{ij}^t \exp\left(-\frac{(z - \mu_j)^2}{2\sigma^2}\right).$$

The nodes are distributed in an equidistant way in  $[-I_{\max}, I_{\max}]$  which allows us to achieve the same resolution over the whole defined range. Note here that  $\mu, \sigma$  depend on the maximum estimated filter response  $I_{\max}$ . The final parameters that we consider are the data term weights  $\lambda^t$  which are constrained to be non-negative, i.e.,  $\lambda^t > 0$ . During training, all constraints on the parameters are realized based on projected gradient methods.

However, although the  $\|k_i^t\|_2 \leq 1$  constraint slightly restricts the scaling ambiguity of the activation functions, it is still not optimal for training. Furthermore, it is challenging to estimate the maximum filter responses as this could be different for the individual kernels and, hence, the maximal resolution of the activation functions cannot be exploited. This suggests

to scale the input  $z$  of the activation function

$$z := \frac{z - \mu_z}{\sqrt{\sigma_z^2 + \epsilon}} \alpha_z$$

where  $\mu_z, \sigma_z^2$  are the mean and variance values of the input  $z$  and  $\epsilon = 10^{-6}$  is a small constant to avoid the division by zero. The parameter  $\alpha_z$  is responsible to re-scale the input range and can be fixed or learned from data. This transformation is known as batch normalization in deep learning literature [113].

### 3.3 Variational Network Training

During the offline training procedure, the goal is to find an optimal parameter set  $\theta = (\theta^0, \dots, \theta^{T-1})$ ,  $\theta^t = (w_{ij}^t, k_i^t, \lambda^t)$  for our proposed VN in Equation (3.6). After training, the parameters  $\theta$  are fixed and we can reconstruct previously unseen data efficiently by forward-propagating the new data through the VN. To set up the training procedure, we minimize a loss function over a set of images  $S$  with respect to the parameters  $\theta$ . The loss function defines the similarity between the reconstructed image  $x^T$  and a clean, artifact-free reference image  $x_{\text{ref}}$ . A common choice for the loss function is the Mean Squared Error (MSE), which is defined along with other similarity measures in Section 3.4. To solve this highly non-convex training problem, we deploy different first-order optimizers such as the Inertial Incremental Proximal Gradient (IIPG) optimizer, which is related to the Inertial Proximal Alternating Linearized Minimization (IPALM) algorithm [188], or the proximal ADAM optimizer with block-preconditioning on the parameters based on the original ADAM optimizer [129]. For algorithmic details on these first-order optimizers we refer to Appendix C and additionally for IIPG to [140]. First-order optimizers require both the loss function value and the gradient with respect to the parameters  $\theta$ . This gradient can be computed by simple back-propagation [147], i.e., applying the chain rule

$$\frac{\partial \mathcal{L}(x^T(\theta), x_{\text{ref}})}{\partial \theta^t} = \frac{\partial x^t}{\partial \theta^t} \cdot \frac{\partial x^{t+1}}{\partial x^t} \cdots \frac{\partial x^T}{\partial x^{T-1}} \cdot \frac{\partial \mathcal{L}(x^T(\theta), x_{\text{ref}})}{\partial x^T}.$$

The derivation of the gradients for the parameters is provided in Appendix D.

A question that arises often is, why we did not deploy the standard ADAM optimizer [129] for our problems, as it is the most common optimizer for stochastic optimization of deep learning problems. We observed during our experiments that either convolution kernels or activation functions were favored, leading to either noisy kernels or activation functions that did not move very far from their initialization. We believe that the different parameter sets such as kernels and influence functions are defined in different domains, which makes it challenging for standard optimization schemes if they are all treated equally. These observation motivated us to continuously work on different optimizers. The IIPG for example has proven convergence rates for deterministic problems, however, the convergence proofs for stochastic, non-convex

problems are still a subject to future work. The IPALM optimizer [188] with line-search allows us to have self-adapted step sizes for the defined parameter blocks. However, as each parameter block is updated in an alternating manner, optimization becomes exhausting if the number of parameter blocks increases. Furthermore, we observed issues in optimization when the batch size becomes small, which resulted in large oscillations of the loss. These observation inspired us to include a preconditioning based on the gradient norm of the parameter blocks in the ADAM optimizer together with a proximal mapping for our specific VN structure. Details about this optimizer can be found in the Appendix C.2.

### 3.4 Loss Functions and Quantitative Evaluation

The success of supervised machine learning approaches depends on many aspects, which are the training data, network architecture, optimizer, and a loss function. The loss function measures the similarity between the reconstructed image and a reference image during training. Consequently, the loss function has a huge impact on the final image quality of the reconstructions. For simplicity, we drop the dependence on the samples in the definitions of the loss functions.

A common choice in deep learning approaches is the MSE

$$\mathcal{L}_{\text{mse}}(x^T(\theta), x_{\text{ref}}) = \frac{1}{2N} \|x^T(\theta) - x_{\text{ref}}\|_2^2,$$

where  $N$  denotes the number of pixels. Although the MSE is easy to optimize, it is known to be not robust against outliers and to result in blurred reconstructions [274] due to its averaging behavior. There are also a number of evaluation measures that are commonly used for quantitative evaluation, which are related to the MSE. The Normalized Root Mean Squared Error (NRMSE) allows for comparison between data that differ in terms of scaling

$$\mathcal{L}_{\text{nrms}}(x^T(\theta), x_{\text{ref}}) = \frac{\|x^T(\theta) - x_{\text{ref}}\|_2}{\|x_{\text{ref}}\|_2}.$$

The Peak Signal-To-Noise Ratio (PSNR) involves the maximum possible intensity value  $I_{\text{max}}$  of the image and is a common measure used for quantitative evaluation

$$\mathcal{L}_{\text{psnr}}(x^T(\theta), x_{\text{ref}}) = 20 \log_{10} \frac{I_{\text{max}} \sqrt{N}}{\|x^T(\theta) - x_{\text{ref}}\|_2}.$$

Obviously, all these measures compare images pixel-by-pixel, however, it is known that these measures represent the human perceptual system poorly [248, 274].

Instead of comparing pixel-wise intensity values, the perceptual-motivated Structural Similarity Index (SSIM) [248] considers local patch statistics. The variant of SSIM which is commonly used in literature consists of a luminance term  $l$  and contrast term  $c$  defined as

follows

$$\begin{aligned} \text{SSIM}(x, x_{\text{ref}}) &= \frac{1}{N} \sum_{i \in \Omega} l(x(i), x_{\text{ref}}(i))^\alpha c(x(i), x_{\text{ref}}(i))^\beta \\ &= \frac{1}{N} \sum_{i \in \Omega} \left( \frac{2\mu_{x(i)}\mu_{x_{\text{ref}}(i)} + c_1}{\mu_{x(i)}^2 + \mu_{x_{\text{ref}}(i)}^2 + c_1} \right)^\alpha \left( \frac{2\sigma_{x(i)}\sigma_{x_{\text{ref}}(i)} + c_2}{\sigma_{x(i)}^2 + \sigma_{x_{\text{ref}}(i)}^2 + c_2} \right)^\beta. \end{aligned}$$

The means  $\mu_{\cdot(i)}$  and standard deviations  $\sigma_{\cdot(i)}$  for each patch around the pixel position  $i$  defined in the domain  $\Omega$  can be modeled by convolving the whole image with a Gaussian filter with standard deviation  $\sigma_G$  or an average filter of a predefined window size, according to the implementation. The parameters  $c_1, c_2$  are user-defined constants and the exponents  $\alpha$  and  $\beta$  are commonly set to 1. Common settings for the SSIM are a Gaussian filter of size 11 along with standard deviation  $\sigma_G$  and  $c_1 = (0.01L)^2, c_2 = (0.03L)^2$  where  $L$  denotes the dynamic range of the underlying images. As the SSIM is defined in the interval  $[-1, 1]$  the final loss function can be written as

$$\mathcal{L}_{\text{ssim}}(x^T(\theta), x_{\text{ref}}) = 1 - \text{SSIM}(x^T(\theta), x_{\text{ref}}).$$

More details on the implementation of the SSIM can be found in Appendix D.2.

As the formulation shows, the outcome of the measure depends on the chosen Gaussian kernel with standard deviation  $\sigma_G$ . Wang et al. [249] proposed a Multi-Scale Structural Similarity Index (MS-SSIM), which evaluates and combines the SSIM values across different resolutions. However, this involves computing a pyramid at different levels which is computationally demanding. As we aim at using the MS-SSIM as loss function, we follow Zhao et al. [274] who evaluate the SSIM for various standard deviations  $\sigma_{G,m}$ , imitating different scales  $m$ , leading to following definition

$$\text{MS-SSIM}(x, x_{\text{ref}}) = \frac{1}{N} \sum_{i \in \Omega} l_M^\alpha(x(i), x_{\text{ref}}(i)) \prod_{m=1}^M c_m^{\beta_m}(x(i), x_{\text{ref}}(i)).$$

For our purpose, we use four different standard deviations  $\sigma_{G,m} \in \{0.5, 1, 2, 4\}$ . Again, the exponents are dropped not only for simplicity but also for computational reasons, as this results in undefined behavior during optimization because of the accelerated implementation of exponents in the used Tensorflow framework. For quantitative evaluation, we use a factor of  $1/M$  for the exponents. Similar to SSIM the full loss function for MS-SSIM can be written as

$$\mathcal{L}_{\text{ms-ssim}}(x^T(\theta), x_{\text{ref}}) = 1 - \text{MS-SSIM}(x^T(\theta), x_{\text{ref}}).$$

Although there exist more complex quantitative error measures such as Feature Similarity Index (FSIM) [156] or HaarPsi [202], not every quantitative similarity measure can be used as a loss function, because differentiability is required for network training, which might not be

the case for more complex quantitative measures.

### 3.5 Deep Regularization for Variational Networks

Previous work investigated trainable FoE regularization that focuses on low-level features. In deep learning, impressive results are achieved by exploiting mid- and high-level features in a deep network, consisting of cascaded convolution and activation layers. However, this commonly requires a huge number of network parameters which makes it difficult to train and a number of strategies have to be followed to ease the training of these architectures. Ronneberger et al. [205] proposed a U-net architecture that showed great success not only for segmentation tasks but also for image-to-image problems. The U-net can be seen a symmetric Convolutional Neural Network (CNN) with a contracting and expanding path, where both information from the downsampling path, modelled by skip connections, and information from the upsampling path are combined. However, as these architectures tend to be huge, an enormous amount of (augmented) training data are needed to train these architectures. To incorporate mid- and high-level features in our VN, we deepen the structure of the Fields-of-Experts regularizer by a series of convolution operators and potential functions, while still fulfilling the requirement of an energy functional. Furthermore, we still aim at keeping *structure* in our models to keep the number of network parameters low instead of stacking more and more layers.

Inspired by the work of Ronneberger et al. [205], we add the responses of different regularization terms consisting of a different number of layers. Let us define following abbreviations to account for the different dimensions of the layers

$$\begin{aligned} L_1 &= \frac{N_x N_{f_1}}{N_{f_{in}}}, \\ L_2 &= \frac{N_x N_{f_2}}{N_{f_{in}} N_{ds}}, \\ L_3 &= \frac{N_x N_{f_3}}{N_{f_{in}} N_{ds}^2}. \end{aligned}$$

The number of input features is denoted by  $N_{f_{in}}$ , the number of features for the individual layers  $l$  is denoted by  $N_{f_l}$ . It is also possible that  $K_l$  involves a downsampling operator where  $N_{ds}$  denotes the downsampling factor. We define the regularizers as

$$\begin{aligned} \mathcal{R}_1[x] &= \langle \rho_1(K_1 x), \mathbf{1} \rangle, & K_1 : \mathbb{K}^{N_x} &\rightarrow \mathbb{K}^{L_1}, \rho_1 : \mathbb{K}^{L_1} \rightarrow \mathbb{K}^{L_1}, \\ \mathcal{R}_2[x] &= \mathcal{R}_1[x] + \langle \rho_2(K_2 \rho_1(K_1 x)), \mathbf{1} \rangle, & K_2 : \mathbb{K}^{L_1} &\rightarrow \mathbb{K}^{L_2}, \rho_2 : \mathbb{K}^{L_2} \rightarrow \mathbb{K}^{L_2}, \\ \mathcal{R}_3[x] &= \mathcal{R}_2[x] + \langle \rho_3(K_3 \rho_2(K_2 \rho_1(K_1 x))), \mathbf{1} \rangle, & K_3 : \mathbb{K}^{L_2} &\rightarrow \mathbb{K}^{L_3}, \rho_3 : \mathbb{K}^{L_3} \rightarrow \mathbb{K}^{L_3}. \end{aligned}$$

Note here that the convolution operators  $K_l$  and potential functions  $\rho_l$  operate on a stack of features  $N_{f_l}$ . In every layer, we exploit the feature responses of previous layers. The responses

of all layers are finally added. Deriving the gradients yields

$$\begin{aligned}\nabla\mathcal{R}_1[x] &= K_1^\top\phi_1(K_1x), \\ \nabla\mathcal{R}_2[x] &= K_1^\top\phi_1(K_1x)\left[1+K_2^\top\phi_2(K_2\rho_1(K_1x))\right], \\ \nabla\mathcal{R}_3[x] &= K_1^\top\phi_1(K_1x)\left[1+K_2^\top\phi_2(K_2\rho_1(K_1x))\left[1+K_3^\top\phi_3(K_3\rho_2(K_2\rho_1(K_1x)))\right]\right].\end{aligned}$$

A graphical visualization of these regularizers up to  $l = 3$  is depicted in Figure 3.2. If we inspect the gradient of the proposed deep regularizer, visualized in Figure 3.3, we observe similarities to the U-net architecture [205]. Most obvious are the contracting and expanding paths as well as residual connections at the different layers. Note here that as it is possible to include a downsampling operator in  $K_l$ , resulting in an adjoint upsampling operator after  $K_l^\top$  in the gradient derivation.

### 3.6 Implementation Details

Nowadays, different frameworks that can be used for deep learning are available and most of them can be used with Python. Examples are Tensorflow<sup>1</sup> developed by Google, PyTorch<sup>2</sup> developed by Facebook, Caffe<sup>3</sup> developed by Berkeley Artificial Intelligence (AI) research and Theano<sup>4</sup> developed at Montreal Institute for Learning Algorithms. All of these frameworks have in common that they were basically implemented in C++, have Graphics Processing Unit (GPU) support based on CUDA and CUDNN and offer a way for automatic differentiation. Our initial variant for VNs was implemented from scratch using Matlab. To have more flexibility, we first switched to Theano, however, when it came to the application of medical data, we constantly ran out of memory which made it impossible to train larger networks, and the support of complex numbers was very limited at that time. Hence, we implemented our own version<sup>5</sup> from scratch using C++/CUDA with CUDNN support, which could be used for various applications in a GPU memory efficient way. However, the limited flexibility would have required tedious software engineering. As Tensorflow was also published at the same time, which offered more flexibility and also more support for complex numbers, we finally moved our VN framework to tensorflow. This explains why the different applications in the preceding chapters have different setups.

<sup>1</sup><http://www.tensorflow.org>

<sup>2</sup><https://pytorch.org/>

<sup>3</sup><http://caffe.berkeleyvision.org/>

<sup>4</sup><http://deeplearning.net/software/theano/>

<sup>5</sup>Special credits go to my colleagues Erich Kobler and Teresa Klatzer.

## Variational Networks for Image Enhancement in Limited-Angle Computed Tomography (CT)

Look at situations from all angles,  
and you will become more open.

---

*Dalai Lama*

This chapter is based on the publication:

K. Hammernik, T. Würfl, T. Pock, and A. Maier. A deep learning architecture for limited-angle computed tomography reconstruction. In *Bildverarbeitung für die Medizin 2017*, pages 92–97, 2017

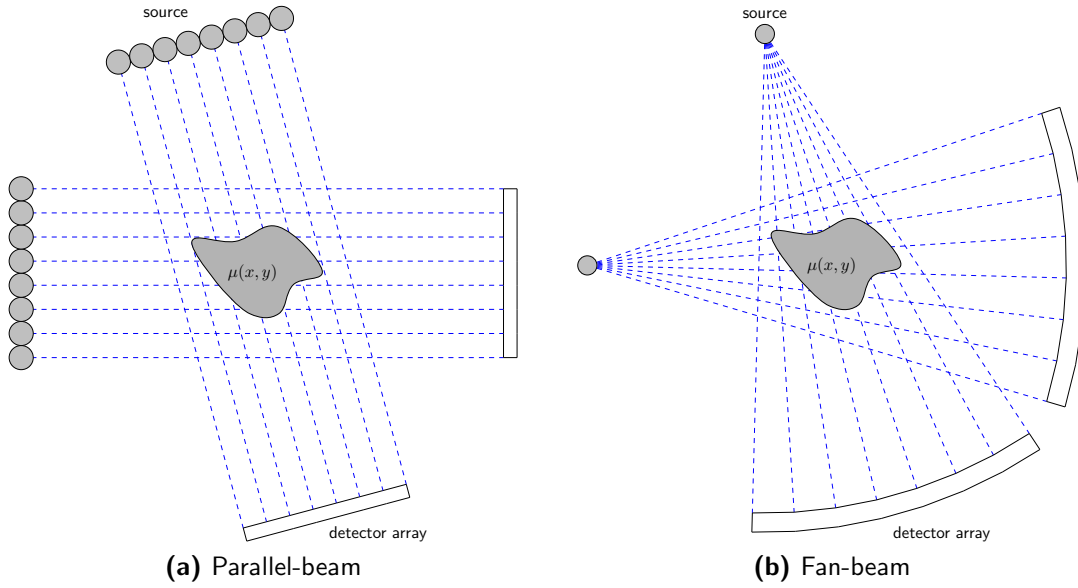
### Contents

---

4.1	Fundamental Principles of CT . . . . .	42
4.2	Artifact Correction in Limited-Angle CT . . . . .	44
4.3	Methods . . . . .	47
4.4	Results . . . . .	49
4.5	Discussion . . . . .	49

---

The general formulation of Variational Networks (VNs) offers a great application potential for both medical image reconstruction and image enhancement. In this first application, VNs are explored for artifact correction in limited-angle CT. First, this chapter outlines the fundamental principles of CT from a signal processing perspective in Section 4.1. Section 4.2 introduces the challenges of limited-angle CT data. Reconstruction of limited-angle CT data is seen as a two-step process, as presented in Section 4.3 and the preceding sections, where in the first step reconstruction is carried out using a Convolutional Neural Network (CNN) approach proposed by Würfl et al. [253], while the main focus is put on the correction of artifacts in image domain.



**Figure 4.1:** Examples for different beam geometries in CT.

## 4.1 Fundamental Principles of CT

Along with Magnetic Resonance Imaging (MRI), CT is an important imaging modality for clinical examinations. The main advantages to MRI are lower costs and faster image acquisition, however, CT scans involve ionizing radiation which poses a potential risk for cancer. In the following, fundamentals of CT image formation and reconstruction are described. For more details on the physics and image formation process of CT, we refer the interested reader to [62, 270].

The term tomography is derived from the Greek words *tomos* (τόμος, slice) and *graphein* (γράφειν, to write), hence, tomography means to image a cross-section. For this purpose, X-rays of an initial intensity  $I_0(s, \beta)$  at a detector position  $(s, \beta)$  are directed at an object at multiple orientations. As the X-rays travel through the object, they are attenuated due to the interior structures of the object and arrive with reduced intensity  $I_d(s, \beta)$  at the detector position  $(s, \beta)$ . Figure 4.1 illustrates source-detector systems for acquisition such as parallel-beam and fan-beam geometries. In the following, we consider only the parallel-beam geometry to understand the fundamental principles of CT reconstruction.

The relationship between the source intensity  $I_0$  and the detected intensity  $I_d$  is described by the Lambert-Beer's law

$$I_d(s, \beta) = I_0(s, \beta) \exp\left(-\int_C \mu(x, y) dl\right) \Rightarrow p(s, \beta) = \ln\left(\frac{I_0(s, \beta)}{I_d(s, \beta)}\right) = \int_{C(s, \beta)} \mu(x, y) dl,$$

i.e., we measure the line-integral over a line  $C$  of the attenuation coefficients  $\mu(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}$



with compact support. The line  $C(t) = (x(t), y(t))$  is parametrized by

$$x(t) = s \cos \beta - t \sin \beta, \quad y(t) = s \sin \beta + t \cos \beta$$

where  $s$  defines the detector position,  $t$  the ray direction, and  $\beta$  the view angle. Using this parametrization, the line integral, also termed projection, can be expressed as

$$p(s, \beta) = \int_{-\infty}^{\infty} \mu(s \cos \beta - t \sin \beta, s \sin \beta + t \cos \beta) dt.$$

This principle is illustrated by the projection of a point source in Figure 4.2, where we see the relationship between  $x$ - $y$  and  $s$ - $t$  plane as well as the angle-dependency on  $\beta$ . The sinogram depicts a set of projection data of different angles  $\beta$ , where we view the projections in the  $s$ - $\beta$  plane. Once the projection data are acquired, the main question is how to recover the attenuation coefficients  $\mu(x, y)$  within the object from the measured beam intensity. To understand the back-projection of the data, we first consider the projection-slice theorem. The basis for the projection-slice theorem is the Fourier Transform (FT)

$$F(u, v) = (\mathcal{F}_{2D}f)(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-j2\pi(ux+vy)} dx dy$$

and inverse FT

$$f(x, y) = (\mathcal{F}_{2D}^*F)(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v) e^{j2\pi(ux+vy)} du dv.$$

The symbols  $\mathcal{F}$  and  $\mathcal{F}^*$  denote the forward and inverse Fourier operator. Consider now the 1D FT of the projection

$$P(w, \beta) = (\mathcal{F}_{1D}p)(w, \beta) = \int_{-\infty}^{\infty} p(s, \beta) e^{-j2\pi ws} ds.$$

This can be rewritten as

$$P(w, \beta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x, y) e^{-j2\pi w(x \cos \beta + y \sin \beta)} dx dy,$$

where we see the relation of the 1D-FT of a projection to the 2D-FT of a slice

$$P(w, \beta) = F(w \cos \beta, w \sin \beta), \quad u = w \cos \beta, \quad v = w \sin \beta.$$

Taking the inverse FT to recover  $\mu(x, y)$  yields

$$\mu(x, y) = \int_0^{2\pi} \int_0^{\infty} F(w \cos \beta, w \sin \beta) e^{j2\pi w(x \cos \beta + y \sin \beta)} w \, dw \, d\beta, \quad (4.1)$$

where the factor  $w$  is the determinant of the Jacobian that arises due to the change from the Cartesian coordinate system  $(u, v)$  into the polar coordinate system  $(w, \beta)$ . As  $p(s, \beta) = p(-s, \beta + \pi)$ , Equation (4.1) can be rewritten as

$$\mu(x, y) = \int_0^{\pi} \int_{-\infty}^{\infty} |w| P(w, \beta) e^{j2\pi w(x \cos \beta + y \sin \beta)} \, dw \, d\beta,$$

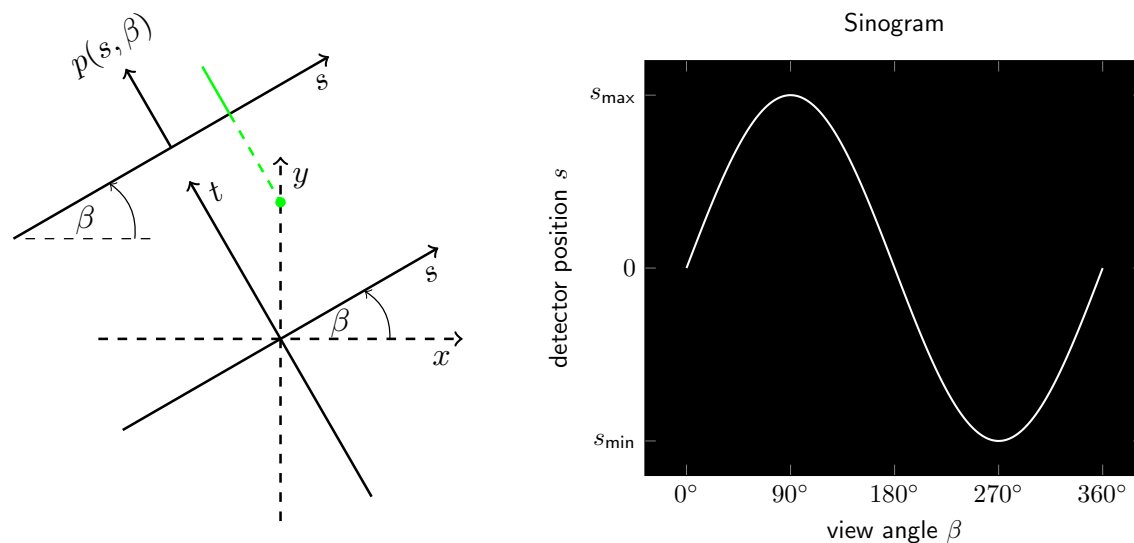
with changed integration limits. This can be interpreted as filtering the FT of the projection data with a ramp-filter  $|w|$  before applying the inverse FT to reconstruct  $\mu(x, y)$ , known as Filtered Back-Projection (FBP).

Another more complex example to show the CT reconstruction principle is illustrated in Figure 4.3. It becomes obvious that a certain number of projection data are required to fully recover an image. The influence on the FBP reconstruction on the number of projections is depicted in Figure 4.4, along with the error to the reference image and the sinogram. If the full angular range is considered with an angle spacing of  $0.5^\circ$ , the image can be exactly reconstructed. Taking only a few projections with equidistant spacing of  $5^\circ$  results in noise-like streaking artifacts in the FBP reconstruction, which means that a much lower radiation dose is required to obtain the projection data. Recently, VNs have shown promising results in dose-reduction by interrupted-beam acquisition and tube-current reduction, which can be formulated as CT-reconstruction and CT-denoising problem, respectively [141]. In this chapter, we focus on artifact-correction of limited-angle CT data in image domain using VNs.

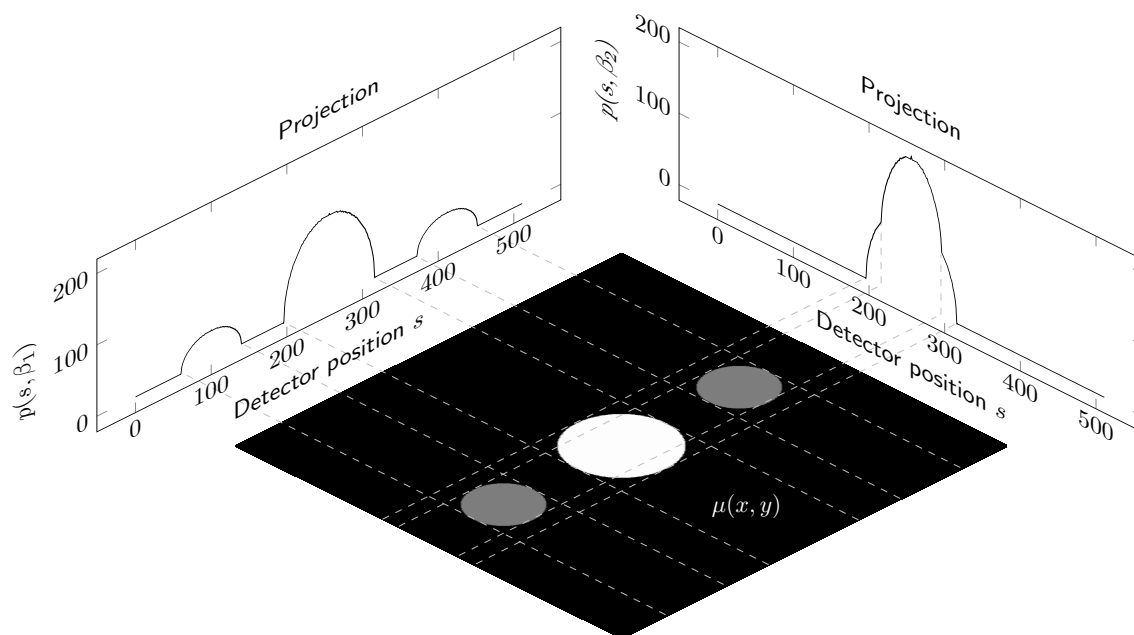
## 4.2 Artifact Correction in Limited-Angle CT

Limited-angle CT is important in some applications where the projection data cannot be acquired over the full angular range. Examples for such configurations are robot assisted scanners in medicine or scanning of very large objects in industrial CT. As limited-angle CT does not acquire data over the full angular range, the projection data are incomplete which results in intensity inhomogeneities as well as streaking artifacts in the image domain, illustrated in the third row of Figure 4.4. Further sources for streaking artifacts are the non-linear attenuation of polychromatic X-rays or inelastic scattering of photons. All these artifacts are corrected with specialized heuristic compensation procedures that tune each step independently.

Many specialized iterative algorithms exist which clearly improve the image quality [109, 110]. A disadvantage of iterative techniques is their high runtime requirement. In contrast, analytical algorithms are less computationally demanding, but typically suffer from intensity

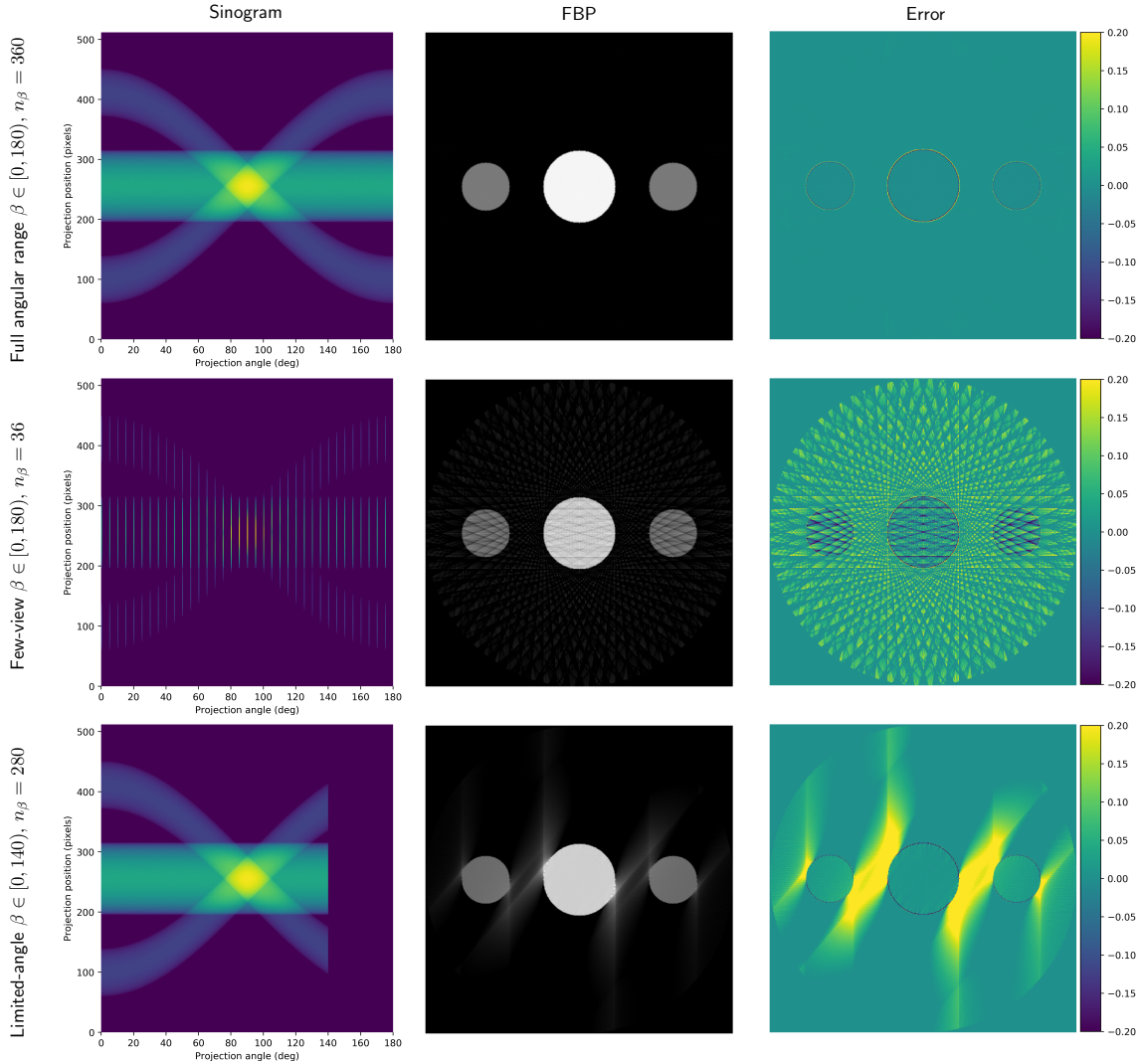


**Figure 4.2:** CT projection principle for a point source. A projection  $p(s, \beta)$  is obtained at different view-angles  $\beta$ . The sinogram represents the projection data in the  $s$ - $\beta$  plane.



**Figure 4.3:** CT projection principle for a circle phantom. A certain number of projections are required at different angles  $\beta$  to identify the exact locations of the circles and, hence, the attenuation coefficients  $\mu(x, y)$  of the image.

inhomogeneities and streaking artifacts in the image domain due to missing projections. To correct for intensity inhomogeneities, Riess et al. [203] use a heuristic scheme based on compensation weights. Würfl et al. [253] reformulate filtered back-projection as a neural network and learn compensation weights for limited-angle CT reconstruction. However, their approach



**Figure 4.4:** Examples for FBP reconstructions for view angles that cover the full angular range with angle spacing  $\Delta\beta = 0.5^\circ$  (first row), few-view reconstruction with  $\Delta\beta = 5^\circ$  (second row) and limited-angle reconstruction with  $\Delta\beta = 0.5^\circ$  (third row). Depending on which projection data are missing, this results in incoherent artifacts for few-view FBP reconstruction or strong coherent streaking artifacts and intensity inhomogeneities for limited-angle FBP reconstruction.

cannot account for the remaining streaking artifacts due to the missing non-linear filtering step.

To correct for remaining streaking artifacts, a number of non-linear filtering methods can be applied. For limited-angle CT, Riess et al. [203] apply a bilateral filter [231] after the compensation of missing projection data. Although there exists a number of generic non-linear filtering methods such as BM3D [50], first-order and higher-order Total Generalized Variation (TGV) [21, 207], they can mainly correct for unstructured Gaussian noise. These models cannot describe the complex image content as they make assumptions on the image

statistics such as piece-wise constancy in the case of the Total Variation (TV) semi-norm. This motivated us to use deep learning approaches that are able to account for coherent noise artifacts to overcome the limitation of generic non-linear filtering methods. In this chapter, we propose a two-step deep learning architecture for limited-angle CT reconstruction. In a first step, compensation weights in the projection domain are learned based on [253] to correct for intensity changes. Inspired by variational image restoration, we formulate a non-linear filtering problem as a VN to eliminate coherent streaking artifacts in the image domain.

### 4.3 Methods

We propose a two-step deep learning architecture for artifact compensation in limited-angle CT. The basic network architecture is illustrated in Figure 4.5. In a first step, we learn the compensation weights to account for missing projection data due to the limited-angle CT acquisition. In a second step, we eliminate streaking artifacts using a VN architecture.

**Step 1: A neural network to learn compensation weights** To correct for intensity inhomogeneities in limited-angle CT, we adapt the network architecture of Würfl et al. [253] to our problem. The input of the neural network is a sinogram with missing angular data, denoted by  $y_{\text{sin}}$ . The network represents the fan-beam reconstruction as

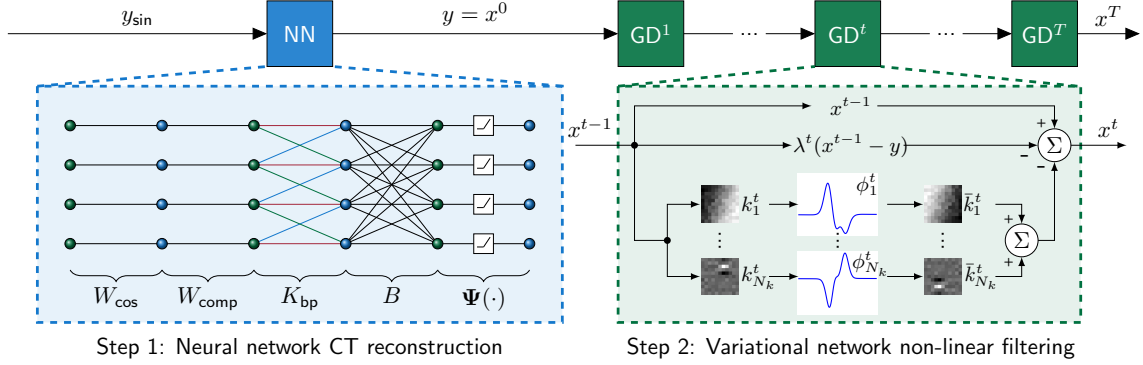
$$y = \Psi (BK_{\text{bp}}W_{\text{comp}}W_{\text{cos}}y_{\text{sin}}) ,$$

where  $B$  denotes the backprojection operator,  $K_{\text{bp}}$  implements filtering with a one-dimensional convolution kernel and the weight operators  $W_{\text{cos}}$  and  $W_{\text{comp}}$  implement element-wise multiplications with cosine weights and compensation weights, respectively. The non-negativity constraint is realized via the operator  $\Psi$ . While the operators  $B$  and  $W_{\text{cos}}$  are fixed, the compensation weights in  $W_{\text{comp}}$  and the one-dimensional reconstruction filter in  $K_{\text{bp}}$  are learned. During a training procedure, we compare the neural network reconstruction  $y$  to a reference  $y_{\text{gt}}$  obtained from full angular data using the Mean Squared Error (MSE)

$$\mathcal{L}_{\text{mse}} (W_{\text{comp}}, K_{\text{bp}}) = \frac{1}{2S} \sum_{i=1}^S \|y - y_{\text{gt}}\|_2^2, \quad (4.2)$$

where  $S$  is the number of training samples. For more details, we refer the interested reader to [253]. After training, the network can be applied to a new sinogram and yields the intensity-corrected reconstruction  $y$ . As this network only corrects artifacts in the projection domain, we now introduce a VN for suppressing the remaining streaking artifacts in the image domain.

**Step 2: A Variational Network to remove streaking artifacts** To remove streaking artifacts in the neural network reconstruction  $y \in \mathbb{R}^{N_x}$ , we learn a non-linear filtering method



**Figure 4.5:** Deep learning architecture for limited-angle CT reconstruction. The first neural network (blue) models FBP and corrects the intensity inhomogeneities in the image domain by learning compensation weights  $W_{\text{comp}}$  in the projection domain. The second VN (green) formulates non-linear filtering as  $T$  unrolled Gradient Descent (GD) steps. In each step  $t$ , the filters  $k_i^t$ , derivative of potential functions  $\phi_i^t$  and the regularization parameter  $\lambda_t$  are learned to remove the remaining streaking artifacts.

using a VN. We seek an optimal image with eliminated streaking artifacts  $x \in \mathbb{R}^{N_x}$ . The variational image restoration problem is given as

$$E(x) = \frac{\lambda}{2} \|x - y\|_2^2 + \sum_{i=1}^{N_k} \langle \rho_i(K_i x), \mathbf{1} \rangle ,$$

where the first term is a data fidelity term that measures the similarity to the intensity-corrected network input  $y$  and the second term is the regularization term that imposes prior knowledge on the image  $x$ . The impact of both terms is regulated by a parameter  $\lambda$ . The regularization term is based on the Fields of Experts (FoE) model where we apply  $N_k$  convolution operators  $K_i : \mathbb{R}^{N_x} \rightarrow \mathbb{R}^{N_x}$ , followed by non-linear functions  $\rho_i : \mathbb{R}^{N_x} \rightarrow \mathbb{R}^{N_x}$  to  $x$  and  $\mathbf{1} \in \mathbb{R}^{N_x}$  is a vector of ones. Plugging the gradient of the variational model into Landweber iterative algorithm Equation (3.3) yields the VN for image restoration

$$x^t = x^{t-1} - \sum_{i=1}^{N_k} (K_i^t)^\top \phi_i^t(K_i^t x^{t-1}) - \lambda^t(x^{t-1} - y), \quad 1 \leq t \leq T. \quad (4.3)$$

In the gradient calculation, we additionally introduce the derivative of potential functions  $\phi_i^t : \mathbb{R}^{N_x} \rightarrow \mathbb{R}^{N_x}$  and transpose convolution operators  $(K_i^t)^\top$ . In a training procedure similar to the neural network reconstruction in Section 4.3, we obtain the convolution kernels  $K_i^t$ , non-linear derivatives of potential functions  $\phi_i^t$  and the regularization parameter  $\lambda^t$  for each of the  $T$  gradient steps by minimizing the MSE as stated in Equation (4.2). The VNs are trained using the Limited-Memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm with stage-wise pre-training for 100 iterations and joint training for 700 iterations according to [46].

Method	PSNR	SSIM
Neural network	34.66±2.07	0.91±0.01
Bilateral filtering ( $\sigma_s = 0.5, \sigma_c = 0.1$ )	29.93±3.61	0.91±0.02
BM3D ( $\sigma = 1.5$ )	34.75±2.09	0.91±0.02
TV ( $\lambda = 300$ )	34.82±2.10	0.91±0.01
Second-order TGV ( $\lambda = 2, \alpha_0 = 0.01, \alpha_1 = 0.02$ )	34.80±2.09	0.91±0.01
VN (kernel size 5)	36.13±2.13	0.93±0.01
VN (kernel size 7)	36.86±2.01	0.94±0.01
VN (kernel size 9)	38.14±2.27	<b>0.95±0.01</b>
VN (kernel size 11)	37.87±1.98	<b>0.95±0.01</b>
VN (kernel size 13)	<b>38.23±2.06</b>	<b>0.95±0.01</b>

**Table 4.1:** Quantitative comparison of non-linear filtering methods along with the used parameter settings. The comparison is performed in terms of PSNR and SSIM (mean  $\pm$  standard deviation) in the Field of View (FoV). The neural network result is the intensity-inhomogeneity corrected output of a first correction step and defines the input to all methods.

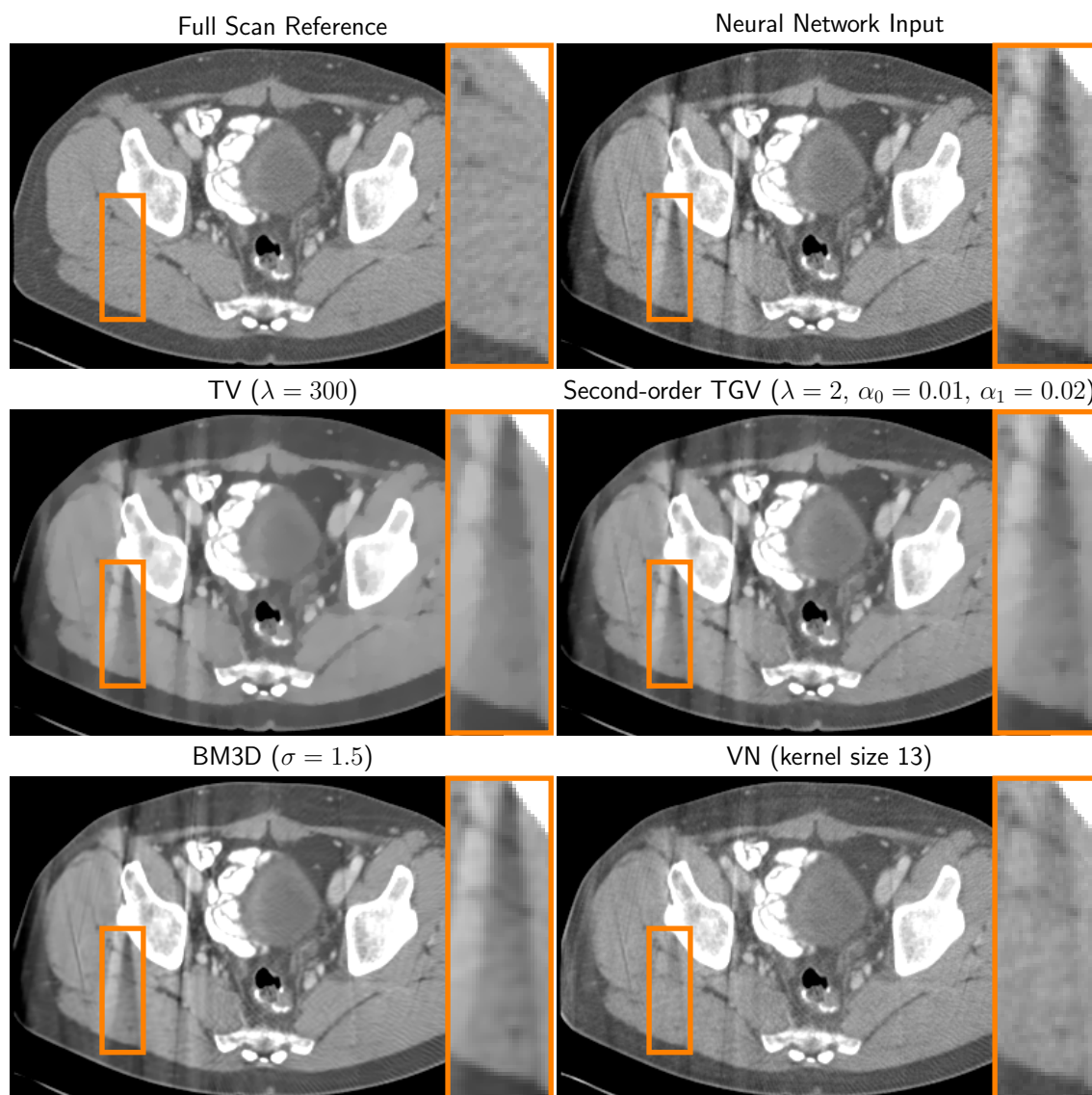
**Experimental Setup** To obtain training data, we simulated 450 fan-beam projections of size  $512 \times 512$  from volumetric datasets of ten different patients. For evaluation, we performed a 5-fold cross validation and split the dataset into 80% training data and 20% validation data. For our VN architecture, we report results for different kernel sizes  $k \in \{5, 7, 9, 11, 13\}$  and fixed the number of filter kernels  $N_k = 24$  and gradient steps  $T = 5$  empirically.

## 4.4 Results

We compared our VN results to bilateral filtering, BM3D, TV and second-order TGV qualitatively and quantitatively. Table 4.1 shows the mean values and standard deviations for Peak Signal-To-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). The parameters for all methods were estimated by grid-search. Figure 4.6 shows the qualitative comparison of the different methods and illustrates that the VN result has less streaking artifacts and appears more natural compared to BM3D. Our deep learning architecture outperforms all methods qualitatively and quantitatively. The best results can be achieved for a filter kernel size of 13.

## 4.5 Discussion

We propose a two-step deep learning architecture to correct for imperfections in limited-angle CT reconstruction due to missing projection data. In a first step, we correct intensity inhomogeneities in the image domain by learning compensation weights in the projection domain. In a second step, we train a VN to learn regularization to remove structured streaking artifacts. Our proposed approach substantially reduces streaking artifacts and outperforms current state-of-the-art non-linear filtering approaches that can mainly deal with unstructured noise. The strength of our proposed model is that it eliminates the need for manual tuning and replaces heuristic compensation steps by data-driven optimization. We see that the non-linear



**Figure 4.6:** Qualitative comparison of different non-linear filtering methods to the full scan reference. The VN reconstruction with kernel size  $k = 13$  shows significantly reduced streaking artifacts compared to BM3D. The neural network result is the intensity-inhomogeneity corrected output of a first correction step and defines the input to all methods.

filtering performs better for larger kernel sizes. However, Table 4.1 indicates that the results for kernel size 11 is worse in terms of PSNR compared to kernel sizes 9 and 13, which might be due to slight instabilities during training with the L-BFGS optimizer. This suggests to use different optimizers as outlined in Section 3.3. As the results suggests, larger kernel sizes are advantageous to deal with structured noise artifacts, as information over a larger FoV can be processed. Due to the large extent of the artifacts, deeper regularizers with increased perceptual field, motivated by CNNs and the U-net architecture [205], could be helpful to further



decrease the artifacts.

In the proposed architecture, reconstruction and artifact-correction are formulated as two different subproblems. Our proposed VN architecture suggests to reformulate this problem into a single problem by including the real reconstruction operator into the data term, including learning of the compensation weights. Therefore, artifacts in the image domain and in the data domain could be treated jointly and, hence, the final reconstruction quality might be improved. Another possibility for joint artifact correction in both domains would be to train the two different network architectures in an end-to-end manner. Further extensions to the proposed setup should also account for more physical effects.



## Variational Networks for 2D Cartesian MR Image Reconstruction

In theory, there is no difference between theory and practice. But, in practice, there is.

---

Walter J. Savitch

This chapter is based on the publications:

K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, 2018, <https://github.com/VLOGroup/mri-variationalnetwork>

F. Knoll, K. Hammernik, E. Kobler, T. Pock, M. P. Recht, and D. 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

### Contents

---

5.1	Fundamental Principles of Magnetic Resonance (MR) Image Reconstruction . . . . .	54
5.2	Data Acquisition . . . . .	69
5.3	Insights into Learning a Variational Network (VN) for Accelerated Magnetic Resonance Imaging (MRI) Data . . . . .	71
5.4	Exploring the Generalization Potential of VNs for Accelerated MR Imaging . . . . .	96
5.5	Influence of Loss Function Design for Accelerated MR Image Reconstruction . . . . .	98

---

5.6 Improved Regularization for Accelerated MR Image Reconstruction . . . . .	101
5.7 Intra-Vendor Reproducibility . . . . .	104
5.8 Conclusion and Outlook . . . . .	104

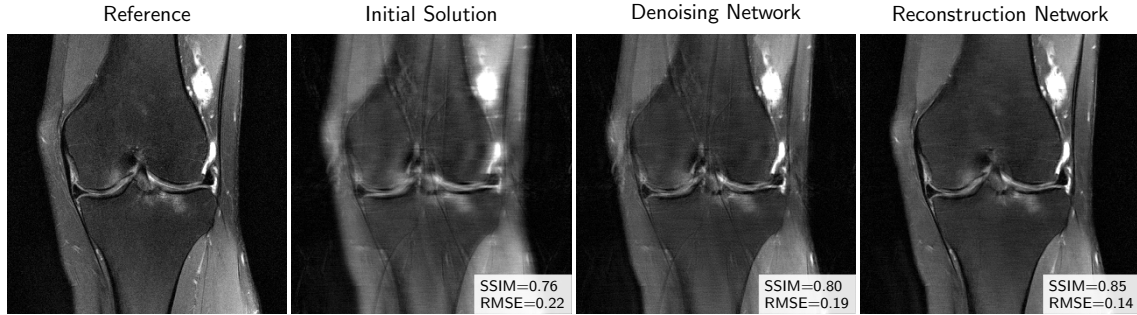
---

In Chapter 4 we formulated a VN for image enhancement. This rises the important question if the acquired raw data should be included in the learning-based reconstruction pipeline. To get a first answer to this question, we performed an experiment with VNs for accelerated MR image reconstruction. Accordingly, we trained one reconstruction VN using the data term  $\mathcal{D}[x, y] = \|Ax - y\|_2^2$ , where  $A$  models the linear MRI forward operator and  $y$  is the acquired MR data. For the second VN, we set the data consistency term to a denoising data term  $\mathcal{D}[x, y] = \|x - x^0\|_2^2$ , where  $x^0$  corresponds to the initial zero-filled solution  $x^0 = A^*y$ . While the only difference between the two VNs is the data consistency term, the number of parameters and structure of the regularizer to handle complex numbers is the same for both networks. Furthermore, the VNs were trained on 10 coronal proton-density-weighted knee datasets with FS according to [84]. Figure 5.1 shows the results for the reconstruction VN and denoising VN on four times undersampled Cartesian data. It clearly demonstrates that the reconstruction VN outperforms the denoising VN for this specific setup in terms of visual impression and the quantitative measures, i.e., Root-Mean-Squared-Error (RMSE) and Structural Similarity Index (SSIM). In the presence of more training data and more complex networks, the denoising network might also learn the appearance of certain structures in the image domain, and might lead to improved results. However, the information content of the measurement data can only be equal for orthogonal transformations in a best case scenario and decreases with every processing step otherwise. These results also suggest that it is beneficial to include the data term in the reconstruction to decrease the complexity of the regularizer.

In this chapter, we first introduce fundamental principles for MR image reconstruction in Section 5.1, including signal generation, practical acquisition strategies, challenges in accelerated data acquisition and Parallel Imaging (PI), which are important to define the MR forward model for static imaging. To explore the potential of the proposed VNs for MR image reconstruction, a clinical knee protocol was used to acquire MR data from a representative patient population, described in Section 5.2. First insights in using VNs for MR image reconstruction, including qualitative and quantitative comparisons to state-of-the-art reconstruction methods as well as a reader study on image quality are presented in Section 5.3. Further experiments consider the generalization potential of the proposed VNs in Section 5.4, the influence of loss functions on learning in Section 5.5 and the influence of the regularization in Section 5.6.

## 5.1 Fundamental Principles of MR Image Reconstruction

MRI offers excellent soft tissue contrast and is the leading imaging modality for a wide range of musculoskeletal, neurological and oncological diseases. However, MRI suffers from long ac-



**Figure 5.1:** Comparison of a reconstruction VN and denoising VN for the reconstruction of four times Cartesian undersampled MRI data of the knee. The two networks were trained with the same configuration [84] and only differed in the data consistency term. The denoising results show a lot of remaining artifacts which can be fully suppressed using the reconstruction VN. The quantitative measures RMSE and SSIM support the observations in image quality.

quisition time, which results in high exam costs and makes the acquisition itself uncomfortable for the patients. Furthermore, the acquired data are prone to artifacts due to patient motion. This section covers important concepts for accelerated MR image reconstruction from a signal processing perspective. Hence, the fundamental physics of MRI is not covered in this thesis. For details on the MR physics, we refer the interested reader to textbooks [17, 24, 155, 173].

### 5.1.1 From Signals to Images

In conventional MRI, we aim at measuring a discretized map of macroscopic magnetization related to the concentration of hydrogen protons  $^1\text{H}$  and the tissue-specific molecular environment of an object that is placed in a strong external magnetic field  $B_0$ . The hydrogen nuclei precess with the so-called *Larmor* frequency  $f_0$  around the direction of the external magnetic field  $B_0$ , following the Larmor equation

$$f_0 = \gamma B_0,$$

where  $\gamma$  is the gyromagnetic ratio which is specific for the underlying nuclei ( $\gamma_{^1\text{H}} = 42.58 \text{ MHz/T}$ ). In the presence of  $B_0$ , the magnetic moments of the individual protons are summed up and form the longitudinal net magnetization. In order to get a signal, the net magnetization has to be pushed out of the equilibrium, which can be achieved by applying a short Radio Frequency (RF) pulse that matches the Larmor frequency of the spins. This phenomenon is known as *resonance*. This results in a transverse magnetization of the precessing protons which induce an electrical voltage in a receiver coil according to Faraday's law of induction. After excitation of the spins, the protons turn back into the equilibrium state and align again with the  $B_0$  field.

Right now, we know about the origin of an MR signal, however, the coil measures a composite signal from all protons, which does not include any spatial information. To identify the correct spatial locations, different encoding schemes are required which are realized by

applying magnetic gradients in different directions. Figure 5.2 shows the encoding process for a regular Spin Echo (SE). In the first step, a slice selection gradient  $G_z$  is applied orthogonal to the desired image plane. We refer the slice direction to  $z$ . This results in a change of the Larmor frequency for different slice positions

$$f_z = \gamma(B_0 + G_z z).$$

Simultaneously, an RF pulse is applied to match a band of Larmor frequencies  $f_z \pm \Delta f_z$  defining a desired slice position with specific thickness  $\Delta z$ . In the second step, a Phase Encoding (PE) gradient is applied in  $y$  direction for a fixed time  $T_y$ . Once the gradient is turned off, the spins precess with the same frequency but different phase  $\varphi(y)$  as a function of the  $y$  position

$$\varphi(y) = \gamma G_y T_y y.$$

In the final readout step, the signal undergoes Frequency Encoding (FE) in  $x$  direction. Hence, the Larmor frequency varies along the  $x$  direction and the protons precess with frequency

$$f_x = \gamma(B_0 + G_x x).$$

In the presence of FE and PE, the signal  $S_q$  can be acquired over time  $t$  in a  $Q$ -channel receive-coil array following [204]

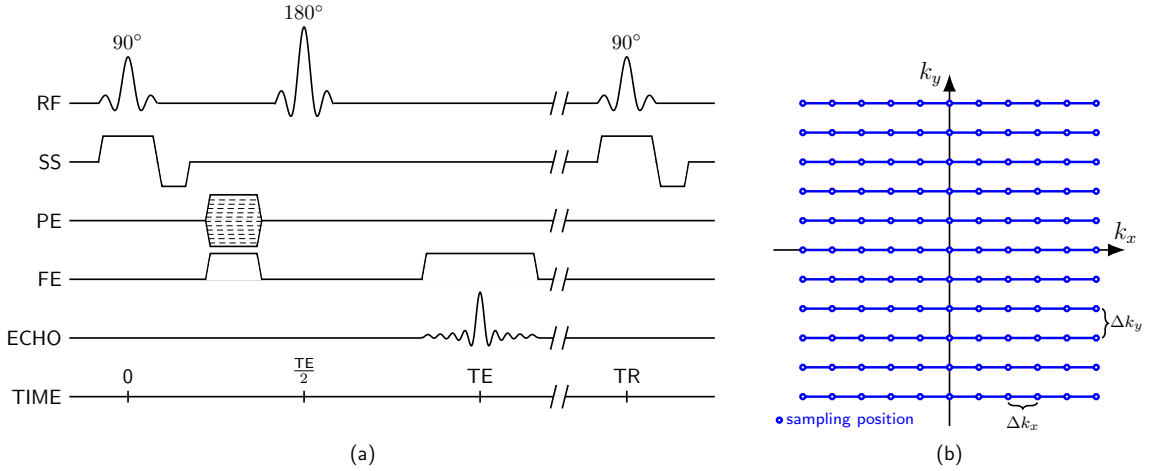
$$S_q(k_x, k_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(x, y) c_q(x, y) e^{-j2\pi(\gamma G_x t x + \gamma G_y T_y y)} dx dy.$$

Here,  $s$  describes the prepared transverse magnetization that arises due to the excited spins affected by the RF pulses. The sensitivity  $c_q$  of the  $q^{\text{th}}$  receiver coil modulates the magnetization depending on the spatial location of the receiver coil which is described by the principle of reciprocity [106, 107, 226]. We see that this equation follows the 2D Fourier Transform (FT)  $\mathcal{F}_{2D}$

$$S_q(k_x, k_y) = (\mathcal{F}_{2D}(s \cdot c_q))(k_x, k_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(x, y) c_q(x, y) e^{-j2\pi(k_x x + k_y y)} dx dy \quad (5.1)$$

with spatial frequencies  $k_x = \gamma G_x t$  and  $k_y = \gamma G_y T_y$ . The symbol

With the described Cartesian sampling strategy, the pulse sequence has to be repeated after a fixed Repetition Time (TR) in order to perform the different PE steps and consequently fill the corresponding lines in  $k$ -space. Hence, it is obvious that the PE is the time consuming part in data acquisition. Another important timing parameter is the Echo Time (TE). The TE defines the time between the center of the slice-selective RF pulse and the maximum peak of the echo signal which represents the  $k$ -space center. Depending on the selection of TE and TR, the transverse magnetization  $s$  is prepared differently and, hence, different tissue contrasts



**Figure 5.2:** Cartesian sampling: (a) Example pulse sequence of a SE and sampling trajectory (b). To identify the correct spatial locations, gradients for slice selection (SS), phase encoding (PE) and frequency encoding (FE) have to be applied. The repetition time (TR) and echo time (TE) influence the overall contrast of the acquired images.

are achieved. The image contrast itself depends on the spin distribution and relaxation effects of the flipped magnetization, i.e., longitudinal (spin-lattice) relaxation, characterized by the  $T_1$  relaxation time, transversal (spin-spin) relaxation, characterized by the  $T_2$  relaxation time, and additional dephasing effects due to locally varying  $B_0$  inhomogeneities characterized by  $T_2^*$ .

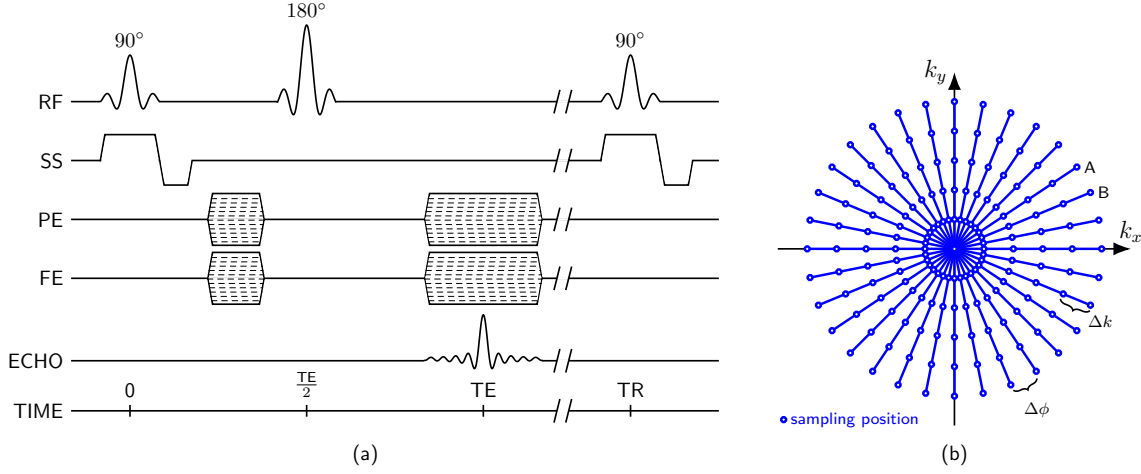
Besides the simple Cartesian sampling trajectory, there exist also other sampling strategies such as radial sampling. Figure 5.3 exemplifies the SE sequence used for magnetization preparation along with the gradient timing for radial sampling trajectories. To sample the lines, known as spokes, for the radial trajectory, PE and FE are switched on and off simultaneously. The spokes can be acquired at different angles  $\phi_n$  by varying the strengths of the gradients  $G_{x,n}$  and  $G_{y,n}$  in  $x$  and  $y$  direction for every  $n^{\text{th}}$  excitation

$$\phi_n = \arctan \frac{G_{y,n}}{G_{x,n}}.$$

The described sampling strategies already illustrate that the theoretically continuous signal has to be sampled during a limited amount of time in practice. However, discrete sampling of a finite signal introduces new challenges which will be covered in the following sections.

### 5.1.2 The Discrete World of $k$ -Space

Obtaining a perfect reconstruction would require to sample the  $k$ -space at an infinite number of locations, which is not feasible in practice. Hence, the MR signal is sampled at a finite number  $N_x$  and  $N_y$  of discrete locations  $k_x = \Delta k_x i_x$ ,  $i_x = -\frac{N_x}{2}, \dots, \frac{N_x}{2} - 1$  and  $k_y = \Delta k_y i_y$ ,  $i_y =$



**Figure 5.3:** Radial sampling: (a) Example pulse sequence of a SE and sampling trajectory (b). To identify the correct spatial locations, gradients for slice selection (SS), phase encoding (PE) and frequency encoding (FE) have to be applied. The repetition time (TR) and echo time (TE) influence the overall contrast of the acquired images.

$-\frac{N_y}{2}, \dots, \frac{N_y}{2} - 1$  within a frequency bandwidth  $k_{x,\max}$  and  $k_{y,\max}$ , leading to the discrete FT

$$S[l_x \Delta k_x, l_y \Delta k_y] = \Delta x \Delta y \sum_{i_x = -\frac{N_x}{2}}^{\frac{N_x}{2} - 1} \sum_{i_y = -\frac{N_y}{2}}^{\frac{N_y}{2} - 1} s[i_x \Delta x, i_y \Delta y] e^{-j l_x \Delta k_x l_y \Delta k_y i_x \Delta x i_y \Delta y}. \quad (5.2)$$

Due to the discrete sampling, the imaged object occurs periodically at a distance  $\text{FOV}_x$ ,  $\text{FOV}_y$ . If we assume that a physical object is bounded by  $\text{FOV}_x$  and  $\text{FOV}_y$ , the replicas do not overlap if the sampling steps  $\Delta k$  fulfill the Nyquist criterion. If the conditions defined in the following equations does not hold, the replicas overlap in image space, resulting in aliasing artifacts, also known as backfolding or wrap-around artifacts

$$\Delta k_x \leq \frac{1}{\text{FOV}_x}, \quad \Delta k_x = \gamma G_x \Delta t,$$

$$\Delta k_y \leq \frac{1}{\text{FOV}_y}, \quad \Delta k_y = \gamma G_y T_y,$$

where  $\Delta t$  is the sampling time interval at readout. By varying the strength of the PE gradient  $G_y$ , different spatial locations  $k_y$  can be addressed. Inversely, the highest spatial frequency  $k_{\max}$  defines the image resolution

$$\Delta x = \frac{1}{2k_{x,\max}},$$

$$\Delta y = \frac{1}{2k_{y,\max}}.$$



The base resolution  $N_x$  defines the matrix size and, hence, the number of encoding steps, resulting in following relations in image space and  $k$ -space

$$N_x = \Delta x \Delta k_x.$$

These relationships between image space and  $k$ -space are visualized in Figure 5.4. For radial sampling, we require assumptions on the angular and the frequency sampling interval  $\Delta\phi$  and  $\Delta k$  [17, 155], illustrated in Figure 5.3b. While  $\Delta k$  is easy to obtain by

$$\Delta k = \gamma G \Delta t \leq \frac{1}{FOV}, \quad G = \sqrt{G_x^2 + G_y^2},$$

obtaining  $\Delta\phi$  is a bit trickier. First, we use the small angle approximation to express the line  $\overline{AB}$  joining  $A$  and  $B$  in Figure 5.3 as

$$\overline{AB} = 2 \sin\left(\frac{\Delta\phi}{2}\right) k_{\max} \approx \Delta\phi k_{\max}.$$

The Nyquist criterion is fulfilled if the distance  $\overline{AB}$  at the edge of  $k$ -space does not exceed the frequency sampling interval  $\Delta k$

$$\Delta\phi \leq \frac{\Delta k}{k_{\max}} \leq \frac{1}{k_{\max} FOV}.$$

An interesting question is how many spokes  $N_\phi$  are required to fulfill the Nyquist criterion. The number of spokes  $N_\phi$  and the number of samples along the trajectory  $N_k$  can be expressed as

$$N_\phi = \frac{\pi}{\Delta\phi} N_k = \frac{\Delta k}{2k_{\max}}.$$

This leads to the relationship

$$N_\phi = \frac{\pi}{2} N_k$$

which states that  $\frac{\pi}{2} \approx 1.57$  times more samples are required than a corresponding Cartesian sampling to fulfill the Nyquist criterion.

Truncation artifacts that occur due to windowing of the infinite signal are another type of artifacts in MR data acquisition. As windowing with a rectangular window in the Fourier domain corresponds to convolving the image with a sinc filter, this results in Gibb's ringing artifacts that occur as parallel lines along sharp edges. Increasing the matrix size or decreasing the Field of View (FoV), which both increase the sampling density, can reduce the truncation artifacts. However, this decreases the voxel size as well as the Signal-to-Noise Ratio (SNR), and increases the acquisition time due to the higher number of PE steps for the increased matrix size. As the sampling density is in general lower in PE direction, the truncation artifacts are

also more prominent in the PE direction.

We see that there are essential connections between base resolutions  $N_x$  and  $N_y$ , spatial resolutions  $\Delta x$  and  $\Delta y$ , selection of the FoV, sampling distances  $\Delta k_x$  and  $\Delta k_y$ , and the maximum sampling frequencies  $k_{x,\max}$  and  $k_{y,\max}$ . Hence, changing a single parameter affects all other parameters as well as the SNR, too. To sample more efficiently, different strategies are considered in practice.

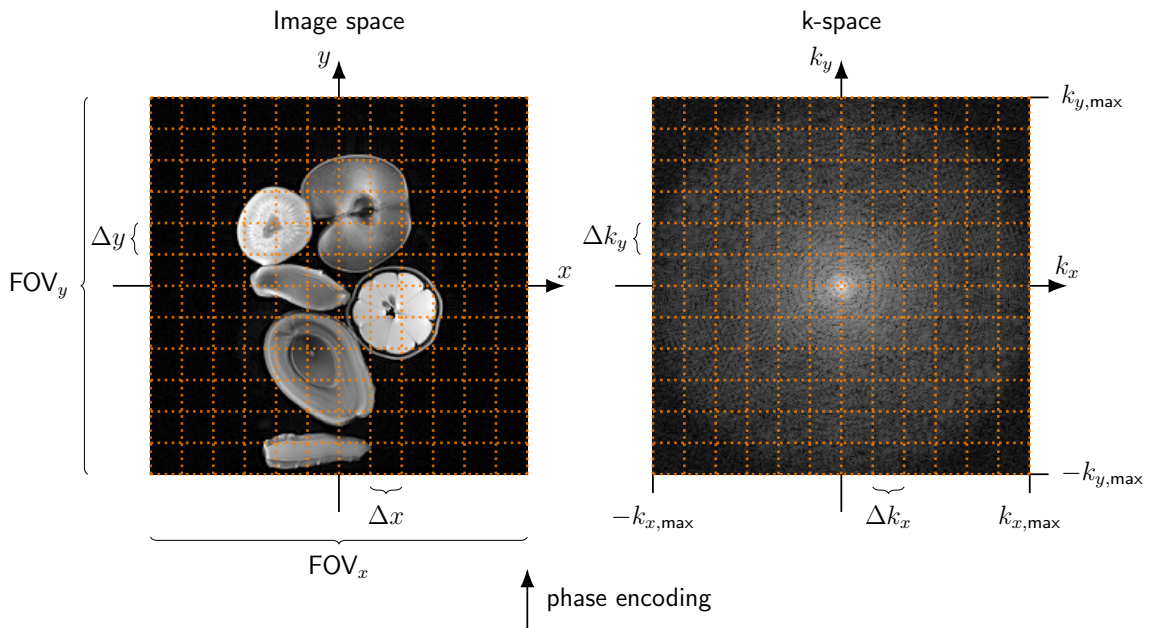
### 5.1.3 Acquisition Strategies in Clinical Practice

In this section, common acquisition strategies in clinical practice are described, which are relevant for modeling our MR forward and adjoint operators. Let us assume the following scenario: We examine different patients and want to obtain images with the same base resolution. Due to the different anatomies of the patients, the anatomy we want to visualize might not fit in the desired FoV. In order to fit the anatomy on the image, the FoV can be adapted for every patient, however, this influences the spatial resolution. Furthermore, increasing the FoV in PE direction or increasing the base resolution also increases the number of PE steps and consequentially increases the acquisition time.

**Readout Oversampling.** One strategy to avoid aliasing artifacts in FE direction is to oversample the  $k$ -space in FE direction. Readout oversampling doubles the number of FE steps and consequentially doubles the FoV at no additional acquisition time costs. For displaying the final images at the desired matrix size, the oversampled part is cropped in image domain. Readout oversampling is implemented on Siemens MR scanners by default.

**Phase Oversampling.** While readout oversampling can be done at no additional time costs, phase oversampling requires additional excitation pulses to increase the FoV in PE direction, which is time consuming. Instead of increasing the FoV and base resolution, to maintain the same spatial resolution, the FoV can be increased only for acquisition, e.g., when the patient's anatomy does not fit into the desired FoV but the base resolution should stay the same. The effect of phase oversampling, which is defined as percentage, is visualized in the second column of Figure 5.5. In a post-processing step, the rectangular FoV is cropped to display only the central part at the given base resolution.

**Phase Resolution.** Reducing the phase resolution, defined as percentage, can be used to decrease the number of PE steps and, hence, acquisition time. In fact, this is realized by lowering the bandwidth  $k_{y,\max}$  as displayed in the third column of Figure 5.5. Lowering  $k_{y,\max}$  influences the spatial resolution  $\Delta y$ , therefore, the acquired voxels are rectangular. To obtain square voxels and increase image resolution, the  $k$ -space is padded with zeros, known as zero filling or Fourier interpolation. No information is gained or lost due to zero-filling, hence, no additional artifacts are introduced.

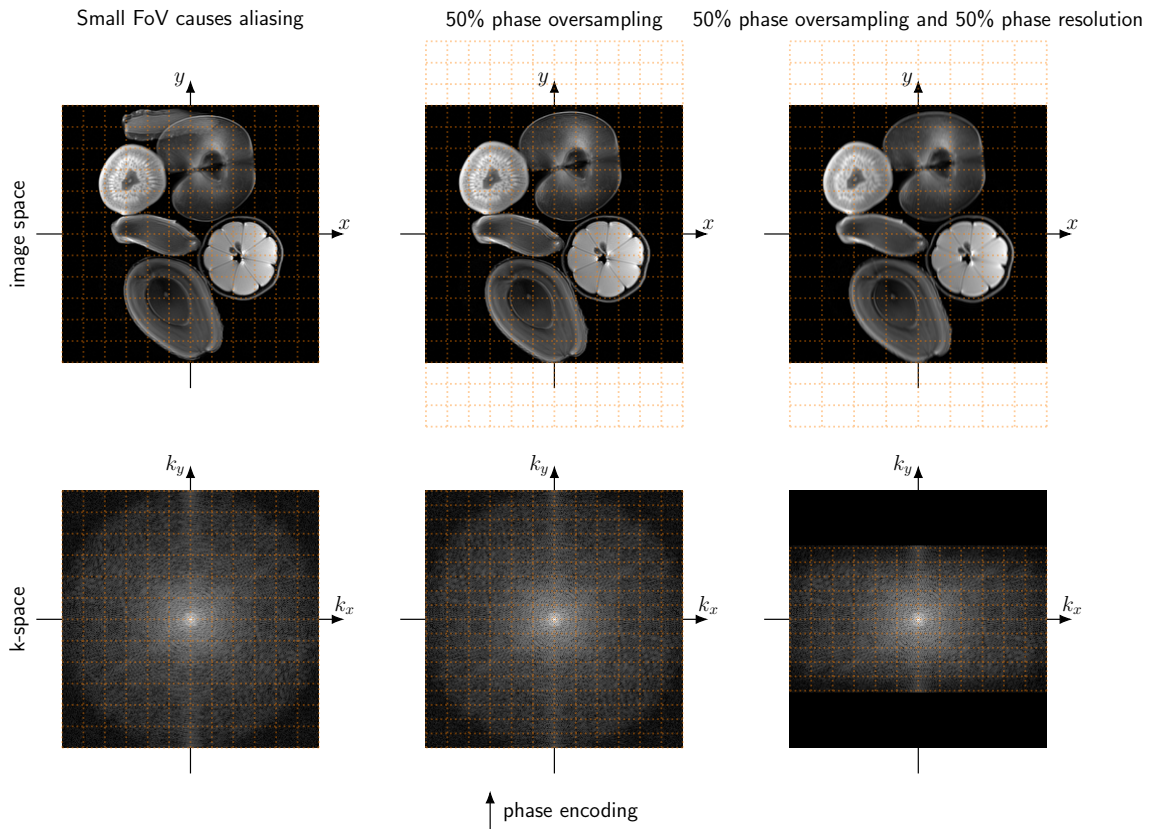


**Figure 5.4:** Visualization of the relationship of  $k$ -space and image space in terms of the FoV, spatial resolutions  $\Delta x$  and  $\Delta y$ , sampling frequencies  $\Delta k_x$  and  $\Delta k_y$ , and maximum  $k$ -space sampling frequencies  $k_{x,max}$  and  $k_{y,max}$ .

### 5.1.4 Accelerated MR Imaging

The major drawbacks of MR imaging are the physical limitations of the acquisition times. As outlined in the previous sections, the FoV as well as the desired resolution are important parameters that determine the number of required PE steps and impact the duration of the scan. Hence, faster acquisition techniques are desired to improve not only patient comfort and compliance, but also decrease, e.g., motion artifacts. Improved hardware concepts such as higher field strengths, stronger gradients and phased array receive coils with multiple elements lead to improved acquisition time and image quality. Due to the increased Specific Absorption Rate (SAR) associated with higher field strengths, physiological phenomena such as nerve stimulations limit further developments in this direction. The introduction of PI [77, 190, 227] led to a breakthrough in accelerated data acquisition. Raw  $k$ -space data is acquired in multiple coil elements as depicted in Figure 5.6 along with the individually reconstructed images in Figure 5.7. When only a subset of data are sampled, characteristic aliasing artifacts occur. To recover an image from these undersampled data, the data redundancy and spatial sensitivity of the individual coil elements are exploited to remove the artifacts. Theoretically, the acceleration potential of PI-based methods is limited by the number of coil elements. Practically, the measured data of the individual coil elements deliver redundant information which limits the acceleration potential as the underlying reconstruction problem becomes ill-posed.

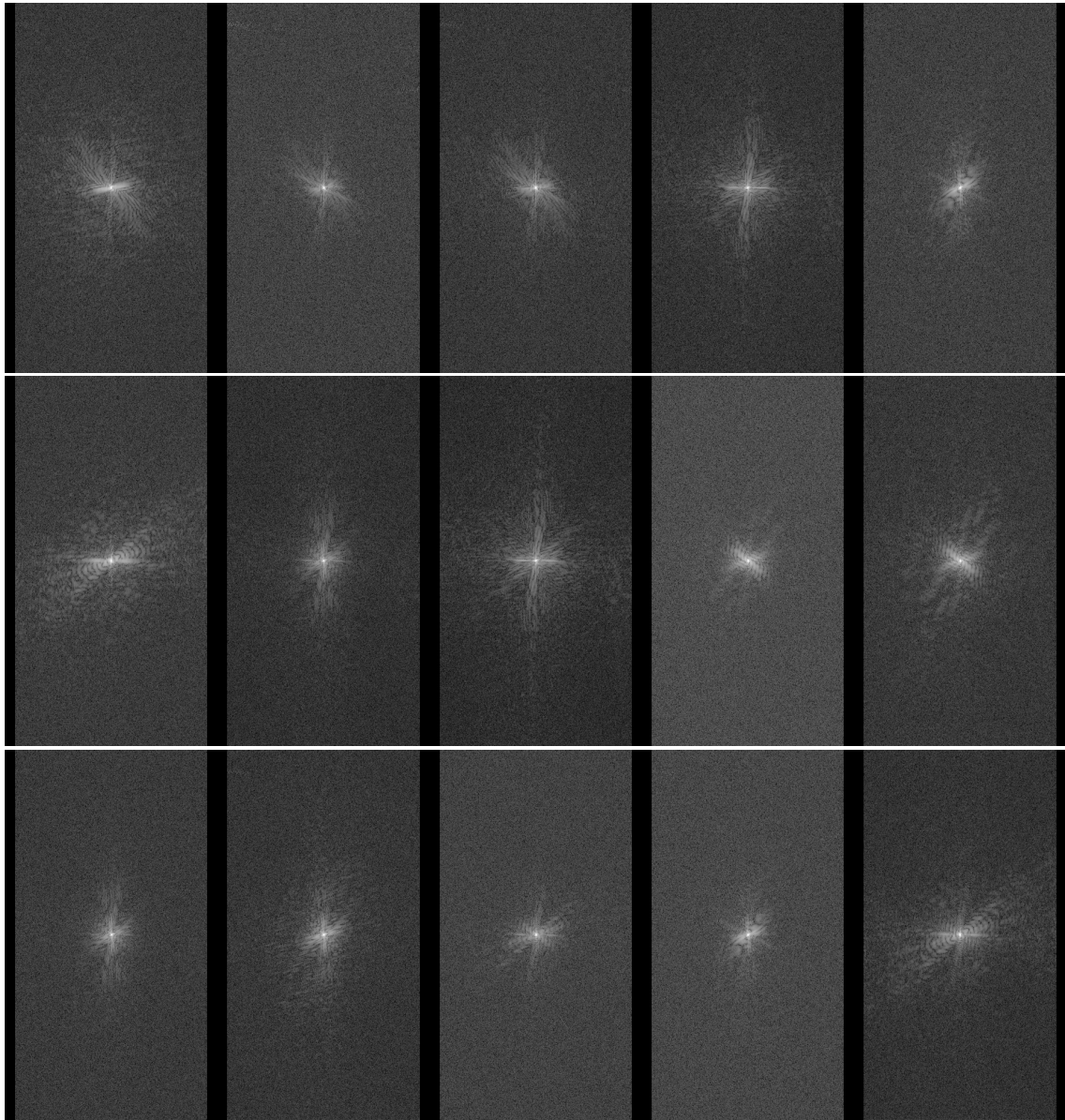
In the clinical setting, an acceleration factor of 2 is used for static 2D Cartesian imaging. Dynamic acquisitions allow for far higher acceleration due to the data redundancy in time



**Figure 5.5:** Reducing the FoV of Figure 5.4 results in aliasing of the turmeric into the apple. This effect can be avoided by oversampling in PE direction, which increases the FoV in this direction, while displaying only the desired part. To decrease acquisition time, the phase resolution can be reduced, resulting in a degraded image quality.

domain. The acceleration potential for 3D imaging is also higher as the number of phase encoding steps can be reduced in two directions. However, an accurate estimate of the coil sensitivity maps is required. As the coil sensitivity maps change based on the coil load, additional strategies have to be carried out. Examples to estimate the coil sensitivity maps are to perform a pre-scan or to acquire additional low-resolution scan lines, also known as ACLs. For reconstruction, the coil sensitivity maps have to be known either *implicitly* as for  $k$ -space methods, or *explicitly* as for image-based methods.

A new era of PI began in 1997 when Sodickson et al. [227] introduced Simultaneous Acquisition of Spatial Harmonics (SMASH), where missing PE lines in  $k$ -space are recovered by exploiting the spatial coil configurations. Further important developments in  $k$ -space methods were proposed by Griswold et al. [77] with Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA). Here, linear, shift-invariant convolution kernels are estimated from scan-specific ACLs in the center of the  $k$ -space, describing the relationship between the individual coil elements. These kernels are then used to recover the missing  $k$ -space data before the inverse FT is applied. In contrast to  $k$ -space methods, Pruessmann et al. [190] introduced



**Figure 5.6:** Acquired  $k$ -space data from a 15-channel knee coil.

Sensitivity Encoding (SENSE) that requires an explicit estimation of coil sensitivity maps (see Figure 5.8). The explicit coil sensitivity maps are applied in image domain to support the unaliasing. While the initial publication focused on Cartesian sampling [190], an extension to arbitrary sampling trajectories was presented in [189]. Pruessmann et al. [190] also introduced the  $g$ -factor map to characterize the ill-posedness of the reconstruction problem due to the redundant data that are measured by the individual coil elements in practice. The geometry factor  $g$  is a spatially varying measure for noise amplification in PI methods and can be determined via a noise covariance matrix of the single coil elements. The geometry factor  $g$  also



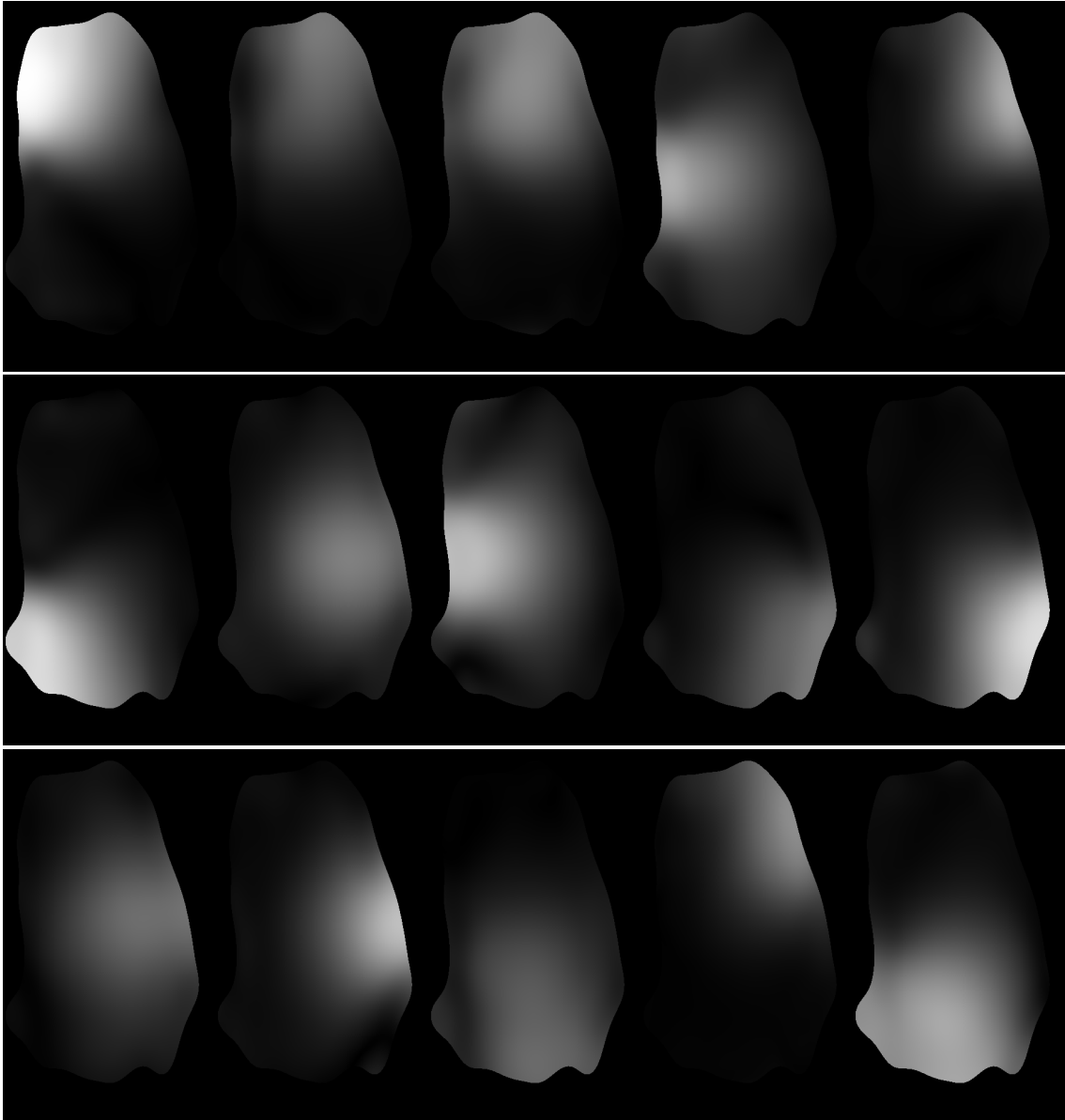
**Figure 5.7:** Reconstructed coil images for fully sampled data acquired with a 15-channel knee coil.

influences the  $\text{SNR}_{\text{pi}}$  of PI methods, together with the acceleration factor  $R$

$$\text{SNR}_{\text{pi}} = \frac{\text{SNR}_{\text{full}}}{\sqrt{Rg}},$$

where  $\text{SNR}_{\text{full}}$  is the SNR of the fully sampled image. For further information on noise estimation and  $g$ -factor estimation, we suggest further reading, e.g., [10, 22].

Indeed,  $k$ -space and image-space methods are treated independently although there exist a lot of connections. With the introduction of ESPIRiT, Uecker et al. [236] showed how to



**Figure 5.8:** Coil sensitivity maps were estimated for each channel using ESPIRiT [236] from 24 Auto-Calibration Lines (ACLs) in the center of the  $k$ -space.

bridge the gap between  $k$ -space and image-space methods and, hence, achieve the advantages of both methods.

In this thesis, we focus on a SENSE-based reconstruction. Therefore, we need the definition of the forward model which is described in Section 5.1.5. A more detailed overview of PI is out-of-scope of this thesis and we refer the interested reader to the original publications or tutorial-/overview papers [18, 216, 235].

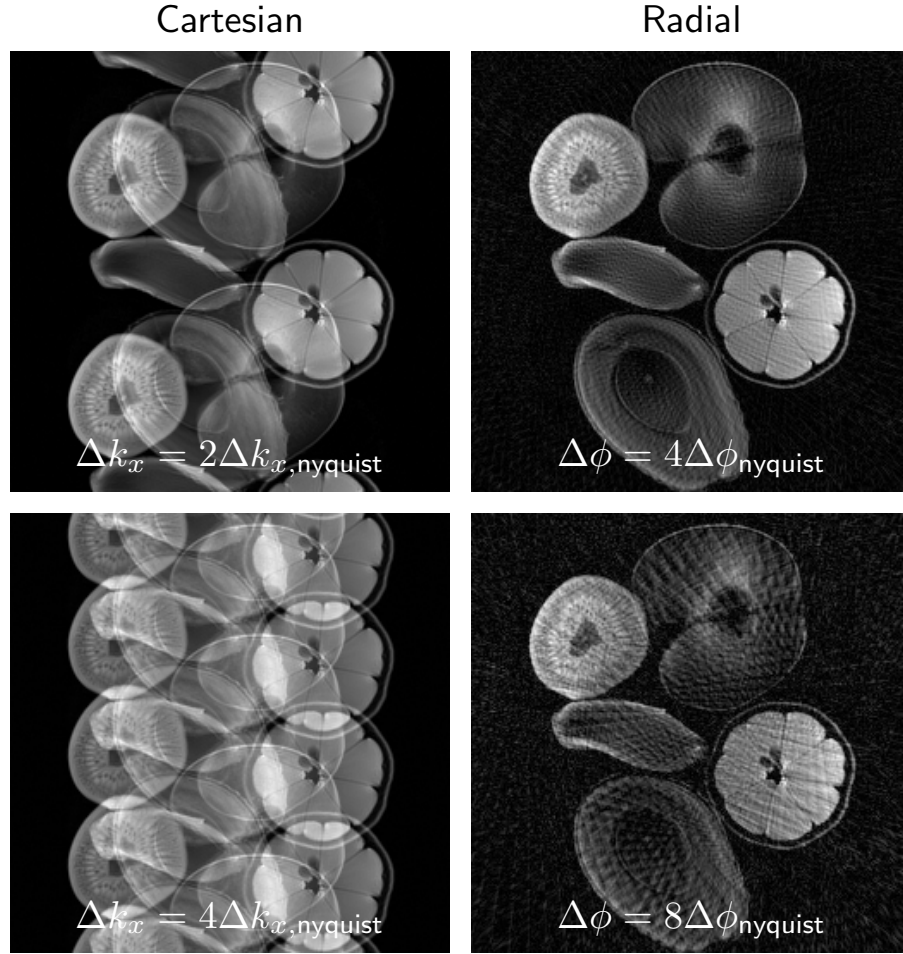
Besides the improved hardware concepts, the acquisition time in MR imaging can be also

reduced by advanced software concepts and reconstruction methods, where Compressed Sensing (CS) [31, 60, 164] led to a substantial improvement additional to PI. A compact overview of CS can be found in [163]. CS relies on three conditions to obtain images from  $k$ -space data sampled below the Nyquist rate [177, 220]. The first CS condition requires a data acquisition protocol for undersampling such that the artifacts become incoherent in a certain transform domain [31, 60]. In MRI, we usually achieve incoherence by random [164] or non-Cartesian sampling trajectories [19], such as radial or spiral sampling schemes. Figure 5.9 depicts how different undersampling schemes impact the aliasing structure in the final reconstruction. Radial sampling, illustrated in the right column, results in characteristic noise-like streaking artifacts. The incoherence of these artifacts fulfills the CS conditions. However, most routine clinical MRI examinations are still based on Cartesian sequences as shown in the left column, resulting in severe coherent aliasing artifacts. Especially in the case of 2D sequences, it can be challenging to fulfill the criteria for incoherence required by CS [105]. The second requirement for CS is that the image to be reconstructed must have a sparse representation in a certain transform domain. Common choices are the Wavelet transform [53, 164] or the Total Variation (TV) semi-norm [19, 132, 133, 207]. In these transform domains, the  $\ell_1$  norm is commonly applied to obtain approximate sparsity. The third CS condition requires a non-linear reconstruction algorithm that balances sparsity in the transform domain against consistency with the acquired undersampled  $k$ -space data. Despite the high promises of CS and its advances in different specialized research areas, its translation to clinical examinations is challenging. Besides the requirements on incoherence, the handcrafted sparsifying transforms are too simple to describe the complex image content of medical images. Furthermore, CS algorithms rely on a number of hyper-parameters that need to be tuned by hand thoroughly, which reduces the generalization potential and is not feasible in clinical practice. Finally, many of the used algorithms suffer from long reconstruction times.

### 5.1.5 Definition of the Static MR Forward Model

In this thesis, we use an MR forward model for Cartesian undersampled data, motivated by the image-based PI method SENSE [190]. The undersampled data is acquired by multiple receiver coils, where each coil is only sensitive in a certain region, which is modeled by coil sensitivity maps. Furthermore, the acquired data includes FE oversampling, as this is a standard behavior of the used scanner hardware. To limit the computational burden in image domain, we include the removal of FE oversampling in the forward model. We aim at reconstructing an image  $x \in \mathbb{C}^{N_x}$  of size  $M \times N$  and  $N_x = MN$  from noisy measurement data  $y \in \mathbb{C}^{N_y}$ ,  $N_y = \widetilde{M}NQ$  acquired with multiple receiver coils  $Q$ . The acquired data  $y$  additionally includes FE oversampling reflected in  $\widetilde{M}$ . While undersampling is performed in PE direction, the oversampling in this direction is not touched by the forward operator  $A$  to avoid additional artifacts. Hence,  $x$  and  $y$  have the same dimension in the PE encoding direction. In contrast, the FE oversampling can be removed without any problems in  $A$ . Hence, the forward operator





**Figure 5.9:** Examples for aliasing of Cartesian and radially undersampled data. While Cartesian undersampling leads to characteristic coherent backfolding artifacts, radially undersampled data results in incoherent streaking artifacts.

$A : \mathbb{C}^{N_x} \rightarrow \mathbb{C}^{N_y}$  is defined as

$$Ax = \begin{pmatrix} m \odot \mathcal{F}(c_1 \odot Bx) \\ \vdots \\ m \odot \mathcal{F}(c_Q \odot Bx) \end{pmatrix}$$

where  $c = [c_1, \dots, c_Q] \in \mathbb{C}^{N_y}$  is the set of coil sensitivities,  $m \in \mathbb{C}^{N_{x,pad}}$  and  $B : \mathbb{C}^{N_x} \rightarrow \mathbb{C}^{N_{x,pad}}$  is the pre-processing operator where  $N_{x,pad} = \widetilde{M}N$ . The symbol  $\odot$  denotes the Hadamard product, i.e., element-wise multiplications. First, the pre-processing operator  $B$  pads the image in FE-direction, which is in our case the  $y$  direction, by zeros to achieve the image of size  $\widetilde{M} \times N$ . Then the coil-sensitivity maps  $c_q$  are applied in image domain. This is followed by a centered FT  $\mathcal{F}$ , scaled by  $\frac{1}{\sqrt{\widetilde{M} \times N}}$ . The undersampling pattern is stored in the sampling

mask  $m$ , which is finally multiplied on the data in the Fourier domain.

We require also the transpose (backward) operator  $A^*(y) : \mathbb{C}^{N_y} \rightarrow \mathbb{C}^{N_x}$  defined as

$$A^*y = B^* \left( \sum_{q=1}^Q c_q^* \odot \mathcal{F}^*(y_q \odot m) \right).$$

The individual coil raw data is masked and transformed to image domain by a centered inverse FT  $\mathcal{F}^*$ , scaled by  $\frac{1}{\sqrt{M \times N}}$ . The individually transformed coil-images are combined with the sensitivity maps  $c_q$  and summed up over the coil dimension  $q$ . Finally, the post-processing operator  $B^* : \mathbb{C}^{N_{x,pad}} \rightarrow \mathbb{C}^{N_x}$  crops the central portion of the image in FE-direction at  $(\frac{nFE}{4} : \frac{3nFE}{4})$  to achieve an image  $x \in \mathbb{C}^{N_x}$ .

Once we have these operators defined, we are able to obtain a SENSE-based reconstruction by solving the least-squares problem Equation (3.2). In cases where computing the pseudo-inverse is infeasible, iterative methods such as the Conjugate Gradient (CG) method can be applied to obtain a solution. For MRI, the CG SENSE was introduced in [189], which is used as a reference method through this chapter. To further stabilize the reconstruction, regularization techniques are applied (see Section 2.1). Throughout this chapter, we compare our approach to second-order Total Generalized Variation (TGV), where we refer the interested reader to [21, 132] for details.

**Sampling Masks.** Basically, any pattern can be used as sampling masks. In our case, we focus on Cartesian sampling. Our raw data is acquired with reduced phase resolution meaning that some of the the acquired raw data are padded with zeros to achieve squared voxels. Naturally, we would setup a sampling mask that is also zero at these position. However, this leads to an undefined behaviour in these regions, because they can be filled with any unwanted values during reconstruction as there is no constraint forcing these regions to zero in the data term. Another example is visualized in Figure 5.10. Here, we use the fully-sampled Stanford 3D FSE Knee dataset<sup>1</sup> which was acquired with a Cartesian sampling pattern. If we extract 2D slices of these data, we observe that the raw data are defined in an ellipsoidal Region Of Interest (ROI) as illustrated in Figure 5.10 (d). If arbitrary random sampling patterns are used to reconstruct these data, the regions outside the ROI have to be set to ones to again force the raw data to zero in these regions. To show the influence, we reconstruct 2D slices of these data with CG SENSE with a high number of CG iterations and TGV. For the sampling mask, we generated a random pattern with a net acceleration factor of 4 according to Lustig et al. [164], including a fully sampled block of size  $24 \times 24$  in the center of  $k$ -space. The reconstructed images are depicted along with the resulting  $k$ -space in Figure 5.10 as well as the used sampling masks. In the case of CG SENSE, we observe larger noise amplification in the reconstructed image if the regions outside the ROI are set to zero in the sampling mask. For TGV, we do not observe a visual difference in the reconstruction. This can be explained

<sup>1</sup><http://mridata.org/list?project=Stanford%20Fullysampled%203D%20FSE%20knees>

Acceleration	Acquisition Time					Total Time
	Coronal PDw	Coronal PDw-FS	Axial T <sub>2</sub> w-FS	Sagittal T <sub>2</sub> w-FS	Sagittal PDw	
Fully Sampled	3:53 min	4:17 min	2:58 min	4:03 min	7:16 min	22:27 min
Clinical IPAT 2	2:41 min	2:51 min	1:38 min	1:58 min	3:48 min	12:56 min
IPAT 3	1:54 min	2:02 min	1:10 min	1:23 min	2:41 min	9:10 min
IPAT 4	1:32 min	1:39 min	0:58 min	1:10 min	2:05 min	7:24 min

**Table 5.1:** Acquisition times for a clinical knee protocol with different acceleration factors.

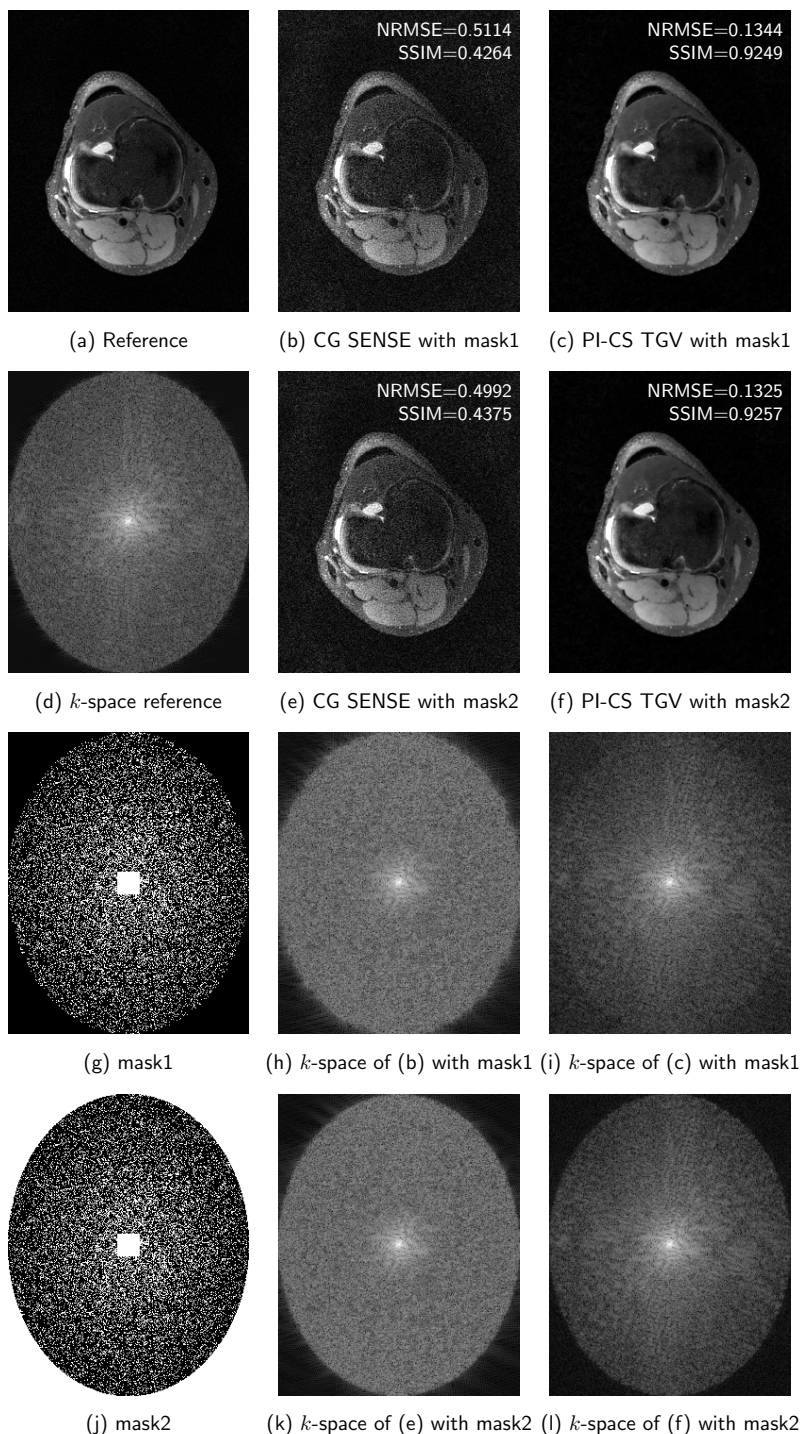
that the TGV regularizer does not allow regions with amplified noise to appear. However, the impact of masking the undefined regions can be best seen in the  $k$ -space of the reconstructed images visualized in Figure 5.10 (h,i,k,l). Here, we clearly see that the undefined regions have to be considered in the sampling mask to achieve valid reconstruction results for all iterative methods, including but not limited to CG SENSE, CS-based or learning-based methods.

## 5.2 Data Acquisition

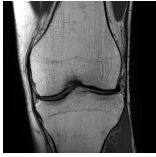

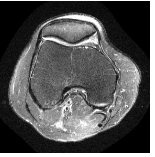
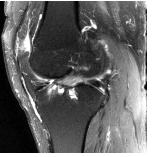
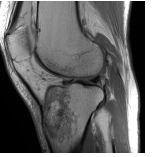
A major goal of our work was to explore the generalization potential of a learning based approach for MRI reconstruction. For this purpose, we used a standard clinical knee protocol for data acquisition with a representative patient population that differed in terms of anatomy, pathology, gender, age and body mass index. The protocol consisted of five 2D Turbo Spin Echo (TSE) sequences that differed in terms of contrast, orientation, matrix size and SNR. For each sequence, we scanned 20 patients on a clinical 3T system (Siemens Magnetom Skyra) at New York University Langone Health, NY, USA, using a 15-channel knee coil. All data were acquired without acceleration, and undersampling was performed retrospectively. In addition, we acquired prospectively accelerated data for one case. The total scan times for different acceleration factors are listed in Table 5.1. The number of acquired slices was chosen individually for each clinical patient exam. The study was approved by our institutional review board. To test the algorithm on different hardware, additional fully sampled and prospectively undersampled data were acquired with the same hardware specifications for one case at Graz University of Technology, Graz, Austria. The sequence parameters along with the details about the scanned patient population are listed in Table 5.2.

Coil sensitivity maps were precomputed from a data block of size  $24 \times 24$  at the center of  $k$ -space using ESPIRiT [236]. For both training and quantitative evaluation, each network reconstruction was compared against a gold standard reference image. We defined this gold standard as the coil-sensitivity combined reconstruction which was obtained by applying the adjoint operator  $A^*$  to the fully sampled  $k$ -space data. The fully sampled raw data were then retrospectively undersampled for both training and testing. Concerning all experiments the acceleration factor, e.g.,  $R = 4$ , is solely the number of skipped phase encoding steps, while the effective acceleration factor is slightly less due to the fully sampled block in the center of  $k$ -space.

With the recent release of the fastmri dataset [268], a large open database is now available



**Figure 5.10:** The fully-sampled raw data is defined by an ellipsoidal region. The sensitivity-combined reconstruction is shown in (a) along with its  $k$ -space in (b). If the mask in (i) is used, the regions are not forced to 0 in the data consistency term. To account for the undefined regions, we set these regions to 1 (j). Although there is not a great visual difference for these examples for CG SENSE (e,h) and TGV (f,i) when using the masks (g,j), it impacts the  $k$ -space greatly. While regions in  $k$ -space are forced to 0 (k,l) when using mask2 (j), these regions are filled with unwanted signal (h,i) when using mask1 (g), which might result in artifacts and noise in the image domain.

Parameter	Clinical Protocol				
	Coronal PDw	Coronal PDw-FS	Axial T <sub>2</sub> w-FS	Sagittal T <sub>2</sub> w-FS	Sagittal PDw
Sample image					
Base resolution	320	320	320	320	384
Field-of-View (FoV)	140 mm	140 mm	140 mm	140 mm	140 mm
Phase encoding direction	R $\gg$ L	R $\gg$ L	R $\gg$ L	H $\gg$ F	H $\gg$ F
Phase resolution	90%	90%	80%	80%	80%
Phase oversampling	15%	15%	50%	100%	100%
Slice thickness	3 mm	3 mm	3mm	3 mm	3 mm
TR	2750 ms	2870 ms	4000 ms	4300 ms	2800 ms
TE	27 ms	33 ms	65 ms	50 ms	27 ms
Fat saturation (FS)	No	Weak	Weak	Weak	No
Turbo factor	4	4	9	11	4
Flip angle	180°	180°	150°	180°	150°
Bandwidth	260 Hz/Px	200 Hz/Px	252 Hz/Px	200 Hz/Px	303 Hz/px
Number of slices	35-42	33-44	33-41	31-40	31-38
Gender	5 F / 15 M	10 F / 10 M	10 F / 10 M	11 F / 9 M	11 F / 9 M
Age	15-76 yrs	30-80 yrs	20-70 yrs	12-73 yrs	15-94 yrs
BMI	20.46-32.94	19.76-33.87	19.20-35.69	18.16-37.31	18.69-35.15

**Table 5.2:** Overview of sequence parameters for the used clinical knee protocol along with details about the scanned patient population.

for the MRI reconstruction community. From the fastmri dataset, we extracted the data acquired with the same hardware and matrix-size parameter as our dataset, resulting in 140 training volumes each for coronal PDw and coronal PDw-FS, and 50 test volumes for these contrasts and performed additional experiments with these data.

### 5.3 Insights into Learning a VN for Accelerated MRI Data

Despite the high promise of CS approaches, most routine clinical MRI examinations are still based on Cartesian sequences. Especially in the case of 2D sequences, it can be challenging to fulfill the criteria for incoherence required by CS [105]. One other obstacle in the way of incorporating CS into some routine clinical routine examinations is the fact that the sparsifying transforms employed in CS applications up to now may be too simple to capture the complex image content associated with biological tissues. This can lead to reconstructions that appear blocky and unnatural, which reduces acceptance by clinical radiologists. A further drawback, not only for CS but for advanced image acquisition and reconstruction methods in general, is the long image reconstruction time typically required for iterative solution of non-linear optimization problems. A final challenge concerns the selection and tuning of hyper-parameters for CS approaches. A poor choice of hyper-parameters leads either to over-regularization, i.e., excessively smooth or unnatural-looking images, or else to images that still show residual under-sampling artifacts. The goal of this section is to demonstrate that, using learning approaches,

we can achieve accelerated and high-quality MR image reconstructions from undersampled data which do not fulfill the usual CS conditions.

In this section, we formulate a VN for accelerated MRI reconstruction of complex-valued multi-channel MR data. We train the VN on a complete clinical protocol for musculoskeletal imaging, evaluating performance for different acceleration factors, and for both regular and pseudo-random Cartesian 2D sampling. Furthermore, we evaluate the image quality of the proposed approach compared to PI-CS using a reader study, examined by trained radiologists. Using clinical patient data, we investigate the capability of the VN approach to preserve unique pathologies that are not included in the training data set.

### 5.3.1 Methods

In MRI reconstruction, we naturally deal with complex numbers. Here, we introduce a mapping to real-valued numbers that we will use throughout this chapter. We define complex images  $\tilde{x}$  of size  $n_x \times n_y = N_x$  as equivalent real images  $x$  as follows:

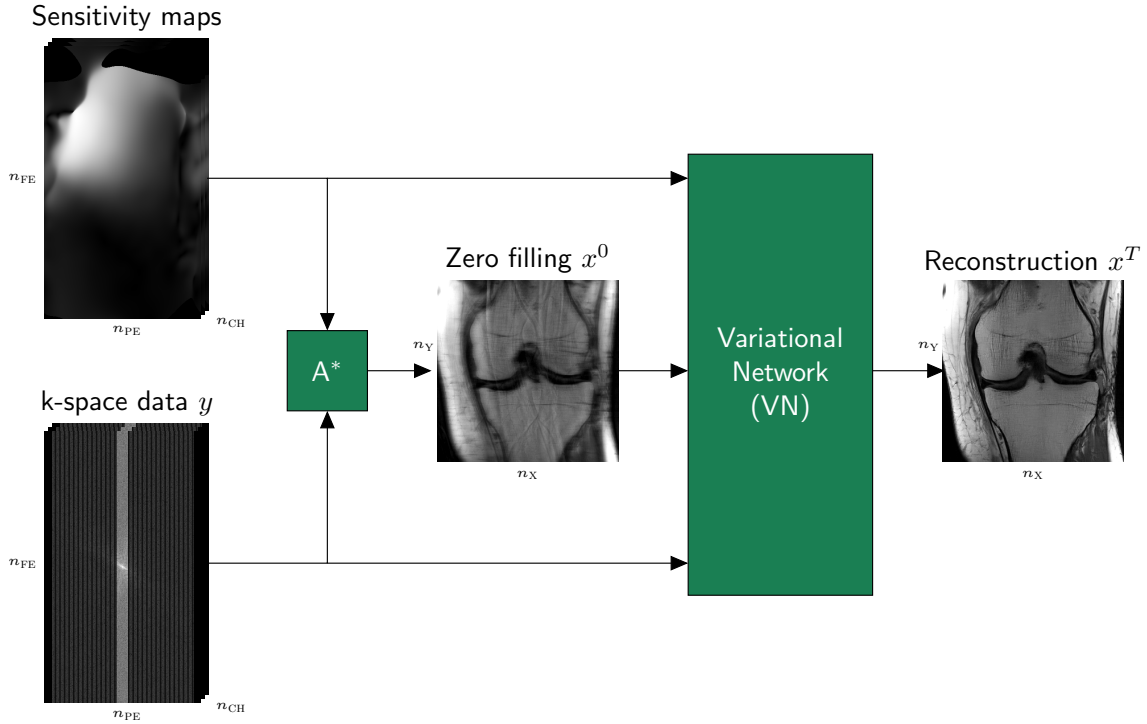
$$\tilde{x} = x_{\text{re}} + jx_{\text{im}} \in \mathbb{C}^{N_x} \Leftrightarrow x = (x_{\text{re}}, x_{\text{im}}) \in \mathbb{R}^{2N_x}.$$

Considering an inverse problem (3.1), we denote the reconstructed image as  $x \in \mathbb{R}^{2N_x}$  and  $y \in \mathbb{R}^{N_y}$  is the given undersampled  $k$ -space data, where missing data are padded by zeros. The linear forward sampling operator  $A$  implements point-wise multiplications with  $Q$  coil sensitivity maps, FTs, and undersampling according to a selected sampling pattern as described in detail in Section 5.1.5. Originally, the operator  $A$  is defined by the mapping  $\mathbb{C}^{N_x} \rightarrow \mathbb{C}^{N_y}$ , but embedding it in our real-valued problem changes the mapping to  $\mathbb{R}^{2N_x} \rightarrow \mathbb{R}^{2N_y}$ . This convolution operator  $K = (K_{\text{re}}, K_{\text{im}}) : \mathbb{R}^{2N_x} \rightarrow \mathbb{R}^{2N_y}$  models convolutions with filter kernels  $k \in \mathbb{R}^{s \times s \times 2}$  of size  $s$

$$Kx = K_{\text{re}}x_{\text{re}} + K_{\text{im}}x_{\text{im}}, \quad x \in \mathbb{R}^{2N_x} \Leftrightarrow x * k = x_{\text{re}} * k_{\text{re}} + x_{\text{im}} * k_{\text{im}}, \quad x \in \mathbb{R}^{n_x \times n_y \times 2}. \quad (5.3)$$

and non-linear potential functions  $\rho(z) = (\rho(z_1), \dots, \rho(z_N))^T : \mathbb{R}^{N_x} \rightarrow \mathbb{R}^{N_x}$ . Using this notation, we end up in a VN structure as illustrated in Figure 5.12. A zero filled solution is computed from the undersampled  $k$ -space data by applying the adjoint operator  $A^*$ . The measured raw data and sensitivity maps, together with the zero filled initializations, are fed into the VN as illustrated in Figure 5.11. The sensitivity maps are used in the operators  $A$  and  $A^*$ , which perform sensitivity-weighted image combination and can also implement other processing steps such as the removal of readout oversampling. While both raw data and operators  $A$  and  $A^*$  are required in every iteration of the VN to implement the gradient of the data term, the gradient of the regularization is only applied in the image domain as depicted in Figure 5.12.

To set up the training procedure, we minimize a loss function over a set of images  $S$  with respect to the VN parameters  $\theta$ . A common choice for the loss function is the Mean Squared



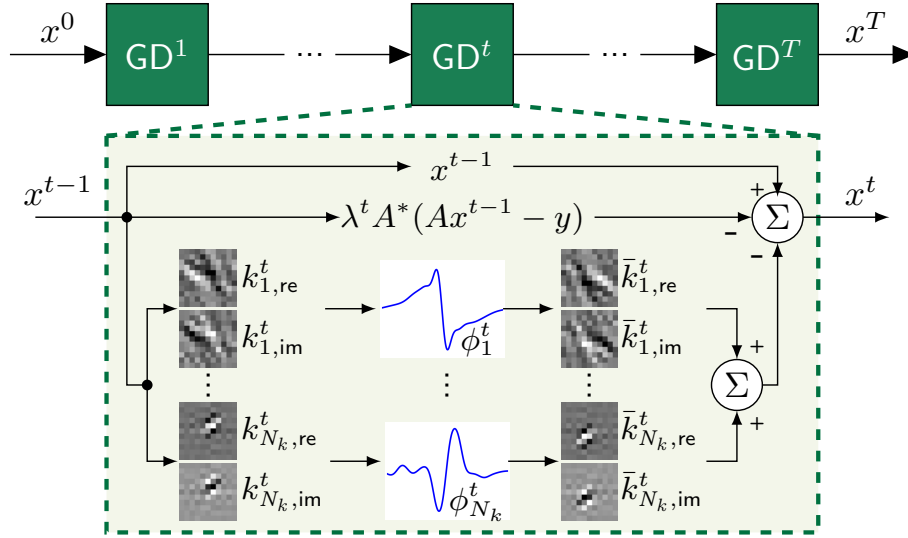
**Figure 5.11:** Proposed image reconstruction pipeline: A zero filled solution is computed from the undersampled  $k$ -space data by applying the adjoint operator  $A^*$ , which involves the application of the sampling mask, inverse FT and coil sensitivity maps. We feed the undersampled  $k$ -space data, coil sensitivity maps and the zero filling solution to the VN to obtain a reconstruction. For simplicity, we show the magnitude images, but all the input and output data of the VN are complex-valued.

Error (MSE). As we are dealing with complex numbers in MRI reconstruction and we typically assess magnitude images, we define the MSE of ( $\epsilon$ -smoothed) absolute values as loss function for training

$$\min_{\theta} \left\{ \mathcal{L}(x^T(\theta), x_{\text{ref}}) = \frac{1}{2S} \sum_{s=1}^S \left\| |x_s^T(\theta)|_{\epsilon} - |x_{\text{ref},s}|_{\epsilon} \right\|_2^2 \right\}, \quad |x|_{\epsilon} = \sqrt{x_{\text{re}}^2 + x_{\text{im}}^2} + \epsilon,$$

where  $|\cdot|_{\epsilon}$  is understood in a point-wise manner. More details on the choice of this loss functions as well as experiments using different loss function is presented in Section 5.5.

**Experimental Setup.** Our experiments differed in contrast, orientation, acceleration factor and sampling pattern. For all our experiments, we pre-normalized the acquired  $k$ -space volumes with  $n_{\text{sl}}$  slices by  $\frac{\sqrt{n_{\text{sl}}10000}}{\|y\|_2}$ . We trained an individual VN for each experiment and kept the network architecture fixed for all experiments. The VN consisted of  $T = 10$  steps. The initial reconstruction  $x_0$  was defined by the zero filled solution. In each iteration  $N_k = 48$  real/imaginary filter pairs of size  $11 \times 11$  were learned. For each of the  $N_k$  filters, the corresponding activation function was defined by  $N_w = 31$  Gaussian radial basis functions (RBFs)



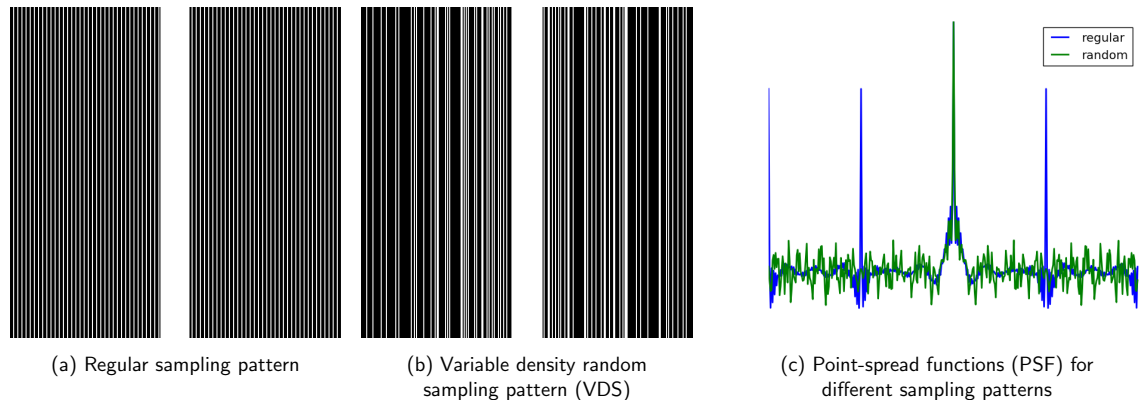
**Figure 5.12:** The VN consists of  $T$  gradient descent steps. To obtain a reconstruction, we feed the undersampled k-space data, coil sensitivity maps and the zero filling solution to the VN. Here, a sample gradient step is depicted in detail. As we are dealing with complex-valued images, we learn separate filters  $k_i^t$  for the real and complex plane. The non-linear activation function  $\phi_i^t$  combines the filter responses of these two feature planes. During a training procedure, the filter kernels, activation functions and data term weights  $\lambda^t$  are learned.

equally distributed between  $[-150, 150]$ . Including the data term weight  $\lambda^t$  in each step, this resulted in a total of 131,050 network parameters.

For optimization, we used the Inertial Incremental Proximal Gradient (IIPG) optimizer described in Appendix C.1. The IIPG optimizer allows us to handle the previously described constraints on the network parameters. We generated a training set for each of the five knee datasets. In each experiment, we used 20 image slices from 10 patients with the same contrast weighting and orientation, which amounts to 200 images, as the training set. For each patient, the central 20 slices were used for training. In fact, each single pixel of these training images provides a training example. In the case of a  $320 \times 320$  matrix, this results in more than 20 million pixels which is orders of magnitudes larger than the number of network parameters. The training set was split into mini batches of size 10. Optimization was performed for 1000 iterations with a step size of  $\eta = 10^{-3}$ .

**Implementations.** The VN approach as well as the reference methods were implemented in C++/CUDA with CUDNN support. We provide Python and Matlab interfaces for testing. Experiments were performed on a system equipped with an Intel Xeon E5-2698 Central Processing Unit (CPU) (2.30GHz) and a single Nvidia Tesla M40 Graphics Processing Unit (GPU). For dictionary learning, we used the Matlab implementation provided by Ravishankar et al. [195] and extended their formulation to be used with our multi-coil sampling operator. This requires to solve Eq. (7) in their work using the conjugate gradient method which additionally





**Figure 5.13:** Cartesian sampling patterns for  $R = r$ : (a) shows a regular clinical pattern and (b) illustrates a VRS pattern. Both patterns have the same amount of ACLs and PE steps. While the regular Cartesian sampling pattern introduces coherent backfolding artifacts, depicted by the blue PSF in (c), the VRS pattern introduces more randomness, depicted by the green PSF in (c).

increases runtime. Source code and data are available online based on Tensorflow<sup>2</sup>.

**Experiments.** In the first step, we investigated whether the learning-based VN approach actually benefits from structured undersampling artifacts due to regular undersampling, or if it performs better with incoherent undersampling artifacts as are typically present in CS applications. We used a regular sampling scheme with fully-sampled  $k$ -space center consisting of 24 ACLs, identical to the vendor implementation of an accelerated TSE sequence on an MR-system. This regular sampling pattern, visualized in Figure 5.13a, results in coherent aliasing artifacts as indicated by the Point Spread Function (PSF) Figure 5.13c. We also generated a Variable Density Random Sampling (VRS) pattern according to Lustig et al. [164], illustrated in Figure 5.13b, in order to introduce more randomness as indicated by the PSF in Figure 5.13c. Both sampling patterns have the same fully-sampled  $k$ -space center and same number of PE steps. We evaluated the acceleration factors  $R \in \{3, 4\}$  for two sequences which differ in contrast and SNR. The second step was to explore the generalization potential with respect to different contrasts and orientations of a clinical knee protocol. In a third step, we performed an experiment with prospectively accelerated data.

**Evaluation.** We tested our algorithm on data from 10 clinical patients per sequence and reconstructed the whole imaged volume for each patient. These cases were not included in the training set, and they also contained pathology not represented in the training set. It is worth noting that the number of slices was different for each patient, depending on the individual optimization of the scan protocol by the MR technologist.

We compared our learning-based VN to the linear PI reconstruction method CG SENSE [190] and a combined PI-CS non-linear reconstruction method based on

<sup>2</sup><https://github.com/VLOGroup/mri-variationalnetwork>

TGV [21, 132]. Additionally, we compared our qualitative results to dictionary learning [195] and provide quantitative measures for the selected cases. However, a full comparison to dictionary learning for all cases is out of scope of this work due to the long runtime requirements (approximately one hour per slice). The forward and adjoint operators for all three reference methods, in particular the coil sensitivity maps, were consistent with our VN approach. All hyper-parameters for CG SENSE and PI-CS TGV such as the number of iterations and regularization parameters were estimated individually by grid search for each sampling pattern, contrast and acceleration factor, such that the MSE of the reconstruction to the gold standard reconstruction was minimized. For dictionary learning, we used the standard parameters as in [195] and estimated the regularization parameter by grid search such that the MSE of the depicted slices was minimized. We assessed the reconstruction results quantitatively in terms of MSE, Normalized Root Mean Squared Error (NRMSE), and SSIM [248] with  $\sigma = 1.5$  on the magnitude images.

In addition to the qualitative and quantitative evaluation, we performed a reader study that compared results from the proposed VN method with results from PI-CS TGV. The 50 test cases from all five sequences were independently reviewed by two fellowship trained musculoskeletal radiologists who were blinded to the MRI reconstruction method. Cases were reviewed in two different sessions, separated by 2 weeks to minimize recall bias. Each session consisted of a random selection of 25 VN and 25 TGV reconstructions. Using a 4-point ordinal scale, reconstructed images were evaluated for sharpness (1: no blurring, 2: mild blurring, 3: moderate blurring, 4: severe blurring), SNR (1: excellent, 2: good, 3: fair, 4: poor), presence of aliasing artifacts (1: none, 2: mild, 3: moderate, 4: severe) and overall image quality (1: excellent, 2: good, 3: fair, 4: poor). Comparisons in terms of image quality scores, averaged over the two readers, were made using a one-sided Wilcoxon signed-rank test. The null hypothesis that PI-CS TGV reconstruction results are equal or better than VN-based results is rejected at significance level  $\alpha = 0.05$  if the resulting p-value of the test is lower than the significance level  $\alpha$ .

### 5.3.2 Results

First, we show results for the retrospective experiments. This is followed by prospective results as well as results for the reader study on image quality in. Finally, we show some examples for learned VN parameters in Section 5.3.2.

**Retrospective Variational Network Reconstructions.** Figure 5.18 displays the impact of acceleration factor  $R = 4$  and sampling patterns for CG SENSE, dictionary learning, PI-CS TGV and our learned VN on coronal PDw images. Additionally, we plot zero filling solutions to illustrate the amount and structure of undersampling artifacts. Difference images to the reference are visualized in Figure 5.19. The reconstruction results for acceleration factor  $R = 3$  along with the difference images are illustrated in Figure 5.14 and Figure 5.15. Residual artifacts and noise amplification can be observed for CG SENSE, in particular for  $R = 4$ .

In case of acceleration factor  $R = 3$ , the PI-CS image appears less noisy than CG SENSE; however, similar undersampling artifacts are present. For  $R = 4$  the PI-CS TGV result contains fewer undersampling artifacts than CG SENSE but small details in the image are already lost.

Dictionary learning leads to an improved removal of undersampling artifacts, resulting in a lower NRMSE than PI-CS TGV for this particular case. The learned VN suppresses these artifacts while still providing sharper and slightly more homogeneous images. Interestingly, dictionary learning as well as the PI-CS TGV and learned VN reconstruction with  $R = 3$  regular sampling perform slightly better than with VRS in terms of intensity homogeneity and sharpness. For acceleration  $R = 4$ , randomness improves the reconstruction results.

Similar observations can be made for coronal PDw-FS scans, as depicted in Figure 5.20. Again, the reconstruction results for acceleration factor  $R = 3$  along with the difference images are illustrated in Figure 5.16 and Figure 5.17. The main difference is that this sequence has a lower SNR compared to the non-FS version. Since additional noise reduces sparsity, the PI-CS TGV reconstructions produce an even more unnatural blocky pattern and contain substantial residual artifacts. The dictionary learning results appear blurrier at image edges and the general reconstruction quality is lowered at this level of SNR, which can best be seen in the error maps in Figure 5.19 and is supported by the quantitative values for this particular slice. Our learned VN is able to suppress these undersampling artifacts and shows improved image quality at this SNR level as well.

All our observations are supported by the quantitative evaluation depicted in Table 5.4 for  $R = 4$  and in Table 5.3 for  $R = 3$ . The wide range in quantitative values over the different sequences illustrates the effect of SNR on the reconstructions. The learned VN reconstructions show superior performance in terms of MSE, NRMSE and SSIM in all cases. Table 5.4 and Table 5.3 also support the qualitative impression that there is no improvement using VRS for  $R = 3$  for PI-CS TGV and VN reconstruction. In contrast, VRS outperforms regular sampling for  $R = 4$  in all coronal cases.

We illustrate results for individual scans with regular sampling of  $R = 4$  for a complete knee protocol, which contains various pathologies, taken from subjects ranging in age from 15 to 57, and anatomical variants, including a pediatric case. In particular, the coronal PDw scan (M32) depicted in Figure 5.18 shows osteoarthritis, most advanced within the lateral tibiofemoral compartment with associated marginal osteophyte formation, indicated by the green bracket. An extruded and torn medial meniscus, indicated by the green arrow, is visible in the coronal PDw-FS scan in Figure 5.20. Additionally, this patient (F57) has broad-based, full-thickness chondral loss within the medial compartment and a subchondral cystic change underlying the medial tibial plateau, as indicated by the green bracket. Further results for different orientations and contrasts are illustrated in Figure 5.22 for regular sampling with  $R = 4$  along with the error maps in Figure 5.23. The sagittal PDw scan illustrates a skeletally immature patient (F15) with almost completely fused tibial physes. A partial tear of the posterior cruciate ligament is visible in the sagittal T<sub>2</sub>w-FS scan M34. A full-thickness chondral defect centered in the medial femoral trochlea (green arrow) is visible on the axial T<sub>2</sub>w-FS scan (F45) on a background of patellofemoral osteoarthritis.

Data set	Method	Regular			Random		
		MSE	NRMSE	SSIM in %	MSE	NRMSE	SSIM in %
Coronal PDw	Zero Filling	17.69±4.05	0.17±0.02	81.61±2.09	17.18±3.96	0.17±0.02	81.00±2.15
	CG SENSE	2.25±0.55	0.12±0.03	90.16±2.15	4.82±1.02	0.15±0.03	85.45±2.45
	PI-CS TGV	1.13±0.30	0.08±0.02	93.85±1.58	1.88±0.43	0.09±0.02	91.22±1.67
	Learning	<b>0.88±0.21</b>	<b>0.07±0.02</b>	<b>94.89±1.45</b>	<b>1.27±0.29</b>	<b>0.08±0.02</b>	<b>93.61±1.48</b>
Coronal PDw-FS	Zero Filling	18.07±3.40	0.21±0.03	77.48±2.81	17.93±3.32	0.21±0.03	77.00±2.78
	CG SENSE	11.05±1.37	0.21±0.03	80.12±3.08	12.73±1.32	0.24±0.04	75.15±4.52
	PI-CS TGV	5.84±0.96	0.17±0.03	82.40±3.84	6.51±0.90	0.17±0.03	81.53±3.73
	Learning	<b>4.86±0.75</b>	<b>0.15±0.03</b>	<b>84.88±3.37</b>	<b>5.26±0.70</b>	<b>0.16±0.03</b>	<b>84.31±3.30</b>
Sagittal T <sub>2</sub> w-FS	Zero Filling	17.11±3.29	0.18±0.02	87.85±1.93	15.09±2.69	0.18±0.02	87.49±1.91
	CG SENSE	4.08±0.92	0.12±0.03	91.81±2.06	6.81±1.38	0.15±0.03	89.14±2.22
	PI-CS TGV	2.63±0.58	0.10±0.02	93.54±2.10	3.44±0.66	0.11±0.02	92.54±2.03
	Learning	<b>2.42±0.49</b>	<b>0.09±0.02</b>	<b>94.26±1.87</b>	<b>2.93±0.56</b>	<b>0.10±0.02</b>	<b>93.57±1.80</b>
Sagittal PDw	Zero Filling	3.61±0.54	0.09±0.01	90.50±1.57	2.94±0.45	0.08±0.01	90.83±1.58
	CG SENSE	0.45±0.09	0.05±0.02	95.42±1.31	0.81±0.13	0.06±0.02	93.46±1.35
	PI-CS TGV	0.31±0.06	<b>0.04±0.01</b>	97.23±1.01	0.47±0.08	<b>0.04±0.01</b>	96.45±1.08
	Learning	<b>0.29±0.05</b>	<b>0.04±0.01</b>	<b>97.53±0.95</b>	<b>0.40±0.07</b>	<b>0.04±0.01</b>	<b>96.89±0.98</b>
Axial T <sub>2</sub> w-FS	Zero Filling	37.46±8.64	0.25±0.01	82.52±1.61	36.05±7.84	0.25±0.01	81.64±1.77
	CG SENSE	15.59±2.85	0.19±0.02	87.41±1.89	20.24±3.40	0.23±0.02	82.59±2.67
	PI-CS TGV	10.32±2.98	0.16±0.03	88.23±2.87	11.81±2.79	0.18±0.03	87.18±2.79
	Learning	<b>7.92±1.93</b>	<b>0.14±0.02</b>	<b>90.78±2.05</b>	<b>9.06±1.90</b>	<b>0.15±0.02</b>	<b>89.82±2.09</b>

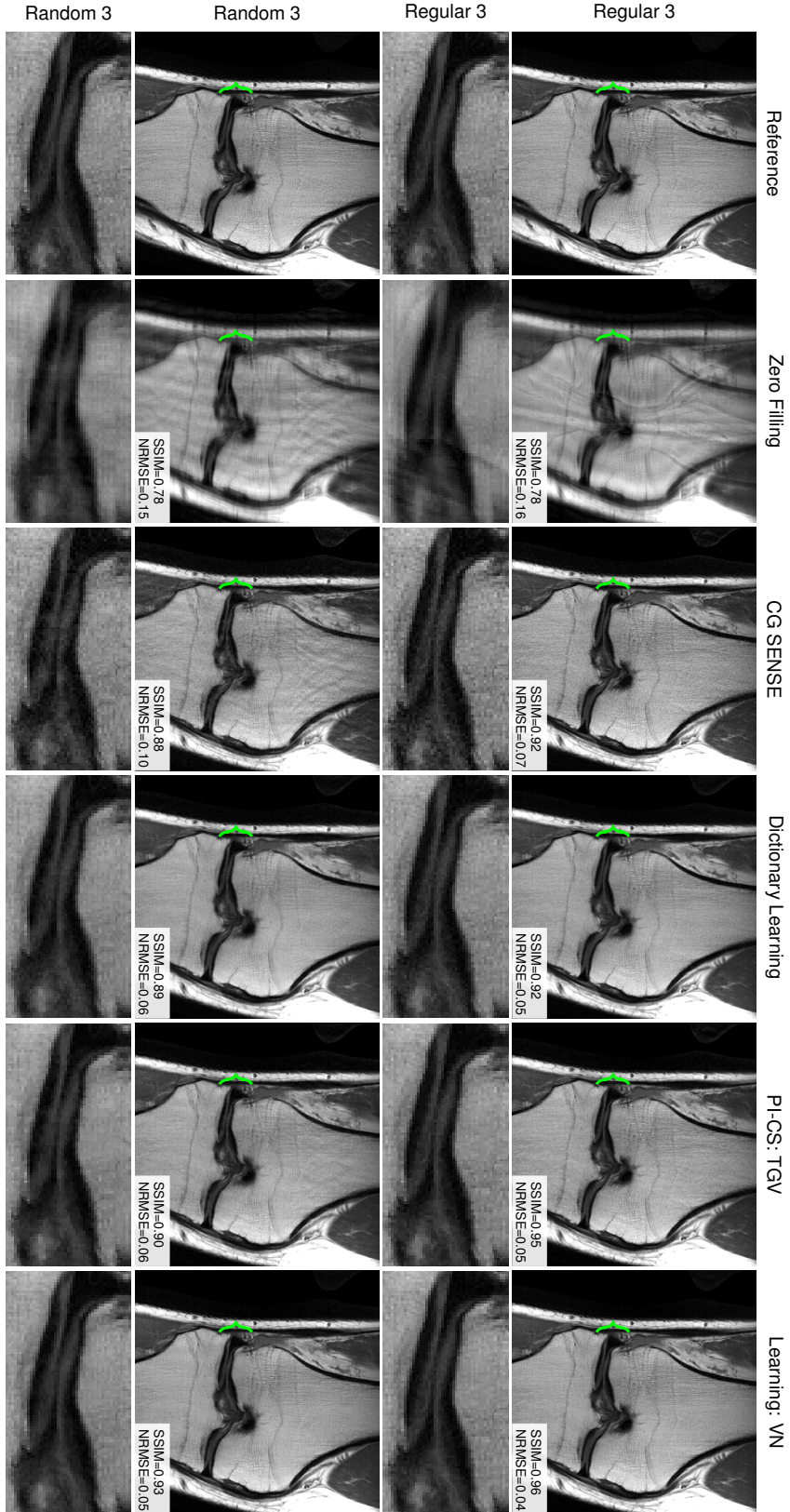
**Table 5.3:** Quantitative evaluation results in terms of MSE, NRMSE, and SSIM for a clinical knee protocol and acceleration factor  $R = 3$  for regular sampling and VRS.

Data set	Method	Regular			Random		
		MSE	NRMSE	SSIM in %	MSE	NRMSE	SSIM in %
Coronal PDw	Zero Filling	19.41±4.43	0.17±0.02	79.00±2.36	15.83±3.68	0.16±0.02	80.64±2.41
	CG SENSE	5.20±0.97	0.16±0.03	84.01±2.21	4.26±0.98	0.15±0.03	85.57±2.29
	PI-CS TGV	2.35±0.40	0.09±0.02	89.80±1.75	1.91±0.45	0.09±0.02	90.36±1.79
	Learning	<b>1.64±0.28</b>	<b>0.08±0.02</b>	<b>92.14±1.68</b>	<b>1.37±0.32</b>	<b>0.08±0.02</b>	<b>92.86±1.63</b>
Coronal PDw-FS	Zero Filling	20.71±4.07	0.23±0.03	73.96±3.04	17.69±3.30	0.22±0.03	75.10±3.17
	CG SENSE	14.55±1.62	0.25±0.05	73.06±4.62	11.79±1.39	0.24±0.04	74.78±4.55
	PI-CS TGV	7.73±1.14	0.19±0.04	79.19±4.14	7.07±1.07	0.18±0.03	79.69±4.09
	Learning	<b>6.49±0.80</b>	<b>0.17±0.03</b>	<b>81.97±3.60</b>	<b>5.81±0.85</b>	<b>0.17±0.03</b>	<b>82.47±3.67</b>
Sagittal T <sub>2</sub> w-FS	Zero Filling	16.66±3.14	0.19±0.03	85.71±2.62	17.35±3.21	0.19±0.03	84.91±2.59
	CG SENSE	6.27±1.62	0.15±0.04	87.86±3.08	9.55±2.11	0.18±0.04	85.06±3.11
	PI-CS TGV	3.39±0.82	<b>0.11±0.03</b>	91.84±2.81	4.76±0.95	0.13±0.03	90.29±2.70
	Learning	<b>2.99±0.68</b>	<b>0.11±0.03</b>	<b>92.83±2.40</b>	<b>3.92±0.81</b>	<b>0.12±0.03</b>	<b>91.85±2.35</b>
Sagittal PDw	Zero Filling	5.17±0.75	0.11±0.01	87.53±1.95	3.32±0.51	0.09±0.01	89.49±1.80
	CG SENSE	0.86±0.15	0.06±0.02	92.74±1.46	1.03±0.16	0.07±0.02	92.37±1.48
	PI-CS TGV	0.49±0.09	0.05±0.01	96.22±1.17	0.64±0.11	<b>0.05±0.01</b>	95.47±1.24
	Learning	<b>0.44±0.07</b>	<b>0.04±0.01</b>	<b>96.64±1.16</b>	<b>0.52±0.09</b>	<b>0.05±0.01</b>	<b>96.07±1.17</b>
Axial T <sub>2</sub> w-FS	Zero Filling	44.57±9.95	0.27±0.02	78.52±1.92	48.03±11.13	0.28±0.02	77.80±1.98
	CG SENSE	23.75±4.56	0.24±0.03	80.30±3.20	31.98±4.88	0.27±0.02	78.87±2.43
	PI-CS TGV	13.65±3.78	0.18±0.03	85.51±3.25	15.30±2.57	0.19±0.02	84.93±2.60
	Learning	<b>10.63±2.48</b>	<b>0.16±0.02</b>	<b>88.46±2.43</b>	<b>12.06±2.13</b>	<b>0.17±0.02</b>	<b>87.74±2.30</b>

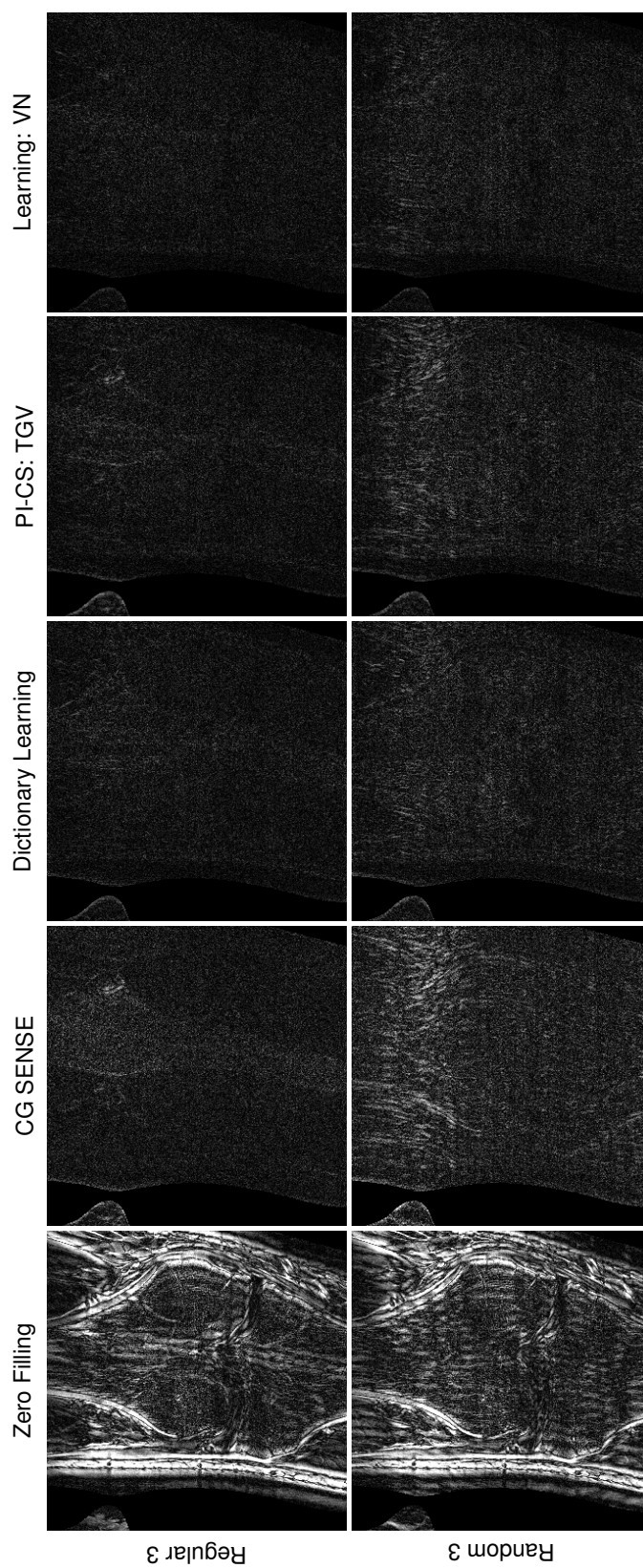
**Table 5.4:** Quantitative evaluation results in terms of MSE, NRMSE, and SSIM for a clinical knee protocol and acceleration factor  $R = 4$  for regular sampling and VRS.

Dataset	Criterion	Reader scores regular		
		PI-CS TGV	Learning	p-value
Coronal PDw	Artifact	3.60±0.57	1.65±0.07	<b>0.0010</b>
	Sharpness/Blur	2.90±0.14	2.15±0.07	<b>0.0234</b>
	SNR	2.60±0.28	1.45±0.21	<b>0.0078</b>
	Overall image quality	3.30±0.14	2.05±0.21	<b>0.0010</b>
Coronal PDw-FS	Artifact	3.95±0.07	2.90±0.42	<b>0.0020</b>
	Sharpness/Blur	3.95±0.07	3.15±0.64	<b>0.0020</b>
	SNR	3.75±0.21	2.90±0.71	<b>0.0049</b>
	Overall image quality	3.95±0.07	3.20±0.57	<b>0.0020</b>
Sagittal T <sub>2</sub> w-FS	Artifact	2.90±0.14	2.80±0.28	0.3750
	Sharpness/Blur	3.40±0.14	2.75±0.21	<b>0.0156</b>
	SNR	3.20±0.28	2.50±0.28	<b>0.0234</b>
	Overall image quality	3.30±0.28	2.75±0.07	<b>0.0078</b>
Sagittal PDw	Artifact	2.10±0.14	2.00±0.14	0.4063
	Sharpness/Blur	2.10±0.14	2.10±0.14	0.6875
	SNR	1.60±0.00	1.50±0.28	0.3828
	Overall image quality	2.20±0.14	2.05±0.07	0.2656
Axial T <sub>2</sub> w-FS	Artifact	3.15±0.07	3.10±0.57	0.5000
	Sharpness/Blur	3.05±0.07	2.95±0.49	0.3750
	SNR	3.10±0.14	2.75±0.49	<b>0.0313</b>
	Overall image quality	3.20±0.14	3.05±0.49	0.2266

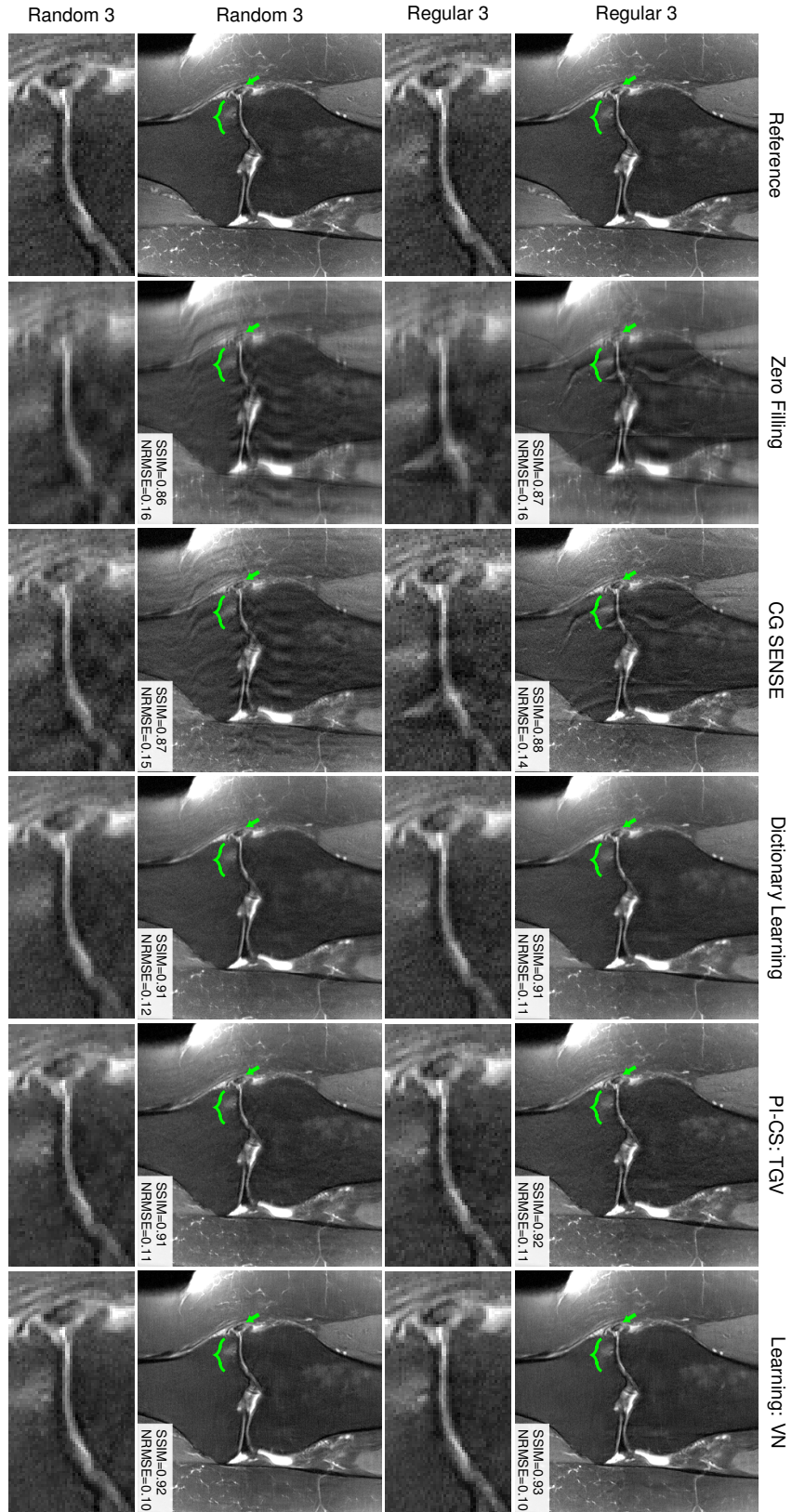
**Table 5.5:** Image quality reader scores for a clinical knee protocol and acceleration factor  $R = 4$  for regular sampling. For the reader scores, we depict the mean values and standard deviations averaged over both readers along with the p-value obtained by the one-sided Wilcoxon signed-rank test. Values that accept the alternative hypothesis with a significance level  $\alpha = 0.05$ , that VN reconstructions have a better quality score, are marked as bold.



**Figure 5.14:** Coronal PDw scan with acceleration  $R = 3$  of a 32-year-old male. The green bracket indicates osteoarthritis. The first and second row depict reconstruction results for regular Cartesian sampling, the third and fourth row depict the same for Quantitative evaluation results in terms of MSE, NRMSE, and SSIM for a clinical knee protocol and acceleration factor  $R = 4$  for regular sampling and VRS. Zoomed views show that the learned VN reconstruction appears slightly sharper than the PI-CS TGV reconstruction. Although dictionary learning can handle artifacts better than PI-CS TGV and produce a visually more appealing results, the quantitative values are slightly worse. For regular sampling, the results illustrate that the VN reconstruction can suppress undersampling artifacts better than CG SENSE and PI-CS TGV, and works on similar lines with dictionary learning. Quantitative values for VN outperform the reference methods for this specific slice. For this acceleration factor of  $R = 3$ , the results based on random sampling appear slightly blurrier than the results based on regular sampling.

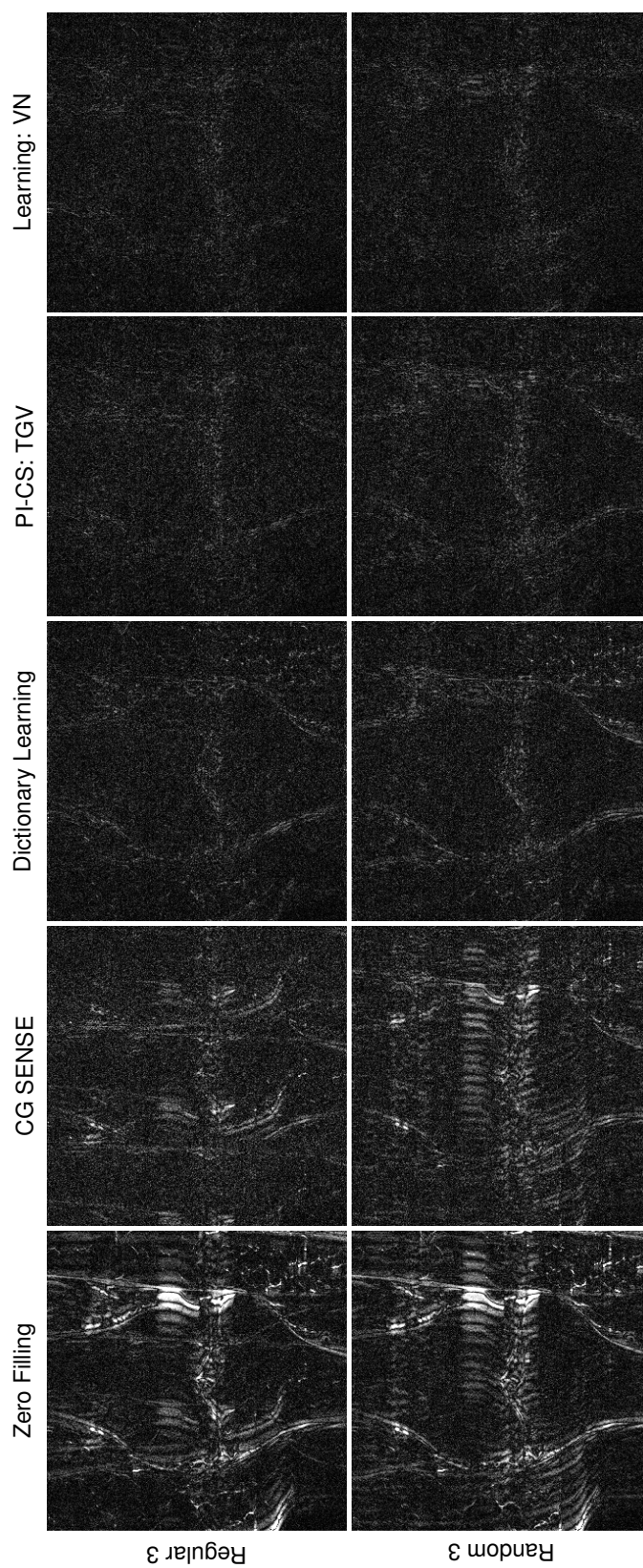


**Figure 5.15:** Difference images to reference image for the reconstructed coronal PDw scans with acceleration  $R = 3$  presented in Figure 5.14. The VN reconstructions show the least error compared to the other methods.

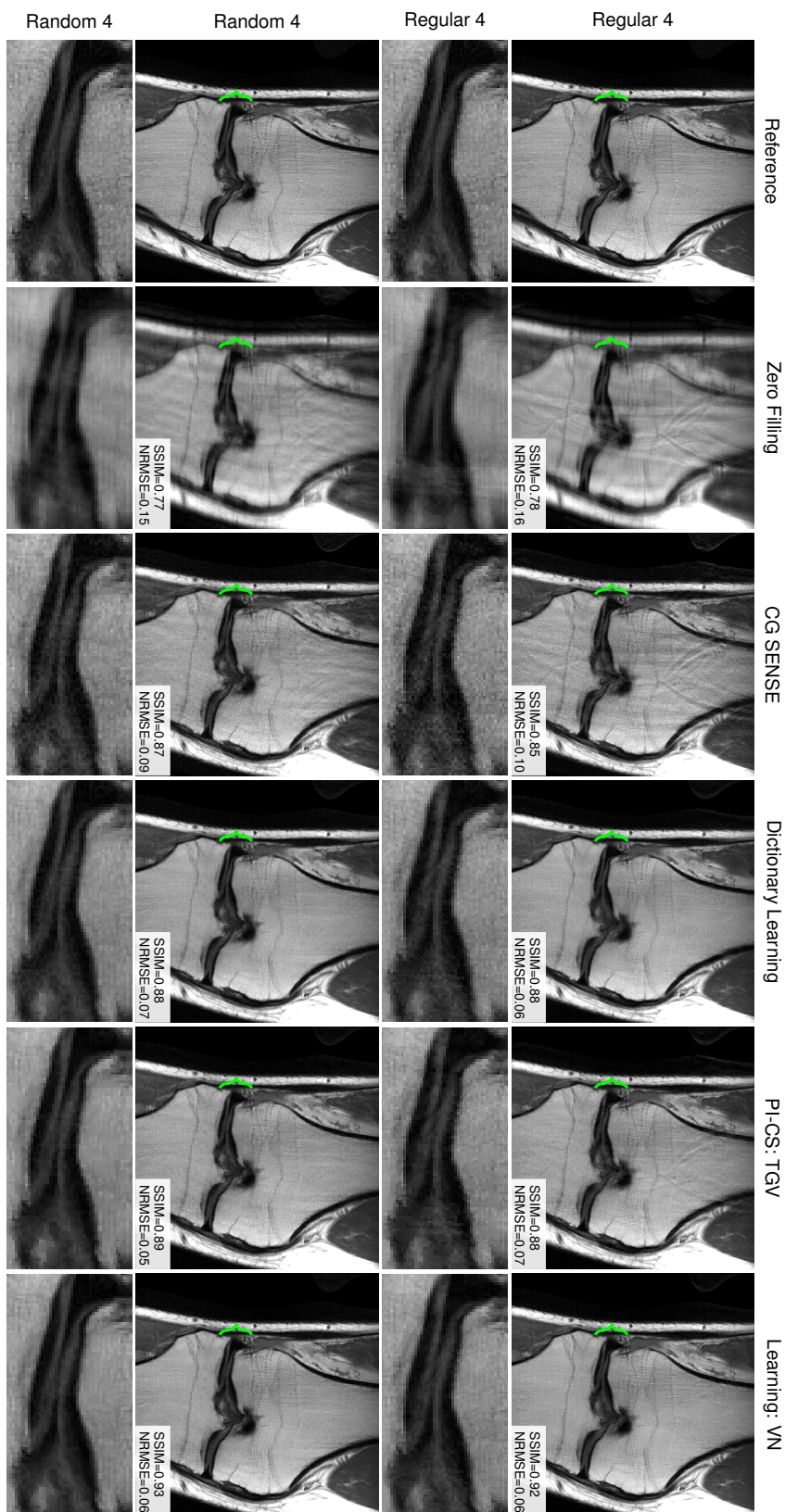


**Figure 5.16:** Coronal PDw-FS scan with acceleration  $R = 3$  of a 57-year-old female. The green bracket indicates broad-based, full-thickness chondral loss and a subchondral cystic change. The green arrow depicts an extruded and torn medial meniscus. The first and second row depict reconstruction results for regular Cartesian sampling, the third and fourth row depict the same for VRS. The zoomed views show that the learned VN reconstruction appears sharper than the PLCS TGV and dictionary learning reconstruction. For regular sampling, the results illustrate that the VN reconstruction can suppress undersampling artifacts better. Again, results based on random sampling appear slightly blurrier than the results based on regular sampling.

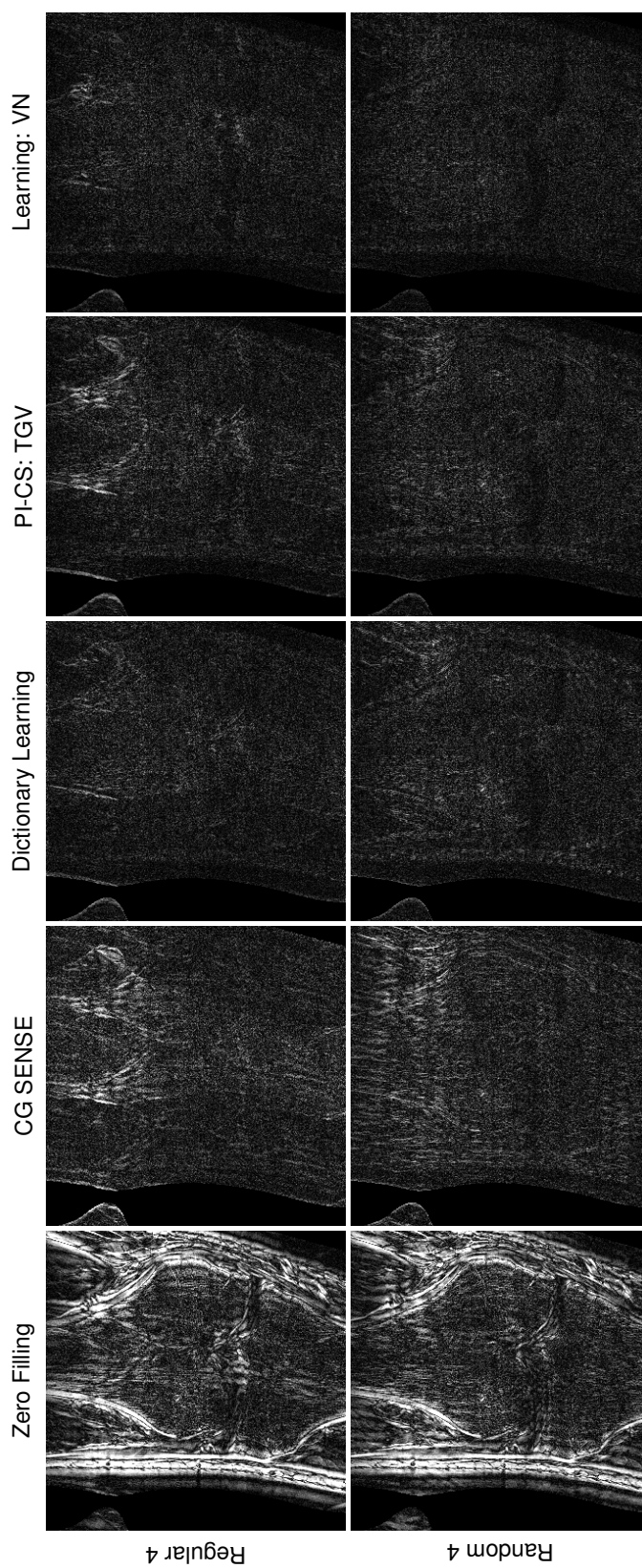




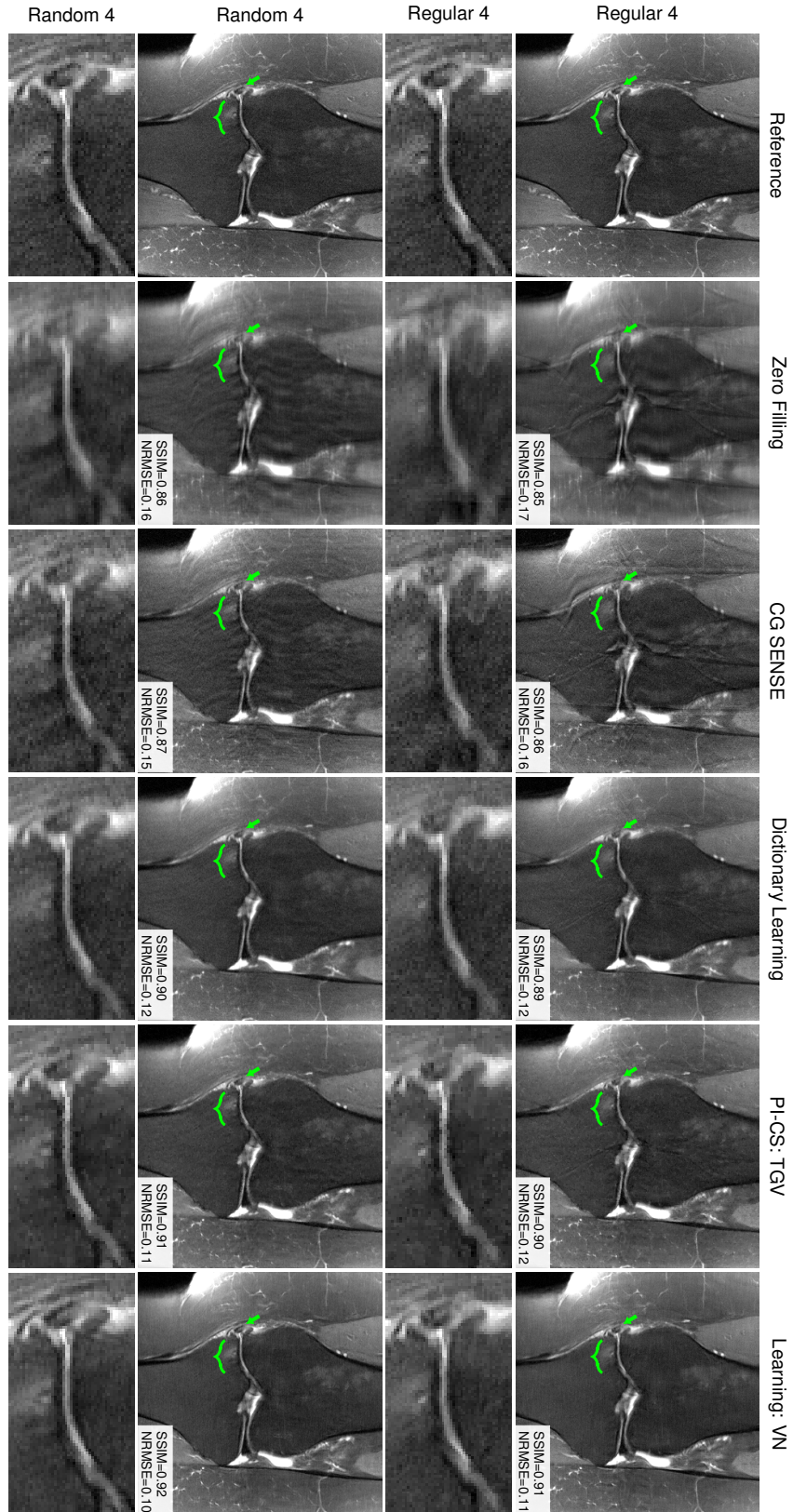
**Figure 5.17:** Difference images to reference image for the reconstructed coronal PDw-FS scans with acceleration  $R = 3$  presented in Figure 5.16. We observe large errors at boundaries for dictionary learning. The VN reconstructions show the least error compared to the other methods.



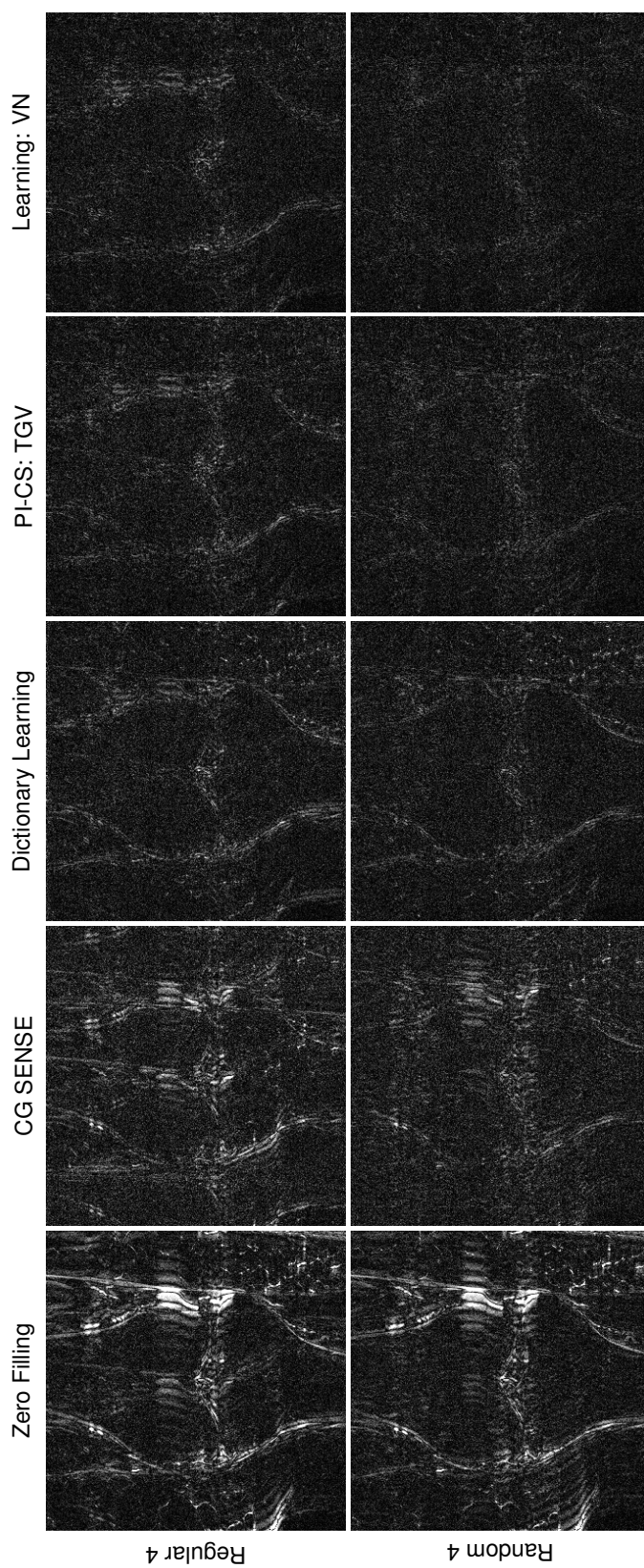
**Figure 5.18:** Coronal PDw scan with acceleration  $R = 4$  of a 32-year-old male. The green bracket indicates osteoarthritis. The first and third row depict reconstruction results for regular Cartesian sampling, the second and fourth row depict the same for VRS. Zoomed views show that the learned VN reconstruction appears slightly sharper than the PI-CS TGV and dictionary learning reconstruction. The dictionary learning and VN reconstruction can significantly suppress artifacts unlike CG SENSE and PI-CS TGV. Results based on random sampling show reduced residual artifacts and slightly increased sharpness in comparison to regular sampling.



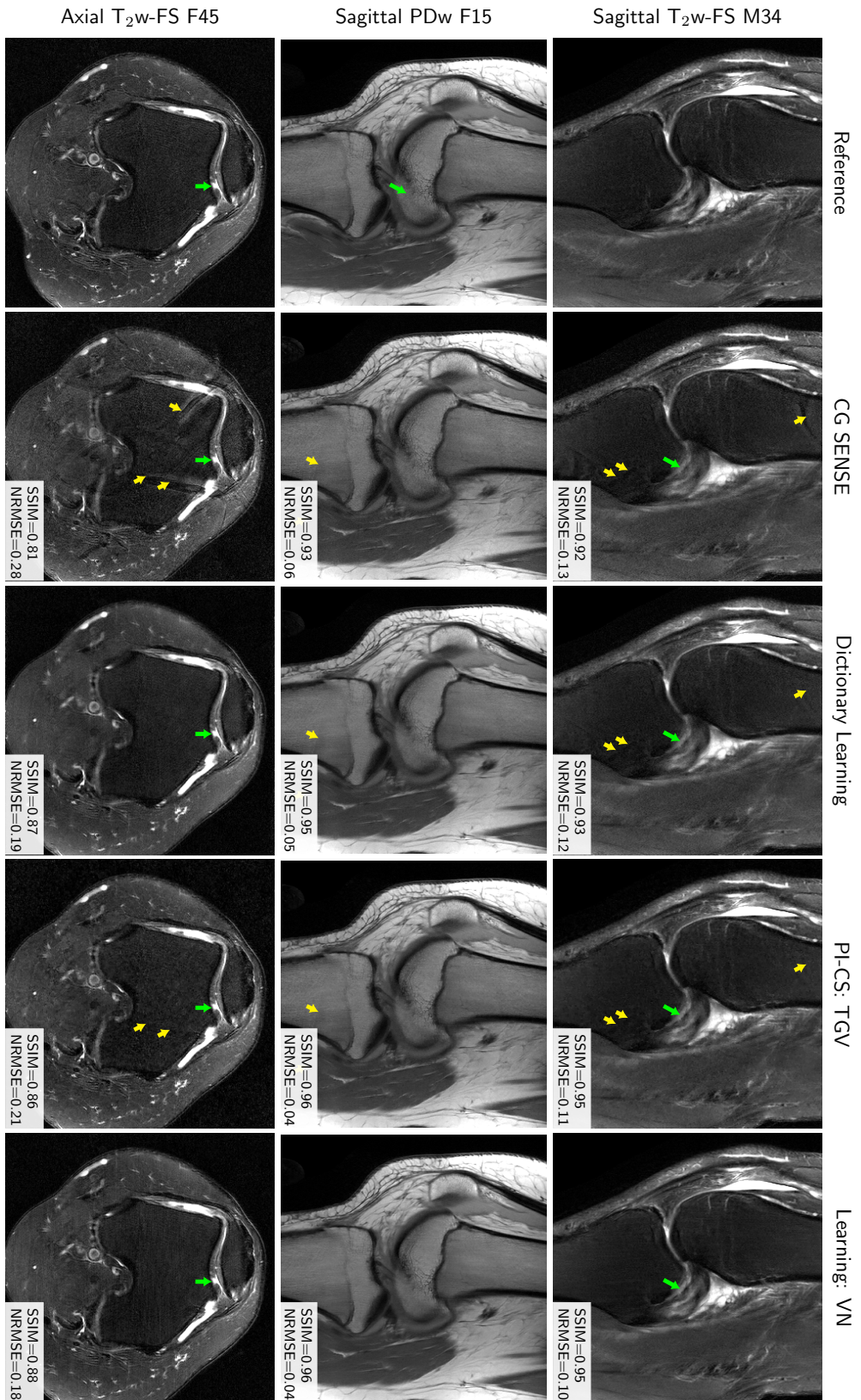
**Figure 5.19:** Difference images to reference image for the reconstructed coronal PDw scans with acceleration  $R = 4$  presented in Figure 5.18. The undersampling artifacts can be clearly observed in the CG SENSE and zero filling results. While TGV has a remaining undersampling artifact for regular sampling, the dictionary learning method can suppress this artifact. However, we observe larger errors at object boundaries in the dictionary learning results. The VN result has the least error compared to the reference methods.



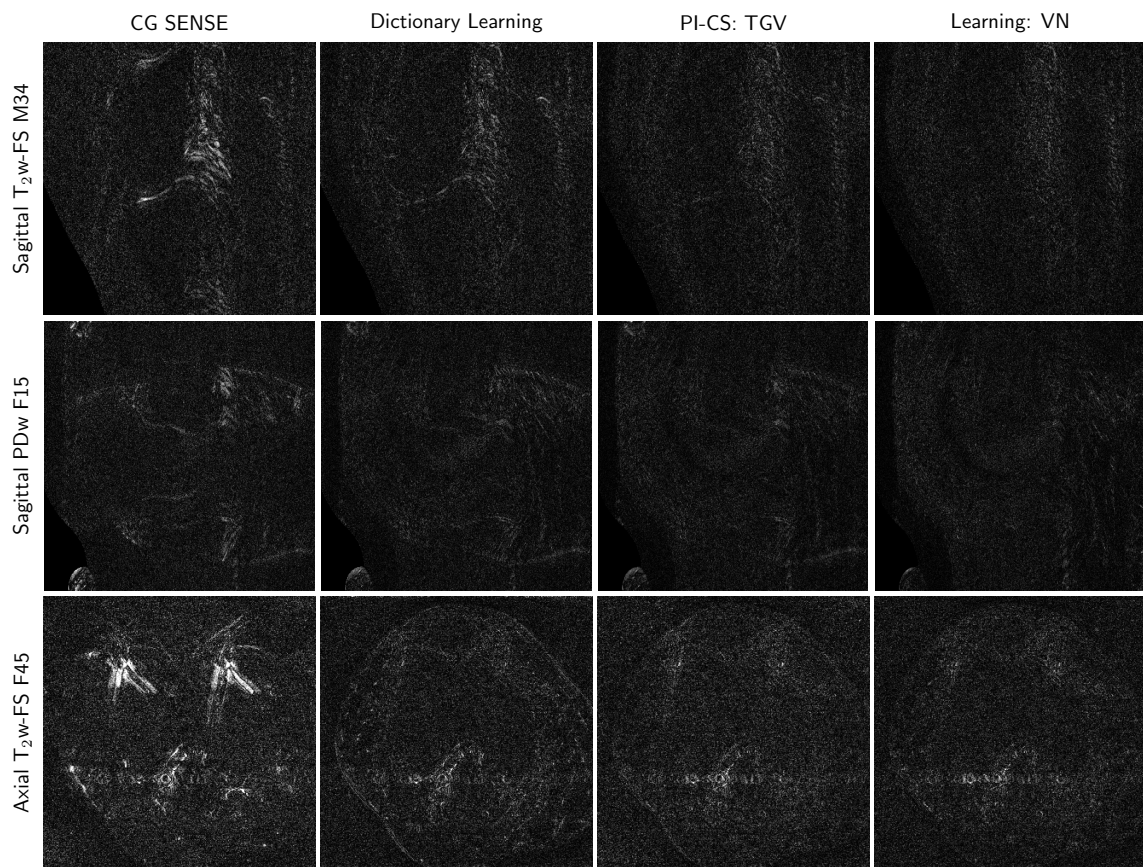
**Figure 5.20:** Coronal PDw-FS scan with acceleration  $R = 4$  of a 57-year-old female. The green bracket indicates broad-based, full-thickness chondral loss and a subchondral cystic change. The green arrow depicts an extruded and torn medial meniscus. The first and second row depict reconstruction results for regular Cartesian sampling, the third and fourth row depict the same for VRS. The zoomed views show that the learned VN reconstruction appears sharper than the PI-CS TGV and dictionary learning reconstruction. The VN reconstruction shows reduced artifacts compared to the other methods. Results based on random sampling show reduced residual artifacts and appear sharper than the results based on regular sampling.



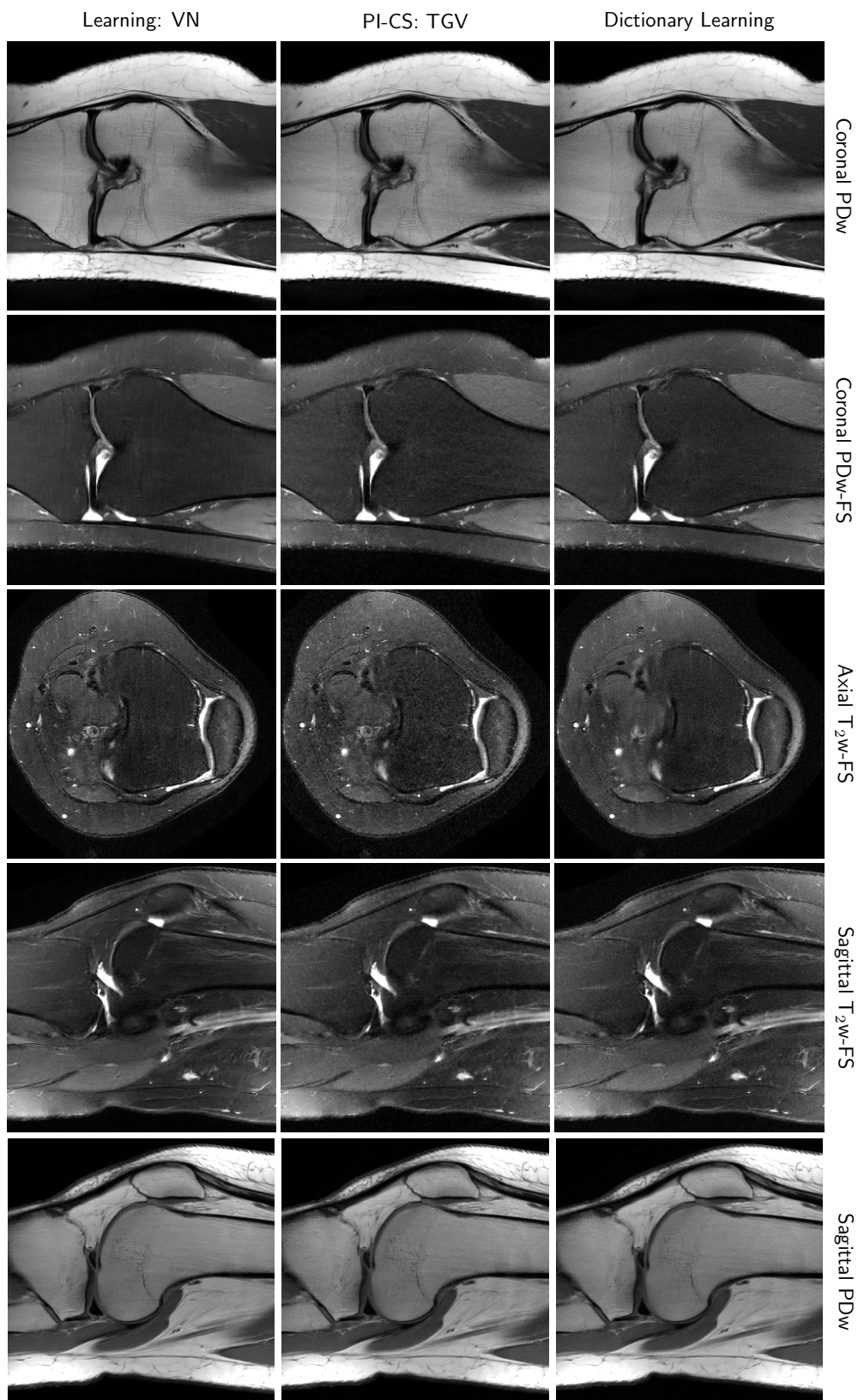
**Figure 5.21:** Difference images to reference image for the reconstructed coronal PDw-FS scans with acceleration  $R = 4$  presented in Figure 5.20. The undersampling artifacts can be clearly observed in the CG SENSE and zero filling results. Both TGV and dictionary learning have residual undersampling artifact for regular sampling. We observe larger errors at object boundaries in the dictionary learning results. The VN result has the least error compared to the reference methods and is able to suppress the undersampling artifacts.



**Figure 5.22:** Reconstruction results for sagittal T<sub>2</sub>w-FS, sagittal PDw and axial T<sub>2</sub>w-FS sequences of a complete knee protocol for acceleration factor  $R = 4$  with regular undersampling. Each sequence here is illustrated with results from a different patient, identified by gender and age (e.g., M50 indicates a 50-year-old male). Pathological cases and a pediatric case are shown for both male and female patients of various ages. Green arrows and brackets indicate pathologies. Yellow arrows show residual artifacts that are visible in the different reconstructions, but not in the learned VN reconstructions.

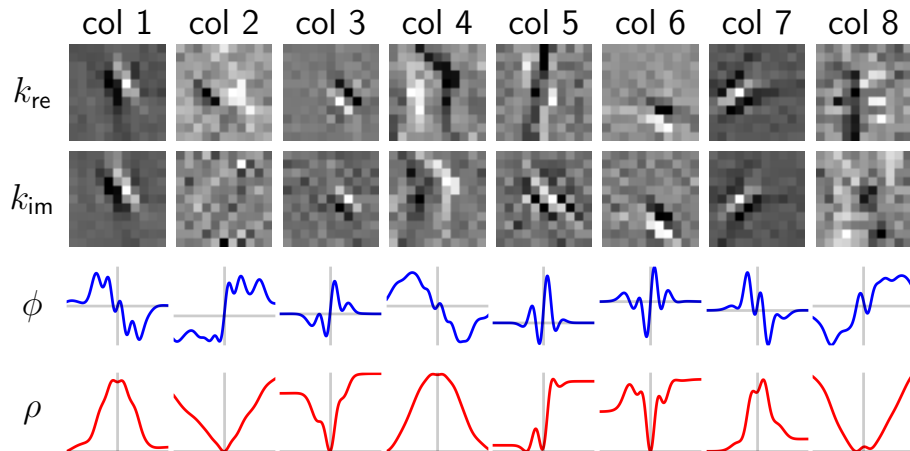


**Figure 5.23:** Difference images for sagittal  $T_2w$ -FS, sagittal PDw and axial  $T_2w$ -FS sequences of a complete knee protocol presented in Figure 5.22.



**Figure 5.24:** Reconstruction results of prospectively undersampled data for regular sampling  $R = 4$ . We show reconstruction results for dictionary learning, PI-CS TGV and our VN for a whole knee protocol of a F27 volunteer. We observe a similar behavior as for the retrospective undersampled data. Dictionary learning and PI-CS TGV perform reasonably well for non-FS scans. While the FS scans appear artificial with a PI-CS TGV reconstruction, we observe a noise pattern in the dictionary learning results, most prominent in the sagittal T<sub>2w</sub>-FS scan. Dictionary learning appears slightly blurrier, which is best seen in the axial slice. The VN reconstructions have less undersampling artifacts and an improved SNR.





**Figure 5.25:** Examples of learned parameters of the VN. Filter kernels for the real  $k_{re}$  and imaginary  $k_{im}$  plane as well as their corresponding activation  $\phi'$  and potential function  $\rho$  are shown. The potential function  $\rho$  was obtained by integrating the activation function  $\phi'$  including an additional integration constant.

The presence of these particular variations, which were not included in the training data set, does not negatively affect the learned VN reconstruction. The reduction of residual aliasing artifacts, marked by yellow arrows, the reduced noise level, and the overall improved image quality lead to an improved depiction of the pathologies when compared to the reference methods. Again, the quality improvement of the learned VN is supported by the quantitative analysis of similarity measures depicted in Table 5.3 and Table 5.4.

**Prospective Variational Network Reconstructions.** The reconstruction results of prospectively undersampled data for regular sampling and acceleration  $R = 4$  are depicted in Figure 5.24. While PI-CS TGV and dictionary learning perform reasonably well for non-FS scans, a noise pattern can be observed in certain regions for dictionary learning and blocky appearance for PI-CS TGV. Our VN reconstructions are more homogeneous and less prone to remaining artifacts.

**Reader Study.** The average scores of the readers together with the p-values of the Wilcoxon signed-rank test are listed in Table 5.5. The mean values of the reader scores indicate that all VN reconstructions have equal or better scores than the PI-CS TGV reconstructions. P-values indicate that the null hypothesis is rejected for most of the sequences for the given significance level  $\alpha$ . Coronal as well as sagittal  $T_2$  VN reconstructions have significantly better image quality than PI-CS TGV. The difference between the individual reconstruction methods for the sagittal PDw case is not significant, which is already obvious in the negligible difference of the qualitative results and quantitative results for this sequence. No significant difference in image quality, except SNR, can be observed for the axial  $T_2w$ -FS scans.

**Variational Network Parameters.** Examples of learned filter kernel pairs for real and imaginary feature planes are plotted along with their corresponding activation and potential functions in Figure 5.25. The potential functions are computed by integrating the learned activation functions, and they can be linked directly to the norms that are used in the regularization terms of traditional CS algorithms. We observe that some are very close to the convex  $\ell_1$  norm used in CS (e.g., the function in the 2nd column), but we can also observe substantial deviations. We can identify functions with log-student-t characteristics and concave functions. Some of the learned filter pairs have the same structure in both the real and imaginary plane while some of them seem to be inverted in the real and imaginary part.

### 5.3.3 Discussion

While deep learning has resulted in clear breakthroughs in Computer Vision, the application of deep learning to medical image reconstruction is just beginning [246]. Initial results for our deep learning image reconstruction approach presented in detail here were first presented at the Annual Meeting of the International Society for Magnetic Resonance in Medicine in May of 2016 [88]. Early attempts to use machine learning for MRI reconstruction were based on dictionary learning [26, 195, 198]. The key difference to our VN approach is that they learn a reconstruction online as a combination of dictionary elements directly from undersampled data, hence, no reference data is required. Although the learned dictionary might be reused, a new optimization problem has to be performed for every new reconstruction, which is computationally demanding. While dictionary learning methods act on patches, which need to be properly combined, and do not involve non-linearities in the combination of dictionary elements, our proposed VN approach directly reconstructs the whole images and learns non-linearities, which are important to enhance or suppress certain filter responses. Wang et al. [244] showed first results using a Convolutional Neural Network (CNN) architecture to define a relationship between zero filled solution and high-quality images based on pseudo-random sampling. The learned network can then be used as regularization in a non-linear reconstruction algorithm.

Yang et al. [263] introduced a network architecture that is based on unrolling the Alternating Direction Method of Multipliers algorithm. They proposed to learn all parameters including image transforms and shrinkage functions for CS-based MRI. Han et al. [96] learned destreaking on Computed Tomography (CT) images and then fine-tuned the learning on MR data to remove streaking from radially undersampled  $k$ -space data. All three approaches used single-coil data, and it remains unclear how they deal with the complex domain of MR images. Kwon et al. [145] introduced a neural network architecture to estimate the unfolding of multi-coil Cartesian undersampled data. Similar to a classic SENSE reconstruction [190], unfolding is performed line-by-line. This restricts the applicability to a fixed matrix size and a particular 1D undersampling pattern. Most recently, Lee et al. [151] used residual learning to train two CNNs to estimate the magnitude and phase images of Cartesian undersampled data.

In this work, we present the first learning-based MRI reconstruction approach for clinical multi-coil data. Our VN architecture combines useful properties of two successful fields: vari-

ational methods and deep learning. We formulate image reconstruction as a variational model and embed this model in a gradient descent scheme, which forms the specific VN structure. The VN was first introduced as a trainable reaction-diffusion model [46] with application to classic image processing tasks [46, 130, 267]. All these tasks are similar in the sense that the data are corrupted by unstructured noise in the image domain. MR image reconstruction presents several substantial differences: complex-valued multi-coil data are acquired in the Fourier domain and transformed into the image domain. This involves the use of coil sensitivity maps and causes distinct artifacts related to the sampling pattern. For our MR image reconstruction task, the optimal design of the VN, such as the number of stages, the number of filters per stage and the kernel size, is currently an open question. Our particular design choice is based on preliminary experiments [88] and, in line with the experiments presented here, delivered consistent results for a wide range of experimental conditions. We also found that the performance of our VN was stable when varying the design of the architecture. In practice, the design of the network is essentially a trade-off between model complexity and training efficiency. For example, the number of RBFs that are used to model the activation functions in a smoothed function approximation, defines the flexibility to approximate arbitrary functions in an accurate way. In our experimental setup as well as in the latest studies on image processing tasks [140], we reduced the number of RBFs compared to the initial work [46] by a half without a loss in performance but with reduced training time.

Our VN structure allows us to visualize the learned parameters, which is non-trivial for classical CNNs [269]. The potential functions are computed by integrating the learned activation functions, and they can be linked directly to the norms that are used in the regularization terms of traditional CS algorithms. Some of the learned filter pairs have the same structure in both the real and imaginary plane while some of them seem to be inverted in the real and imaginary part. In general, the filters in both the real and imaginary part represent different (higher-order) derivative filters of various scales and orientations, similar to Gabor filters [54, 68]. Handcrafted Gabor filters have been successfully used in image processing [115], and learning-based approaches [142] report similar filters. It has also been shown that these types of filters have a strong relation to the human perceptual system [180].

Some of the learned potential functions in Figure 5.25 are very close to the convex  $\ell_1$  norm used in CS (e.g., the function in the 3rd column), but we can also observe substantial deviations. We can identify functions with log-student-t characteristics also used in [206]. Indeed, non-convex functions of log-student-t type introduce more sparsity than, e.g., the convex  $\ell_1$ -norm and are reported to fit the statistics of natural images better than the  $\ell_1$ -norm [119]. Potential functions like those in columns 1, 4 and 7 have been associated with image sharpening in the literature [278].

Designing filters and functions is not a trivial task. Using learning-based approaches provides a way to tune these parameters such that they are adapted to specific types of image features and artifact properties. The strength of our algorithm are the trainable activation functions which stands in contrast to other deep learning approaches that use fixed activation functions such as Rectified Linear Units or sigmoid functions. Hence, instead of adding more

and more layers and creating deeper networks, we introduce more structure and flexibility in the individual layers, which might help to reduce the overall complexity of the network. As shown in [140] for image denoising and non-blind deblurring, fixing the activation functions to less flexible, e.g., convex, functions might also lead to a decrease in performance for our application.

Compared to convex  $\ell_1$  minimization where we can understand the characteristics and artifacts of hand-crafted filters and potential functions, learning-based methods are often considered to be black-boxes, which are difficult to interpret. While we cannot claim insight into the properties of the model and the resulting images to the same degree of a simpler model like TV, one of the key strengths of our proposed VN is the motivation by a generalized, trainable variational model. To gain an understanding of what the VN learns, we first inspect the intermediate outputs of the gradient descent steps of our VN. We observe successive low-pass and high-pass filtering, and note that the prevalence of undersampling artifacts decreases after each single iteration. A continuous improvement over the iterations does not occur because our training is designed such that the result after the last gradient step is optimal in terms of the error metric chosen for evaluation. Although it would be possible to train the VN for progressive improvement, this would reduce the flexibility of the algorithm for adjusting the learned parameters during the training procedure.

In any iterative CS approach, every reconstruction is handled as an individual optimization problem. This is a fundamental difference to our proposed data-driven VN. In our VN approach, we perform the computationally expensive optimization as an offline pre-computation step to learn a set of parameters for a small fixed number of iterations. In our experiments, one training took approximately four days for on a single graphics card. Once the VN is trained, the application to new data is extremely efficient, because no new optimization problem has to be solved and no additional parameters have to be selected. In our experiments, the VN reconstruction took only 193 ms for one slice. In comparison, the reconstruction time for zero filling was 11 ms, for CG SENSE with 6 iterations 75 ms and for PI-CS TGV with 1000 primal-dual iterations [132] 11.73 s on average. Thus, the online VN reconstruction using the learned parameters for the fixed number of iterations does not affect the hard time constraints during a patient exam. However, PI-CS TGV requires hundreds of iterations to obtain a convergent solution, 1000 iterations were used for grid search to guarantee a converged solution.

Our VN is individually trained for different sampling patterns, reflected in the forward and adjoint operators. We do not learn a global mapping between undersampled  $k$ -space and reconstruction, but how to enhance local structures, while ensuring consistency to the acquired  $k$ -space data. First results towards learning a general regularizer, that could be applied for any sampling pattern, were recently presented at the annual meeting of ISMRM in 2017 [87]: We showed that a network trained for regular sampling patterns can be used for reconstruction of randomly sampled data, but a network trained for randomly sampled data is not capable of removing coherent undersampling artifacts, which indicates that the dependency of sampling patterns is required to train the regularizer. However, the systematic performance evaluation for a wide range of sampling patterns is beyond the scope of this particular manuscript, and

will be the target of future work. We will not only explore joint training of various sampling patterns, acceleration factors and sequences, but also the application of VN reconstruction to non-Cartesian sampling, dynamic and multi-parametric data.

The reconstruction quality of all methods does not only rely on the sampling pattern, but also on other parameters. Larger filter sizes, such as the  $11 \times 11$  filters used in our VN architecture, provide the possibility to capture more efficiently the characteristic backfolding artifacts of Cartesian undersampled data, which are spread over several pixels. This stands in contrast to models like TV or TGV that are based on gradient filters in a small neighborhood (e.g., only forward differences in the  $x$  and  $y$  direction are considered). To suppress artifacts with PI-CS TGV, the regularization parameters must be chosen in such a way that the remaining image appears over-smoothed, and fine details are lost. Even though the piece-wise affine prior model of TGV is more complex than the piece-wise constant prior model of TV, the images appear artificial, especially if MR images with low SNR are reconstructed. Dictionary learning involves also larger filter kernels and works reasonably well for data with high SNR, reconstructions of low SNR data contain lots of noisy regions and blurry edges.

The image quality reader study confirms our quantitative and qualitative observations for regular sampling of  $R = 4$ . In general, the image quality of the FS sequences was rated lower than for the non-FS sequences for both VN and PI-CS TGV. The difference between the two types of sequences is the baseline SNR, which is much lower for the FS sequences. It is well known that in all CS-based methods, the best performance can be achieved in the case of a high baseline SNR and incoherent artifacts. The presented experiments demonstrate that if the corruption of the reconstructed images is dominated by noise, performance of both CS and VN reconstruction drops. If the baseline SNR drops to a level where the noise has a higher impact than aliasing artifacts, the VN concentrates on denoising instead of undersampling artifact removal. In addition, some of our results show residual artifacts, most prominent in the axial sequences. The source of these artifacts is residual aliasing and Gibbs' ringing. These residual methods are present in all our reconstructions and not unique for our VN.

While radiologists learn throughout their career to distinguish certain patterns in images such as artifacts, we have to reflect the quality of learning in our presented approach by not only choosing the right architecture but also a suitable similarity measure. As demonstrated by our evaluation, quantitative scores are not always on par with image quality readings by radiologists. The used MSE for training compares pixel-wise differences and is likely not optimal for representing similarity to artifact-free reference reconstructions. Future investigations will also involve the choice of different error metrics or the investigation of Generative Adversarial Networks (GANs) [74] for training.

## 5.4 Exploring the Generalization Potential of VNs for Accelerated MR Imaging

MRI offers a great flexibility during data acquisition with respect to numerous acquisition parameters that can be adapted individually for each acquired sequence. Furthermore, different institutions have different protocols with varying imaging parameters for the same clinical exam. Therefore, one of the major questions regarding the success of learning-based approaches in practice is how the trained architectures generalize with respect to changes in the data, because re-training a network for each individual protocol is simply infeasible in clinical practice. Hence, we want to further explore open questions such as how many samples are needed to train a network or how sensitive is the learned network to changes in the input data. We focus in this section on the influence of the number of training samples and the influence with respect to baseline SNR and present results for an increased patient population compared to [136]. Further experiments on the influence of image contrast and SNR along with results on transfer learning can be found in [136].

### 5.4.1 Methods

For the following experiments, we use the same VN architecture as presented in Section 5.3.1 with  $N_k = 24$  filter kernels and additional batch normalization as outlined in Section 3.2, where the scaling parameter is learned. The network is trained with the Proximal ADAM optimizer with block-preconditioning (see Appendix C.2) for 300 epochs using a batch size of 10. The batch normalization update is turned after 50 epochs. The learning rate is set to  $4e-3$  and exponential decay (decay rate=0.75, decay steps=30 epochs) is used to decrease the learning rate during optimization, along with the ADAM parameters  $\beta_1 = 0.9, \beta_2 = 0.999$ . Note here that the maximum number of iterations for the optimizer depends on the number of samples. As we fix the number of epochs, we guarantee that each sample is seen equally often during training, independent of the sample size.

As training and test data we use a subset of the fully sampled raw data from the fastmri dataset [268] which were acquired using the same hardware and matrix-size parameters as our coronal PDw and PDw-FS knee data (see Table 5.2). These data do not only differ in image contrast but also in the baseline SNR as discussed in Section 5.3. Training is performed on the central 20 slices of each case and data is normalized with respect to the maximal value in the zero-filled solution. The data is retrospectively undersampled using a regular Cartesian sampling scheme with acceleration factor  $R = 4$ . The trainings are evaluated jointly for all datasets as well as individually for both datasets using the NRMSE, Peak Signal-To-Noise Ratio (PSNR) and SSIM. The following trainings are performed to study the influence of the number of training cases  $n$  and the heterogeneity of the training data:

1. Coronal PDw,  $n = 10$  as studied in Section 5.3
2. Coronal PDw,  $n = 140$

Training Data	Test Data								
	Coronal PDw ( $n = 20$ )			Coronal PDw-FS ( $n = 20$ )			Joint coronal ( $n = 40$ )		
	NRMSE	PSNR	SSIM	NRMSE	PSNR	SSIM	NRMSE	PSNR	SSIM
Coronal PDw ( $n = 10$ )	0.0530±0.0059	41.94±1.35	0.9655±0.0069	0.1604±0.0245	36.07±2.26	0.8631±0.0478	0.1067±0.0565	39.01±3.48	0.9143±0.0616
Coronal PDw ( $n = 140$ )	<b>0.0498±0.0057</b>	<b>42.50±1.33</b>	<b>0.9688±0.0066</b>	0.1518±0.0234	36.48±2.25	0.8730±0.0451	0.1008±0.0538	39.49±3.53	0.9209±0.0578
Coronal PDw-FS ( $n = 10$ )	0.0599±0.0051	40.23±1.33	0.9530±0.0088	0.1075±0.0110	39.02±1.75	0.9247±0.0217	0.0837±0.0253	39.63±1.67	0.9389±0.0218
Coronal PDw-FS ( $n = 140$ )	0.0551±0.0049	41.12±1.33	0.9611±0.0075	<b>0.1048±0.0110</b>	<b>39.32±1.78</b>	<b>0.9295±0.0211</b>	0.0799±0.0263	40.22±1.80	0.9453±0.0224
Joint coronal ( $n = 20$ )	0.0537±0.0050	41.52±1.35	0.9634±0.0072	0.1076±0.0115	39.11±1.80	0.8631±0.0478	0.0806±0.0284	40.32±2.00	0.9447±0.0249
Joint coronal ( $n = 280$ )	0.0507±0.0050	42.20±1.35	0.9684±0.0065	0.1055±0.0112	39.30±1.80	0.9291±0.0215	<b>0.0781±0.0287</b>	<b>40.75±2.15</b>	<b>0.9487±0.0253</b>

**Table 5.6:** Quantitative results in terms of the NRMSE, PSNR and SSIM for generalization experiments using VNs. Joint training of both coronal knee datasets performed similarly well as individual training of the single contrasts. Applying a VN to a contrast that was not seen during training resulted in a dropped performance.

3. Coronal PDw-FS,  $n = 10$  as studied in Section 5.3
4. Coronal PDw-FS,  $n = 140$
5. Coronal PDw and PDw-FS,  $n = 20$
6. Coronal PDw and PDw-FS,  $n = 280$

## 5.4.2 Results

We first assess the generalization potential of our proposed VNs with respect to contrast, accompanied with different levels of SNR qualitatively, depicted in Figure 5.26. Unsurprisingly, the best results were achieved when a VN was applied to test data from the same sequence as it was trained on. If we apply a VN on different data, we basically observe similar behavior as for CS-based approaches. If a VN is trained on PDw-FS data with low SNR and applied to PDw data with higher SNR, the reconstructed images appear slightly blurred and contain some residual artifacts. In contrast, if a VN is trained on PDw data with high SNR and applied to PDw-FS data with lower SNR, a substantial amount of noise is still present in the reconstructed images. If both coronal knee datasets with different levels of SNR are trained jointly, the VNs perform on similar lines as the individually trained networks. The quantitative values supporting our qualitative observations are presented in Table 5.6. These values indicate that joint training of both coronal knee datasets performed similarly well as training for the individual contrast.

Second, we studied the influence of the number of training samples on the final reconstruction quality. While almost no difference in terms of image quality can be seen in Figure 5.26, residual artifacts are slightly reduced and the images appear slightly more homogeneous when browsing through the whole test dataset and comparing the images directly on the screen. This observation is again supported by Figure 5.26. Improvements in all quantitative measures can be observed if the number of training samples is pushed by  $14\times$ . Pushing the number of training samples improved image quality for all data.

### 5.4.3 Discussion

In our experiments, we focused on two coronal knee datasets that differed both in contrast and SNR. We observed a substantial reduction of image quality if the training and test data differed. Indeed, our results can be related to observations from CS approaches. While in our VN, the data term weights  $\lambda$ , the regularizer as well as the step sizes of the gradient descent scheme are learned from data, the number of iterations for an iterative CS approach [189] and the regularization parameter [164] have to be manually tuned for the individual datasets. The choice of these parameters have an influence on the resolution, residual aliasing artifacts as well as noise amplification due to the g-factor. While a set of CS hyper-parameters tuned for PDw-FS data leads to over-smoothed reconstructions, CS hyper-parameters tuned for PDw data with high SNR leads to noisy reconstructions with residual artifacts of low SNR PDw-FS data. Our VN reconstructions show exactly the same behavior for these two image contrasts. These observations raise the question if the VN is sensitive to both the image contrast and level of SNR. This question was studied in more detail in [136] and results showed that the critical parameter in training is SNR. Low SNR PDw-FS data showed comparable performance on VNs individually trained for PDw-FS data and PDw with additionally added noise, that matches the SNR of PDw-FS data. The results for training both coronal knee datasets jointly perform similarly as the individual trainings. This again poses an interesting question: If we again think of CS-based approaches, the regularizer is constant, but the regularization parameter is adapted to the individual levels of SNR. For the VNs, the learned regularization parameters are the same for PDw and PDw-FS data, hence, different levels of SNR. This indicates that the regularization term might compensate for the difference in SNR.

The presented experiments towards the generalization potential of VNs pose an interesting starting point to use VNs as a discovery tool and to explore relations between CS-based approaches and VNs more deeply. Specifically, it would be interesting to explore the performance of a convex VN as studied by [140] for medical applications in future work. Another interesting point is how we can include prior knowledge about the SNR level into the proposed VN architecture, hence, to train a generalized VN which is independent of image contrast and level of SNR, without increasing the complexity of regularizers.

## 5.5 Influence of Loss Function Design for Accelerated MR Image Reconstruction

One important ingredient that impacts the success of deep learning-based approaches is not only a suitable network architecture but also a meaningful loss function for training. It is common to train deep networks with a simple squared  $L_2$  norm. In a simple experiment, we show that it might be already beneficial to consider magnitude images instead of complex-valued images for the evaluation of the loss function. In this experiment, we perform CG SENSE reconstructions of PDw and PDw-FS scans and assess the reconstruction quality quantitatively in terms of the RMSE on both complex-valued and magnitude images, denoted as  $MSE_{\text{complex}}$



and  $\text{MSE}_{\text{abs}}$ , as illustrated in Figure 5.27. We observe that the minimum  $\text{MSE}_{\text{complex}}$  value is achieved after a lower number of iterations as the minimum  $\text{MSE}_{\text{abs}}$  value. This observation holds for both investigated contrasts, i.e., high and low SNR data. When assessing the qualitative results in Figure 5.28 for  $R = 3$ , the results for a different number of iterations makes a huge difference especially for low SNR data. While the CG SENSE result after 2 iterations is still blurry and corrupted by artifacts, the result after 3 iterations gets sharper and artifacts are reduced. This observation motivates to evaluate the magnitude images in the loss function. Therefore, we validate in this section if this observation also holds for VN reconstructions trained with loss functions on complex-valued and magnitude images.

While the  $L_2$  norm is simple to compute, it, however, suffers from a number of drawbacks: The  $L_2$  norm compares pixel-wise differences, which stands in stark contrast of how the human perceptual system works, and it is not robust against outliers [248, 274]. These observations suggest to use different loss functions for training, e.g., the SSIM, which takes local patch statistics into account. In this section, we will also investigate the impact of perceptual-based loss functions such as the SSIM and Multi-Scale Structural Similarity Index (MS-SSIM) on VN reconstructions.

### 5.5.1 Methods

The experiments were conducted with the same VN architecture as presented in Section 5.4.1. We used  $N_k = 24$  filter kernels of size  $11 \times 11$  and  $T = 10$  stages. We added batch normalization and additionally learn the scaling parameter (see Section 3.2). For training, we used the Proximal ADAM optimizer with block-preconditioning presented in Appendix C.2 with parameters  $\beta_1 = 0.9, \beta_2 = 0.999$ . We ran the optimizer for 300 epochs with learning rate  $4e-3$  and exponential decay (decay rate=0.75, decay steps=30 epochs). The batch normalization was turned off after 50 epochs. For training and testing, we used the fastmri dataset [268] and extracted the data which were acquired with the same hardware and matrix-size parameters as our initial coronal data presented in Table 5.2. The raw data were globally normalized with the maximum value of the zero-filled solution. We used the central 20 slices for training of each case to avoid slices with too much background. Evaluation was performed on the whole volume.

We performed retrospective Cartesian undersampling for  $R = 4$ . The trainings were performed individually for each coronal contrast, PDw and PDw-FS, for a varying number of datasets  $n \in \{10, 140\}$ . For quantitative evaluation, we report the same error measures as used as loss function along with the NRMSE and PSNR values. The goal of our experiments is to compare the influence of different pixel-based and patch-based loss functions. Therefore, we performed experiments with following loss functions:

1. Pixel-based:  $\text{MSE}_{\text{complex}}$
2. Pixel-based:  $\text{MSE}_{\text{abs}}$
3. Patch-based: SSIM with a Gaussian kernel with standard deviation  $\sigma = 1.5$

Dataset	Training Loss Function	MSE <sub>complex</sub> in 10 <sup>-5</sup>	MSE <sub>abs</sub> in 10 <sup>-5</sup>	SSIM in 1	MS-SSIM in 1	NRMSE in 1	PSNR in dB
Coronal PDw ( $n = 10$ )	MSE <sub>complex</sub>	<b>13.82±3.60</b>	9.37±2.52	0.9625±0.0074	0.9755±0.0051	<b>0.0522±0.0054</b>	41.86±1.33
	MSE <sub>abs</sub>	14.21±3.64	<b>9.20±2.42</b>	0.9655±0.0069	0.9763±0.0051	0.0530±0.0059	<b>41.94±1.35</b>
	MS-SSIM	14.77±3.70	9.69±2.46	<b>0.9673±0.0068</b>	<b>0.9764±0.0052</b>	0.0541±0.0062	41.70±1.42
	SSIM $\sigma = 1.5$	15.22±3.76	9.98±2.50	0.9669±0.0068	0.9757±0.0052	0.0550±0.0065	41.57±1.42
Coronal PDw ( $n = 140$ )	MSE <sub>complex</sub>	<b>12.32±3.37</b>	8.36±2.32	0.9643±0.0074	0.9775±0.0049	<b>0.0493±0.0052</b>	42.37±1.31
	MSE <sub>abs</sub>	12.54±3.39	<b>8.10±2.21</b>	0.9688±0.0066	0.9785±0.0047	0.0498±0.0057	<b>42.50±1.33</b>
	MS-SSIM	12.65±3.31	8.36±2.18	0.9705±0.0063	<b>0.9790±0.0047</b>	0.0500±0.0057	42.35±1.38
	SSIM $\sigma = 1.5$	12.83±3.33	8.44±2.20	<b>0.9708±0.0063</b>	0.9787±0.0047	0.0504±0.0056	42.31±1.37
Coronal PDw-FS ( $n = 10$ )	MSE <sub>complex</sub>	4.60±1.26	2.72±0.77	0.9160±0.0222	0.9401±0.0173	0.1079±0.0102	38.83±1.70
	MSE <sub>abs</sub>	<b>4.56±1.24</b>	<b>2.59±0.72</b>	0.9247±0.0217	0.9423±0.0174	<b>0.1075±0.0110</b>	<b>39.02±1.75</b>
	MS-SSIM	4.61±1.23	2.61±0.70	0.9272±0.0215	<b>0.9429±0.0176</b>	0.1082±0.0113	38.98±1.78
	SSIM $\sigma = 1.5$	4.68±1.25	2.63±0.71	<b>0.9278±0.0214</b>	0.9423±0.0177	0.1091±0.0113	38.95±1.77
Coronal PDw-FS ( $n = 140$ )	MSE <sub>complex</sub>	<b>4.33±1.26</b>	2.49±0.71	0.9223±0.0221	0.9436±0.0171	<b>0.1046±0.0108</b>	39.20±1.75
	MSE <sub>abs</sub>	<b>4.33±1.21</b>	<b>2.42±0.67</b>	0.9295±0.0211	0.9447±0.0171	0.1048±0.0110	<b>39.32±1.78</b>
	MS-SSIM	4.40±1.22	2.47±0.67	0.9301±0.0209	<b>0.9451±0.0170</b>	0.1057±0.0112	39.24±1.78
	SSIM $\sigma = 1.5$	4.45±1.22	2.47±0.67	<b>0.9312±0.0208</b>	0.9448±0.0171	0.1063±0.0112	39.24±1.78

**Table 5.7:** Quantitative results for VNs trained with different loss functions for  $R = 4$ . The VNs were trained individually for the coronal contrasts and  $n = 10$  and  $n = 140$ , for acceleration  $R = 4$ . The results indicate that the training was successful in most cases and VNs trained with a specific loss functions get the best score in the same measure for testing.

4. Patch-based: MS-SSIM with standard deviations  $\sigma \in \{0.5, 1, 2, 4\}$

## 5.5.2 Results

Table 5.7 illustrates the quantitative results of our experiments. The results show that the training and testing perform as expected and the quantitative scores are best for the measure it was trained for. Small exceptions can be observed for a smaller training data size  $n = 10$  for PDw between SSIM and MS-SSIM and PDw-FS between MSE<sub>complex</sub> and MSE<sub>abs</sub>. However, the differences between the quantitative and also qualitative results visualized in Figure 5.29 are marginal. A general trend towards a single loss function is hard to see and varies for the individual slices. While SSIM and MS-SSIM shows subtle improvements in terms of texture and visibility of tiny structures as depicted by the green arrow in Figure 5.29, one also observes more residual undersampling artifacts. The largest quantitative and qualitative difference can be observed between reconstructions trained with MSE<sub>complex</sub> and MSE<sub>abs</sub>. Reconstructions trained with MSE<sub>abs</sub> appear sharper and less blurry than reconstructions trained with MSE<sub>complex</sub>.

## 5.5.3 Discussion

In this section, we studied the influence of various loss functions for training our VNs. The first motivation was to compute the loss function on magnitude images instead of complex-valued images, illustrated in Figure 5.27 and Figure 5.28. Indeed, our results verified our expectations and the VN trained with MSE<sub>abs</sub> yields more visually appealing and less blurred results. While the largest gain was achieved when switching from MSE<sub>complex</sub> to MSE<sub>abs</sub> as a loss function, the differences to SSIM and MS-SSIM are subtle. While some regions appear slightly more textured when training with SSIM, the images also contain more residual artifacts than MSE<sub>abs</sub>. Using MS-SSIM as a loss function yields results which appear qualitatively between SSIM and MSE<sub>abs</sub> and might be a compromise between these two loss functions. When comparing

$MSE_{\text{complex}}$  and SSIM, we definitely see an improvement towards the SSIM as loss function which is similar to observations in [274]. However, we do not see great differences between  $MSE_{\text{complex}}$ , SSIM and MS-SSIM as reported by [274] for image denoising. With current loss functions, one has to weight the importance of slightly more textured images and visibility of residual artifacts. The proposed  $MSE_{\text{abs}}$  loss function provides most stable results for our experiments and is less expensive to compute than SSIM and MS-SSIM. However, more advanced loss functions should be considered in future work that reflect the human perceptual system better than currently available loss functions, especially in the case of low SNR data where the fully-sampled reference itself is noisy. This yields also into the direction if the loss should be evaluated on better features, e.g., low-level features of the VGG net as suggested in [149], including an adversarial approach [89, 149] or finding new ways of learning a loss function.

## 5.6 Improved Regularization for Accelerated MR Image Reconstruction

The last sections have shown promising results for MR image reconstruction using a Fields of Experts (FoE) regularizer that maps convolutions of complex numbers to a real-valued convolution response Equation (5.3). A first question that might arise is why the convolution of a complex-valued image maps to the real domain  $K : \mathbb{C} \rightarrow \mathbb{R}$ . Very early experiments motivated us to use this setting as there was no obvious advantage to use complex-valued convolutions in terms of visually improved image quality for one of our first settings. As training improved due to a more stable architecture and optimization schemes, we resume our research into the direction of more robust and improved regularizers.

### 5.6.1 Methods

Indeed, the idea of using complex convolutions and complex activation functions is not new [16, 70, 103, 153, 176, 211, 212, 237] and has also attracted the deep learning community [232, 234]. Indeed, this topic should be noted as phase information might get lost if the complex information is not treated adequately, although it is important for certain tasks, e.g., speech processing [222]. Reichert and Serre [201] introduced a biological motivation by neuronal synchrony that complex-valued networks could be used to build more versatile representations of data. In our case, we study the influence of complex-valued convolutions for MR image reconstruction. The complex-valued convolution reads as

$$x * k = (x_{\text{re}} * k_{\text{re}} - x_{\text{im}} * k_{\text{im}}) + j(x_{\text{im}} * k_{\text{re}} + x_{\text{re}} * k_{\text{im}}).$$

Compared to Equation (5.3), it is obvious that the number of parameter stays the same and the number of convolution operations is doubled, which increases the training time drastically. As the output of the convolution operation is now complex, the next question is how the

activation functions  $\phi$  are modeled. There exists different strategies such as applying the same real-valued activation functions  $\phi_{\text{re}}$  to both real and imaginary part of a complex number separately [176]

$$\phi(z) := \phi_{\text{re}}(z_{\text{re}}) + j\phi_{\text{re}}(z_{\text{im}}).$$

Again, this operation does not require additional parameters. In literature, also other activation functions were proposed to cope with complex-valued signals. Georgiou and Koutsougeras [70] introduced an activation function that only influences the magnitude of a complex-valued signal while keeping the phase of this signal constant. Virtue et al. [242] proposed a similar activation function that preserves the magnitude while modifying the phase of the complex-valued signal, termed complex cardioid. In this section, we also study the influence of learning complex-valued activation functions, which we define as

$$\phi(z) := \phi_{\text{re}}(z_{\text{re}}) + j\phi_{\text{im}}(z_{\text{im}}).$$

Here, we apply individual activation functions to the real and imaginary part of  $z$ . With this formulation, the number of parameters increases by  $N_k \cdot N_w$  in each stage  $t$  where  $N_k$  is the number of filter kernels and  $N_w$  defines the number of RBFs.

The previous modifications of the regularizer are still based on FoE regularizer. Although the first intuition would be to simply stack more iterations to improve the reconstruction results, this might not be useful when the operators  $A$  and  $A^*$  are expensive to compute. Hence, further developments consider the structure of the regularizer itself. While the strength of common deep learning approaches is to extract complex features across a number of layers, our FoE regularizer acts on a single level. To incorporate more complex features, we extend the single-level FoE regularizer to a deeper regularizer as described in Section 3.5.

### 5.6.2 Experimental Setup

We performed experiments with the same data and optimizer settings as in Section 5.5.1. For the regular VN, we performed  $T = 10$  gradient steps, where in each of these steps  $N_k = 24$  filter kernels of size  $11 \times 11$  were learned. To model the activation functions, we used a weighted combination of  $N_w = 31$  RBFs. To study the influence of complex-valued activations and convolutions, we performed the following experiments: While the number of parameters does not change when switching from regular to complex-valued convolutions, additional parameters for the activation function are required when conducting experiments with complex-valued activations. For complex activations, batch normalization is conducted for the real and imaginary filter response separately. However, the number of convolution operations doubles when performing complex-valued convolutions instead of regular convolutions. For the deep regularization, we use two layers with  $N_{k_1} = 24$  of size  $7 \times 7$  and  $N_{k_2} = 48$  of size  $5 \times 5$ . For the second layer, the results are downsampled using a strided convolution with a Gaussian filter ( $\sigma = 1.8$ ). This setup results in following parameter configurations:

Dataset	Model	MSE <sub>complex</sub> in $10^{-5}$	MSE <sub>abs</sub> in $10^{-5}$	SSIM in 1	MS-SSIM in 1	NRMSE in 1	PSNR in dB
Coronal PDw ( $n = 140$ )	regular	12.54±3.39	8.10±2.21	0.9688±0.0066	0.9785±0.0047	0.0498±0.0057	42.50±1.33
	complex convolution	11.90±3.20	7.73±2.10	0.9704±0.0063	0.9795±0.0046	0.0485±0.0055	42.70±1.32
	complex activation	11.90±3.19	7.71±2.08	0.9706±0.0063	0.9795±0.0046	0.0485±0.0055	42.71±1.33
	deep	<b>11.08±3.03</b>	<b>7.17±1.95</b>	<b>0.9728±0.0061</b>	<b>0.9807±0.0044</b>	<b>0.0468±0.0053</b>	<b>43.03±1.36</b>
Coronal PDw-FS ( $n = 140$ )	regular	4.33±1.21	2.42±0.67	0.9295±0.0211	0.9447±0.0171	0.1048±0.0110	39.32±1.78
	complex convolution	4.23±1.16	2.36±0.64	0.9306±0.0208	0.9458±0.0168	0.1036±0.0109	39.42±1.78
	complex activation	4.26±1.20	2.39±0.66	0.9298±0.0210	0.9454±0.0170	0.1039±0.0110	39.38±1.79
	deep	<b>4.22±1.15</b>	<b>2.34±0.63</b>	<b>0.9313±0.0208</b>	<b>0.9461±0.0168</b>	<b>0.1035±0.0108</b>	<b>39.47±1.78</b>

**Table 5.8:** Quantitative results for VNs trained with different regularizers. The VNs were trained individually for the coronal contrasts from  $n = 140$  cases. Quantitative improvements can be achieved for other regularization techniques than the regular FoE regularizer.

1. Regular: 6577 parameters / stage
2. Complex convolution: 6577 parameters / stage
3. Complex activation: 7345 parameters / stage
4. Deep regularization: 33457 parameters / stage

### 5.6.3 Results

The quantitative results are summarized in Table 5.8. The results indicate that when using other regularizers than the regular single-level FoE regularizer, all quantitative measures can be improved. While there is a difference between complex-valued convolution and a regular regularizer, further degrees of freedom in the activation function did not lead to any further improvements. The largest improvements can be achieved by the deep regularizer. Although the quantitative improvements appear minor, there is a substantial difference between the regular and deep regularizer as illustrated in Figure 5.30. The deep regularizer yields a better artifact suppression than the regular FoE regularizer and results in more homogeneous images with an increased detectability of details. This figure also shows that the regularizer based on complex-valued convolutions is slightly improved, however, the difference to the reconstruction using complex-valued activations is hardly visible. In some regions, the regularizer based on complex-valued convolutions appear slightly sharper than the regular FoE regularizer.

### 5.6.4 Discussion

In this section, we gained first insights in using different kinds of regularizers. As MR images are complex-valued it might be obvious to use either complex-valued convolutions or complex-valued activations. The results indicated that there is a slight improvement when using complex-valued convolutions, but no substantial improvement could be observed for complex-valued activations. However, compared to the regular FoE regularizer, the complex-valued convolutions require double as many convolution operations which substantially increases training time. For our specific training data using TSE knee sequences and the specified VN structure, the general impact of these types of regularization is minor, however, they

might be considered for different MRI applications. The greatest impact on the results showed the deep regularizer. While the number of parameters is increased, the results appear very promising in terms of further reduced artifacts and detectability of details. These first results using deep regularization offer a great potential for further research as more complex features can be considered.

## 5.7 Intra-Vendor Reproducibility

All previous experiments were performed on data that were acquired at NYU Langone Health. One natural question that appears is if the results are reproducible across the hardware of the same vendor: How does the network trained on data from NYU perform on data that were acquired at Graz University of Technology? To get a first answer to this question, we acquired both fully-sampled and prospectively undersampled data from a F29 volunteer. A full knee protocol was acquired on a 3T Siemens Magnetom Skyra and a 15 channel knee coil with identical sequence parameter settings as described in Table 5.2. The acquired data were reconstructed using the trained VNs of Section 5.3 for retrospective and prospective Cartesian undersampling of  $R = 4$ . The reconstructed images are depicted in Figure 5.31. We observe similar behaviour in terms of reconstruction quality as reported in Section 5.3. No obvious differences in image quality can be observed when reconstructing the images with the same scanner hardware at a different institution.

## 5.8 Conclusion and Outlook

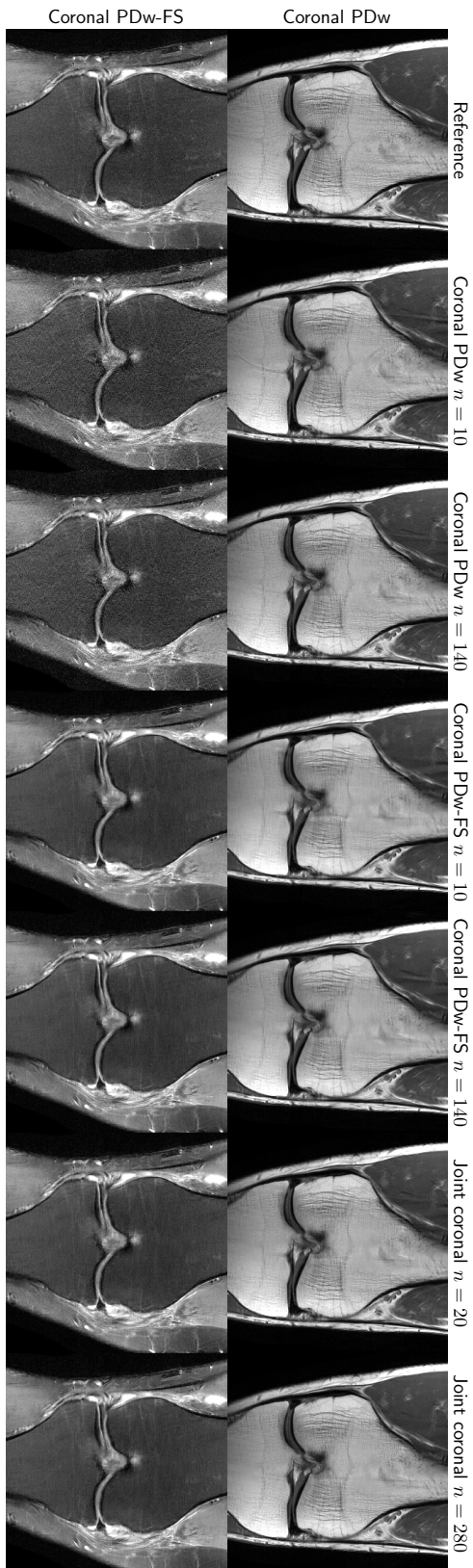
In this chapter, we provided an overview of using VNs for static MR image reconstruction. Our main focus was set on the data design where we intended to stay close to a clinical setting. We trained our VNs on real clinical data of patients undergoing knee exams. Furthermore, we proposed the first deep learning-based approach for PI, which is standard for MR acquisitions. To validate the proposed VN, we did not only perform retrospective experiments but tested it also on prospectively accelerated data. This data greatly varies in terms of SNR. Our results indicate that low SNR images are more challenging to reconstruct, because the loss function favors blurry instead of sharp solutions. Also, the presented loss function variations did not result in a visual improvement of these data. Hence, an important topic in future work is to cope with the great variability in SNR.

A question that arises here is: How far can we go with acceleration? In our experiments, we showed results up to a undersampling factor of  $R = 4$ . We did not go further with acceleration because we intended to stay close to a clinically realistic acceleration potential. The less data are available, the more likely it is that learning-based approaches invent new structures and can only make guesses on the missing information. In the worst medical cases, pathologies would be either missed or invented. This is especially crucial when using GANs. They are known to produce realistic results, however, there is no guarantee that the results are clinically valid.

Thaler et al. [230] gained insights on the influence of both necessary data and adversarial loss functions for a successful reconstruction.

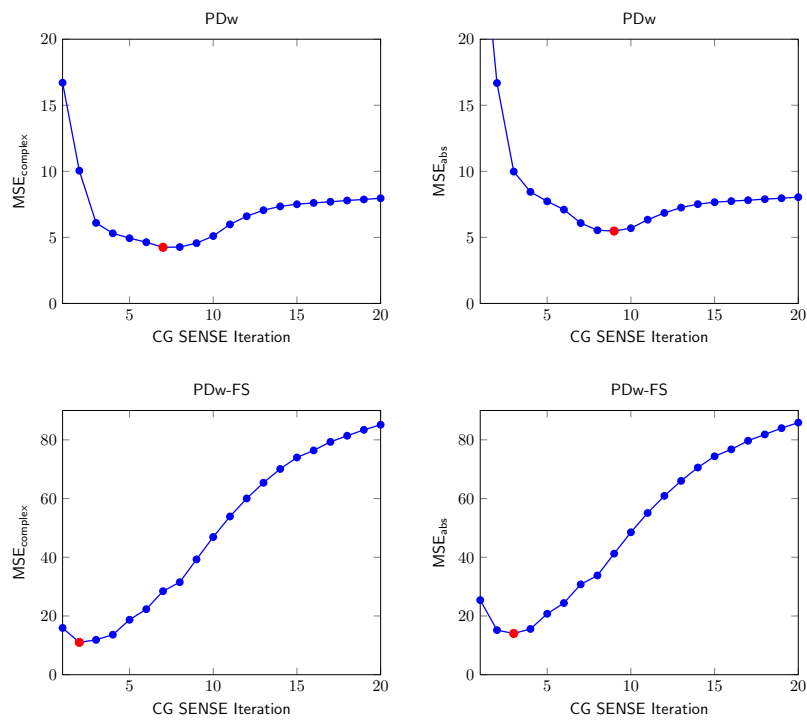
Our presented observations were done for a specific hardware and sequence setting. This does not mean that these observations are valid for other data, too. The presented experiments might become more important when dealing with data that are more sensitivity to the phase information, so they should definitely be considered in future work. When using, e.g., Gradient Echo (GRE) sequences that are more sensitive to phase than TSE sequences, or for quantitative imaging. Although we expected a greater visual improvement of using complex-valued convolutions and activations already for our TSE sequences, the impact on GRE sequences might be larger. However, the quantitative results for complex-valued convolutions and activations were slightly improved. For our setting, a deep regularizer provides a great potential to further improve reconstruction results.

The aim of this chapter was not to present a perfectly designed and optimized network architecture, but to give an overview over potential improvements and pitfalls in data preparation, network and loss function design. Hence, it should provide a starting point for various directions in future work. One open question is still robust data normalization. As MR data has no quantitative meaning, normalization has to be done in either  $k$ -space or image space. To streamline implementations we performed normalization to  $[0,1]$  in image domain based on the zero filling reconstruction slice-by-slice. We believe that more robust, global normalization might further improve the reconstruction quality. Furthermore, a suitable choice for the network architecture, loss function and optimizer has to be made. Especially choosing a proper loss function that represents trained radiologist's eyes will greatly impact the reconstruction results. We also believe that deeper regularization can better deal with the structured aliasing artifacts of Cartesian undersampling schemes. However, as training data are limited imposing structure as presented with our deep regularizer helps keeping the number of network parameter feasible and provides a way to train networks with limited training data.

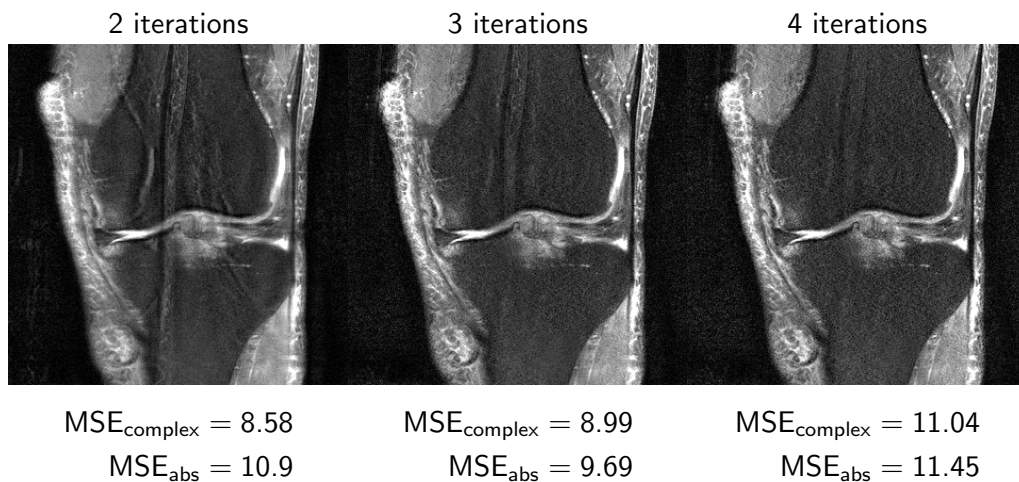


**Figure 5.26:** Results for the generalization potential of the proposed VNs. The VNs were trained for different datasets and number of samples  $n$ , denoted in the column captions. After training, the VNs were applied to both coronal PDw and PDw-FS scans of two different subjects. VNs trained for both coronal knee datasets perform similarly as trained for the individual contrasts. VNs trained for a wrong level of SNR results in a similar behaviour as CS-based methods were a wrong regularization parameter is used, hence, over-smoothing or noise amplification.

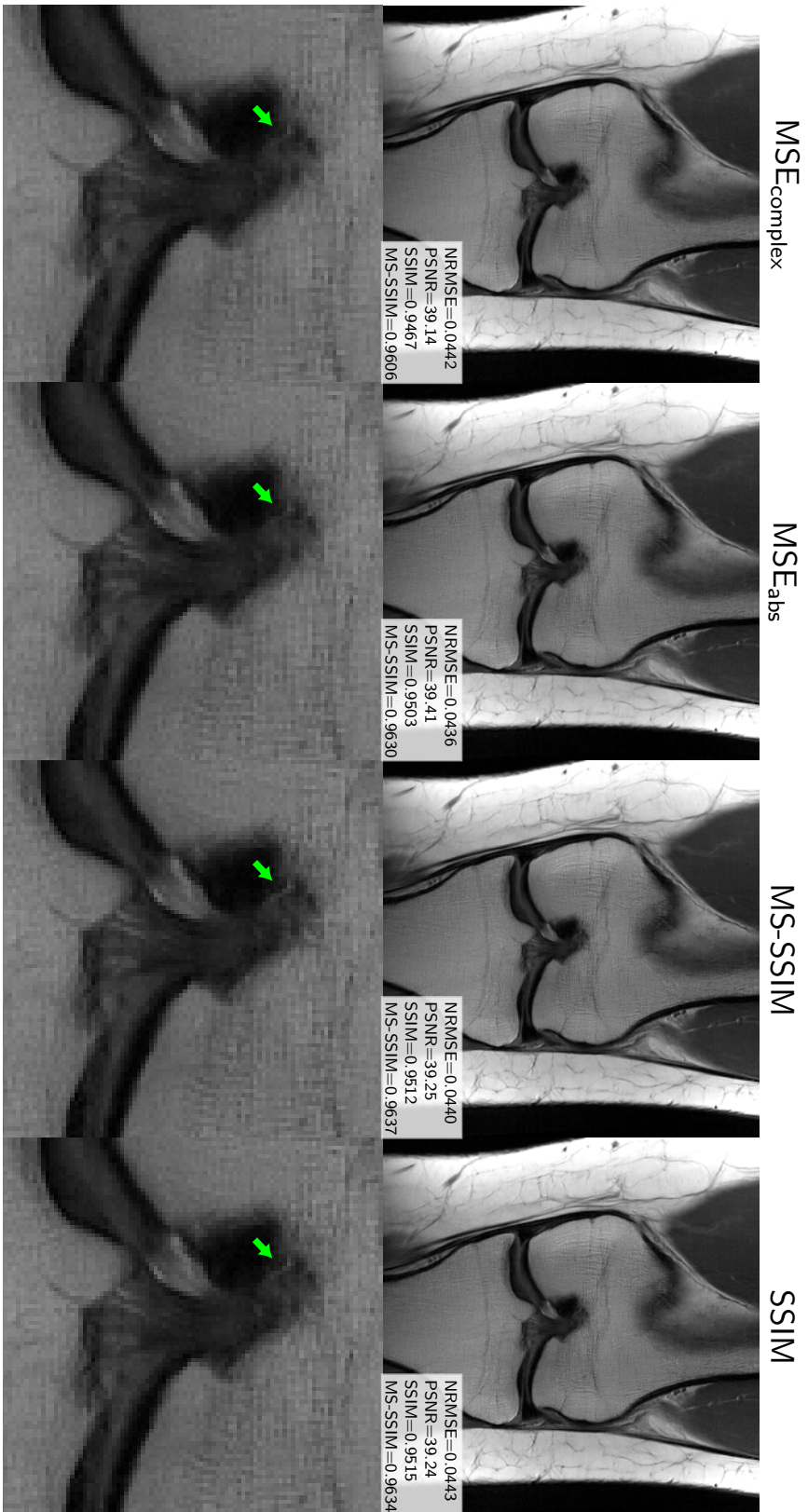




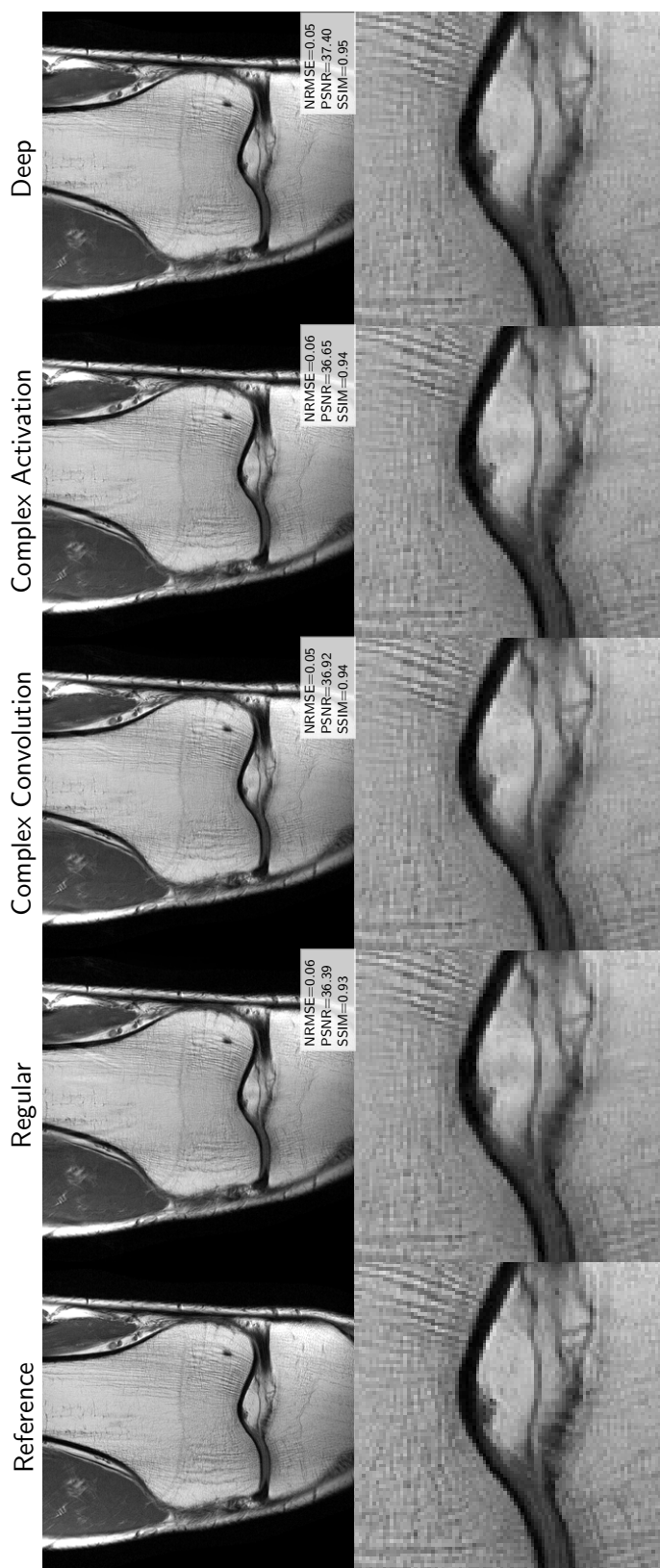
**Figure 5.27:** Reconstruction errors depending on the number of CG SENSE iterations for different contrasts at  $R = 4$ . The optimal number of iterations differs for the  $MSE_{\text{complex}}$  and  $MSE_{\text{abs}}$  values, marked by the red dot.



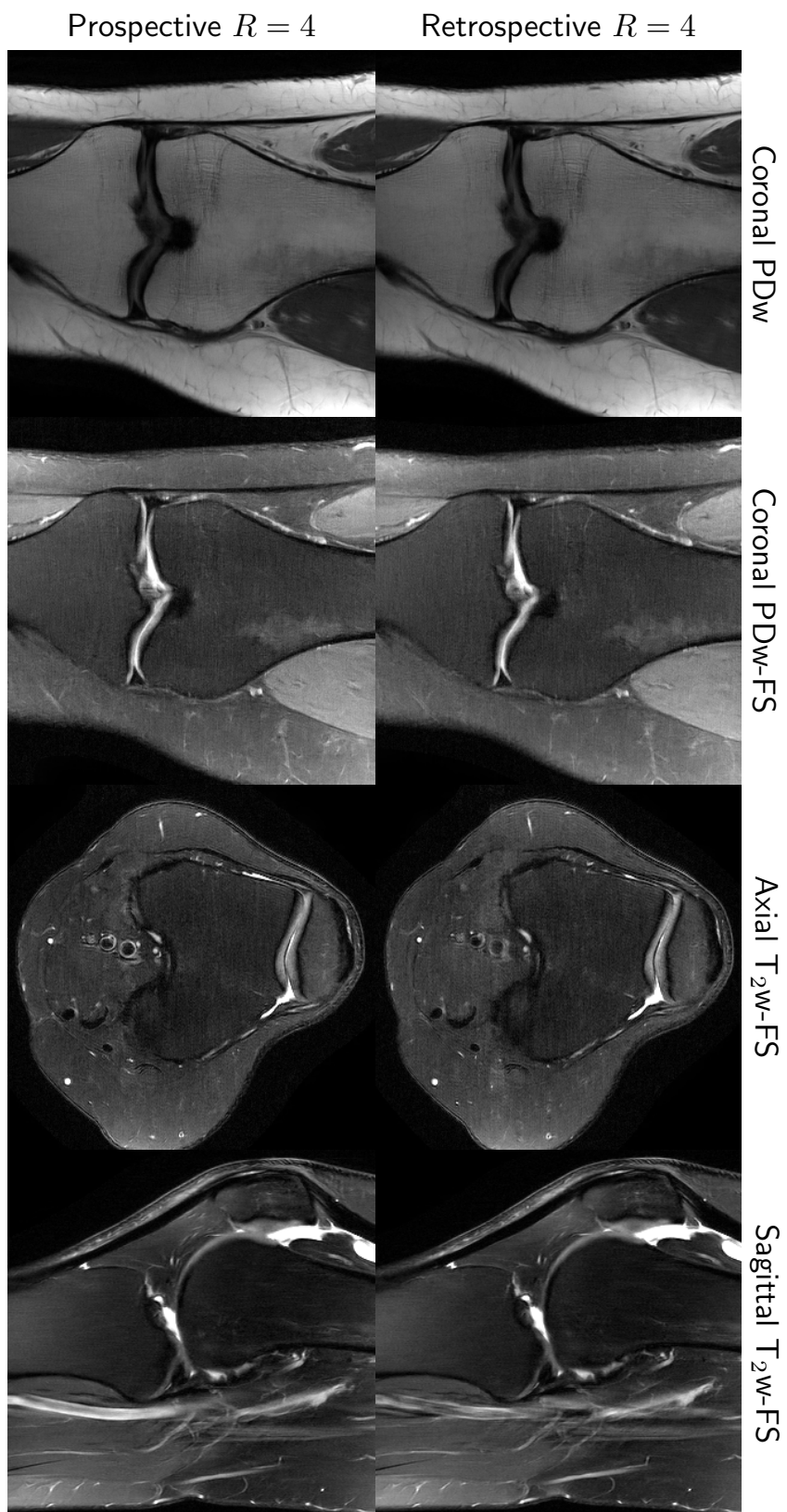
**Figure 5.28:** CG SENSE reconstructions for a different number of iterations for PDw-FS data at  $R = 3$ . The quantitative values for  $MSE_{\text{complex}}$  and  $MSE_{\text{abs}}$  differ. The result after 3 iterations is visually more pleasant than the result for 2 iterations, motivating the use of  $MSE_{\text{abs}}$ .



**Figure 5.29:** VN reconstructions of a coronal PDw scan for  $R = 4$ . The VNs were trained with different loss functions. A slight difference between MSE<sub>abs</sub>, SSIM and MS-SSIM can be seen in the detailed view where a tiny structure indicated by the green arrow, appears most prominent in the SSIM and MS-SSIM reconstructions. The MSE<sub>complex</sub> appears blurriest compared to the other reconstructions.



**Figure 5.30:** VN reconstruction results using different types of regularization for a coronal PDw scan and  $R = 4$ . The regularizer based on complex-valued convolutions yields slightly improved results compared to the regular FoE regularizer, while complex-valued activations could not improve the results substantially. The best results yields the deep regularizer with decreased artifacts and more homogeneous images compared to the regular FoE regularizer, supported by the detailed view.



**Figure 5.31:** Intra-vendor reproducibility: Retrospective and prospective VN reconstructions for  $R = 4$  (F29). The data were acquired with the same hardware and parameter settings at a different institution than the data the VNs were trained with. No visible difference in terms of image quality can be observed.

## Variational Networks for Dynamic Magnetic Resonance (MR) Image Reconstruction

On ne voit bien qu'avec le cœur.  
L'essentiel est invisible pour les yeux.

*Antoine de Saint-Exupéry*

This chapter involves following abstract, which is accepted for ISMRM 2019:  
K. Hammernik, M. Schloegl, R. Stollberger, and T. Pock. Dynamic Multicoil Reconstruction using Variational Networks. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, page to appear, 2019

### Contents

---

<b>6.1</b>	<b>Fundamentals of Accelerated Cardiac MR Image Reconstruction . . . . .</b>	<b>112</b>
<b>6.2</b>	<b>Methods . . . . .</b>	<b>113</b>
<b>6.3</b>	<b>Results . . . . .</b>	<b>116</b>
<b>6.4</b>	<b>Discussion . . . . .</b>	<b>117</b>

---

Recent developments in deep learning for accelerated MR image reconstruction have shown improved results over Compressed Sensing (CS)-based approaches for both static [84] and dynamic [100, 193, 213] imaging. However, the acquisition of ground-truth data is the key ingredient for the success of these approaches. Generation of dynamic MR ground-truth data is challenging. Cardiac CINE imaging for example requires data acquisition in breath-hold, which can take around 25 seconds in order to collect fully-sampled data. This is already infeasible for many patients undergoing clinical routine examinations, hence, Parallel Imaging (PI) with a

two-fold acceleration is used as an essential standard. Cardiac CINE acquisitions are gated by an Electrocardiography (ECG) signal, which impedes the acquisition of patients with arrhythmia. Furthermore, a trade-off between temporal resolution and duration of breath-hold has to be made to acquire fully-sampled ground-truth data. The number of acquisition frames also depends on the current heart rate of the imaged patient.

In this chapter, we give first insights on how Variational Networks (VNs) can be used for dynamic CINE imaging. Therefore, we introduce the dynamic forward model and extend the VN formulation to spatio-temporal regularization. The approach is tested on different sampling patterns, acceleration factors and contrasts. Furthermore, we shortly discuss the importance of high-quality ground-truth data for the success of deep learning-based approaches.

## 6.1 Fundamentals of Accelerated Cardiac MR Image Reconstruction

The main challenge of dynamic acquisitions is to achieve balanced spatial and temporal resolution, while avoiding other motion artifacts such as respiratory motion. To achieve this balance, accelerated data acquisition strategies are advantageous. In dynamic MR, PI with special undersampling strategies allow us to exploit the data redundancy in both spatial and temporal domain. To account for the different time dimensions, extensions to common image-based and  $k$ -space PI techniques were proposed in [23, 108, 125, 233].

While it is challenging to fulfill the CS requirements in 2D static imaging, such as the incoherence of undersampling artifacts in Cartesian sampling, there exist different possibilities to fulfill these conditions in dynamic Magnetic Resonance Imaging (MRI). The sparsity condition can be fulfilled by temporal Fourier Transform (FT) [120] or spatio-temporal regularization based on Wavelets or temporal finite differences [3, 66, 158, 165].

After the seminal work of Candès et al. [30], Low-Rank plus Sparse (L+S) decomposition [181] became increasingly popular to reconstruct dynamic MR images. In this approach, a reconstruction problem is formulated that decomposes the reconstruction into a static low-rank component and a sparse component, containing the dynamics of the image. Image decomposition is also exploited by Infimal Convolution (IC)-type regularization [104], which additionally allows for a separation of components at various temporal dynamics. Schloegl et al. showed that Infimal-Convolution-Total-Generalized-Variation (ICTGV) yields robust reconstructions of various dynamic MRI applications [215].

In contrast to fixed regularization schemes, adaptive learning techniques such as dictionary learning were proposed in the context of dynamic MRI [26], where local patches are approximated by dictionary atoms. However, these techniques suffer from a huge memory requirement and extensive computational demand in the range of several hours for a single single-coil case which limits the practical applicability. Recent advances in deep learning enable not only to adapt to the underlying training data but offer also high reconstruction speed. However, most of current learning-based approaches for dynamic MRI [98, 193, 213] neglect the PI component

and work only on single-coil data. VNs have shown promising results for multi-coil reconstruction in static imaging as shown in Chapter 5, hence, we introduce a natural extension of VNs with spatio-temporal regularization to interact with dynamic multi-coil data.

## 6.2 Methods

For dynamic imaging, an additional time component is introduced, which makes reconstruction approaches more computationally demanding as compared to static 2D imaging. This limits also the design of appropriate network architectures, which should sometimes fit on a single Graphics Processing Unit (GPU) and be trained in a feasible time window. This section further provides details on the setup of the proposed VN architecture along with details on the acquired data and experimental design.

### 6.2.1 A Variational Network for Dynamic Applications

The basic theory of VNs was described in Chapter 3. There exist different approaches to formulate VNs, ranging from reaction-diffusion processes [46], over Landweber regularization [84] to an unrolled incremental Gradient Descent (GD) scheme [140]. Let us define here the VN by an unrolled incremental GD scheme with  $J$  components and a fixed number of iterations  $T$

$$x^t = x^{t-1} - \sum_{j=0}^J \nabla h_j(x^{t-1}), \quad 1 \leq t \leq T$$

to obtain a spatio-temporal reconstruction  $x \in \mathbb{C}^{N_x N_y N_t}$  of size  $N_x \times N_y \times N_t$ , where  $N_x$  and  $N_y$  denote the resolution in  $x$  and  $y$  direction and  $N_t$  denotes the number of time frames. The term  $\sum_{j=0}^J h_j(x^{t-1})$  is defined by a generalized CS model for dynamic MR image reconstruction, including a data-fidelity

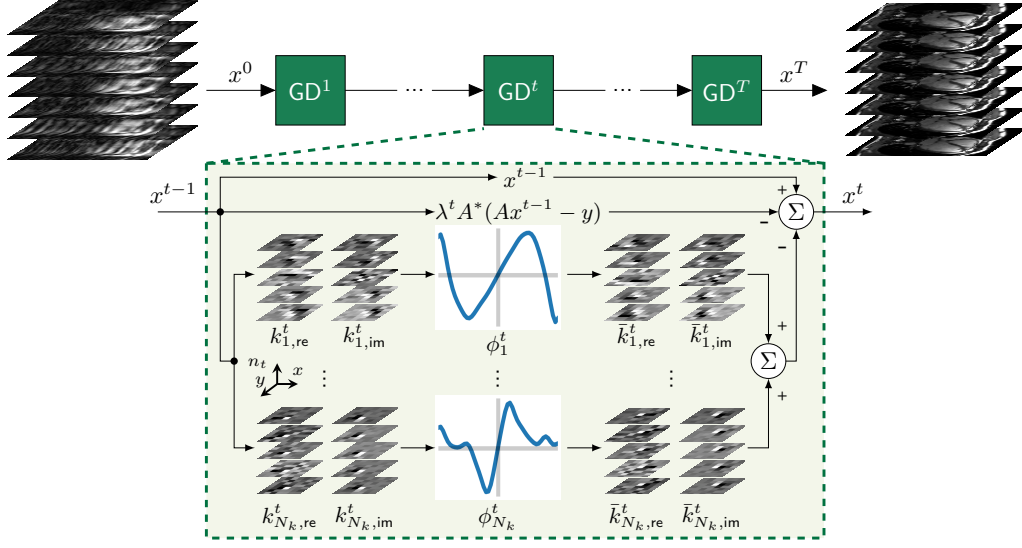
$$h_0(x) = \frac{\lambda}{2} \|Ax - y\|_2^2, \quad (6.1)$$

where  $\lambda > 0$ , and a regularization term  $h_j(x)$ ,  $j = 1, \dots, J$ . The data-fidelity  $h_0(x)$  enforces consistency of the reconstruction  $x$  to the acquired raw data  $y \in \mathbb{C}^{N_x N_y N_t N_c}$  for  $N_c$  coils. The dynamic multi-coil forward operator  $A$ , i.e.,

$$A : x = (x_{n_t})_{n_t=1, \dots, N_t} \rightarrow (\mathcal{F}_t [c_q x_{n_t}])_{q=1, \dots, N_c; n_t=1, \dots, N_t},$$

involves coil-sensitivity profiles  $c_q$ , that are assumed to be static in time, and Fourier transforms with temporally varying sampling masks  $\mathcal{F}_t$ . For the regularization part, we use a Fields of Experts (FoE) regularizer that acts on the spatio-temporal 2D+t volume

$$h_j(x) = \langle \mathbf{1}, \rho_j(k_j * x) \rangle, \quad j = 1, \dots, J.$$



**Figure 6.1:** VN for 2D+t multi-coil MR image reconstruction, defined as a sequence of GD steps. Prior information such as 2D+t filter kernels  $k$ , non-linear activation functions  $\phi$  and data term weights  $\lambda$  are learned from pairs of undersampled multi-coil raw data  $y$  and fully sampled reference images in an end-to-end manner.  $x^0$  is defined by the initial time-averaged solution

Here,  $k_j \in \mathbb{C}^{s \times s \times s_t}$  denote spatio-temporal complex-valued filter kernels of size  $s \times s \times s_t$  and  $\rho : \mathbb{C}^{N_x N_y N_t} \rightarrow \mathbb{R}^{N_x N_y N_t}$  model non-linear potential functions. In this work, we define the convolution operation as

$$x * k = (x_{re} * k_{re} - x_{im} * k_{im}) + j(x_{im} * k_{re} + x_{re} * k_{im}).$$

When computing the gradient of the variational model Equation (6.1), we receive the VN structure as depicted in Figure 6.1. Due to the gradient computation, we now involve the gradient of potential functions, termed activation functions  $\phi = \rho' : \mathbb{C}^{N_x N_y N_t} \rightarrow \mathbb{C}^{N_x N_y N_t}$  which are defined in a complex-valued setting as

$$\phi(z) := \phi_{re}(z_{re}) + j\phi_{im}(z_{im}).$$

While Section 5.6 showed only subtle improvements, initial experiments for dynamic MRI motivated the use of this kind of regularization here. The input to the VN illustrated in Figure 6.1 is defined by the temporal average of the zero filled reconstruction.

### 6.2.2 Data Acquisition

We acquired retrospectively gated CINE cardiac data from five healthy volunteers in breath-hold using a 3T scanner (Siemens Magnetom Skyra) and a spine-/body-coil. The number of active coil elements varied for the individual acquisitions between 26-34 channels. The acquired datasets consisted of one two-chamber (2CH), four-chamber (4CH) and Left Ventricular

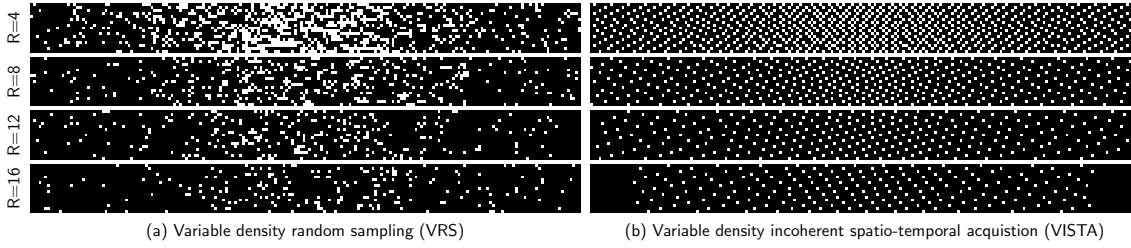


Outflow Tract (LVOT)-view as well as four Short Axis (SA) views. The SA scans consisted of one basal, two mid-ventricular and one apical view. This setup results in a total number of 35 datasets. Two different sequences with a matrix size of  $192 \times 192$ , voxel size  $1.8 \times 1.8 \times 6\text{mm}^3$  and temporal resolution  $\Delta t \sim 50\text{ms}$  were acquired with following acquisition parameters: (1) FLASH TR/TE/FA=5.8ms/3.16ms/12° and (2) bSSFP TR/TE/FA=3.9ms/1.72ms/40°. The number of time frames varied for the individual contrasts and volunteers between 17-33 frames. From the acquired multi-coil data, coil sensitivity maps were estimated using an iterative variational approach presented in [214]. The gold standard reconstruction is obtained by Sensitivity Encoding (SENSE) [190] reconstruction of the acquired fully-sampled  $k$ -space data. For the individual experiments, we retrospectively undersampled the fully-sampled  $k$ -space data.

### 6.2.3 Experimental Setup and Evaluation

For our experiments, we investigate the influence of different pseudo-random Cartesian sampling patterns, which were simulated at various acceleration rates  $R \in \{4, 8, 12, 16\}$ . A common sampling pattern for both static and dynamic imaging is Variable Density Random Sampling (VRS) [164], depicted in Figure 6.2a. Although the degree of incoherence is high for VRS, these patterns suffer from several limitations due to large gaps in the  $k$ -space as well as the reduced sampling density for high frequencies [9], resulting in degraded reconstructions. In contrast, Variable Density Incoherent Spatio-Temporal Acquisition (VISTA) [9], illustrated in Figure 6.2b, samples all frequencies in  $k$ -space in a more controlled way, leading to an improved image quality even at high acceleration factors. For training, we generated a pool of 100 sampling masks for both VRS and VISTA patterns at the different acceleration factors, which increases the variability of undersampling artifacts and avoids that the VNs are learned for a specific undersampling pattern. The sampling patterns are chosen randomly at each training iteration.

We trained individual VNs for each contrast and sampling pattern. The architecture of the VNs stayed constant for the individual experiments. The VNs consisted of  $T = 18$  gradient descent steps, where in each of these steps  $N_k = 24$  filters of size  $7 \times 7 \times 5$  were learned together with the additional scaling parameter of the batch normalization (decay=0.999,  $\epsilon = 1e-3$ ) explained in Section 3.2. The filter kernels were constrained to the unit norm ball and to have zero-mean. The activation functions consisted of a weighted combination of  $N_w = 31$  Gaussian radial basis functions (RBFs). Together with the filter kernels and activation functions, the data term weights  $\lambda$  are learned, which were initialized with  $\lambda_0 = 2$ . For training, we used the Proximal ADAM optimizer with block-preconditioning with the ADAM parameters  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$  and step size  $\nu = 2e-2$ . The optimizer is described in more detail in Appendix C.2. We trained the VN for 10000 iterations using a batch size of 1. The update of the means and standard deviations of the batch normalization is turned off after 1000 iterations. We used cosine annealing [162] to decrease the learning rate during training. The VNs are initialized with the time averaged solution. A number of 28 datasets from 4 volunteers were used for training. During training, the number of time frames is fixed  $N_t = 17$  due to GPU memory constraints.



**Figure 6.2:** Different  $k$ - $t$  sampling patterns of size  $192 \times 17$  for dynamic MR image reconstruction: (a) illustrates VRS and (b) illustrates VISTA at different acceleration rates. The sampling in Phase Encoding (PE) direction is denoted by the horizontal axis, while the sampling in time is represented by the vertical axis.

The consecutive frames are chosen randomly during training. For testing, we evaluated all acquired time frames because less GPU memory is required as no gradient calculations are carried out. We compare our VN results to the combined PI-CS method ICTGV [215], which showed to outperform L+S [181] for dynamic CINE MR data. The data were reconstructed with the ICTGV parameters  $\alpha_0/\alpha_1 = 1/2$ ,  $\beta(t_1) = 4$ ,  $\beta(t_2) = 0.5$ ,  $\gamma_{1,2}(s) = 0.5$ , which were obtained via grid search. All reconstructions were evaluated quantitatively with respect to the gold standard by means of the Normalized Root Mean Squared Error (NRMSE), Peak Signal-To-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) on 7 datasets from another volunteer, which were not used for training. Evaluation was performed in a region of  $124 \times 124 \times N_t$  that contained mainly the moving heart.

### 6.3 Results

We show the performance of the learned VN on different anatomies and datasets. All figures illustrate the reconstructions compared to the reference and ICTGV along with the difference images and the used sampling pattern. Figure 6.3 illustrates the results for a bSSFP 2CH view for VISTA sampling at  $R = 8$ . From a visual perspective, ICTGV and VN perform on similar lines, while the difference maps show a decreased error for the VN reconstruction. The results for a FLASH 4CH view based on VISTA sampling at  $R = 16$  are depicted in Figure 6.8. While we can observe a blocky pattern in the ICTGV result, the VN appears smoother with slightly less error in the difference maps. For a bSSFP SA scan at  $R = 12$ , we compare VISTA sampling in Figure 6.4 and VRS sampling in Figure 6.5. These images show that VISTA sampling yields improved results compared to VRS sampling for both ICTGV and VN reconstructions. When comparing ICTGV and VN reconstructions, we observe a similar behaviour as for the other results in the spatial domain. In the temporal domain, some regions appear to have higher error in the VN reconstructions. Similar observations can be made when we compare VISTA sampling in Figure 6.6 and VRS in Figure 6.7 for a FLASH SA scan at  $R = 12$ . Again, the VN reconstructions appear smoother and less blocky than the ICTGV reconstructions.

Quantitative results for VISTA and VRS sampling at acceleration rates  $R \in \{4, 8, 12, 16\}$

Pattern	$R$	Model	$\text{NRMSE}_{\text{complex}}$ in 1	$\text{NRMSE}_{\text{abs}}$ in 1	PSNR in dB	SSIM in 1
VISTA	4	ICTGV	0.0556±0.0179	0.0399±0.0126	45.13±4.44	0.9710±0.0115
		VN	<b>0.0501±0.0107</b>	<b>0.0354±0.0073</b>	<b>45.99±3.78</b>	<b>0.9783±0.0052</b>
	8	ICTGV	0.0836±0.0258	0.0626±0.0203	41.23±4.46	0.9450±0.0155
		VN	<b>0.0738±0.0125</b>	<b>0.0531±0.0083</b>	<b>42.39±3.43</b>	<b>0.9606±0.0087</b>
	12	ICTGV	0.1001±0.0242	0.0754±0.0178	39.46±3.93	0.9298±0.0147
		VN	<b>0.0933±0.0118</b>	<b>0.0678±0.0079</b>	<b>40.22±3.09</b>	<b>0.9457±0.0090</b>
	16	ICTGV	0.1145±0.0236	0.0866±0.0158	38.18±3.56	0.9183±0.0162
		VN	<b>0.1084±0.0147</b>	<b>0.0788±0.0085</b>	<b>38.92±3.04</b>	<b>0.9355±0.0113</b>
VRS	4	ICTGV	0.0783±0.0274	0.0575±0.0194	41.98±4.51	0.9555±0.0176
		VN	<b>0.0673±0.0121</b>	<b>0.0483±0.0080</b>	<b>43.23±3.51</b>	<b>0.9680±0.0072</b>
	8	ICTGV	0.1053±0.0270	0.0789±0.0189	39.05±3.91	0.9332±0.0164
		VN	<b>0.0964±0.0147</b>	<b>0.0697±0.0094</b>	<b>40.00±3.22</b>	<b>0.9475±0.0099</b>
	12	ICTGV	0.1231±0.0251	0.0935±0.0178	37.51±3.59	0.9194±0.0158
		VN	<b>0.1212±0.0188</b>	<b>0.0890±0.0141</b>	<b>37.90±3.36</b>	<b>0.9320±0.0110</b>
	16	ICTGV	0.1452±0.0218	0.1106±0.0141	35.99±2.98	0.9074±0.0165
		VN	<b>0.1444±0.0131</b>	<b>0.1065±0.0077</b>	<b>36.27±2.63</b>	<b>0.9116±0.0118</b>

**Table 6.1:** Quantitative results for the bSSFP dataset for VISTA and VRS at different acceleration rates  $R \in \{4, 8, 12, 16\}$ . The VN outperforms ICTGV in all cases.

are illustrated in Table 6.1 for the bSSFP dataset and in Table 6.2 for the FLASH dataset. The VN reconstructions outperform the ICTGV reconstruction in all cases for all evaluated similarity measures. We also observe that VISTA sampling outperforms VRS sampling at the same acceleration rates.

Besides example reconstruction results, we plot a subset of learned VN parameters in Figure 6.9. The image shows different learned activation functions  $\phi$  along with their potential functions  $\rho$ . The potential functions have a shape similar to a log-student-t function in the first row and a quadratic function in the second row. The filter kernels are based on various derivatives both in spatial and temporal domain. While the filter kernels in the first row mainly act in the spatial domain, temporal derivatives are considered by the filter kernels depicted in the second row.

## 6.4 Discussion

In this chapter, we presented a first application of VNs to dynamic multi-coil MR data. We showed the efficiency of the proposed approach compared to a state-of-the-art PI-CS method ICTGV. In terms of image quality, the differences between ICTGV and VN reconstructions are hard to see for the bSSFP dataset. The differences between ICTGV and VN reconstructions are more obvious for the FLASH dataset which has a lower Signal-to-Noise Ratio (SNR) compared to the bSSFP dataset. Specifically, the ICTGV reconstructions appear blocky, which is a typical

Pattern	$R$	Model	$\text{NRMSE}_{\text{complex}}$ in 1	$\text{NRMSE}_{\text{abs}}$ in 1	PSNR in dB	SSIM in 1
VISTA	4	ICTGV	0.0701±0.0139	0.0487±0.0098	40.28±2.76	0.9070±0.0507
		VN	<b>0.0560±0.0060</b>	<b>0.0390±0.0040</b>	<b>42.10±2.00</b>	<b>0.9386±0.0198</b>
	8	ICTGV	0.0873±0.0161	0.0611±0.0110	38.27±2.55	0.8801±0.0547
		VN	<b>0.0746±0.0074</b>	<b>0.0518±0.0054</b>	<b>39.64±1.94</b>	<b>0.9103±0.0272</b>
	12	ICTGV	0.0977±0.0140	0.0698±0.0093	37.07±2.22	0.8641±0.0458
		VN	<b>0.0868±0.0084</b>	<b>0.0612±0.0061</b>	<b>38.18±1.87</b>	<b>0.8930±0.0304</b>
16	ICTGV	0.1067±0.0129	0.0768±0.0082	36.21±2.03	0.8523±0.0436	
	VN	<b>0.0990±0.0096</b>	<b>0.0700±0.0066</b>	<b>37.01±1.86</b>	<b>0.8769±0.0328</b>	
VRS	4	ICTGV	0.0789±0.0150	0.0555±0.0107	39.12±2.67	0.9008±0.0504
		VN	<b>0.0661±0.0062</b>	<b>0.0464±0.0042</b>	<b>40.57±1.89</b>	<b>0.9275±0.0215</b>
	8	ICTGV	0.1003±0.0128	0.0723±0.0089	36.76±2.26	0.8704±0.0476
		VN	<b>0.0895±0.0086</b>	<b>0.0633±0.0060</b>	<b>37.88±1.87</b>	<b>0.8955±0.0271</b>
	12	ICTGV	0.1204±0.0174	0.0891±0.0139	34.97±2.26	0.8494±0.0472
		VN	<b>0.1096±0.0132</b>	<b>0.0775±0.0087</b>	<b>36.14±2.01</b>	<b>0.8735±0.0353</b>
16	ICTGV	0.1380±0.0208	0.1036±0.0166	33.69±2.56	0.8354±0.0457	
	VN	<b>0.1336±0.0208</b>	<b>0.0950±0.0157</b>	<b>34.42±2.08</b>	<b>0.8521±0.0353</b>	

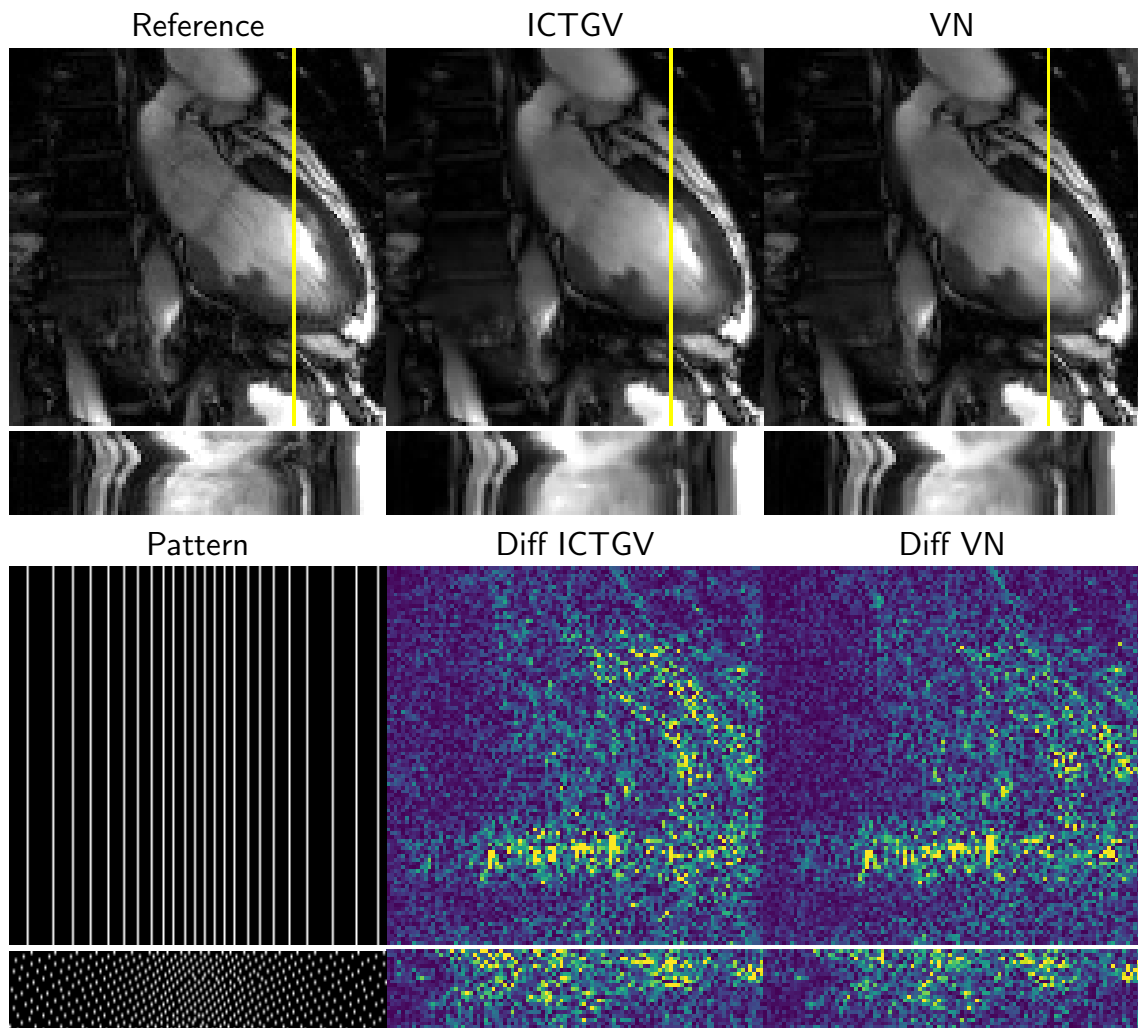
**Table 6.2:** Quantitative results for the FLASH dataset for VISTA and VRS at different acceleration rates  $R \in \{4, 8, 12, 16\}$ . The VN outperforms ICTGV in all cases.

behaviour of this method on low SNR data. In contrast, the FLASH reconstructions with the VN appear smooth, similar to observations of low SNR data in the static case presented in Chapter 5. The quantitative values indicate that the VN reconstructions outperform the ICTGV reconstructions in all cases for all sampling patterns. Our comparison of VISTA and VRS is similar to observations in [9]. VISTA sampling outperforms VRS due to the controlled and optimally distributed sampling of high- and low frequency data across time-frames.

While we see improvements of the presented VN approach in terms of image quality, there also exists an obvious potential in the design of temporal regularization. The cross-sectional views in the presented figures show a slightly higher error density for VN reconstructions compared to ICTGV reconstructions in certain regions. Here, different regularization strategies could be considered to improve the temporal regularization.

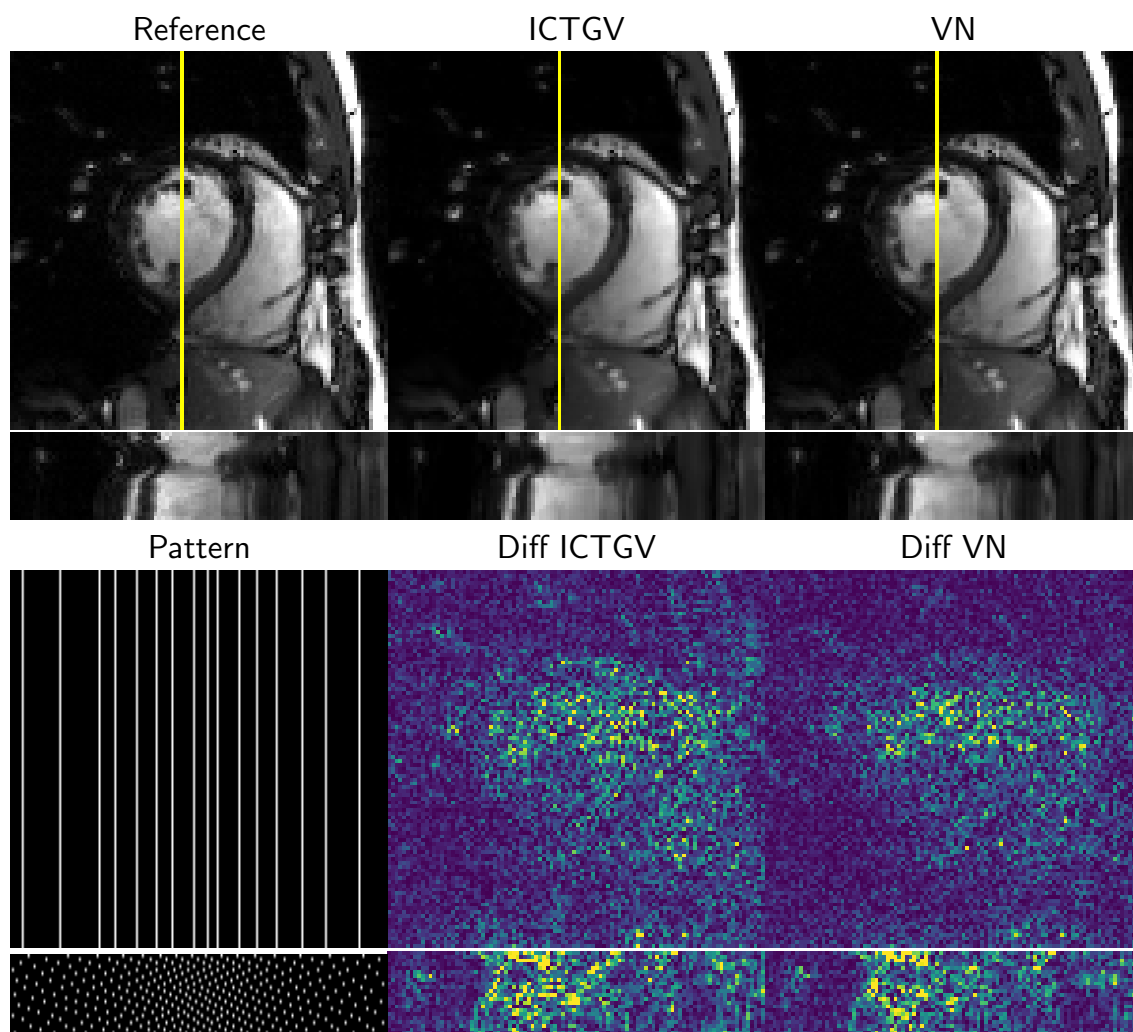
Besides improved image quality, the major advantage of using VNs for image reconstruction is the reduced runtime compared to PI-CS ICTGV. While the reconstruction time for ICTGV is  $\sim 45$ s, VN reconstructions can be obtained in  $\sim 7$ s, which is almost  $7\times$  faster than the considered state-of-the-art approach.

All our experiments were conducted on CINE cardiac MR data of 5 healthy volunteers. In general, CINE cardiac imaging is an important modality in dynamic multi-coil MR image reconstruction. However, it poses an interesting challenge for learning-based approaches to generate high-quality reference data due to limitations in breath-hold capabilities and signal preparation. For bSSFP data, typical acquisition artifacts such as banding might impact the

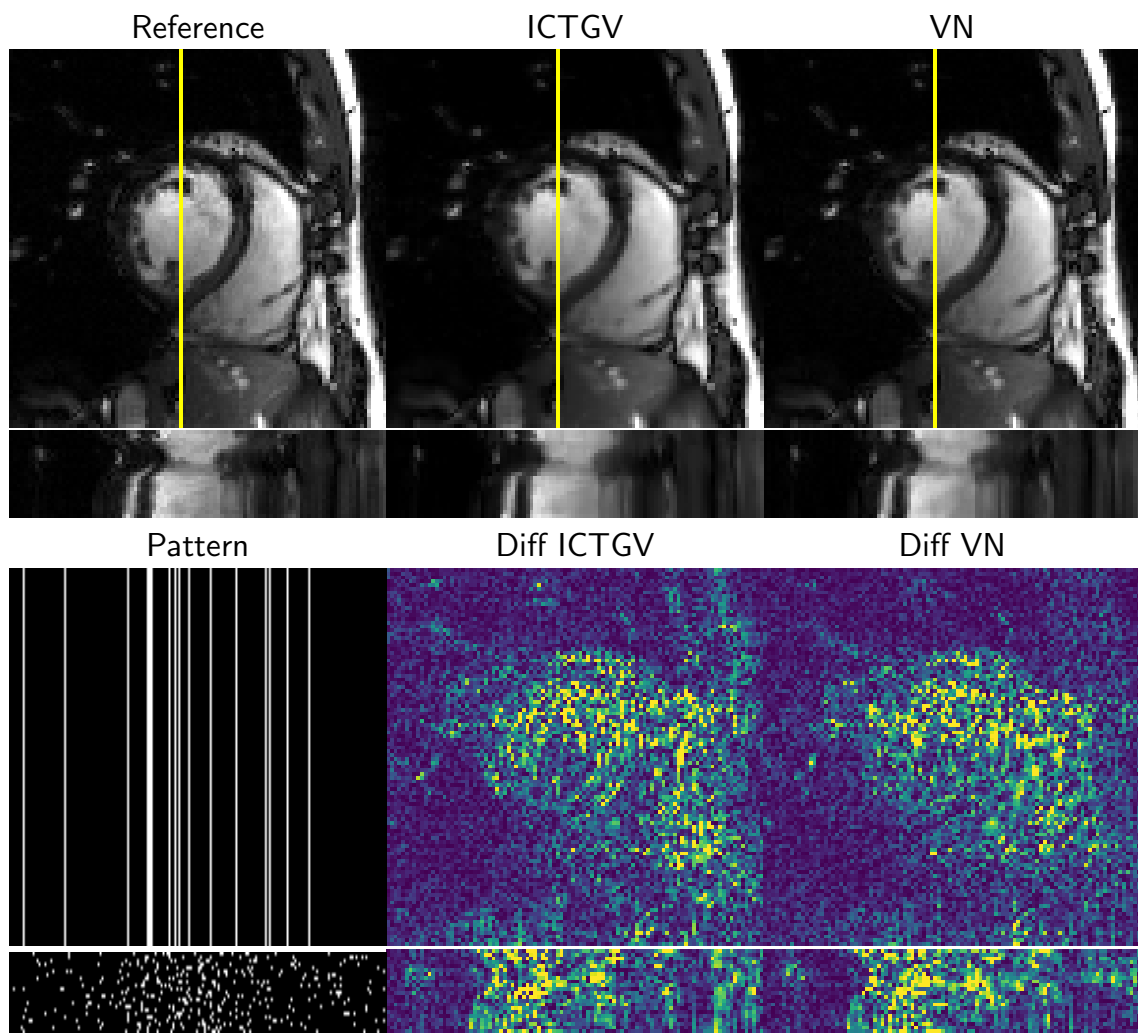


**Figure 6.3:** bSSFP: 2CH view for VISTA sampling at  $R = 8$ . The first row shows the reference, ICTGV and VN reconstruction. The second row illustrates the sampling pattern and difference images of ICTGV and VN to the reference.

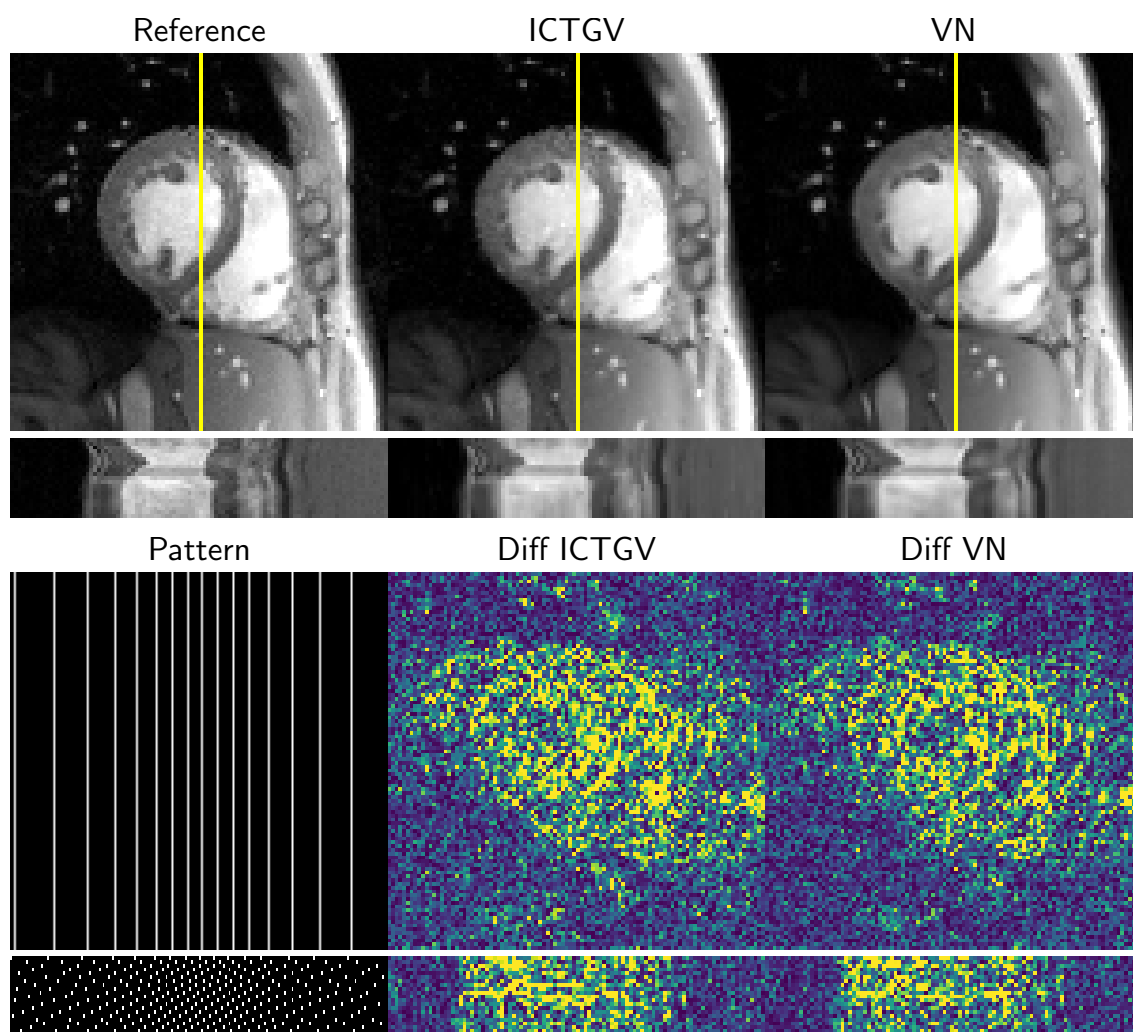
quality of learning. In contrast, FLASH contains less acquisition artifacts, but is characterized by a poorer SNR, leading to an unnatural behaviour of ICTGV and slightly smoothed results of the VN. In future work, we want to further explore the impact of low-quality training data and artifacts in the training data. Furthermore, we will consider more data to train the VNs. We will also investigate Dynamic Contrast Enhanced (DCE) data or angiography data, where a major challenge is the acquisition of reference data. Hence, further solutions are required such as generating artificial, numerical phantoms for training or discovering further solutions in unsupervised learning.



**Figure 6.4:** bSSFP: SA view for VISTA sampling at  $R = 12$ . The first row shows the reference, ICTGV and VN reconstruction. The second row illustrates the sampling pattern and difference images of ICTGV and VN to the reference.

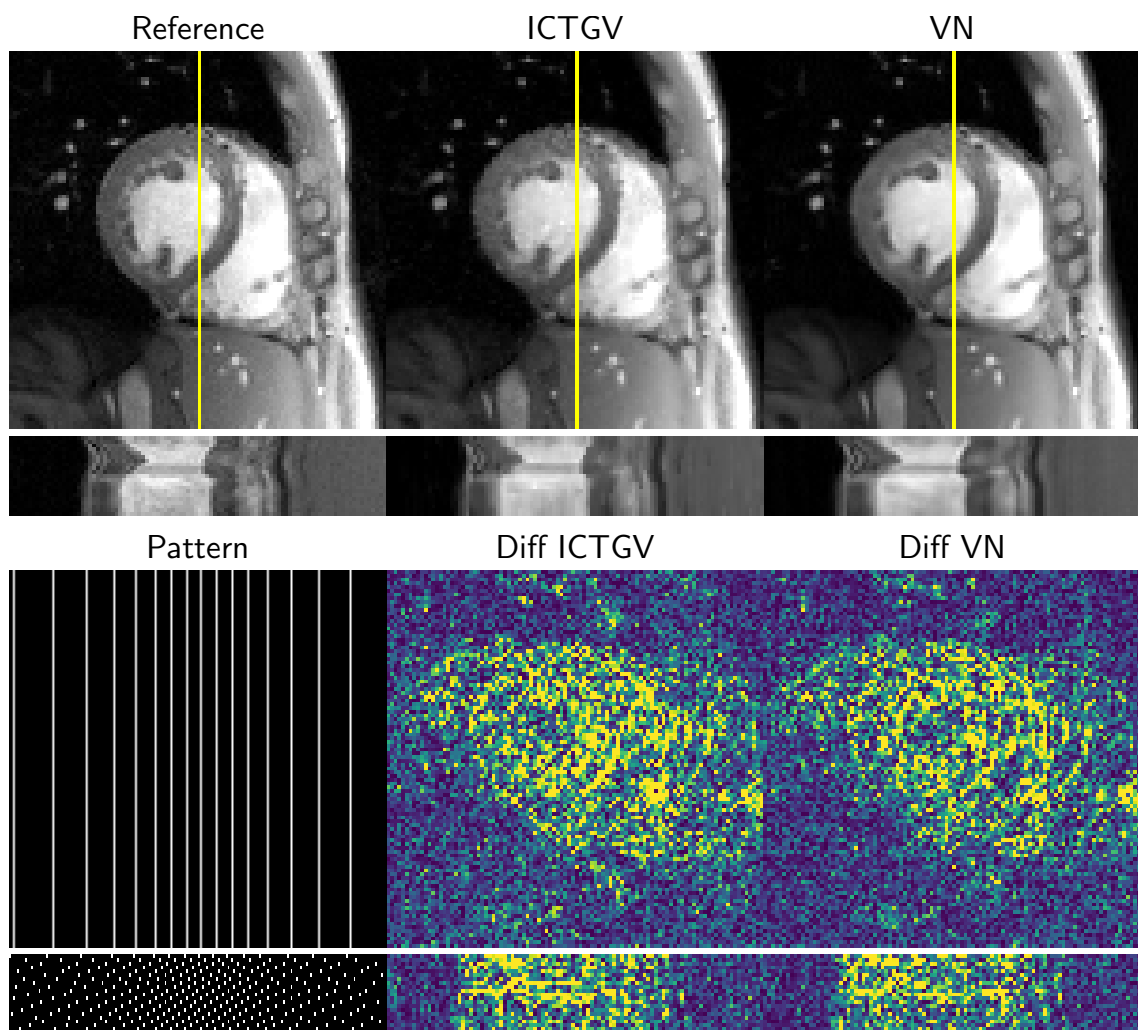


**Figure 6.5:** bSSFP: SA view for VRS sampling at  $R = 12$ . The first row shows the reference, ICTGV and VN reconstruction. The second row illustrates the sampling pattern and difference images of ICTGV and VN to the reference.

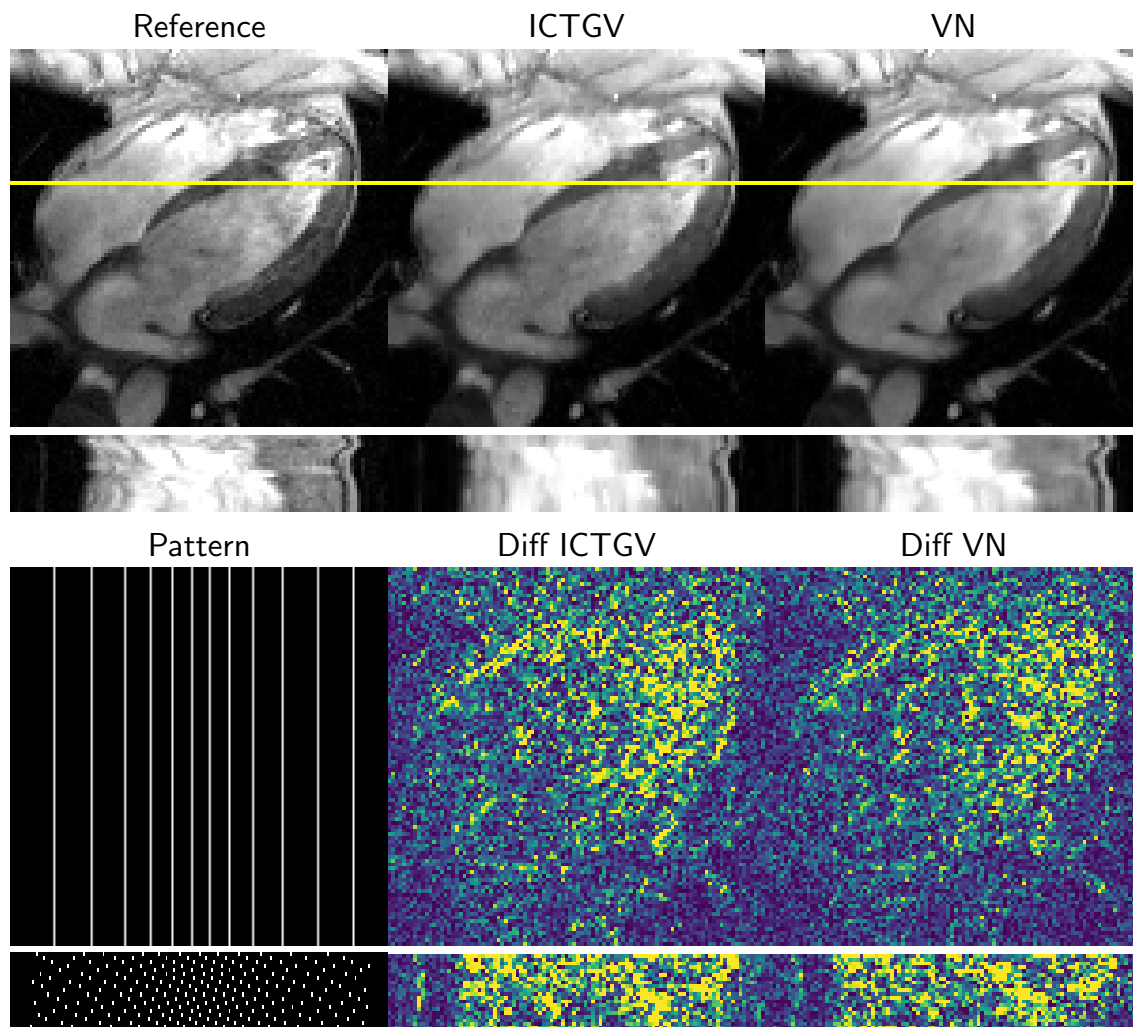


**Figure 6.6:** FLASH: SA view for VISTA sampling at  $R = 12$ . The first row shows the reference, ICTGV and VN reconstruction. The second row illustrates the sampling pattern and difference images of ICTGV and VN to the reference.

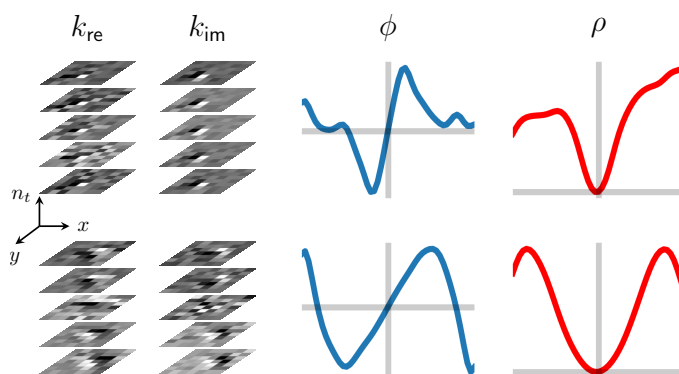




**Figure 6.7:** FLASH: SA view for VRS sampling at  $R = 12$ . The first row shows the reference, ICTGV and VN reconstruction. The second row illustrates the sampling pattern and difference images of ICTGV and VN to the reference.



**Figure 6.8:** FLASH: 4CH view for VRS sampling at  $R = 16$ . The first row shows the reference, ICTGV and VN reconstruction. The second row illustrates the sampling pattern and difference images of ICTGV and VN to the reference.



**Figure 6.9:** Learned VN parameters for dynamic CINE MRI.

## Conclusion and Outlook

Wherever you go, go with all your heart.

---

*Confucius*

While machine learning for classification and segmentation has been established for medical tasks in the last years, the application to medical image reconstruction is rather new. In 2016, first promising research results started a new era for novel and exciting developments in the field of medical image reconstruction. During the past two years, many different learning-based approaches were proposed for medical image *reconstruction*, even though many Convolutional Neural Networks (CNNs) architectures were applied for *image enhancement*, while others considered the raw measurement data. Based on the mathematical properties of the individual approaches, we provided a broad overview of the recent developments for image reconstruction in computer vision and medical imaging. It is noticeable that the network architectures that regard reconstruction as image enhancement are mainly based on the U-net architecture [205] containing a huge number of model parameters. Huge model architectures require a great amount of training data to avoid over-fitting which is hardly feasible in clinical practice. Furthermore, the consistency to the raw measurement data is lost during image enhancement, which, however, provide important information for the reconstruction procedure. In cases where the forward model is unknown, a direct mapping from the raw measurement data to the image might have a great potential. For example, the approaches that learn a direct mapping might be useful in applications where the forward imaging model is only known with some degree of uncertainty. In contrast, learning a fully iterative approach has several advantages: We often have prior knowledge about the acquisition process and physics of the model available. This allows us to incorporate this prior knowledge in the reconstruction process and enforce *data consistency* in various variants. As this prior knowledge greatly impacts the reconstruction process, already small and compact network architectures with a comparably low number of network parameters results in promising reconstruction results of high quality. This also reduces the amount of training data and the risk of over-fitting.

Our mission statement is to use prior knowledge and impose *structure* to our models. This led to the concept of Variational Networks (VNs) that combine two successful fields: deep learning and variational models. Due to the close connection to variational models, we gain first insights into the meaning of the learned parameters. The learned Fields of Experts (FoE) regularizer allows us to draw first connections to already known properties of filter kernels and potential functions. Furthermore, the proposed VN can be seen as a global framework, which can easily be adapted to various applications for computer vision and medical imaging, ranging from image enhancement to image reconstruction or segmentation. We believe that by using structured models, we are able to compete with approaches that consist of potentially too many parameters, while only requiring little training data.

For our applications, we focused on supervised learning for a variety of medical applications. Most of these applications have in common that the reconstructions contain characteristic coherent artifacts which cannot be suppressed by classical iterative methods. In these cases, learning-based approaches can overcome typical limitations of classical approaches, including parameter tuning, handcrafted feature engineering and long reconstruction times. Our results for enhancing limited-angle Computed Tomography (CT) images and Cartesian undersampled Magnetic Resonance (MR) knee imaging proved to yield high quality results compared to other state-of-the-art methods. As the coherent artifacts spread over the whole image, more complex regularizers with an increased perceptive field might be beneficial to further increase the reconstruction quality and suppress artifacts. The presented deep regularization provides a promising outlook in this direction and will be investigated further in future work. In the case of dynamic MR imaging, more emphasis will be given to a more balanced regularization between spatial and temporal domain, e.g., by using different composite regularizers or Infimal Convolution (IC)-type regularization.

Major focus was given to design a clinically useful Parallel Imaging (PI) solution using VNs for MR image reconstruction, i.e., using the raw multi-coil data. In our case, we use an image-based PI method that requires explicit estimation of coil sensitivity maps, which are currently pre-computed. To avoid potential errors in this pre-computation step, we might estimate the coil sensitivities and reconstruction jointly or use setups that are independent of an explicit estimation of coil sensitivity maps.

In this thesis, we also observed that besides a suitable network architecture and optimizer, the loss function is a key ingredient for successful, high quality reconstruction. However, most of the commonly used similarity measures, including but not limited to Normalized Root Mean Squared Error (NRMSE), Peak Signal-To-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), reflect the human observations poorly, especially if the reference data itself is noisy. This observation poses the future challenge to develop novel loss functions that reflect the radiologists' eyes and to deal with noisy measurement data and reference data. Adversarial approaches might provide a starting point to this issue, however, special care should be taken in the case of medical imaging to verify that no artificial structures are introduced. Hence, a thorough clinical evaluation is required.

Clinical evaluation is not only required on image quality but also more importantly on diag-

nostic content. To allow for a potential translation of the proposed VNs, we focused specifically on providing a clinical useful environment in our experiments. In a first reader study on image quality for clinical MR knee imaging we experienced improved results for the proposed VN compared to classical Compressed Sensing (CS) approaches for four times Cartesian under-sampled data. Although the image quality is reduced compared to the reference images and some residual artifacts still remain, the open question is if this impacts the diagnostic value of the learning-based reconstruction compared to a clinically accelerated protocol. This question directs future research to large-scale clinical reader studies as well as improved quantitative measures that allow for visually improved image reconstructions. Due to the increased research in the area of medical image reconstruction, the public availability of databases will be a key aspect in the future of medical imaging. This will allow for better training of architectures and also provide a basis for comparing different approaches which was hardly possible because every institution used its own dataset for evaluation. Training data is still the key to successfully train all kinds of network architectures based on supervised learning. In some applications such as dynamic MR imaging it is challenging to acquire suitable training data. This greatly raises the potential for unsupervised learning methods in future work to learn representations from only corrupted measurement data.

Even though large benchmarks exist, a major goal should not be to bet benchmarks in terms of quantitative numbers, but to come up with realistic solutions that can be adopted to clinical routine exams. However, we often see only the tip of the iceberg as in Figure 7.1 which are amazingly looking results. Hence, the goal of this thesis was not to provide an highly optimized network architecture with fine-tuned parameter settings, but to provide a broad overview over the importance and potential pitfalls of data acquisition, data setup, network as well as loss function design and optimization.



**Figure 7.1:** The iceberg theory of deep learning. Original image by ©Ralph A. Clevenger/CORBIS. Modified for educational purpose only.





## List of Acronyms

ACL	Auto-Calibration Line
ADMM	Alternating Direction Method of Multipliers
AI	Artificial Intelligence
ALOHA	Annihilating Filter-Based Low-Rank Hankel Structured Matrix Completion Approach
ANN	Artificial Neural Network
AUTOMAP	Automated Transform by Manifold Approximation
BP	Backprojection
CG	Conjugate Gradient
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CS	Compressed Sensing
CT	Computed Tomography
DCE	Dynamic Contrast Enhanced
DDR	Digitally Reconstructed Radiograph
DICOM	Digital Imaging and Communications in Medicine
DTI	Diffusion Tensor Imaging
ECG	Electrocardiography
EPI	Echo Planar Imaging
FBP	Filtered Back-Projection
FE	Frequency Encoding
FoE	Fields of Experts
FoV	Field of View
FSIM	Feature Similarity Index
FT	Fourier Transform
GAN	Generative Adversarial Network
GD	Gradient Descent

---

GPU	Graphics Processing Unit
GRAPPA	Generalized Autocalibrating Partially Parallel Acquisitions
GRE	Gradient Echo
IC	Infimal Convolution
ICTGV	Infimal-Convolution-Total-Generalized-Variation
IIPG	Inertial Incremental Proximal Gradient
IPALM	Inertial Proximal Alternating Linearized Minimization
ISTA	Iterative Shrinkage and Thresholding Algorithm
KSAE	K-sparse Autoencoder
L+S	Low-Rank plus Sparse
L-BFGS	Limited-Memory Broyden-Fletcher-Goldfarb-Shanno
l.s.c.	lower-semicontinuous
LISTA	Learned Iterative Shrinkage and Thresholding Algorithm
LVOT	Left Ventricular Outflow Tract
MAE	Mean Absolute Error
MAP	Maximum-A-Posteriori
MAR	Metal Artifact Correction
MD	Medical Doctor
MLP	Multi Layer Perceptron
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MS-SSIM	Multi-Scale Structural Similarity Index
MSE	Mean Squared Error
NLM	Non-Local Means
NRMSE	Normalized Root Mean Squared Error
OMP	Orthogonal Matching Pursuit
PAT	Photoacoustic Tomography
PD	Primal-Dual
PE	Phase Encoding
PET	Positron Emission Tomography
PG	Proximal Gradient
PI	Parallel Imaging
PSF	Point Spread Function
PSNR	Peak Signal-To-Noise Ratio
RBF	Gaussian radial basis function
ReLU	Rectified Linear Unit
RF	Radio Frequency
RMSE	Root-Mean-Squared-Error
ROI	Region Of Interest
SA	Short Axis



SAR	Specific Absorption Rate
SE	Spin Echo
SENSE	Sensitivity Encoding
SMASH	Simultaneous Acquisition of Spatial Harmonics
SNR	Signal-to-Noise Ratio
SPECT	Single Photon Emission Computed Tomography
SQS	Separable Quadratic Surrogate
SSIM	Structural Similarity Index
SVD	Singular Value Decomposition
TE	Echo Time
TGV	Total Generalized Variation
TR	Repetition Time
TSE	Turbo Spin Echo
TV	Total Variation
VISTA	Variable Density Incoherent Spatio-Temporal Acquisition
VN	Variational Network
VRS	Variable Density Random Sampling
wGAN	Wasserstein Generative Adversarial Network
ZTE	Zero Echo Time





## List of Publications

My work at the Institute of Computer Graphics and Vision led to the following peer-reviewed publications. The list includes also publications that are currently under review.

### Book Chapters

K. Hammernik and F. Knoll. Machine Learning for Image Reconstruction. In D. Rueckert, G. Fichtinger, and S. K. Zhou, editors, *Handbook of Medical Image Computing and Computer Assisted Intervention*. Elsevier, 2018 (under review)

### Patents

F. Knoll, K. Hammernik, T. Pock, and D. K. Sodickson. System, Method And Computer-Accessible Medium For Learning An Optimized Variational Network For Medical Image Reconstruction, 2017

### Journal Publications

K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, 2018, <https://github.com/VLOGroup/mri-variationalnetwork>

F. Knoll, K. Hammernik, E. Kobler, T. Pock, M. P. Recht, and D. 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

J. Yao, J. E. Burns, D. Forsberg, A. Seitel, A. Rasoulian, P. Abolmaesumi, K. Hammernik, M. Urschler, B. Ibragimov, R. Korez, T. Vrtovec, I. Castro-Mateos, J. M. Pozo, A. F. Frangi, R. M. Summers, and S. Li. A Multi-Center Milestone Study of Clinical Vertebral CT Segmentation. *Computerized Medical Imaging and Graphics*, 49:16–28, 2016

E. M. Littringer, S. Zellnitz, K. Hammernik, V. Adamer, H. Friedl, and N. A. Urbanetz. Spray Drying of Aqueous Salbutamol Sulfate Solutions Using the Nano Spray Dryer B-90 - The Impact of Process Parameters on Particle Size. *Drying Technology*, 31(12):1346–1353, 2013

## Conference Contributions

### 2019

K. Hammernik, M. Schloegl, R. Stollberger, and T. Pock. Dynamic Multicoil Reconstruction using Variational Networks. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, page to appear, 2019

M. V. Zibetti, A. Sharafi, K. Hammernik, F. Knoll, and R. R. Regatte. Variational Networks for Accelerating Biexponential 3D-T1 $\rho$  Mapping of Knee Cartilage. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, page to appear, 2019

### 2018

F. Thaler, K. Hammernik, C. Payer, M. Urschler, and D. Stern. Computed Tomography Slice Reconstruction from a Limited Number of Digitally Reconstructed Radiographs Using Wasserstein Generative Adversarial Networks. In *Machine Learning for Medical Image Reconstruction*, 2018

E. Kobler, M. Muckley, B. Chen, F. Knoll, K. Hammernik, T. Pock, D. Sodickson, and R. Otazo. Variational Deep Learning for Low-Dose Computed Tomography. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6687–6691, 2018

K. Hammernik, E. Kobler, T. Pock, M. Recht, D. K. Sodickson, and F. Knoll. Variational Adversarial Networks for Accelerated MR Image Reconstruction. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, page 1091, 2018

F. Knoll, K. Hammernik, E. Kobler, T. Pock, M. P. Recht, and D. K. Sodickson. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, 2018

F. Knoll, K. Hammernik, E. Kobler, T. Pock, D. K. Sodickson, and M. P. Recht. Analysis of the influence of deviations between training and test data in learned image reconstruction. In *ISMRM Workshop on Machine Learning*, 2018

M. V. Zibetti, A. Sharafi, K. Hammernik, F. Knoll, and R. R. Regatte. Comparing Learned Variational Networks and Compressed Sensing for T1 $\rho$  Mapping of Knee Cartilage. In *ISMRM Workshop on Machine Learning, Part II*, 2018

## 2017

K. Hammernik, T. Würfl, T. Pock, and A. Maier. A deep learning architecture for limited-angle computed tomography reconstruction. In *Bildverarbeitung für die Medizin 2017*, pages 92–97, 2017

K. Hammernik, F. Knoll, D. Sodickson, and T. Pock. On the Influence of Sampling Pattern Design on Deep Learning-Based MRI Reconstruction. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, page 644, 2017

K. Hammernik, F. Knoll, D. Sodickson, and T. Pock. L2 or Not L2: Impact of Loss Function Design for Deep Learning MRI Reconstruction. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, page 687, 2017

K. Hammernik, T. Pock, and R. Nuster. Variational photoacoustic image reconstruction with spatially resolved projection data. *Proceedings of SPIE, Photons Plus Ultrasound: Imaging and Sensing*, 10064:100643I, 2017

F. Knoll, K. Hammernik, E. Garwood, A. Hirschmann, L. Rybak, M. Bruno, T. Block, J. Babb, T. Pock, D. Sodickson, and M. Recht. Accelerated Knee Imaging Using a Deep Learning Based Reconstruction. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, page 645, 2017

E. Kobler, T. Klatzer, K. Hammernik, and T. Pock. Variational networks: Connecting variational methods and deep learning. In *German Conference on Pattern Recognition*, pages 281–293, 2017, <https://github.com/VLOGroup/denoising-variationalnetwork>

T. Klatzer, D. Soukup, E. Kobler, K. Hammernik, and T. Pock. Trainable Regularization for Multi-frame Superresolution. In *German Conference on Pattern Recognition*, pages 90–100, 2017

**2016**

K. Hammernik, F. Knoll, D. K. Sodickson, and T. Pock. Learning a Variational Model for Compressed Sensing MRI Reconstruction. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, number 24, page 1088, 2016

T. Klatzer, K. Hammernik, P. Knöbelreiter, and T. Pock. Learning Joint Demosaicing and Denoising Based on Sequential Energy Minimization. In *IEEE International Conference on Computational Photography*, pages 1–11, 2016

M. Urschler, K. Hammernik, T. Ebner, and D. Štern. Automatic Intervertebral Disc Localization and Segmentation in 3D MR Images Based on Regression Forests and Active Contours. In *Computational Methods and Clinical Applications for Spine Imaging*, pages 130–140. Springer International Publishing, 2016

**2015**

K. Hammernik, T. Ebner, D. Stern, M. Urschler, and T. Pock. Vertebrae Segmentation in 3D CT Images based on a Variational Framework. In *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*, Lecture Notes in Computational Vision and Biomechanics, pages 227–233. Springer, 2015

For network training, we consider the following optimization problem

$$\begin{aligned} & \min_{\theta} \mathcal{L}(x^T(\theta), x_{\text{ref}}) \\ & \theta = (\theta^1, \dots, \theta^T), \quad \theta^t = (w_{ij}^t, k_i^t, \lambda^t), \quad 1 \leq t \leq T \\ \text{s.t. } & \theta^t \in \mathcal{C} = \left\{ \lambda^t \geq 0, \xi^\top k_i^t = 0 \forall i, \|k_i^t\|_2 \leq 1 \forall i \right\}. \end{aligned} \quad (\text{C.1})$$

To solve this non-convex training problem, we exploit different first-order optimizers to obtain parameters updates  $\theta^t = (w_{ij}^t, k_i^t, \lambda^t)$  for the  $m^{\text{th}}$  iteration. To realize additional constraints on the parameters, we perform the projections during optimization

$$(\lambda^{m+1}, k^{m+1}) = \text{proj}_{\mathcal{C}}^{\eta}(\tilde{\lambda}^{m+1}, \tilde{k}^{m+1}).$$

As the constraints do not depend on each other, we can consider the projections independently. To realize the non-negativity constraint on the data term weights  $\lambda^{m+1}$ , the parameter update  $\tilde{\lambda}^{m+1}$  is clamped at zero

$$\lambda^{m+1} = \max(0, \tilde{\lambda}^{m+1}).$$

For the projection onto the filter kernel constraints, we first subtract the means  $\xi^\top \tilde{k}^{m+1}$  from the current kernel parameter estimates and then project the kernel onto the unit-sphere

$$\begin{aligned} \tilde{k}_{\xi}^{m+1} &= \tilde{k}^{m+1} - \xi^\top \tilde{k}^{m+1}, \\ k^{m+1} &= \frac{\tilde{k}_{\xi}^{m+1}}{\|\tilde{k}_{\xi}^{m+1}\|_2}. \end{aligned}$$

**Algorithm 8** Inertial Incremental Proximal Gradient (IIPG)**Input:** Split training set  $\mathcal{S}$  into  $N_B$  mini batches  $\mathcal{B}$  such that  $\mathcal{S} = \bigcup_{b=1}^{N_B} \mathcal{B}_b$ **Choose:** Step size  $\eta$ , number of epochs  $N_E$ , initial parameters  $\theta^0$ **for**  $e = 1, \dots, N_E$  **do**

$$\beta_e = \frac{e-1}{e+2}$$

**for**  $b = 1, \dots, N_B$  **do** $m = N_B(e-1) + b$  // Current iterate

$$\hat{\theta}^{m+1} = \theta^m + \beta_e(\theta^m - \theta^{m-1})$$
 // Over-relaxation

$$g^{m+1} = \frac{\partial \mathcal{L}(x^T(\hat{\theta}^{m+1}), x_{\text{ref}})}{\partial \theta}$$
 // Compute gradient on current mini batch  $\mathcal{B}_b$

$$\tilde{\theta}^{m+1} = \hat{\theta}^{m+1} - \eta g^{m+1}$$
 // Compute gradient step

$$\theta^{m+1} = \text{proj}_{\mathcal{C}}^{\eta}(\tilde{\theta}^{m+1})$$
 // Compute projections

**end for****end for**

## C.1 Inertial Incremental Proximal Gradient (IIPG)

For our work in [84, 136], we used the Inertial Incremental Proximal Gradient (IIPG) optimizer. This IIPG variant of projected gradient descent is related to the Inertial Proximal Alternating Linearized Minimization (IPALM) algorithm [188]. The whole sequence generated by IPALM is guaranteed to converge to a stationary point in the non-convex non-stochastic case under certain constraints on the step size and inertial parameters. The analysis for the stochastic version is left to future research. In the IIPG Algorithm 8, the parameter updates are calculated on a single mini batch. First, we perform over-relaxation where we set a over-relaxation constant  $\beta_e$  depending on the current epoch  $e$  to achieve moderate acceleration. Second, we compute the gradient with respect to the parameters on the current mini batch which yields a new parameter update  $\tilde{\theta}^{m+1}$  for the current iteration  $m$ . Finally, the parameter update is projected onto the defined constraint if defined.

## C.2 Proximal ADAM with Block-Preconditioning

The ADAM optimizer [129] is a popular and stable algorithm for stochastic optimization and commonly used to solve deep learning problems. It involves first-order and second-order moment estimates  $v_1$  and  $v_2$ , which are used to scale the effective step size of the individual parameters. While  $v_1$  involves the gradient of the parameters,  $v_2$  is computed as the square of the parameter gradients, which results in a point-wise scaling of the parameters  $v_1/\sqrt{v_2}$ . When observing problem Equation (C.1) we see that we are dealing with different blocks of parameters that additionally require a proximal mapping of a function  $G$ , which are in our case



**Algorithm 9** Proximal ADAM with Block-Preconditioning**Choose:** Step size  $\eta$ , number of iterations  $N_{\max}$ , initial parameters  $\theta^0$ **Choose:** Exponential decay rates for the moment estimates  $\beta_1, \beta_2 \in [0, 1)$ **for**  $m = 1, \dots, N_{\max}$  **do** $g^m = \nabla_{\theta} \mathcal{L}(x^T(\theta^{m-1}), x_{\text{ref}})$  // Compute gradient on current mini batch $v_1^m = \beta_1 v_1^{m-1} + (1 - \beta_1) g^m$  // Update of the biased first moment estimateCompute block-wise gradient norms to define the scaling  $g_{\text{bl}}$  according to Equation (C.3) $v_2^m = \beta_2 v_2^{m-1} + (1 - \beta_2) g_{\text{bl}}^m$  // Update of the second moment estimate $\hat{v}_1^m = v_1^m / (1 - \beta_1^m)$  // Bias-correction of the first moment estimate $\hat{v}_2^m = v_2^m / (1 - \beta_2^m)$  // Bias-correction of the second moment estimate $\hat{\theta}^m = \theta^{m-1} - (\eta v_1^m) / (\sqrt{\hat{v}_2^m} + \epsilon)$  // Parameter update $\theta^m = \text{proj}_{\mathcal{C}}^{\eta / (\sqrt{\hat{v}_2^m} + \epsilon)}(\hat{\theta}^m)$ **end for**projections, in a defined norm  $D$ :

$$x^* \in \arg \min_{x \in \mathbb{K}^{N_x}} G(x) + \frac{1}{2} \|x - \hat{x}\|_D^2 \quad (\text{C.2})$$

Inspired by [187, 188], we aim at performing block-wise scaling of the gradients which therefore do not influence our projection calculations. Hence, we define the block-wise scaling for  $B$  parameter blocks as

$$g_{\text{bl}}^m = \text{diag} \left( \|g_1^m\|_2^2, \dots, \|g_1^m\|_2^2, \|g_2^m\|_2^2, \dots, \|g_2^m\|_2^2, \dots, \|g_B^m\|_2^2, \dots, \|g_B^m\|_2^2 \right)^\top \quad (\text{C.3})$$

where the elements  $g_{\text{bl}}^m$  are repeated according to the size of the  $b^{\text{th}}$  parameter block. This results in the proximal ADAM with block-preconditioning algorithm shown in Algorithm 9.





## Gradient Calculations

### D.1 Variational Network Parameters

In every gradient step  $t$ , we seek the derivatives with respect to the parameters  $\theta^t = (w_{ij}^t, k_i^t, \lambda^t)$  of a loss function  $\mathcal{L}(x^T(\theta), x_{\text{ref}})$ . For simplicity, we drop the dependency of  $x^T$  on the parameters  $\theta$  and the subscript  $s$  and show the calculations only for a single training example. The gradient steps are given as

$$x^t = x^{t-1} - \sum_{i=1}^{N_k} (K_i^t)^\top \phi_i^t (K_i^t x^{t-1}) - \lambda^t A^* (Ax^{t-1} - y), \quad 1 \leq t \leq T$$

The derivatives with respect to the parameters  $\theta^t$  are obtained by back-propagation [147]

$$\frac{\partial \mathcal{L}(x^T(\theta), x_{\text{ref}})}{\partial \theta^t} = \frac{\partial x^t}{\partial \theta^t} \cdot \underbrace{\frac{\partial x^{t+1}}{\partial x^t} \cdots \frac{\partial x^T}{\partial x^{T-1}}}_{e^t} \cdot \frac{\partial \mathcal{L}(x^T(\theta), x_{\text{ref}})}{\partial x^T}$$

The reconstruction error of the  $t^{\text{th}}$  gradient step is given by  $\frac{\partial \mathcal{L}(x^T(\theta), x_{\text{ref}})}{\partial x^t} = e^t$ .

**Derivative of the Loss Function.** First, we require the gradient of the loss function  $\mathcal{L}$  with respect to the reconstruction  $x^T$  defined as  $e^T$ . It is computed as

$$\frac{\partial \mathcal{L}(x^T(\theta), x_{\text{ref}})}{\partial x^T} = e^T$$

Different loss function along with their derivatives are listed in Appendix D.2.

**Derivative of the Data Term Weights  $\lambda^t$ .** The derivative of the reconstruction  $x^T$  wrt. to  $\lambda^t \in \mathbb{R}$  for the  $t^{\text{th}}$  gradient step is expressed as:

$$\frac{\partial \mathcal{L}(x^T(\theta), x_{\text{ref}})}{\partial \lambda^t} = \frac{\partial x^t}{\partial \lambda^t} \frac{\partial \mathcal{L}(x^T(\theta), x_{\text{ref}})}{\partial x^t} = \langle -A^*(Ax^{t-1} - y), e^t \rangle.$$

**Derivative of the Activation Functions  $\phi_i^t$ .** A single activation function  $\phi_i^t(z) = (\phi_i^t(z_1), \dots, \phi_i^t(z_N)) : \mathbb{R}^N \rightarrow \mathbb{R}^N$  is defined by a weighted combination of  $N_w$  Gaussian radial basis functions (RBFs)

$$\phi_i^t(z_l) = \sum_{j=1}^{N_w} w_{ij}^t \exp\left(-\frac{(z_l - \mu_j)^2}{2\sigma^2}\right), \quad l = 1, \dots, N, \quad w_{ij}^t \in \mathbb{R}.$$

This can be rewritten in a matrix-vector notation

$$\phi_i^t(z) = \begin{pmatrix} \phi_i^t(z_1) \\ \vdots \\ \phi_i^t(z_N) \end{pmatrix} = \begin{bmatrix} \exp\left(-\frac{(z_1 - \mu_1)^2}{2\sigma^2}\right) & \dots & \exp\left(-\frac{(z_1 - \mu_{N_w})^2}{2\sigma^2}\right) \\ \vdots & \ddots & \vdots \\ \exp\left(-\frac{(z_N - \mu_1)^2}{2\sigma^2}\right) & \dots & \exp\left(-\frac{(z_N - \mu_{N_w})^2}{2\sigma^2}\right) \end{bmatrix} \begin{pmatrix} w_{i1}^t \\ \vdots \\ w_{iN_w}^t \end{pmatrix} = M_i^t(z) w_i^t.$$

During training, we learn the weights  $w_i^t \in \mathbb{R}^{N_w}$  and express its gradient as

$$\begin{aligned} \frac{\partial \mathcal{L}(x^T(\theta), x_{\text{ref}})}{\partial w_i^t} &= \frac{\partial x^t}{\partial w_i^t} \frac{\partial \mathcal{L}(x^T(\theta), x_{\text{ref}})}{\partial x^t} = \\ &= -\frac{\partial}{\partial w_i^t} \left\{ (K_i^t)^\top M_i^t (K_i^t x^{t-1}) w_i^t \right\} e^t = - (M_i^t (K_i^t x^{t-1}))^\top K_i^t e^t. \end{aligned}$$

**Derivative of the Intermediate Reconstructions  $x^{t-1}$ .** Further gradients with respect to the reconstructions from intermediate steps are given as

$$\frac{\partial x^t}{\partial x^{t-1}} = I - \sum_{i=1}^{N_k} (K_i^t)^\top \text{diag}((\phi^t)'(K_i^t x^{t-1})) K_i^t - \lambda^t A^* A$$

where  $I$  denotes the identity matrix. This also requires the second derivative  $(\phi^t)'_i(z)$  of the potential functions  $\rho(z)^t$ , which is expressed as

$$(\phi^t)'_i(z) = \begin{bmatrix} -\frac{(z_1 - \mu_1)}{\sigma^2} \exp\left(-\frac{(z_1 - \mu_1)^2}{2\sigma^2}\right) & \dots & -\frac{(z_1 - \mu_{N_w})}{\sigma^2} \exp\left(-\frac{(z_1 - \mu_{N_w})^2}{2\sigma^2}\right) \\ \vdots & \ddots & \vdots \\ -\frac{(z_N - \mu_1)}{\sigma^2} \exp\left(-\frac{(z_N - \mu_1)^2}{2\sigma^2}\right) & \dots & -\frac{(z_N - \mu_{N_w})}{\sigma^2} \exp\left(-\frac{(z_N - \mu_{N_w})^2}{2\sigma^2}\right) \end{bmatrix} w_i^t$$

**Derivative of the Filter Kernels  $k_i^t$ .** To compute the derivative with respect to the filter kernels  $k_i^t$  we have to introduce further relationships between our given parameters. The

convolution can be defined as matrix-vector multiplication

$$k_i^t * x^{t-1} \Leftrightarrow K_i^t x^{t-1} = X^{t-1} k_i^t$$

where the matrix  $X^{t-1} : \mathbb{R}^{2s^2} \rightarrow \mathbb{R}^N$  is a suitably shifted representation of the image  $x^{t-1}$  and  $k_i^t \in \mathbb{R}^{2s^2}$  is the vectorized filter kernel. The gradient step also involves rotated filter kernels  $\bar{k}_i^t$  due to the transpose operation of the kernel matrix  $(K_i^t)^\top$ . As we want to calculate the derivative with respect to  $k_i^t$  and not to their rotated version, we introduce a rotation matrix  $R : \mathbb{R}^{2s^2} \rightarrow \mathbb{R}^{2s^2}$  that has the same effect as the transpose operation

$$\bar{k}_i^t = R k_i^t.$$

The convolution can be rewritten as

$$(K_i^t)^\top \phi_i^t(K_i^t x^{t-1}) = \tilde{\phi}_i^t(K_i^t x^{t-1}) \bar{k}_i^t = \tilde{\phi}_i^t(K_i^t x^{t-1}) R k_i^t$$

where  $\tilde{\phi}_i^t(K_i^t x^{t-1}) : \mathbb{R}^N \rightarrow \mathbb{R}^{2s^2}$  is a suitable matrix representation of  $\phi_i^t(K_i^t x^{t-1})$ . Applying the product rule yields following expression for the kernel derivative

$$\begin{aligned} \frac{\partial (K_i^t)^\top \phi_i^t(K_i^t x^{t-1})}{\partial k_i^t} &= \frac{\partial \phi_i^t(K_i^t x^{t-1})}{\partial k_i^t} K_i^t + \frac{\partial k_i^t}{\partial k_i^t} \left[ \tilde{\phi}_i^t(K_i^t x^{t-1}) R \right]^\top = \\ &= (X^{t-1})^\top \text{diag}(\phi^t)'_i(K_i^t x^{t-1}) K_i^t + R^\top \tilde{\phi}_i^t(K_i^t x^{t-1}). \end{aligned}$$

The full derivative may be expressed as

$$\begin{aligned} \frac{\partial \mathcal{L}(x^T(\theta), x_{\text{ref}})}{\partial k_i^t} &= \frac{\partial x^t}{\partial k_i^t} \frac{\partial \mathcal{L}(x^T(\theta), x_{\text{ref}})}{\partial x^t} = \\ &= - \left[ (X^{t-1})^\top \text{diag}(\phi^t)'_i(K_i^t x^{t-1}) K_i^t + R^\top \tilde{\phi}_i^t(K_i^t x^{t-1}) \right] e^t. \end{aligned}$$

## D.2 Loss functions

The choice of a proper loss function has high impact on the training performance. We introduced the Mean Squared Error (MSE) as a loss function. We can also use other pixel-wise loss functions such as the Mean Absolute Error (MAE), or take into account higher-order derivatives. We expect also an improvement of image quality by taking a perceptual model such as the SSIM or its extension to the Multi-Scale Structural Similarity Index (MS-SSIM). In the following, we describe loss functions and derive the gradients. These functions can be easily integrated and evaluated in our variational network. For simplicity, we drop the dependence on the samples in the definitions of the loss functions.

**MSE.** A common loss function is the MSE where we penalize the squared pixel difference

$$\mathcal{L}_{\text{mse}}(x^T(\theta), x_{\text{ref}}) = \frac{1}{2N} \|x^T(\theta) - x_{\text{ref}}\|_2^2.$$

where  $N$  denotes the number of pixels. Its derivative is given as

$$\frac{\partial \mathcal{L}_{\text{mse}}(x^T(\theta), x_{\text{ref}})}{\partial x^T(\theta)} = e^T = \frac{1}{N} (x^T(\theta) - x_{\text{ref}}).$$

**MAE.** Another possibility is to penalize the absolute difference in pixel intensities

$$\mathcal{L}_{\text{mae}}(x^T(\theta), x_{\text{ref}}) = \frac{1}{N} |x^T(\theta) - x_{\text{ref}}|,$$

with derivative

$$\frac{\partial \mathcal{L}_{\text{mae}}(x^T(\theta), x_{\text{ref}})}{\partial x^T(\theta)} = e^T = \frac{1}{N} \text{sgn}(x^T(\theta) - x_{\text{ref}}),$$

where the  $\text{sgn}$  function is applied in an element-wise manner.

**MSE on  $\epsilon$ -smoothed magnitude images (Magnetic Resonance Imaging (MRI) only).**

In the presence of complex-valued images as apparent in MRI, we observed that computing the MSE on magnitude images leads to more visually appealing results. We define this loss as

$$\frac{\partial \mathcal{L}_{\text{mse-abs}}(x^T(\theta), x_{\text{ref}})}{\partial x^T(\theta)} = \frac{1}{2N} \| |x^T(\theta)|_{\epsilon} - |x_{\text{ref}}|_{\epsilon} \|_2^2, \quad |x|_{\epsilon} = \sqrt{x_{\text{re}}^2 + x_{\text{im}}^2 + \epsilon},$$

where  $|\cdot|_{\epsilon}$  is understood in a point-wise manner. Its derivative is computed as

$$\frac{\partial \mathcal{L}_{\text{mse-abs}}(x^T(\theta), x_{\text{ref}})}{\partial x^T(\theta)} = e^T \Leftrightarrow e_l^T = \frac{x_l^T}{|x_l^T|_{\epsilon}} (|x_l^T|_{\epsilon} - |x_{\text{ref},l}|_{\epsilon}), \quad l = 1, \dots, N.$$

**SSIM.** Instead of looking at pixel-wise measures introduced above, we can also use perceptual models as a loss function. One example for a perceptual model is the SSIM which evaluates patch statistics

$$\mathcal{L}_{\text{ssim}}(x^T(\theta), x_{\text{ref}}) = 1 - \text{SSIM}(x^T(\theta), x_{\text{ref}})$$

with

$$\begin{aligned} \text{SSIM}(x, x_{\text{ref}}) &= \frac{1}{N} \sum_{i \in \Omega} l(x(i), x_{\text{ref}}(i)) c(x(i), x_{\text{ref}}(i)) \\ &= \frac{1}{N} \sum_{i \in \Omega} \frac{2\mu_{x(i)}\mu_{x_{\text{ref}}(i)} + c_1}{\mu_{x(i)}^2 + \mu_{x_{\text{ref}}(i)}^2 + c_1} \frac{2\sigma_{x(i)}\sigma_{x_{\text{ref}}(i)} + c_2}{\sigma_{x(i)}^2 + \sigma_{x_{\text{ref}}(i)}^2 + c_2}, \end{aligned}$$

where  $c_1$  and  $c_2$  are some user-defined constants,  $N$  is the total number of pixels. The SSIM consists of a luminance term  $l$  and a contrast term  $c$  and involves means  $\mu_{\cdot(i)}$  and standard deviations  $\sigma_{\cdot(i)}$  for each patch around the current pixel position  $i$  in a given domain  $\Omega$ . Now, let us define the SSIM loss over the whole image instead of single patches. We calculate mean images  $\mu_{x^T}, \mu_g$  as Gaussian weighted average images where  $K_{\sigma_m} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  denotes the Gaussian convolution operator with standard deviation  $\sigma_m$  and  $x^T, x_{\text{ref}} \in \mathbb{R}^n$  are images represented as column vectors

$$\mu_{x^T} = K_{\sigma_m} x^T, \quad \mu_{x_{\text{ref}}} = K_{\sigma_m} x_{\text{ref}}.$$

The covariance and variance are represented in a similar way. We also introduce the notation  $D_{\mu_{x_{\text{ref}}}} = \text{diag } \mu_{x_{\text{ref}}}, D_{x_{\text{ref}}} = \text{diag } x_{\text{ref}}, D_{\mu_{x^T}} = \text{diag } \mu_{x^T}, D_{x^T} = \text{diag } x^T$  and

$$\begin{aligned} \sigma_{x^T, x_{\text{ref}}} &= K_{\sigma_m} (x_{\text{ref}} \odot x^T) - \mu_{x_{\text{ref}}} \odot \mu_{x^T} = K_{\sigma_m} D_{x_{\text{ref}}} x^T - D_{\mu_{x_{\text{ref}}}} \mu_{x^T}, \\ \sigma_{x^T}^2 &= K_{\sigma_m} (x^T \odot x^T) - \mu_{x^T} \odot \mu_{x^T} = K_{\sigma_m} D_{x^T} x^T - D_{\mu_{x^T}} \mu_{x^T}, \\ \sigma_{x_{\text{ref}}}^2 &= K_{\sigma_m} (x_{\text{ref}} \odot x_{\text{ref}}) - \mu_{x_{\text{ref}}} \odot \mu_{x_{\text{ref}}} = K_{\sigma_m} D_{x_{\text{ref}}} x_{\text{ref}} - D_{\mu_{x_{\text{ref}}}} \mu_{x_{\text{ref}}}. \end{aligned}$$

The operator  $\odot$  denotes the element-wise multiplications or Hadamard product. We can now rewrite the loss for the whole image

$$\text{SSIM}(x^T, x_{\text{ref}}) = \frac{1}{N} \mathbf{1}^\top [d_1 \oslash d_2 \odot d_3 \oslash d_4] = 1 - \frac{1}{N} \mathbf{1}^\top [D_1 D_2^{-1} D_3 d_4^{-1}], \quad (\text{D.1})$$

where

$$\begin{aligned} d_1 &= (2\mu_{x^T} \odot \mu_g + \mathbf{1}c_1), \\ d_2 &= (\mu_{x^T} \odot \mu_{x^T} + \mu_{x_{\text{ref}}} \odot \mu_{x_{\text{ref}}} + \mathbf{1}c_1), \\ d_3 &= (2\sigma_{x^T, x_{\text{ref}}} + \mathbf{1}c_2), \\ d_4 &= (\sigma_{x^T} \odot \sigma_{x^T} + \sigma_{x_{\text{ref}}} \odot \sigma_{x_{\text{ref}}} + \mathbf{1}c_2). \end{aligned}$$

Hence, the terms for luminance  $l$  and contrast  $c$  are given as

$$\begin{aligned} l &= d_1 \oslash d_2, \\ c &= d_3 \oslash d_4. \end{aligned}$$

Note that we denote here the element-wise inverse by  $^{-1}$  and  $\oslash$  represents element-wise divisions. Upper case  $D_i$  denote diagonal matrices  $D_i = \text{diag } d_i$ . Let us revise diagonal matrix properties and some basic derivative rules. For diagonal matrices  $D_i$  it holds that

$$\begin{aligned} D_1 D_2 &= D_2 D_1, \\ D_1^\top &= D_1. \end{aligned}$$

For derivatives wrt. to vectors, we need following rules

$$\begin{aligned}\frac{\partial a^\top y}{\partial x} &= \frac{\partial y}{\partial x} a, \\ \frac{\partial Ax}{\partial x} &= A^\top, \\ \frac{\partial Ay}{\partial y} &= \frac{\partial y}{\partial x} A^\top,\end{aligned}$$

where  $y$  is a vector dependent on  $x$ ,  $a$  is a vector independent of  $x$  and  $A$  denotes a matrix independent of  $x$ .

Using the chain rule and properties of matrices, we estimate the loss as follows

$$\begin{aligned}\frac{\partial \text{SSIM}(x^T, x_{\text{ref}})}{\partial x^T} &= \frac{1}{N} \left[ \frac{\partial l}{\partial x^T} C + \frac{\partial c}{\partial x^T} L \right] \mathbf{1} \\ &= \frac{1}{N} \left[ \frac{\partial d_1}{\partial x^T} D_2^{-1} D_3 D_4^{-1} + \frac{\partial d_2^{-1}}{\partial x^T} D_1 D_3 D_4^{-1} + \frac{\partial d_3}{\partial x^T} D_1 D_2^{-1} D_4^{-1} + \frac{\partial d_4^{-1}}{\partial x^T} D_1 D_2^{-1} D_3 \right] \mathbf{1}.\end{aligned}$$

We compute now the derivatives for  $d_i$

$$\begin{aligned}\frac{\partial d_1}{\partial x^T} &= 2 \left[ D_{\mu_{x_{\text{ref}}}} K_{\sigma_m} \right]^\top = 2 K_{\sigma_m}^\top D_{\mu_{x_{\text{ref}}}}, \\ \frac{\partial d_2^{-1}}{\partial x^T} &= \frac{\partial d_1}{\partial u} \frac{\partial d_2^{-1}}{\partial d_2} = -\frac{\partial d_2}{\partial x^T} D_2^{-2} = -2 \left[ D_{\mu_{x^T}} K_{\sigma_m} \right]^\top D_2^{-2} = -2 K_{\sigma_m}^\top D_{\mu_{x^T}} D_2^{-2}, \\ \frac{\partial d_3}{\partial x^T} &= 2 D_{x_{\text{ref}}} K_{\sigma_m}^\top - 2 K_{\sigma_m}^\top D_{\mu_{x_{\text{ref}}}}, \\ \frac{\partial d_4^{-1}}{\partial x^T} &= - \left( 2 D_{x^T} K_{\sigma_m}^\top - 2 K_{\sigma_m}^\top D_{\mu_{x^T}} \right) D_4^{-2}.\end{aligned}$$

Combining these derivatives with (D.1) leads to the final partial derivative of the loss wrt. the stage outputs  $x^T$

$$\begin{aligned}\frac{\partial \mathcal{L}_{\text{SSIM}}(\theta)}{\partial x^T} = e^T &= -\frac{2}{N} \left[ K_{\sigma_m}^\top \left( D_{\mu_{x_{\text{ref}}}} D_2^{-1} D_3 D_4^{-1} - D_{\mu_{x^T}} D_1 D_2^{-2} D_3 D_4^{-1} \right. \right. \\ &\quad \left. \left. - D_{\mu_{x_{\text{ref}}}} D_1 D_2^{-1} D_4^{-1} + D_{\mu_{x^T}} D_1 D_2^{-1} D_3 D_4^{-2} \right) \right. \\ &\quad \left. + D_{x_{\text{ref}}} K_{\sigma_m}^\top D_1 D_2^{-1} D_4^{-1} - D_{x^T} K_{\sigma_m}^\top D_1 D_2^{-1} D_3 D_4^{-2} \right] \mathbf{1}.\end{aligned}$$

**MS-SSIM.** Once we have obtained the gradients for the SSIM, the gradients for MS-SSIM can be obtained easily. Similar to SSIM, the MS-SSIM loss is defined as

$$\mathcal{L}_{\text{ms-ssim}}(x^T(\theta), x_{\text{ref}}) = 1 - \text{MS-SSIM}(x^T(\theta), x_{\text{ref}})$$



where

$$\text{MS-SSIM}(x, x_{\text{ref}}) = \frac{1}{N} \sum_{i \in \Omega} l_M(x(i), x_{\text{ref}}(i)) \prod_{m=1}^M c_m(x(i), x_{\text{ref}}(i)).$$

To model multiple scales, we compute the contrast terms for varying standard deviations  $\sigma_m$  according to [274]. In our case, we use four different standard deviations  $\sigma_{G,m} \in [0.5, 1, 2, 4]$ . Again, we view this problem on the whole image  $l_M, c_m$  instead of patches as presented for SSIM. The capital letters  $L_M, C_m$  denote the diagonal matrix representation of the vectors  $l_M, c_m$ . This leads to following derivative for the MS-SSIM wrt. the stage outputs  $x^T$

$$\frac{\partial \mathcal{L}_{\text{ms-ssim}}(x^T(\theta), x_{\text{ref}})}{\partial x^T(\theta)} = e^T = -\frac{1}{N} \left[ \frac{\partial l_M}{\partial x^T} + \sum_{n=0}^M \frac{\partial c_n}{\partial x^T} C_n^{-1} L \right] \prod_{m=1}^M C_m \mathbf{1}.$$



## Bibliography

- [1] J. Adler and O. Öktem. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems*, 33(12):1–24, 2017, [https://github.com/adler-j/learned\\_gradient\\_tomography](https://github.com/adler-j/learned_gradient_tomography). (Cited on pages 17, 18, 19).
- [2] J. Adler and O. Öktem. Learned Primal-Dual Reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1322–1332, 2018, [https://github.com/adler-j/learned\\_primal\\_dual](https://github.com/adler-j/learned_primal_dual). (Cited on pages 17, 18, 19, 20).
- [3] G. Adluru, S. P. Awate, T. Tasdizen, R. T. Whitaker, and E. V. DiBella. Temporally constrained reconstruction of dynamic cardiac perfusion MRI. *Magnetic Resonance in Medicine*, 57(6):1027–1036, 2007. (Cited on page 112).
- [4] G. Adluru, T. Tasdizen, M. C. Schabel, and E. V. Dibaba. Reconstruction of 3D dynamic contrast-enhanced magnetic resonance imaging using nonlocal means. *Journal of Magnetic Resonance Imaging*, 32(5):1217–1227, 2010. (Cited on page 16).
- [5] G. Adluru, T. Tasdizen, R. Whitaker, and E. DiBella. Improving undersampled MRI reconstruction using non-local means. In *IEEE International Conference on Pattern Recognition*, pages 4000–4003, 2010. (Cited on page 16).
- [6] H. K. Aggarwal, M. P. Mani, and M. Jacob. Model Based Image Reconstruction Using Deep Learned Priors (Modl). In *IEEE International Symposium on Biomedical Imaging*, pages 671–674, 2018. (Cited on pages 17, 18).
- [7] H. K. Aggarwal, M. P. Mani, and M. Jacob. MoDL: Model Based Deep Learning Architecture for Inverse Problems. *IEEE Transactions on Medical Imaging*, 38(2):394–405, 2019, <https://github.com/hkaggarwal/modl>. (Cited on pages 17, 18).
- [8] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006. (Cited on page 10).
- [9] R. Ahmad, H. Xue, S. Giri, Y. Ding, J. Craft, and O. P. Simonetti. Variable density incoherent spatiotemporal acquisition (VISTA) for highly accelerated cardiac MRI. *Magnetic Resonance in Medicine*, 74(5):1266–78, 2014. (Cited on pages 115, 118).
- [10] S. Aja-Fernández, G. Vegas-Sánchez-Ferrero, and A. Tristán-Vega. Noise estimation in parallel MRI: GRAPPA and SENSE. *Magnetic Resonance Imaging*, 32(3):281–290, 2014. (Cited on page 64).
- [11] D. Allman, A. Reiter, and M. A. Bell. Photoacoustic Source Detection and Reflection Artifact Removal Enabled by Deep Learning. *IEEE Transactions on Medical Imaging*, 37(6):1464–1477, 2018, <https://github.com/derekallman/Photoacoustic-FasterRCNN>. (Cited on page 23).

- [12] S. Antholzer, M. Haltmeier, and J. Schwab. Deep Learning for Photoacoustic Tomography from Sparse Data. *Inverse Problems in Science and Engineering*, pages 1–19, 2018. (Cited on page 23).
- [13] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein Generative Adversarial Networks. In *Proceedings of the International Conference on Machine Learning*, pages 214–223, 2017. (Cited on page 27).
- [14] S. M. Assari, H. Idrees, and M. Shah. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *Proceedings of the European Conference on Computer Vision*, pages 694–711, 2016. (Cited on page 27).
- [15] M. Benning and M. Burger. Modern Regularization Methods for Inverse Problems. *Acta Numerica*, 27:1–111, 2018. (Cited on page 14).
- [16] N. Benvenuto and F. Piazza. On the Complex Backpropagation Algorithm. *IEEE Transactions on Signal Processing*, 40(4):967–969, 1992. (Cited on page 101).
- [17] M. A. Bernstein, K. F. King, and X. J. Zhou. *Handbook of MRI pulse sequences*. Academic Press, 2004. (Cited on pages 55, 59).
- [18] M. Blaimer, F. Breuer, M. Mueller, R. M. Heidemann, M. A. Griswold, and P. M. Jakob. SMASH, SENSE, PILS, GRAPPA: how to choose the optimal method. *Topics in Magnetic Resonance Imaging*, 15(4):223–36, 2004. (Cited on page 65).
- [19] K. T. Block, M. Uecker, and J. Frahm. Undersampled Radial MRI with Multiple Coils. Iterative Image Reconstruction using a Total Variation Constraint. *Magnetic Resonance in Medicine*, 57(6):1086–1098, 2007. (Cited on pages 8, 66).
- [20] S. Boyd, N. Parikh, B. P. E Chu, and J. Eckstein. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011. (Cited on page 16).
- [21] K. Bredies, K. Kunisch, and T. Pock. Total Generalized Variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010. (Cited on pages 8, 46, 68, 76).
- [22] F. A. Breuer, S. A. Kannengiesser, M. Blaimer, N. Seiberlich, P. M. Jakob, and M. A. Griswold. General formulation for quantitative G-factor calculation in GRAPPA reconstructions. *Magnetic Resonance in Medicine*, 62(3):739–746, 2009. (Cited on page 64).
- [23] F. A. Breuer, P. Kellman, M. A. Griswold, and P. M. Jakob. Dynamic autocalibrated parallel imaging using temporal GRAPPA (TGRAPPA). *Magnetic Resonance in Medicine*, 53(4):981–985, 2005. (Cited on page 112).

- [24] R. W. Brown, Y.-C. N. Cheng, E. M. Haacke, M. R. Thompson, and R. Venkatesan, editors. *Magnetic Resonance Imaging*. John Wiley & Sons Ltd, Chichester, UK, 2014. (Cited on page 55).
- [25] A. Buades, B. Coll, and J.-M. Morel. Non-Local Means Denoising. *Image Processing On Line*, 1:208–212, 2011. (Cited on page 16).
- [26] J. Caballero, A. N. Price, D. Rueckert, and J. V. Hajnal. Dictionary learning and time sparsity for dynamic MR data reconstruction. *IEEE Transactions on Medical Imaging*, 33(4):979–994, 2014. (Cited on pages 10, 92, 112).
- [27] C. Cai, C. Wang, Y. Zeng, S. Cai, D. Liang, Y. Wu, Z. Chen, X. Ding, and J. Zhong. Single-shot T2 mapping using overlapping-echo detachment planar imaging and a deep convolutional neural network. *Magnetic Resonance in Medicine*, 80(5):2202–2214, 2018. (Cited on page 22).
- [28] T. T. Cai and L. Wang. Orthogonal Matching Pursuit for Sparse Signal Recovery With Noise. *IEEE Transactions on Information Theory*, 57(7):4680–4688, 2011. (Cited on page 10).
- [29] L. Calatroni, C. Cao, J. Carlos De los Reyes, C.-B. Schönlieb, and T. Valkonen. Bilevel approaches for learning of variational imaging models. In M. Bergounioux, G. Peyré, C. Schnörr, J.-B. Caillaud, and T. Haberorn, editors, *Variational Methods: In Imaging and Geometric Control*, pages 252–290. De Gruyter, Berlin, Boston, 2016. (Cited on pages 12, 15).
- [30] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3):1–37, 2011. (Cited on page 112).
- [31] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006. (Cited on pages 2, 8, 66).
- [32] E. Cha, E. Y. Kim, and J. C. Ye. k-Space Deep Learning for Parallel MRI: Application to Time-Resolved MR Angiography. *arXiv preprint arXiv:1805.03779*, pages 1–11, 2018. (Cited on page 24).
- [33] A. Chambolle, R. De Vore, Nam-Yong Lee, and B. Lucier. Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Transactions on Image Processing*, 7(3):319–335, 1998. (Cited on page 17).
- [34] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011. (Cited on pages 16, 19, 20, 29).

- [35] A. Chambolle and T. Pock. An Introduction to Continuous Optimization for Imaging. *Acta Numerica*, 25:161–319, 2016. (Cited on pages 8, 19).
- [36] A. S. Chaudhari, Z. Fang, F. Kogan, J. Wood, K. J. Stevens, E. K. Gibbons, J. H. Lee, G. E. Gold, and B. A. Hargreaves. Super-resolution musculoskeletal MRI using deep learning. *Magnetic Resonance in Medicine*, 80(5):2139–2154, 2018. (Cited on page 22).
- [37] B. Chen, K. Xiang, Z. Gong, J. Wang, and S. Tan. Statistical Iterative CBCT Reconstruction Based on Neural Network. *IEEE Transactions on Medical Imaging*, 37(6):1511–1521, 2018, <https://github.com/HUST-Tan/Deblur-CBCT>. (Cited on page 16).
- [38] F. Chen, V. Taviani, I. Malkiel, J. Y. Cheng, J. I. Tamir, J. Shaikh, S. T. Chang, C. J. Hardy, J. M. Pauly, and S. S. Vasanawala. Variable-Density Single-Shot Fast Spin-Echo MRI with Deep Learning Reconstruction by Using Variational Networks. *Radiology*, pages 1–8, 2018. (Cited on page 28).
- [39] H. Chen, Y. Zhang, Y. Chen, J. Zhang, W. Zhang, H. Sun, Y. Lv, P. Liao, J. Zhou, and G. Wang. LEARN: Learned Experts’ Assessment-Based Reconstruction Network for Sparse-Data CT. *IEEE Transactions on Medical Imaging*, 37(6):1333–1347, 2018, <https://github.com/maybe198376/LEARN>. (Cited on page 20).
- [40] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang. Low-Dose CT With a Residual Encoder-Decoder Convolutional Neural Network. *IEEE Transactions on Medical Imaging*, 36(12):2524–2535, 2017, <https://github.com/maybe198376/Low-Dose-CT-With-a-Residual-Encoder-Decoder-Convolutional-Neural-Network>. (Cited on page 22).
- [41] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, and J. Zhou. Low-Dose CT Denoising With Convolutional Neural Network. In *IEEE International Symposium on Biomedical Imaging*, pages 2–5, 2017. (Cited on page 22).
- [42] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. In *International Conference on Learning Representations*, pages 1–14, 2015. (Cited on page 1).
- [43] Y. Chen and T. Pock. Trainable Nonlinear Reaction Diffusion: A Flexible Framework for Fast and Effective Image Restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1256–1272, 2017. (Cited on pages 3, 33).
- [44] Y. Chen, T. Pock, R. Ranftl, and H. Bischof. Revisiting loss-specific training of filter-based MRFs for image restoration. In *German Conference on Pattern Recognition*, pages 271–281, 2013. (Cited on pages 14, 15).

- [45] Y. Chen, R. Ranftl, and T. Pock. Insights into analysis operator learning: From patch-based sparse models to higher order MRFs. *IEEE Transactions on Image Processing*, 23(3):1060–1072, 2014. (Cited on pages 14, 15).
- [46] Y. Chen, W. Yu, and T. Pock. On learning optimized reaction diffusion processes for effective image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5261–5269, 2015. (Cited on pages 3, 14, 17, 19, 20, 21, 23, 33, 48, 93, 113).
- [47] Z. Chen, H. Qi, S. Wu, Y. Xu, and L. Zhou. Few-view CT reconstruction via a novel non-local means algorithm. *Physica Medica*, 32(10):1276–1283, 2016. (Cited on page 16).
- [48] I. Y. Chun and J. A. Fessler. Convolutional Dictionary Learning: Acceleration and Convergence. *IEEE Transactions on Image Processing*, 27(4):1697–1712, 2018. (Cited on page 10).
- [49] B. Claus, Y. Jin, L. Gjestebj, G. Wang, and B. De Man. Metal-Artifact Reduction Using Deep-Learning Based Sinogram Completion: Initial Results. In *International Conference on Fully 3D Image Reconstruction*, pages 631–635, 2017. (Cited on page 23).
- [50] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising with block-matching and 3D filtering. In *SPIE Electronic Imaging*, page 606414, 2006. (Cited on pages 16, 46).
- [51] S. U. H. Dar and T. Çukur. A Transfer-Learning Approach for Accelerated MRI using Deep Neural Networks. *arXiv preprint arXiv:1710.02615*, 2017. (Cited on page 26).
- [52] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004. (Cited on page 17).
- [53] I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, 36(5):961–1005, 1990. (Cited on pages 8, 66).
- [54] J. G. Daugman. Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-Dimensional Visual Cortical Filters. *Journal of the Optical Society of America*, 2(7):1160–1169, 1985. (Cited on page 93).
- [55] J. C. De Los Reyes, C. B. Schönlieb, and T. Valkonen. The structure of optimal parameters for image restoration problems. *Journal of Mathematical Analysis and Applications*, 434(1):464–500, 2016. (Cited on pages 12, 15).

- [56] J. C. De los Reyes, C. B. Schönlieb, and T. Valkonen. Bilevel Parameter Learning for Higher-Order Total Variation Regularisation Models. *Journal of Mathematical Imaging and Vision*, 57(1):1–25, 2017.
- [57] J. C. De Los Reyes and C. B. Schönlieb. Image denoising: Learning the noise model via nonsmooth PDE-constrained optimization. *Inverse Problems and Imaging*, 7(4):1183–1214, 2013. (Cited on pages 12, 15).
- [58] S. Diamond, V. Sitzmann, F. Heide, and G. Wetzstein. Unrolled Optimization with Deep Priors. *arXiv preprint arXiv:1705.08041*, pages 1–11, 2017. (Cited on page 17).
- [59] C. B. Do and A. Y. Ng. Transfer learning for text classification. In *Advances in Neural Information Processing Systems*, pages 299–306, 2005. (Cited on page 26).
- [60] D. L. Donoho. Compressed Sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006. (Cited on pages 2, 8, 66).
- [61] A. Dosovitskiy, P. Fischer, E. Ilg, P. Häusser, C. Hazirbas, V. Golkov, P. van der Smagt, D. Cremers, and T. Brox. FlowNet: Learning Optical Flow with Convolutional Networks. In *IEEE International Conference on Computer Vision*, pages 2758–2766, 2015. (Cited on page 1).
- [62] O. Dössel. *Bildgebende Verfahren in der Medizin*. Springer, 2000. (Cited on page 42).
- [63] A. Effland, M. Hölzel, T. Klatzer, E. Kobler, J. Landsberg, L. Neuhäuser, T. Pock, and M. Rumpf. Variational networks for joint image reconstruction and classification of tumor immune cell interactions in melanoma tissue sections. In *Bildverarbeitung für die Medizin 2018*, pages 334–340, 2018. (Cited on page 21).
- [64] E. M. Eksioğlu. Decoupled Algorithm for MRI Reconstruction Using Nonlocal Block Matching Model: BM3D-MRI. *Journal of Mathematical Imaging and Vision*, 56(3):430–440, 2016, [http://web.itu.edu.tr/eksioglu/pubs/BM3D\\_MRI\\_toolbox.zip](http://web.itu.edu.tr/eksioglu/pubs/BM3D_MRI_toolbox.zip). (Cited on page 16).
- [65] T. Eo, Y. Jun, T. Kim, J. Jang, H. J. Lee, and D. Hwang. KIKI-net: Cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. *Magnetic Resonance in Medicine*, 80(5):2188–2201, 2018. (Cited on page 21).
- [66] L. Feng, M. B. Srichai, R. P. Lim, A. Harrison, W. King, G. Adluru, E. V. R. Dibella, D. K. Sodickson, R. Otazo, and D. Kim. Highly accelerated real-time cardiac cine MRI using k-t SPARSE-SENSE. *Magnetic Resonance in Medicine*, 70(1):64–74, 2013. (Cited on page 112).
- [67] C. R. Floyd. An artificial neural network for SPECT image reconstruction. *IEEE Transactions on Medical Imaging*, 10(3):485–7, 1991. (Cited on pages 1, 24).



- [68] D. Gabor. Theory of Communication, 1946. (Cited on page 93).
- [69] L. A. Gatys, A. S. Ecker, and M. Bethge. Image Style Transfer Using Convolutional Neural Networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2414–2423, 2016. (Cited on page 4).
- [70] G. M. Georgiou and C. Koutsougeras. Complex Domain Backpropagation. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 39(5):330–334, 1992. (Cited on pages 101, 102).
- [71] K. Gong, J. Guan, K. Kim, X. Zhang, J. Yang, Y. Seo, G. El Fakhri, J. Qi, and Q. Li. Iterative PET Image Reconstruction Using Convolutional Neural Network Representation. *IEEE Transactions on Medical Imaging*, 38(3):675–685, 2018, <https://github.com/guanjiahui/IterativeCNN>. (Cited on page 16).
- [72] K. Gong, J. Yang, K. Kim, G. El Fakhri, Y. Seo, and Q. Li. Attenuation correction for brain PET imaging using deep neural network based on Dixon and ZTE MR images. *Physics in Medicine and Biology*, 63(12):1–13, 2018. (Cited on page 25).
- [73] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. (Cited on page 1).
- [74] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014. (Cited on pages 27, 95).
- [75] B. Gözcü, R. K. Mahabadi, Y. H. Li, E. Ilicak, T. Çukur, J. Scarlett, and V. Cevher. Learning-Based Compressive MRI. *IEEE Transactions on Medical Imaging*, 37(6):1394–1406, 2018, <https://lions.epfl.ch/lb-csmri>. (Cited on page 25).
- [76] K. Gregor and Y. Lecun. Learning Fast Approximations of Sparse Coding. In *Proceedings of the International Conference on Machine Learning*, pages 399–406, 2010. (Cited on page 17).
- [77] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, 47(6):1202–1210, 2002. (Cited on pages 2, 61, 62).
- [78] J. Gu and J. C. Ye. Multi-Scale Wavelet Domain Residual Learning for Limited-Angle CT Reconstruction. *arXiv preprint arXiv:1703.01382*, 2017. (Cited on page 23).
- [79] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville. Improved Training of Wasserstein GANs. *arXiv preprint arXiv:1704.00028*, 2017. (Cited on page 27).

- [80] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. McCann, and M. Unser. CNN-Based Projected Gradient Descent for Consistent CT Image Reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1440–1453, 2018, <https://github.com/harshit-gupta-epfl/CNN-RPGD>. (Cited on page 16).
- [81] J. Hadamard. Sur les problemes aux derivees partielles et leur signification physique. *Princeton University Bulletin*, 13(13):49–52, 1902. (Cited on page 7).
- [82] I. Häggström, C. R. Schmidlein, G. Campanella, and T. J. Fuchs. DeepPET: A deep encoder-decoder network for directly solving the PET reconstruction inverse problem. *arXiv preprint arXiv:1804.07851*, 2018. (Cited on page 24).
- [83] K. Hammernik, T. Ebner, D. Stern, M. Urschler, and T. Pock. Vertebrae Segmentation in 3D CT Images based on a Variational Framework. In *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*, Lecture Notes in Computational Vision and Biomechanics, pages 227–233. Springer, 2015.
- [84] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, 2018, <https://github.com/VLOGroup/mri-variationalnetwork>. (Cited on pages 14, 17, 19, 20, 21, 26, 54, 55, 111, 113, 138).
- [85] K. Hammernik and F. Knoll. Machine Learning for Image Reconstruction. In D. Rueckert, G. Fichtinger, and S. K. Zhou, editors, *Handbook of Medical Image Computing and Computer Assisted Intervention*. Elsevier, 2018.
- [86] K. Hammernik, F. Knoll, D. Sodickson, and T. Pock. L2 or Not L2: Impact of Loss Function Design for Deep Learning MRI Reconstruction. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, page 687, 2017.
- [87] K. Hammernik, F. Knoll, D. Sodickson, and T. Pock. On the Influence of Sampling Pattern Design on Deep Learning-Based MRI Reconstruction. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, page 644, 2017. (Cited on page 94).
- [88] K. Hammernik, F. Knoll, D. K. Sodickson, and T. Pock. Learning a Variational Model for Compressed Sensing MRI Reconstruction. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, number 24, page 1088, 2016. (Cited on pages 92, 93).
- [89] K. Hammernik, E. Kobler, T. Pock, M. Recht, D. K. Sodickson, and F. Knoll. Variational Adversarial Networks for Accelerated MR Image Reconstruction. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, page 1091, 2018. (Cited on page 101).

- [90] K. Hammernik, T. Pock, and R. Nuster. Variational photoacoustic image reconstruction with spatially resolved projection data. *Proceedings of SPIE, Photons Plus Ultrasound: Imaging and Sensing*, 10064:100643I, 2017.
- [91] K. Hammernik, M. Schloegl, R. Stollberger, and T. Pock. Dynamic Multicoil Reconstruction using Variational Networks. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, page to appear, 2019.
- [92] K. Hammernik, T. Würfl, T. Pock, and A. Maier. A deep learning architecture for limited-angle computed tomography reconstruction. In *Bildverarbeitung für die Medizin 2017*, pages 92–97, 2017. (Cited on pages 23, 26, 28).
- [93] Y. S. Han, J. Yoo, and J. C. Ye. Deep Residual Learning for Compressed Sensing CT Reconstruction via Persistent Homology Analysis. *arXiv preprint arXiv:1611.06391*, 2016. (Cited on page 23).
- [94] Y. Han and J. C. Ye. Framing U-Net via Deep Convolutional Framelets: Application to Sparse-View CT. *IEEE Transactions on Medical Imaging*, 37(6):1418–1429, 2018, <https://github.com/hanyoseob/framing-u-net>. (Cited on page 23).
- [95] Y. Han and J. C. Ye. k-Space Deep Learning for Accelerated MRI. *arXiv preprint arXiv:1805.03779*, pages 1–11, 2018. (Cited on page 24).
- [96] Y. Han, J. Yoo, H. H. Kim, H. J. Shin, K. Sung, and J. C. Ye. Deep learning with domain adaptation for accelerated projection-reconstruction MR. *Magnetic Resonance in Medicine*, 80(3):1189–1205, 2018. (Cited on pages 22, 26, 92).
- [97] M. Hanke, A. Neubauer, and O. Scherzer. A Convergence Analysis of the Landweber Iteration for Nonlinear Ill-Posed Problems. *Numerische Mathematik*, 72(1):21–37, 1995. (Cited on pages 8, 16, 32).
- [98] A. Hauptmann, S. Arridge, F. Lucka, V. Muthurangu, and J. A. Steeden. Real-time cardiovascular MR with spatio-temporal artifact suppression using deep learning-proof of concept in congenital heart disease. *Magnetic Resonance in Medicine*, 81(2):1143–1156, 2019. (Cited on page 112).
- [99] A. Hauptmann, B. Cox, F. Lucka, N. Huynh, M. Betcke, P. Beard, and S. Arridge. Approximate k-Space Models and Deep Learning for Fast Photoacoustic Reconstruction. In *Proceedings of International Workshop on Machine Learning for Medical Image Reconstruction*, pages 103–111, 2018. (Cited on page 19).
- [100] A. Hauptmann, F. Lucka, M. Betcke, N. Huynh, J. Adler, B. Cox, P. Beard, S. Ourselin, and S. Arridge. Model-Based Learning for Accelerated, Limited-View 3-D Photoacoustic Tomography. *IEEE Transactions on Medical Imaging*, 37(6):1382–1393, 2018, [https://github.com/asHauptmann/3DPAT\\_DGD](https://github.com/asHauptmann/3DPAT_DGD). (Cited on pages 18, 19, 21, 111).

- [101] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. (Cited on pages 17, 22).
- [102] F. Heide, M. Steinberger, Y.-t. Tsai, O. Gallo, W. Heidrich, K. Egiazarian, J. Kautz, D. Reddy, and K. Pulli. FlexISP : A Flexible Camera Image Processing Framework. *ACM Transactions on Graphics*, 33(6):1–13, 2014. (Cited on page 16).
- [103] A. Hirose. Continuous complex-valued back-propagation learning. *Electronics Letters*, 28(20):1854–1855, 1992. (Cited on page 101).
- [104] M. Holler and K. Kunisch. On Infimal Convolution of TV-Type Functionals and Applications to Video and Image Reconstruction. *SIAM Journal on Imaging Sciences*, 7(4):2258–2300, 2014. (Cited on page 112).
- [105] K. G. Hollingsworth. Reducing Acquisition Time in Clinical MRI by Data Undersampling and Compressed Sensing Reconstruction. *Physics in Medicine and Biology*, 60(21):R297–R322, 2015. (Cited on pages 3, 9, 66, 71).
- [106] D. I. Hoult. The principle of reciprocity in signal strength calculations - A mathematical guide. *Concepts in Magnetic Resonance*, 12(4):173–187, 2000. (Cited on page 56).
- [107] D. Hoult and R. Richards. The signal-to-noise ratio of the nuclear magnetic resonance experiment. *Journal of Magnetic Resonance (1969)*, 24(1):71–85, 1976. (Cited on page 56).
- [108] F. Huang, J. Akao, S. Vijayakumar, G. R. Duensing, and M. Limkeman. k-t GRAPPA: Ak-space implementation for dynamic MRI with high reduction factor. *Magnetic Resonance in Medicine*, 54(5):1172–1184, 2005. (Cited on page 112).
- [109] Y. Huang, O. Taubmann, X. Huang, V. Haase, G. Lauritsch, and A. Maier. A New Scale Space Total Variation Algorithm for Limited Angle Tomography. In *Proceedings of the 4th International Meeting on Image Formation in X-Ray Computed Tomography*, pages 149–152, 2016. (Cited on page 44).
- [110] Y. Huang, O. Taubmann, X. Huang, V. Haase, G. Lauritsch, and A. Maier. A New Weighted Anisotropic Total Variation Algorithm For Limited Angle Tomography. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 585–588, 2016. (Cited on page 44).
- [111] C. M. Hyun, H. P. Kim, S. M. Lee, S. Lee, and J. K. Seo. Deep learning for undersampled MRI reconstruction. *Physics in Medicine and Biology*, 63(13), 2018, [https://github.com/hpkim0512/Deep\\_MRI\\_Unet](https://github.com/hpkim0512/Deep_MRI_Unet). (Cited on page 22).

- [112] S. J. Inati, J. D. Naegele, N. R. Zwart, V. Roopchansingh, M. J. Lizak, D. C. Hansen, C.-Y. Liu, D. Atkinson, P. Kellman, S. Kozerke, H. Xue, A. E. Campbell-Washburn, T. S. Sørensen, and M. S. Hansen. ISMRM Raw data format: A proposed standard for MRI raw datasets. *Magnetic Resonance in Medicine*, 77(1):411–421, 2017. (Cited on page 26).
- [113] S. Ioffe and C. Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *Proceedings of the International Conference on Machine Learning*, pages 448–456, 2015. (Cited on page 36).
- [114] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2017-Janua, pages 5967–5976, 2017. (Cited on page 27).
- [115] A. K. Jain and F. Farrokhnia. Unsupervised Texture Segmentation using Gabor Filters. *Pattern Recognition*, 24(12):1167–1186, 1990. (Cited on page 93).
- [116] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. (Cited on pages 25, 26).
- [117] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser. Deep Convolutional Neural Network for Inverse Problems in Imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017, <https://github.com/panakino/FBPCConvNet>. (Cited on page 23).
- [118] K. H. Jin and M. Unser. 3D BPCConvNet to Reconstruct Parallel MRI. In *IEEE International Symposium on Biomedical Imaging*, pages 361–364, 2018, <https://github.com/panakino/3dbpconv>. (Cited on page 22).
- [119] Jिंगgang Huang and D. Mumford. Statistics of natural images and models. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 541–547, 1999. (Cited on pages 14, 15, 93).
- [120] H. Jung, K. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye. k-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI. *Magnetic Resonance in Medicine*, 61(1):103–116, 2009. (Cited on page 112).
- [121] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis*, 36:61–78, 2017, <https://github.com/Kamnitsask/deepmedic>. (Cited on page 1).
- [122] E. Kang, W. Chang, J. Yoo, and J. C. Ye. Deep Convolutional Framelet Denosing for Low-Dose CT via Wavelet Residual Network. *IEEE Transactions on Medical Imaging*,

- 37(6):1358–1369, 2018, [https://github.com/eunh/low\\_dose\\_CT](https://github.com/eunh/low_dose_CT). (Cited on page 22).
- [123] E. Kang, J. Min, and J. C. Ye. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. *Medical Physics*, 44(10):e360–e375, 2017, [https://github.com/eunh/low\\_dose\\_CT](https://github.com/eunh/low_dose_CT). (Cited on page 22).
- [124] D. Karras, M. Reczko, V. Mertzios, D. Graveron-Demilly, D. van Ormondt, and R. Papademetriou. Neural network reconstruction of MR images from noisy and sparse k-space samples. In *Proceedings of the International Conference on Signal Processing*, pages 2115–2118, 2000. (Cited on page 1).
- [125] P. Kellman, F. H. Epstein, and E. R. McVeigh. Adaptive sensitivity encoding incorporating temporal filtering (TSENSE). *Magnetic Resonance in Medicine*, 45(5):846–852, 2001. (Cited on page 112).
- [126] B. Kelly, T. P. Matthews, and M. A. Anastasio. Deep Learning-Guided Image Reconstruction from Incomplete Data. *arXiv preprint arXiv:1709.00584*, 2017, <http://arxiv.org/abs/1709.00584>. (Cited on page 16).
- [127] J. P. Kerr and E. B. Bartlett. Neural network reconstruction of single-photon emission computed tomography images. *Journal of Digital Imaging*, 8(3):116–126, 1995. (Cited on page 1).
- [128] K. Kim, D. Wu, K. Gong, J. Dutta, J. H. Kim, Y. D. Son, H. K. Kim, G. El Fakhri, and Q. Li. Penalized PET Reconstruction Using Deep Learning Prior and Local Linear Fitting. *IEEE Transactions on Medical Imaging*, 37(6):1478–1487, 2018. (Cited on page 27).
- [129] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*, 2014. (Cited on pages 36, 138).
- [130] T. Klatzer, K. Hammernik, P. Knöbelreiter, and T. Pock. Learning Joint Demosaicing and Denoising Based on Sequential Energy Minimization. In *IEEE International Conference on Computational Photography*, pages 1–11, 2016. (Cited on pages 20, 93).
- [131] T. Klatzer, D. Soukup, E. Kobler, K. Hammernik, and T. Pock. Trainable Regularization for Multi-frame Superresolution. In *German Conference on Pattern Recognition*, pages 90–100, 2017.
- [132] F. Knoll, K. Bredies, T. Pock, and R. Stollberger. Second order total generalized variation (TGV) for MRI. *Magnetic Resonance in Medicine*, 65(2):480–491, 2011. (Cited on pages 8, 66, 68, 76, 94).

- [133] F. Knoll, C. Clason, K. Bredies, M. Uecker, and R. Stollberger. Parallel imaging with nonlinear reconstruction using variational penalties. *Magnetic Resonance in Medicine*, 67(1):34–41, 2012. (Cited on page 66).
- [134] F. Knoll, K. Hammernik, E. Garwood, A. Hirschmann, L. Rybak, M. Bruno, T. Block, J. Babb, T. Pock, D. Sodickson, and M. Recht. Accelerated Knee Imaging Using a Deep Learning Based Reconstruction. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, page 645, 2017. (Cited on page 29).
- [135] F. Knoll, K. Hammernik, E. Kobler, T. Pock, M. P. Recht, and D. K. Sodickson. Assessment of the generalization of learned image reconstruction and the potential for transfer learning. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, 2018.
- [136] F. Knoll, K. Hammernik, E. Kobler, T. Pock, M. P. Recht, and D. 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. (Cited on pages 20, 26, 96, 98, 138).
- [137] F. Knoll, K. Hammernik, E. Kobler, T. Pock, D. K. Sodickson, and M. P. Recht. Analysis of the influence of deviations between training and test data in learned image reconstruction. In *ISMRM Workshop on Machine Learning*, 2018.
- [138] F. Knoll, K. Hammernik, T. Pock, and D. K. Sodickson. System, Method And Computer-Accessible Medium For Learning An Optimized Variational Network For Medical Image Reconstruction, 2017.
- [139] P. Knoll, S. Mirzaei, A. Müllner, T. Leitha, K. Koriska, H. Köhn, and M. Neumann. An artificial neural net and error backpropagation to reconstruct single photon emission computerized tomography data. *Medical Physics*, 26(2):244–248, 1999. (Cited on page 1).
- [140] E. Kobler, T. Klatzer, K. Hammernik, and T. Pock. Variational networks: Connecting variational methods and deep learning. In *German Conference on Pattern Recognition*, pages 281–293, 2017, <https://github.com/VLOGroup/denoising-variationalnetwork>. (Cited on pages 14, 19, 20, 21, 23, 34, 36, 93, 94, 98, 113).
- [141] E. Kobler, M. Muckley, B. Chen, F. Knoll, K. Hammernik, T. Pock, D. Sodickson, and R. Otazo. Variational Deep Learning for Low-Dose Computed Tomography. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6687–6691, 2018. (Cited on pages 20, 23, 44).

- [142] A. Krizhevsky, I. Sutskever, and H. Geoffrey E. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012. (Cited on pages 1, 93).
- [143] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok. ReconNet: Non-Iterative Reconstruction of Images from Compressively Sensed Measurements. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 449–458, 2016. (Cited on page 24).
- [144] K. Kunisch and T. Pock. A Bilevel Optimization Approach for Parameter Learning in Variational Models. *SIAM Journal on Imaging Sciences*, 6(2):938–983, 2013. (Cited on page 15).
- [145] K. Kwon, D. Kim, and H. Park. A parallel MR imaging method using multilayer perceptron. *Medical Physics*, 44(12):6209–6224, 2017. (Cited on pages 22, 92).
- [146] L. Landweber. An Iteration Formula for Fredholm Integral Equations of the First Kind. *American Journal of Mathematics*, 73(3):615–624, 1951. (Cited on pages 16, 32).
- [147] Y. LeCun, L. Bottou, G. Orr, and K. R. Müller. Efficient BackProp. In *Lecture Notes in Computer Science*, volume 1524, pages 5–50. 1998. (Cited on pages 11, 36, 141).
- [148] Y. LeCun, Y. Bengio, and G. Hinton. Deep Learning. *Nature*, 521(7553):436–444, 2015. (Cited on page 1).
- [149] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi Twitter. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4681–4690, 2017. (Cited on pages 27, 101).
- [150] D. Lee, J. Yoo, S. Tak, and J. Ye. Deep Residual Learning for Accelerated MRI using Magnitude and Phase Networks. *IEEE Transactions on Biomedical Engineering*, 65(9):1985–1995, 2018. (Cited on page 22).
- [151] D. Lee, J. Yoo, and J. C. Ye. Deep artifact learning for compressed sensing and parallel MRI. *arXiv preprint arXiv:1703.01120*, 2017. (Cited on pages 22, 92).
- [152] J. Lee, Y. Han, and J. C. Ye. k-Space Deep Learning for Reference-free EPI Ghost Correction. *arXiv preprint arXiv:1806.00153*, 2018. (Cited on page 24).
- [153] H. Leung and S. Haykin. The Complex Backpropagation Algorithm. *IEEE Transactions on Signal Processing*, 39(9):2101–2104, 1991. (Cited on page 101).
- [154] H. Li and K. Mueller. Low-Dose CT Streak Artifacts Removal using Deep Residual Neural Network. In *Proceedings of the International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, pages 3–6, 2017. (Cited on page 22).



- [155] Z.-P. Liang and P. C. Lauterbur. *Principles of magnetic resonance imaging : a signal processing perspective*. SPIE Optical Engineering Press, 2000. (Cited on pages 55, 59).
- [156] Lin Zhang, Lei Zhang, Xuanqin Mou, and D. Zhang. FSIM: A Feature Similarity Index for Image Quality Assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, 2011. (Cited on page 38).
- [157] E. M. Littringer, S. Zellnitz, K. Hammernik, V. Adamer, H. Friedl, and N. A. Urbanetz. Spray Drying of Aqueous Salbutamol Sulfate Solutions Using the Nano Spray Dryer B-90 - The Impact of Process Parameters on Particle Size. *Drying Technology*, 31(12):1346–1353, 2013.
- [158] J. Liu, A. Lefebvre, M. O. Zenge, M. Schmidt, E. Mueller, and M. S. Nadar. 2D bSSFP real-time cardiac CINE-MRI: compressed sensing featuring weighted redundant Haar Wavelet regularization in space and time. *Journal of Cardiovascular Magnetic Resonance*, 15(S1):P49, 2013. (Cited on page 112).
- [159] Y. Liu, K. K. Gadepalli, M. Norouzi, G. Dahl, T. Kohlberger, S. Venugopalan, A. S. Boyko, A. Timofeev, P. Q. Nelson, G. Corrado, J. Hipp, L. Peng, and M. Stumpe. Detecting Cancer Metastases on Gigapixel Pathology Images. *arXiv preprint arXiv:1703.02442*, 2017. (Cited on page 1).
- [160] Y. Liu, T. Kohlberger, M. Norouzi, G. E. Dahl, J. L. Smith, A. Mohtashamian, N. Olson, L. H. Peng, J. D. Hipp, and M. C. Stumpe. Artificial Intelligence-Based Breast Cancer Nodal Metastasis Detection. *Archives of Pathology & Laboratory Medicine*, pages arpa.2018–0147–OA. (Cited on page 2).
- [161] K. Lønning, P. Putzky, and M. Welling. Recurrent Inference Machines for Accelerated MRI Reconstruction. In *International Conference on Medical Imaging with Deep Learning*, pages 1–11, 2018. (Cited on page 21).
- [162] I. Loshchilov and F. Hutter. SGDR: Stochastic Gradient Descent with Warm Restarts. *arXiv preprint arXiv:1608.03983*, 2016. (Cited on page 115).
- [163] M. Lustig, D. Donoho, J. Santos, and J. Pauly. Compressed Sensing MRI. *IEEE Signal Processing Magazine*, 25(2):72–82, 2008. (Cited on page 66).
- [164] M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007. (Cited on pages 8, 66, 68, 75, 98, 115).
- [165] M. Lustig, J. M. Santos, D. L. Donoho, and J. M. Pauly. k-t SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, page 2420, 2006. (Cited on page 112).

- [166] X. F. Ma, M. Fukuhara, and T. Takeda. Neural network CT image reconstruction method for small amount of projection data. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 449(1-2):366–377, 2000. (Cited on page 1).
- [167] A. Makhzani and B. Frey. Winner-take-all autoencoders. In *Advances in Neural Information Processing Systems*, pages 2791–2799, 2015. (Cited on page 11).
- [168] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley. Least Squares Generative Adversarial Networks. In *IEEE International Conference on Computer Vision*, pages 2794–2802, 2017. (Cited on page 27).
- [169] M. Mardani, E. Gong, J. Y. Cheng, J. Pauly, and L. Xing. Recurrent generative adversarial neural networks for compressive imaging. In *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 1–5, 2017. (Cited on pages 18, 27).
- [170] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanaawala, G. Zaharchuk, L. Xing, and J. M. Pauly. Deep Generative Adversarial Neural Networks for Compressive Sensing (GANCS) MRI. *IEEE Transactions on Medical Imaging*, 38(1):167–179, 2019, <https://github.com/gongenhao/GANCS>. (Cited on pages 18, 27).
- [171] M. T. McCann, K. H. Jin, and M. Unser. Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine*, 34(6):85–95, 2017. (Cited on page 1).
- [172] S. McDonagh, B. Hou, A. Alansary, O. Oktay, K. Kamnitsas, M. Rutherford, J. V. Hajnal, and B. Kainz. Context-sensitive super-resolution for fast fetal magnetic resonance imaging. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 10555, pages 116–126, 2017. (Cited on page 22).
- [173] D. W. McRobbie. *MRI from picture to proton*. Cambridge University Press, 2003. (Cited on page 55).
- [174] T. Meinhardt, M. Moeller, C. Hazirbas, and D. Cremers. Learning Proximal Operators: Using Denoising Networks for Regularizing Inverse Imaging Problems. In *IEEE International Conference on Computer Vision*, pages 1799–1808, 2017. (Cited on page 16).
- [175] M. Nikolova, M. K. Ng, S. Zhang, and W.-K. Ching. Efficient Reconstruction of Piecewise Constant Images Using Nonsmooth Nonconvex Minimization. *SIAM Journal on Imaging Sciences*, 1(1):2–25, 2008. (Cited on page 14).
- [176] T. Nitta. An extension of the back-propagation algorithm to complex numbers. *Neural Networks*, 10(8):1391–1415, 1997. (Cited on pages 101, 102).

- [177] H. Nyquist. Certain Topics in Telegraph Transmission Theory. *Transactions of the American Institute of Electrical Engineers*, 47(2):617–644, 1928. (Cited on pages 8, 66).
- [178] P. Ochs, R. Ranftl, T. Brox, and T. Pock. Techniques for Gradient Based Bilevel Optimization with Nonsmooth Lower Level Problems. *Journal of Mathematical Imaging and Vision*, 56(2):175–194, 2016. (Cited on page 15).
- [179] O. Oktay, W. Bai, M. Lee, R. Guerrero, K. Kamnitsas, J. Caballero, A. de Marvao, S. Cook, D. O’Regan, and D. Rueckert. Multi-input Cardiac Image Super-Resolution Using Convolutional Neural Networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 246–254, 2016. (Cited on page 22).
- [180] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images, 1996. (Cited on page 93).
- [181] R. Otazo, E. Candès, and D. K. Sodickson. Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components. *Magnetic Resonance in Medicine*, 73(3):1125–1136, 2015. (Cited on pages 112, 116).
- [182] H. S. Park, S. M. Lee, H. P. Kim, and J. K. Seo. CT sinogram-consistency learning for metal-induced beam hardening correction. *Medical Physics*, 45(12):5376–5384, 2018. (Cited on page 23).
- [183] J. Park, D. Hwang, K. Y. Kim, S. K. Kang, Y. K. Kim, and J. S. Lee. Computed tomography super-resolution using deep convolutional neural networks. *Physics in Medicine & Biology*, 63(14):145011, 2018. (Cited on page 23).
- [184] P. Paschalis, N. D. Giokaris, A. Karabarbounis, G. K. Loudos, D. Maintas, C. N. Papanicolaos, V. Spanoudaki, C. Tsoumpas, and E. Stiliaris. Tomographic image reconstruction using Artificial Neural Networks. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 527(1-2):211–215, 2004. (Cited on page 1).
- [185] G. Peyré and J. M. Fadili. Learning Analysis Sparsity Priors. In *Sampta’11*, 2011. (Cited on pages 12, 15).
- [186] C.-H. Pham, A. Ducournau, R. Fablet, and F. Rousseau. Brain MRI super-resolution using deep 3D convolutional networks. In *IEEE International Symposium on Biomedical Imaging*, pages 197–200, 2017. (Cited on page 22).
- [187] T. Pock and A. Chambolle. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *International Conference on Computer Vision*, pages 1762–1769, 2011. (Cited on page 139).

- [188] T. Pock and S. Sabach. Inertial Proximal Alternating Linearized Minimization (iPALM) for Nonconvex and Nonsmooth Problems. *SIAM Journal on Imaging Sciences.*, 9(4):1756–1787, 2016. (Cited on pages 29, 36, 37, 138, 139).
- [189] K. P. Pruessmann, M. Weiger, P. Börnert, and P. Boesiger. Advances in sensitivity encoding with arbitrary k-space trajectories. *Magnetic Resonance in Medicine*, 46(4):638–651, 2001. (Cited on pages 63, 68, 98).
- [190] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–962, 1999. (Cited on pages 2, 22, 61, 62, 63, 66, 75, 92, 115).
- [191] P. Putzky and M. Welling. Recurrent Inference Machines for Solving Inverse Problems. *arXiv preprint arXiv:1706.04008*, 2017. (Cited on page 21).
- [192] H. Qi, Z. Chen, S. Wu, Y. Xu, and L. Zhou. Iterative image reconstruction using modified non-local means filtering for limited-angle computed tomography. *Physica Medica*, 32(9):1041–1051, 2016. (Cited on page 16).
- [193] C. Qin, J. Schlemper, J. Caballero, A. N. Price, J. V. Hajnal, and D. Rueckert. Convolutional recurrent neural networks for dynamic MR image reconstruction. *IEEE Transactions on Medical Imaging*, 38(1):280–290, 2019. (Cited on pages 17, 18, 111, 112).
- [194] T. M. Quan, T. Nguyen-Duc, and W.-K. Jeong. Compressed Sensing MRI Reconstruction using a Generative Adversarial Network with a Cyclic Loss. *IEEE Transactions on Medical Imaging*, 37(6):1488–1497, 2018, <https://github.com/tmquan/RefineGAN>. (Cited on page 27).
- [195] S. Ravishankar and Y. Bresler. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Transactions on Medical Imaging*, 30(5):1028–1041, 2011. (Cited on pages 10, 74, 76, 92).
- [196] S. Ravishankar and Y. Bresler. Learning Sparsifying Transforms. *IEEE Transactions on Signal Processing*, 61(5):1072–1086, 2013. (Cited on page 10).
- [197] S. Ravishankar and Y. Bresler. Efficient Blind Compressed Sensing Using Sparsifying Transforms with Convergence Guarantees and Application to Magnetic Resonance Imaging. *SIAM Journal on Imaging Sciences*, 8(4):2519–2557, 2015. (Cited on page 10).
- [198] S. Ravishankar and Y. Bresler. Data-Driven Learning of a Union of Sparsifying Transforms Model for Blind Compressed Sensing. *IEEE Transactions on Computational Imaging*, 2(3):294–309, 2016. (Cited on pages 10, 92).

- [199] S. Ravishankar, A. Lahiri, C. Blocker, and J. A. Fessler. Deep Dictionary-Transform Learning for Image Reconstruction. In *IEEE International Symposium on Biomedical Imaging*, pages 1208–1212, 2018. (Cited on pages 10, 21).
- [200] S. Ravishankar, B. Wen, and Y. Bresler. Online Sparsifying Transform Learning - Part I: Algorithms. *IEEE Journal of Selected Topics in Signal Processing*, 9(4):625–636, 2015. (Cited on page 10).
- [201] D. P. Reichert and T. Serre. Neuronal Synchrony in Complex-Valued Deep Networks. *arXiv preprint arXiv:1312.6115*, 2013. (Cited on page 101).
- [202] R. Reisenhofer, S. Bosse, G. Kutyniok, and T. Wiegand. A Haar wavelet-based perceptual similarity index for image quality assessment. *Signal Processing: Image Communication*, 61:33–43, 2018. (Cited on page 38).
- [203] C. Riess, M. Berger, H. Wu, M. Manhart, R. Fahrig, and A. Maier. TV or not TV? That is the question. In *Proceedings of the International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, pages 341–344, 2013. (Cited on pages 45, 46).
- [204] P. B. Roemer, W. A. Edelstein, C. E. Hayes, S. P. Souza, and O. M. Mueller. The NMR phased array. *Magnetic Resonance in Medicine*, 16(2):192–225, 1990. (Cited on page 56).
- [205] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 234–241, 2015. (Cited on pages 39, 40, 50, 125).
- [206] S. Roth and M. J. Black. Fields of experts. *International Journal of Computer Vision*, 82(2):205–229, 2009. (Cited on pages 14, 33, 93).
- [207] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear Total Variation Based Noise Removal Algorithms. *Physica D*, 60(1-4):259–268, 1992. (Cited on pages 8, 32, 46, 66).
- [208] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen. Improved Techniques for Training GANs. In *Advances in Neural Information Processing Systems*, pages 2234–2242, 2016. (Cited on page 27).
- [209] K. G. Samuel and M. F. Tappen. Learning optimized MAP estimates in continuously-valued MRF models. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 477–484, 2009. (Cited on pages 12, 14, 15).
- [210] C. M. Sandino and J. Y. Cheng. Deep convolutional neural networks for accelerated dynamic magnetic resonance imaging. In *Stanford University CS231N, Course project*. (Cited on page 22).

- [211] S. Scardapane, M. Scarpiniti, D. Comminiello, and A. Uncini. Learning activation functions from data using cubic spline interpolation. In *Neural Advances in Processing Nonlinear Dynamic Signals. WIRN 2017. Smart Innovation, Systems and Technologies*, volume 102, pages 73–83. 2019. (Cited on page 101).
- [212] S. Scardapane, S. Van Vaerenbergh, A. Hussain, and A. Uncini. Complex-valued Neural Networks with Non-parametric Activation Functions. *arXiv preprint arXiv:1802.08026*, 2018. (Cited on page 101).
- [213] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert. A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction. *IEEE Transactions on Medical Imaging*, 37(2):491–503, 2018, <https://github.com/js3611/Deep-MRI-Reconstruction>. (Cited on pages 17, 18, 26, 111, 112).
- [214] M. Schloegl, M. Holler, K. Bredies, and R. Stollberger. A Variational Approach for Coil-Sensitivity Estimation for Undersampled Phase-Sensitive Dynamic MRI Reconstruction. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, page 3692, 2015. (Cited on page 115).
- [215] M. Schloegl, M. Holler, A. Schwarzl, K. Bredies, and R. Stollberger. Infimal convolution of total generalized variation functionals for dynamic MRI. *Magnetic Resonance in Medicine*, 78(1):142–155, 2017. (Cited on pages 112, 116).
- [216] S. O. Schoenberg, O. Dietrich, and M. F. Reiser, editors. *Parallel Imaging in Clinical MR Applications*. Medical Radiology. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. (Cited on page 65).
- [217] J. Schwab, S. Antholzer, R. Nuster, and M. Haltmeier. DALnet: High-resolution photoacoustic projection imaging using deep learning. *arXiv preprint arXiv:1801.06693*, 2018. (Cited on page 23).
- [218] M. Seitzer, G. Yang, J. Schlemper, O. Oktay, T. Würfl, V. Christlein, T. Wong, R. Mohiaddin, D. Firmin, J. Keegan, D. Rueckert, and A. Maier. Adversarial and Perceptual Refinement for Compressed Sensing MRI Reconstruction. In *Machine Learning for Medical Image Reconstruction*, 2018. (Cited on page 27).
- [219] H. Shan, Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, L. Sun, W. Cong, and G. Wang. 3-D Convolutional Encoder-Decoder Network for Low-Dose CT via Transfer Learning From a 2-D Trained Network. *IEEE Transactions on Medical Imaging*, 37(6):1522–1534, 2018, <https://github.com/hmshan/CPCE-3D>. (Cited on page 22).
- [220] C. E. Shannon. Communication in the Presence of Noise. *Proceedings of the Institute of Radio Engineers*, 37(1):10–21, 1949. (Cited on pages 8, 66).

- [221] C. Shen, Y. Gonzalez, L. Chen, S. B. Jiang, and X. Jia. Intelligent Parameter Tuning in Optimization-Based Iterative CT Reconstruction via Deep Reinforcement Learning. *IEEE Transactions on Medical Imaging*, 37(6):1430–1439, 2018. (Cited on page 14).
- [222] G. Shi, M. M. Shanechi, and P. Aarabi. On the importance of phase in human speech recognition. *IEEE Transactions on Audio, Speech and Language Processing*, 14(5):1867–1874, 2006. (Cited on page 101).
- [223] J. Shi, Q. Liu, C. Wang, Q. Zhang, S. Ying, and H. Xu. Super-resolution reconstruction of MR image with a novel residual learning network algorithm. *Physics in Medicine & Biology*, 63(8):085011, 2018. (Cited on page 22).
- [224] O. Shitrit and T. Riklin Raviv. Accelerated magnetic resonance imaging by adversarial neural network. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 10553 LNCS, pages 30–38. Springer, Cham, 2017. (Cited on page 27).
- [225] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference for Learning Representations*, pages 1–14, 2015. (Cited on pages 27, 28).
- [226] D. K. Sodickson, R. Lattanzi, M. Vaidya, G. Chen, D. S. Novikov, C. M. Collins, and G. C. Wiggins. The Optimality Principle for MR signal excitation and reception: New physical insights into ideal radiofrequency coil design. *arXiv preprint arXiv:1808.02087*, 2018. (Cited on page 56).
- [227] D. K. Sodickson and W. J. Manning. Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays. *Magnetic Resonance in Medicine*, 38(4):591–603, 1997. (Cited on pages 2, 61, 62).
- [228] R. Tanno, D. E. Worrall, A. Gosh, E. Kaden, S. N. Sotiropoulos, A. Criminisi, and D. Alexander. Bayesian Image Quality Transfer with CNNs: Exploring Uncertainty in dMRI Super-Resolution. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, 2017. (Cited on page 22).
- [229] F. Thaler, K. Hammernik, C. Payer, M. Urschler, and D. Stern. Computed Tomography Slice Reconstruction from a Limited Number of Digitally Reconstructed Radiographs Using Wasserstein Generative Adversarial Networks. In *Machine Learning for Medical Image Reconstruction*, 2018. (Cited on pages 24, 27).
- [230] F. Thaler, C. Payer, and D. Stern. Volumetric Reconstruction from a Limited Number of Digitally Reconstructed Radiographs Using CNNs. In *Proceedings of the OAGM Workshop 2018*, pages 13–19, 2018. (Cited on pages 24, 105).
- [231] C. Tomasi and R. Manduchi. Bilateral Filtering for Gray and Color Images. In *International Conference on Computer Vision*, pages 839–846, 1998. (Cited on page 46).

- [232] C. Trabelsi, O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal. Deep Complex Networks. 2017, [https://github.com/ChihebTrabelsi/deep\\_complex\\_networks](https://github.com/ChihebTrabelsi/deep_complex_networks). (Cited on page 101).
- [233] J. Tsao, P. Boesiger, and K. P. Pruessmann. k-t BLAST and k-t SENSE: Dynamic MRI with high frame rate exploiting spatiotemporal correlations. *Magnetic Resonance in Medicine*, 50(5):1031–1042, 2003. (Cited on page 112).
- [234] M. Tygert, J. B. Estrach, S. Chintala, Y. LeCun, S. Piantino, and A. Szlam. A mathematical motivation for complex-valued convolutional networks. *Neural computation*, 28(5):815–825, 2016. (Cited on page 101).
- [235] M. Uecker. Parallel Magnetic Resonance Imaging. 2015. (Cited on page 65).
- [236] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig. ESPIRiT – An Eigenvalue Approach to Autocalibrating Parallel MRI: Where SENSE meets GRAPPA. *Magnetic Resonance in Medicine*, 71(3):990–1001, 2014. (Cited on pages 64, 65, 69).
- [237] A. Uncini, L. Vecchi, P. Campolucci, and Francesco. Complex-valued neural networks with adaptive spline activation function for digital radio links nonlinear equalization. *IEEE Transactions on Signal Processing*, 47(2):505–514, 1999. (Cited on page 101).
- [238] M. Urschler, K. Hammernik, T. Ebner, and D. Štern. Automatic Intervertebral Disc Localization and Segmentation in 3D MR Images Based on Regression Forests and Active Contours. In *Computational Methods and Clinical Applications for Spine Imaging*, pages 130–140. Springer International Publishing, 2016.
- [239] T. Valkonen. A primal-dual hybrid gradient method for nonlinear operators with applications to MRI. *Inverse Problems*, 30(5):055012, 2014. (Cited on pages 19, 20).
- [240] C. Van Chung, J. C. De los Reyes, and C. B. Schönlieb. Learning optimal spatially-dependent regularization parameters in total variation image denoising. *Inverse Problems*, 33(7):074005, 2017. (Cited on pages 12, 15).
- [241] S. V. Venkatakrisnan, C. A. Bouman, and B. Wohlberg. Plug-and-Play priors for model based reconstruction. In *IEEE Global Conference on Signal and Information Processing*, pages 945–948, 2013. (Cited on page 16).
- [242] P. Virtue, S. X. Yu, and M. Lustig. Better than real: Complex-valued neural nets for MRI fingerprinting. *IEEE International Conference on Image Processing*, 2017-Sept:3953–3957, 2018. (Cited on page 102).
- [243] V. Vishnevskiy, S. J. Sanabria, and O. Goksel. Image Reconstruction via Variational Network for Real-Time Hand-Held Sound-Speed Imaging. In *Machine Learning for Medical Image Reconstruction (MLMIR)*, 2018. (Cited on pages 17, 20).



- [244] G. Wang. A perspective on deep imaging. *IEEE Access*, 4:8914–8924, 2016. (Cited on pages 1, 92).
- [245] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler. Image Reconstruction is a New Frontier of Machine Learning. *IEEE Transactions on Medical Imaging*, 37(6):1289–1296, 2018. (Cited on pages 1, 29).
- [246] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, and D. Liang. Accelerating magnetic resonance imaging via deep learning. In *IEEE International Symposium on Biomedical Imaging*, pages 514–517, 2016. (Cited on pages 15, 92).
- [247] S. Wang, S. Fidler, and R. Urtasun. Proximal Deep Structured Models. In *Advances in Neural Information Processing Systems (NIPS)*, pages 865–873, 2016. (Cited on page 17).
- [248] 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, 2004. (Cited on pages 27, 37, 76, 99).
- [249] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multi-scale structural similarity for image quality assessment. *IEEE Asilomar Conference on Signals, Systems and Computers*, 2:9–13, 2003. (Cited on page 38).
- [250] B. Wohlberg. Efficient algorithms for convolutional sparse representations. *IEEE Transactions on Image Processing*, 25(1):301–315, 2016. (Cited on page 10).
- [251] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Isgum. Generative Adversarial Networks for Noise Reduction in Low-Dose CT. *IEEE Transactions on Medical Imaging*, 36(12):2536–2545, 2017. (Cited on pages 22, 27).
- [252] D. Wu, K. Kim, G. El Fakhri, and Q. Li. Iterative Low-dose CT Reconstruction with Priors Trained by Artificial Neural Network. *IEEE Transactions on Medical Imaging*, 36(12):2479–2486, 2017, <https://github.com/wudufan/KSAERecon>. (Cited on page 11).
- [253] T. Würfl, F. C. Ghesu, V. Christlein, and A. Maier. Deep Learning Computed Tomography. In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 432–440, 2016. (Cited on pages 24, 41, 45, 47).
- [254] T. Würfl, M. Hoffmann, V. Christlein, K. Breininger, Y. Huang, M. Unberath, and A. K. Maier. Deep Learning Computed Tomography: Learning Projection-Domain Weights From Image Domain in Limited Angle Problems. *IEEE Transactions on Medical Imaging*, 37(6):1454–1463, 2018, <https://github.com/ma0ho/Deep-Learning-Cone-Beam-CT>. (Cited on page 24).

- [255] L. Xiang, Y. Qiao, D. Nie, L. An, W. Lin, Q. Wang, and D. Shen. Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. *Neurocomputing*, 267:406–416, 2017. (Cited on page 23).
- [256] S. Xie, X. Zheng, Y. Chen, L. Xie, J. Liu, Y. Zhang, J. Yan, H. Zhu, and Y. Hu. Artifact Removal using Improved GoogLeNet for Sparse-view CT Reconstruction. *Scientific Reports*, 8(1):6700, 2018. (Cited on page 23).
- [257] J. Xu, E. Gong, J. Pauly, and G. Zaharchuk. 200x Low-dose PET Reconstruction using Deep Learning. *arXiv preprint arXiv:1712.04119*, 2017. (Cited on page 23).
- [258] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang. Low-dose X-ray CT reconstruction via dictionary learning. *IEEE Transactions on Medical Imaging*, 31(9):1682–97, 2012. (Cited on page 10).
- [259] K. Yan, X. Wang, L. Lu, and R. M. Summers. DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *Journal of Medical Imaging*, 5(03):1, 2018, <https://nihcc.app.box.com/v/DeepLesion>. (Cited on page 25).
- [260] B. Yang, L. Ying, and J. Tang. Artificial Neural Network Enhanced Bayesian PET Image Reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1297–1309, 2018. (Cited on page 23).
- [261] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, and D. Firmin. DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1310–1321, 2017. (Cited on pages 22, 27).
- [262] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang. Low-Dose CT Image Denoising Using a Generative Adversarial Network With Wasserstein Distance and Perceptual Loss. *IEEE Transactions on Medical Imaging*, 37(6):1348–1357, 2018. (Cited on pages 22, 27).
- [263] Y. Yang, J. Sun, H. Li, and Z. Xu. ADMM-Net: A Deep Learning Approach for Compressive Sensing MRI. In *Advances in Neural Information Processing Systems*, pages 10–18, 2017, <https://github.com/yangyan92/Deep-ADMM-Net>. (Cited on pages 17, 92).
- [264] J. Yao, J. E. Burns, D. Forsberg, A. Seitel, A. Rasoulian, P. Abolmaesumi, K. Hammernik, M. Urschler, B. Ibragimov, R. Korez, T. Vrtovec, I. Castro-Mateos, J. M. Pozo, A. F. Frangi, R. M. Summers, and S. Li. A Multi-Center Milestone Study of Clinical Vertebral CT Segmentation. *Computerized Medical Imaging and Graphics*, 49:16–28, 2016.

- [265] D. H. Ye, G. T. Buzzard, M. Ruby, and C. A. Bouman. Deep Back Projection for Sparse-View CT Reconstruction. *arXiv preprint arXiv:1807.02370*, 2018. (Cited on page 24).
- [266] J. C. Ye, Y. Han, and E. Cha. Deep Convolutional Framelets: A General Deep Learning Framework for Inverse Problems. *SIAM Journal in Imaging Sciences*, 11(2):991–1048, 2018. (Cited on pages 22, 24).
- [267] W. Yu, S. Heber, and T. Pock. Learning Reaction-Diffusion Models for Image Inpainting. In *German Conference on Pattern Recognition*, pages 356–367. Springer, 2015. (Cited on pages 20, 93).
- [268] J. Zbontar, F. Knoll, A. Sriram, M. J. Muckley, M. Bruno, A. Defazio, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana, Z. Zhang, M. Drozdal, A. Romero, M. Rabbat, P. Vincent, J. Pinkerton, D. Wang, N. Yakubova, E. Owens, C. L. Zitnick, M. P. Recht, D. K. Sodickson, and Y. W. Lui. fastMRI: An Open Dataset and Benchmarks for Accelerated MRI. *arXiv preprint arXiv:1811.08839*, 2018. (Cited on pages 26, 69, 96, 99).
- [269] M. D. Zeiler and R. Fergus. Visualizing and Understanding Convolutional Networks. In *European Conference on Computer Vision*, pages 818–833. Springer International Publishing, 2014. (Cited on page 93).
- [270] G. L. Zeng. *Medical Image Reconstruction*, volume 53. Higher Education Press, 2010. (Cited on page 42).
- [271] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. (Cited on page 1).
- [272] Y. Zhang and H. Yu. Convolutional Neural Network Based Metal Artifact Reduction in X-Ray Computed Tomography. *IEEE Transactions on Medical Imaging*, 37(6):1370–1381, 2018, <https://github.com/yanbozhang007/CNN-MAR.git>. (Cited on page 23).
- [273] Z. Zhang, X. Liang, X. Dong, Y. Xie, and G. Cao. A Sparse-View CT Reconstruction Method Based on Combination of DenseNet and Deconvolution. *IEEE Transactions on Medical Imaging*, 37(6):1407–1417, 2018. (Cited on page 23).
- [274] H. Zhao, O. Gallo, I. Frosio, and J. Kautz. Loss Functions for Image Restoration with Neural Networks. *IEEE Transactions on Computational Imaging*, 3(1):47 – 57, 2016. (Cited on pages 27, 37, 38, 99, 101, 147).
- [275] X. Zheng, S. Ravishankar, Y. Long, and J. A. Fessler. PWLS-ULTRA: An Efficient Clustering and Learning-Based Approach for Low-Dose 3D CT Image Reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1498–1510, 2018. (Cited on page 10).

- [276] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487–492, 2018. (Cited on page 24).
- [277] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In *IEEE International Conference on Computer Vision*, pages 2223–2232, 2017, <https://junyanz.github.io/CycleGAN/>. (Cited on page 27).
- [278] S. C. Zhu and D. Mumford. Prior learning and gibbs reaction-diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(11):1236–1250, 1997. (Cited on page 93).
- [279] M. V. Zibetti, A. Sharafi, K. Hammernik, F. Knoll, and R. R. Regatte. Comparing Learned Variational Networks and Compressed Sensing for  $T1\rho$  Mapping of Knee Cartilage. In *ISMRM Workshop on Machine Learning, Part II*, 2018.
- [280] M. V. Zibetti, A. Sharafi, K. Hammernik, F. Knoll, and R. R. Regatte. Variational Networks for Accelerating Biexponential 3D- $T1\rho$  Mapping of Knee Cartilage. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, page to appear, 2019.