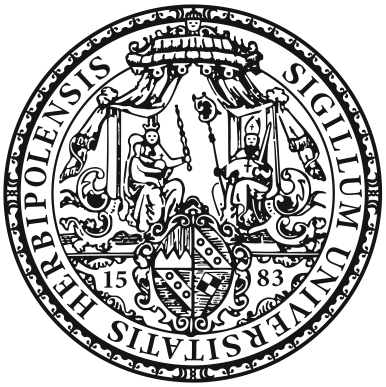
Doctoral Thesis



Variational networks in magnetic resonance imaging - Application to spiral cardiac MRI and investigations on image quality

Variational Networks in der Magnetresonanztomographie - Anwendung auf spirale Herzbildgebung und Untersuchungen zur Bildqualität

Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg
Section Biomedicine

submitted by

Jonas Kleineisel

from

Würzburg, Germany

Würzburg, 2023

Submitted on: _____

Members of the Thesis Committee

Chairperson:	Prof. Dr. Rüdiger Pryss
Primary Supervisor:	Prof. Dr. Tobias Wech
Supervisor (Second):	Prof. Dr. Michael Laßmann
Supervisor (Third):	Prof. Dr. Alfio Borzì
Supervisor (Fourth):	Dr. Johannes Tran-Gia

Date of Public Defence: _____

Date of Receipt of Certificates: _____

Contents

1 Summary

Acceleration is a central aim of clinical and technical research in magnetic resonance imaging (MRI) today, with the potential to increase robustness, accessibility and patient comfort, reduce cost, and enable entirely new kinds of examinations. A key component in this endeavor is image reconstruction, as most modern approaches build on advanced signal and image processing. Here, deep learning (DL)-based methods have recently shown considerable potential, with numerous publications demonstrating benefits for MRI reconstruction. However, these methods often come at the cost of an increased risk for subtle yet critical errors. Therefore, the aim of this thesis is to advance DL-based MRI reconstruction, while ensuring high quality and fidelity with measured data. A network architecture specifically suited for this purpose is the variational network (VN). To investigate the benefits these can bring to non-Cartesian cardiac imaging, the first part presents an application of VNs, which were specifically adapted to the reconstruction of accelerated spiral acquisitions. The proposed method is compared to a segmented exam, a U-Net and a compressed sensing (CS) model using qualitative and quantitative measures. While the U-Net performed poorly, the VN as well as the CS reconstruction showed good output quality. In functional cardiac imaging, the proposed real-time method with VN reconstruction substantially accelerates examinations over the gold-standard, from over 10 to just 1 minute. Clinical parameters agreed on average.

Generally in MRI reconstruction, the assessment of image quality is complex, in particular for modern non-linear methods. Therefore, advanced techniques for precise evaluation of quality were subsequently demonstrated. With two distinct methods, resolution and amplification or suppression of noise are quantified locally in each pixel of a reconstruction. Using these, local maps of resolution and noise in parallel imaging (GRAPPA), CS, U-Net and VN reconstructions were determined for MR images of the brain. In the tested images, GRAPPA delivers uniform and ideal resolution, but amplifies noise noticeably. The other methods adapt their behavior to image structure, where different levels of local blurring were observed at edges compared to homogeneous areas, and noise was suppressed except at edges. Overall, VNs were found to combine a number of advantageous properties, including a good trade-off between resolution and noise, fast reconstruction times, and high overall image quality and fidelity of the produced output. Therefore, this network architecture seems highly promising for MRI reconstruction.

2 Zusammenfassung

Eine Beschleunigung des Bildgebungsprozesses ist heute ein wichtiges Ziel von klinischer und technischer Forschung in der Magnetresonanztomographie (MRT). Dadurch könnten Robustheit, Verfügbarkeit und Patientenkomfort erhöht, Kosten gesenkt und ganz neue Arten von Untersuchungen möglich gemacht werden. Da sich die meisten modernen Ansätze hierfür auf eine fortgeschrittene Signal- und Bildverarbeitung stützen, ist die Bildrekonstruktion ein zentraler Baustein. In diesem Bereich haben Deep Learning (DL)-basierte Methoden in der jüngeren Vergangenheit bemerkenswertes Potenzial gezeigt und eine Vielzahl an Publikationen konnte deren Nutzen in der MRT-Rekonstruktion feststellen. Allerdings besteht dabei das Risiko von subtilen und doch kritischen Fehlern. Daher ist das Ziel dieser Arbeit, die DL-basierte MRT-Rekonstruktion weiterzuentwickeln, während gleichzeitig hohe Bildqualität und Treue der erzeugten Bilder mit den gemessenen Daten gewährleistet wird. Eine Netzwerkarchitektur, die dafür besonders geeignet ist, ist das Variational Network (VN). Um den Nutzen dieser Netzwerke für nicht-kartesische Herzbildgebung zu untersuchen, beschreibt der erste Teil dieser Arbeit eine Anwendung von VNs, welche spezifisch für die Rekonstruktion von beschleunigten Akquisitionen mit spiralen Auslesetrajektorien angepasst wurden. Die vorgeschlagene Methode wird mit einer segmentierten Rekonstruktion, einem U-Net, und einem Compressed Sensing (CS)-Modell anhand von qualitativen und quantitativen Metriken verglichen. Während das U-Net schlecht abschneidet, zeigen die VN- und CS-Methoden eine gute Bildqualität. In der funktionalen Herzbildgebung beschleunigt die vorgeschlagene Echtzeit-Methode mit VN-Rekonstruktion die Aufnahme gegenüber dem Goldstandard wesentlich, von etwa zehn zu nur einer Minute. Klinische Parameter stimmen im Mittel überein.

Die Bewertung von Bildqualität in der MRT-Rekonstruktion ist im Allgemeinen komplex, vor allem für moderne, nichtlineare Methoden. Daher wurden anschließend forgeschrittene Techniken zur präsizen Analyse von Bildqualität demonstriert. Mit zwei separaten Methoden wurde einerseits die Auflösung und andererseits die Verstärkung oder Unterdrückung von Rauschen in jedem Pixel eines untersuchten Bildes lokal quantifiziert. Damit wurden lokale Karten von Auflösung und Rauschen in Rekonstruktionen durch Parallele Bildgebung (GRAPPA), CS, U-Net und VN für MR-Aufnahmen des Gehirns berechnet. In den untersuchten Bildern zeigte GRAPPA gleichmäßig eine ideale Auflösung, aber merkliche Rauschverstärkung. Die anderen Methoden verhalten sich lokal

unterschiedlich je nach Struktur des untersuchten Bildes. Die gemessene lokale Unschärfe unterschied sich an den Kanten gegenüber homogenen Bildbereichen, und Rauschen wurde überall außer an Kanten unterdrückt. Insgesamt wurde für VNs eine Kombination von verschiedenen günstigen Eigenschaften festgestellt, unter anderem ein guter Kompromiss zwischen Auflösung und Rauschen, schnelle Laufzeit, und hohe Qualität und Datentreue der erzeugten Bilder. Daher erscheint diese Netzwerkarchitektur als ein äußerst vielversprechender Ansatz für MRT-Rekonstruktion.

3 Introduction

As an essential component of modern medicine and the life sciences, medical imaging is of supreme importance to public health today. Starting with the discovery of X-rays, and their subsequent applications for manifold purposes including diagnostics, the field has been steadily evolving and growing. Today, medical imaging includes a large variety of methods, among which are on one hand X-ray based techniques like conventional radiography and computed tomography (CT), and on the other hand ultrasound, MRI, and many more. While conventional radiography and CT enable fast and high-quality imaging, they expose the subject to ionizing radiation, which is associated with health risks and therefore a major drawback. Magnetic resonance imaging on the other hand is not harmful. It relies solely on the interactions of electro-magnetic fields with human tissue, which do not cause permanent damage if applied properly. Furthermore, due to the unique mechanisms through which the signals are produced, a wide variety of contrasts, in particular for soft tissues, can be achieved in the created images. For this reason, MRI represents the gold standard for many clinical queries, e.g. in the brain or the heart. However, it also comes with disadvantages: Compared to the process of creating signals in other imaging techniques, in magnetic resonance (MR), the entire scanning procedure is generally rather time-consuming. To some extent, this is simply a result of the physical mechanisms of magnetic resonance imaging, where a complex sequence of operations has to be played out to create signals and encode their spatial location. There are several aspects why this is problematic. First, a lengthy scan can be a significant burden for severely sick patients, or may even be entirely impossible due to their condition. Also, with increasing scan times more patient motion tends to occur, which can corrupt the acquired data, in some cases beyond diagnostic use. Furthermore, the acquisition and maintenance of MRI scanners is extremely expensive. Therefore, utilizing the available scan time as efficiently as possible is of economic interest. Lastly, for some clinical queries, it is necessary to observe dynamic processes in the body. This poses additional challenges, as it requires a certain temporal resolution, limiting the available scan time. For all of these reasons, ever since MRI was conceived as a technique for medical imaging, a central focus of research has been the acceleration of the procedure, while at the same time preseving or even improving image quality.

In the early days of MRI, most progress has been made by improving hardware. The

applied magnetic field strengths were continuously increased, which enabled higher signal levels. Also, gradient systems and pulse sequences were refined throughout. A major advancement was the development and widespread adoption of phased coil arrays. For reconstructing data acquired with these, parallel imaging (PI) algorithms were developed. By lowering the amount of acquired data below the Nyquist rate, an acceleration of the scanning procedure is achieved. This is known as undersampling. If these data are then reconstructed in a conventional manner, the consequence are artifacts in the resulting images. But by using the information that is encoded by multiple receiver coil channels, PI algorithms are able to correct these artifacts, and obtain a high-quality reconstruction. More recently, further advances in signal and image processing were made, along with an increase in the available computing power. This made more complex image processing algorithms feasible, opening new possibilities for improving MRI. An important example is the application of compressed sensing (CS) to MR, which made even higher undersampling rates, and therefore acceleration possible. Under certain conditions, it identifies and reduces redundancies in more complex domains to lower the amount of data that is needed for high-quality reconstructions further.

In recent years, the world of image processing has seen drastic developments. Machine learning, and in particular deep learning (DL) is rapidly penetrating the field, and challenges the state-of-the-art with ever new approaches outperforming previous methods. Advancements in this area often happen simultaneously on the computer hardware and software side: More powerful hardware makes more complex models with ever-higher parameter counts possible, while at the same time the latest models and algorithms are developed with powerful, next-generation hardware in mind. Some of the ideas of deep learning are already decades-old, but only now find their way to widespread adoption, as availability of computing power increases and new ideas for applications are conceived. Consequently, deep learning has also been applied to magnetic resonance imaging. The number and variety of promising possibilities for applications of machine learning in MRI is immense. Included are essentially all aspects of medical imaging, from AI-assissted sequence planning [1] to image reconstruction [2–5], post-processing [6–8] and classification of diseases [9, 10]. While all of these and many more are certainly worthwhile approaches to contribute to, in the thesis at hand the focus is exclusively on applications of deep learning to image reconstruction.

While the idea of applying artificial neural networks for reconstruction of medical images is not new [11], this approach only recently became feasible. Following a surge of interest in deep learning in other areas of image processing [12], the first modern applications of DL in MRI reconstruction started to appear around the middle of the 2010s [13]. Initially, the used network architectures were adapted from other fields, consisting of fully connected and convolutional layers. In 2018, several publications that later became widely influential were presented [2, 3, 14, 15]. What these have in common is that they

proposed new and innovative network architectures, that where specifically adapted for reconstruction of undersampled MRI data. Also, they showed a clear benefit of these architectures over the previous state-of-the-art. Soon after, the number and variety of approaches for exploiting DL in MRI reconstruction skyrocketed [5]. In particular, the concept of unrolling gradient descent iterations [2, 14, 15] proved to be successful and was later adapted in various forms [16–18]. Using the fact that MR reconstruction can be cast as an inverse problem, it can be interpreted as a combination of DL with compressed sensing. The key feature of this class of networks is that learned operators are combined with data consistency mechanisms, to enforce fidelity of the network output with the measured data. In the context of the described recent developments, it seems clear that deep learning-based methods in general bear a considerable potential for increasing reconstruction performance. However, an increasing number of researchers also consider how this comes at the cost of potential for critical errors. The key problem here is that unlike previous methods, DL-based ones have the potential to manufacture an output that appears error-free to a human observer, but actually misrepresent anatomy [18, 19].

A natural question to ask is therefore whether DL-based reconstruction models can be designed in such a way, that the frequency of these errors is minimized. As discussed in the previous paragraph, one approach to this are neural networks of unrolled gradient descent type, which incorporate data consistency mechanisms to this end. While multiple variants exist [14, 15], the focus here lies on variational networks [2]. Whether and - if so - how well this concept is suited for robust MRI reconstruction is a key research question that is to be investigated in this thesis. In particular, it is of interest how VNs perform not only in relation to previously established CS models, but also compared to other DL-based reconstruction methods without data consistency, as this allows drawing conclusions about what benefits these mechanisms bring. Commonly, these evaluations are based on global scores for image quality, like the mean-squared error with respect to some reference. While these measures can provide some coarse information on quality, a lot of nuance is inevitably lost when reducing an entire image to a single number. Therefore, a second central question that progress should be made toward in the following is how to better evaluate image quality. Methods to this end, like point-spread functions for analysis of resolution, or the g-factor for assessment of noise have existed for a long time, but the analysis of new, non-linear reconstruction methods requires further development. This may also lead to interesting insights about current reconstruction methods.

As discussed above, there are manifold benefits to faster scans in MRI. In dynamic imaging however, fast acquisitions are a prerequisite for applying MRI altogether, since observing dynamic processes requires a certain temporal resolution. In functional imaging of the heart, each frame of a dynamic series has to be acquired in no more than 50ms, to obtain an adequate temporal resolution. Since this is out of reach with conventional Cartesian sampling patterns and sufficiently high spatial resolution, "gating"

(also termed a "segmented cine") is usually used, where data from several heartbeats are combined into a single, virtual depiction of the beating heart. However, this requires the patient to hold breath several times during an exam, which can be strenuous, and is prone to errors. Recording data in free breathing would eliminate these problems, but it requires accelerated acquisitions. One way to achieve these are non-Cartesian sampling patterns, which are not yet widely used in clinical practice. This may be partly due to imperfections in the gradient systems, which degrade image quality. To overcome these, correction strategies [20–22] have recently been proposed. A second challenge is image reconstruction. DL-based methods have lately shown considerable potential in MRI reconstruction, but robust DL-based reconstruction methods for non-Cartesian sampling patterns are limited so far [23–29]. To this end, the first part presents a novel VN model for reconstruction of non-Cartesian data in functional imaging of the heart. This serves to investigate how a reconstruction based on VNs can be combined with spiral sampling patterns, and what benefits this can bring to functional imaging of the heart. To assess the suitability of the VN model for this task in relation to previous methods, the proposed VN reconstruction method is compared to a low-rank plus sparse model, a U-Net and the current clinical gold standard, a cine exam with a Cartesian balanced steady-state free precession (bSSFP) sequence. On these methods, extensive evaluation is performed, by comparing common image quality metrics as well as the clinical performance for functional imaging.

For contributing to an improvement in robustness and quality in DL-based MRI reconstruction in general, the second part of this thesis is dedicated to presenting and applying methods for assessing the quality of MR reconstructions. A generalized concept of point-spread functions is used for gaining insight into resolution properties, and a pseudo multiple-replica method is applied for assessing noise. These methods are tested on a suite of common modern reconstruction methods, including VNs, to understand differences and similarities in their behavior. This further allows drawing some rough general conclusions about the reconstruction methods, and also demonstrates how any other advanced method can be evaluated systematically and what problems can arise. A more critical view of image quality may therefore be promoted, which ultimately could improve the development of new reconstruction methods and the interpretation of MR images in general. An additional application of VNs to MRI reconstruction is considered in the summarizing discussion, where the use of 3D VNs for reconstruction of undersampled magnetic resonance cholangiopancreatography is shortly described.

To sum up, this thesis contributes to clinical MRI by advancing DL-based reconstruction with a focus on ensuring high image quality and data fidelity. It proposes a robust reconstruction method based on a variational network for a reliable approach to fast functional imaging of the heart with spiral trajectories, and demonstrates advanced techniques for analyzing different aspects of image quality in modern reconstruction methods.

4 Real-time cardiac MRI using an undersampled spiral k-space trajectory and a reconstruction based on a variational network

The contents of this chapter are adapted from the paper [30]

"Real-time cardiac MRI using an undersampled spiral k-space trajectory and a reconstruction based on a variational network"

by

Jonas Kleineisel[1], Herbert Köstler[1], Julius F. Heidenreich[1], Philipp Eirich[1,2], Bernhard Petritsch[1], Thorsten A. Bley[1], Nils Petri[3], Tobias Wech[1]

who are affiliated with the following institutions:

[1] Department of Diagnostic and Interventional Radiology

[2] Comprehensive Heart Failure Center

[3] Department of Internal Medicine I

at the University Hospital Würzburg in Würzburg, Germany.

The original article was published in *Magnetic Resonance in Medicine* in volume 88 on pages 2167 through 2178 and online (*https://doi.org/10.1002/mrm.29357*)

Previous results have been published as a conference abstract [31].

4.1 Abstract

Purpose: Cardiac MRI represents the gold standard to determine myocardial function. However, the current clinical standard protocol, a segmented Cartesian acquisition, is time-consuming and can lead to compromised image quality in the case of arrhythmia or dyspnea. In this article, a machine learning–based reconstruction of undersampled spiral k-space data is presented to enable free breathing real-time cardiac MRI with high image quality and short reconstruction times.

Methods: Data were acquired in free breathing with a 2D spiral trajectory corrected by the gradient system transfer function. Undersampled data were reconstructed by a variational network (VN), which was specifically adapted to the non-Cartesian sampling pattern. The network was trained with data from 11 subjects. Subsequently, the imaging technique was validated in 14 subjects by quantifying the difference to a segmented reference acquisition, an expert reader study, and by comparing derived volumes and functional parameters with values obtained using the current clinical gold standard, a segmented cine exam.

Results: The scan time for the entire heart was below 1 min. The VN reconstructed data in about 0.9 s per image, which is considerably shorter than conventional model-based approaches. The VN furthermore performed better than a U-Net and comparable to a low-rank plus sparse model in terms of achieved image quality. Functional parameters agreed, on average, with reference data.

Conclusions: The proposed VN method enables real-time cardiac imaging with both high spatial and temporal resolution in free breathing and with short reconstruction times.

4.2 Purpose

In clinical practice, MR-based determination of cardiac function is achieved through segmented Cartesian read-outs acquired over several heartbeats. By means of a simultaneously logged electrocardiogram, data can be processed into a series of fully sampled images, representing one pseudo RR-interval. However, combining data from multiple heartbeats requires a breath-hold for each slice or slice group, which renders the procedure time-consuming and strenuous for the patient. Furthermore, electrocardiogram gating may be corrupted for arrhythmic heartbeats, which frequently results in severe motion artifacts.

To waive the need for electrocardiogram and/or respiratory gating on the one hand, and to maintain the high temporal and spatial resolution on the other hand, many efforts have been made to considerably accelerate the procedure toward real-time imaging [32] in the past. Parallel imaging (PI) [33–35] and compressed sensing (CS) [36–41] ap-

proaches were applied to reconstruct undersampled data for this purpose. Moreover, non-Cartesian sampling was used to acquire data more efficiently [42] and more suitable for the application of model-based acceleration techniques [43, 44]. Because those trajectories are also more susceptible to inaccuracies of the gradient system, correction methods such as those based on a gradient system transfer function were additionally suggested [20, 21].

In the recent past, manifold approaches based on machine learning have been described to further push acquisition and reconstruction speed of MRI [2, 14, 18, 45, 46]. By shifting the optimization workload to an offline training procedure, these have shown potential to reconstruct data faster than previously used compressed-sensing approaches, while also modeling the reconstruction process more effectively [4]. Initially introduced for Cartesian k-space trajectories [2, 14, 18, 46–49], applications to non-Cartesian data have been presented subsequently [23–29].

A specific machine learning–based family of approaches, the so-called unrolled convolutional neural networks [2, 14, 15, 50], can be interpreted as a machine learning equivalent to classical compressed sensing: Regularizing layers are alternated with data consistency steps within a closed network architecture, which can be trained in an end-to-end fashion. Compared with straight-forward encoder–decoder convolutional neural networks (CNNs) like U-Nets [51], which do not as strictly enforce consistency with acquired k-space data, the problem of overfitting is reduced [4]. With respect to CS, learned models can reduce the dimensionality of data more adequately. Moreover, less "iterations" (i.e., cascades [45]) are typically required, and the convolutions of the regularization stage can be processed very efficiently on a GPU. Therefore, unrolled convolutional neural networks provide MR reconstructions in much shorter run times.

To exploit the aforementioned advantages for cardiac MRI, we implemented a VN for the reconstruction of undersampled spiral MR data. The network was subsequently trained and evaluated for real-time measurement of cardiac function in free breathing.

4.3 Methods

4.3.1 Spiral real-time k-space trajectory

Spiral trajectory

All data of this study were acquired on a 3T MR-system (MAGNETOM Prisma, Siemens Healthcare GmbH, Erlangen, Germany) using an in-house developed spoiled gradient echo pulse sequence with a spiral trajectory, which was presented recently [21]. The study was approved by the local ethics committee under IRB approval code 203/18sc following the application on 20.09.2018, and written informed consent was obtained from each participant. The non-Archimedean spiral waveforms were designed using a freely available software package [52]. Ten subsequently acquired spiral arms were equally dis-

tributed in k-space (i.e. with an angle-increment of $\frac{2\pi}{10}$). This represents one moderately undersampled frame of the real time series (see Figure 1). Acquiring one spiral arm takes TR = 4.96 ms, and thus one undersampled real-time frame can be acquired in slightly below 50 ms. Due to limits in peripheral nerve stimulation, which depend on the subject, the maximum gradient slew rate and amplitude had to be adjusted individually for each measurement. Therefore, we give ranges in the following, since the field of view (FOV) and resolution varies slightly between subjects. The FOV corresponds to $130 - 139$ mm close to the k-space center and decreases towards $44 - 47$ mm at k_{max}. The maximum values for gradient slew rates and amplitude were set to $S_{max} = 149.5$ mT/m/ms and $G_{max} = 36$ mT/m. Further parameters were: TE = 0.84 ms, dwell-time = 2.2 µs, flip-angle = 15°, in-plane spatial resolution = $1.34 - 1.46$ mm, slice thickness = 8 mm, phased array coil with 30-34 elements.

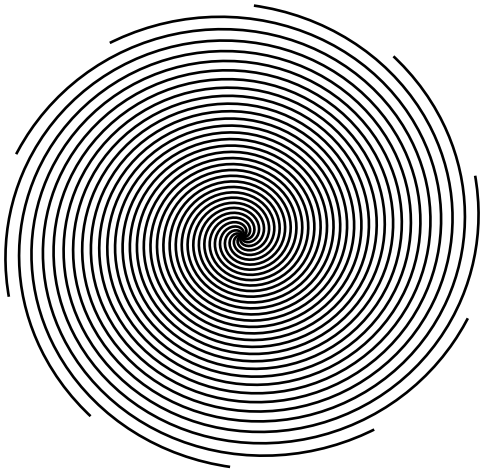


Figure 1: Shape of the spiral k-space trajectory. It consists of ten equally distributed non-Archimedean spiral arms. The sampling density is highest near the center of k-space and decreases towards the periphery.

Pattern rotation during series of real-time frames

A real-time depiction of the beating heart consists of a succession of real-time frames. However, there are two distinct benefits to varying the sampling trajectory over the course of a measurement: First, if combining the data from all real-times frames results in a fully sampled temporal mean image, this can be used for estimating coil sensitivity maps, which are necessary for a parallel imaging reconstruction. Second, if the sampling pattern is varied in the right way across the cardiac cycle, it is possible to bin the data such that one can also generate a segmented reconstruction. This requires the subject to hold their breath during data acquisition, usually with one breath-hold per slice. To

achieve these two objectives, we applied the following sampling pattern:

We first acquired n frames in some fixed orientation, where n is chosen such that $n \cdot 50$ ms slightly exceeds the time of one RR-interval of the subject. We usually took $n \approx 20$. Then, we rotated the entire pattern by an angle, such that the largest gaps in k-space were filled and again acquired a group of n frames, repeated up to eight times. These rotations consequently result in orientations of $0, \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}$, where the rotation angle is given in relation to the first group of frames as a fraction of $\frac{2\pi}{10}$. Combining all frames in the measurement then results in up to 80 equidistantly spaced spiral arms.

As a spiral reference standard for computing sensitivity maps and segmented reconstruction as described in the following, we used 40 equidistantly spaced spiral arms, which results in $\text{FOV}_{center} \approx 540$ mm and $\text{FOV}_{periph} \approx 180$ mm. Since the size of the observed objects is usually in the range of 50 cm, this is not technically fully sampled in the entirety of k-space. However, the moderate undersampling only in the periphery is not perceivable visually, therefore 40 siral arms were deemed sufficient. We refer to these reconstructions as 'spiral reference' in the following. Our real-time frames consequently have an undersampling factor of $R = 4$ with respect to this reference.

If we acquired at least four groups, we have sufficient data for computing sensitivity maps for the series. This was done separately for each slice using the adaptive approach published in [53].

Furthermore, if the subject held their breath during the acquisition, we can generate a segmented reconstruction as follows: Using the k-space center as self-gating signal, we partition each heartbeat into 20 heart phases. Then, each real-time frame gets assigned to that heart phase, where the temporal center point of the real-time frame acquisition is closest to the temporal center of the heart phase. If the duration of each group matches the duration of one heart beat, this results in exactly one frame per heart phase. If we acquire for at least four cardiac cycles, this gives at least 40 equidistant spiral arms in each heart phase, for longer acquisitions up to 80 equidistant spiral arms. Without triggering, the groups do not always have the same frequency as the heart beat. For the breath-hold measurements designated for segmented reconstruction, we therefore usually acquired 8-10 groups (the rotation pattern repeats after 8 groups). Since the heart rate usually varies, acquiring more groups than minimally needed increases the chance of having at least 40 equidistant spiral arms per heart phase. For free breathing measurements where we were only interested in sensitivity maps and real-time reconstructions, we only acquired between 4 and 6 groups.

If one is not interested in segmented reconstructions, the rotation pattern can be chosen more freely. In the original publication [21], the trajectory was rotated by an angle of $\frac{2\varphi}{10}$ between each frame, where $\varphi = \pi(3 - \sqrt{5})$ is the golden angle. Since φ is irrational, no two frames then have the same orientation, while the specific angle results in the gaps in k-space getting filled efficiently. However, a disadvantage is that since

every real-time frame has a different trajectory, undersampling artifacts are changing with every frame. If artifacts remain after applying a reconstruction method, these can be more disruptive. This method and trajectory will be referred to by its acronym 'CRISPI', which was introduced by the original publication [21].

4.3.2 Training, validation and evaluation dataset

For real-time cardiac imaging, obtaining a adequate dataset for training a neural network is challenging, because it is impossible to obtain fully sampled reference images in real time. This is the reason why we need accelerated acquisitions, after all. Therefore, one has to use reference images from other sources.

Clinical routine commonly uses segmented Cartesian bSSFP sequences. Since our spiral acquisitions are based on a spiral spoiled gradient echo sequence, these can not be used for training, as the contrasts differ. Instead, we acquired segmented spiral scans in breath-hold, as detailed in section 4.3.1. Since the segmented frames are combined together from at least four real-time frames in a similar phase in the cardiac cycle, these real-time frames can all be used as training data with the same target image. In this way, we generated 5462 training dataset items from acquisitions of 3 healthy volunteers with between 9 and 13 slices per subject and 168 real-time frames per slice.

Alternatively, one can use some other method of reconstructing undersampled data to generate reference images. This has the profound disadvantage that the network can not overcome any shortcomings of that method, as its output is given as training target. However, the reference images are at least proper real-time frames. We used data acquired with the CRISPI protocol [21] and reconstructed with the same low-rank plus sparse (LRS) model as in the original publication. The images were from 6 healthy volunteers and 2 clinical patients in both free breathing and breath-hold with 11 to 14 slices per subject and 47 real-time frames per slice.

By using a combination of training reference images from these two sources, the intent is that the network can learn the specific favorable properties from each group, e.g. the real-time temporal resolution from the CS reconstructions, while not getting degraded by the adverse properties. The training dataset thus consisted of in total 13334 spiral real-time frames.

For validation during training, we used an additional measurement of a healthy volunteer in free breathing and breath hold with the CRISPI protocol, in total 960 real-time frames with LRS reconstructions as references. Though unlikely to make a big difference, it can nevertheless be seen as a shortcoming that no segmented frames were used in the validation dataset.

We collected an additional dataset for evaluation of the network after training was finished. In the context of artificial neural networks this is sometimes also referred to as the test dataset. For these data, there is no need for reference reconstructions. We did

however reconstruct the data with several other methods for facilitating a comparison between methods, which is discussed in later sections. The dataset is comprised of scans of two healthy volunteers and twelve clinical patients in free breathing. We used a fast version of the pattern described in section 4.3.1, by acquiring 60 real-time frames in 3 groups per slice. This results in a total scan time of approximately 1 minute for all slices (usually around 12). To enable comparison, we also collected Cartesian reference scans from clinical routine. As it is common in clinical practice, the details of the acquisitions vary. They were based on bSSFP in 12 cases and on spoiled gradient echo in 2 cases. For the bSSFP scans, measurement parameters were in the following ranges: $46° \leq$ flip-angle $\leq 50°$, $2.84\,\text{ms} \leq \text{TR} \leq 3.02\,\text{ms}$, $1.42\,\text{ms} \leq \text{TE} \leq 1.51$ ms while the ones based on spoiled gradient echo had a flip-angle of 12°, TE = 2.89 ms and TR = 5.7 ms. For all scans, slice thickness was 7 or 8 mm and in-place resolution between 1.3 and 1.8 mm. All were segmented acquisitions with electrocardiogram (ECG)-triggering and 20 or 25 frames per heart phase. By taking into account the subjects' heart rates during measurement, one can calculate the temporal resolution to be between 26 and 56 ms.

4.3.3 The variational network model

MRI reconstruction as a variational problem

Consider the following view on MRI reconstruction:

Say real-world anatomical information in the form of ${}^1H$ density together with T_1 and T_2 information in space is represented by some image $x \in \mathbb{C}^n, n \in \mathbb{N}$. Then with some plausible assumptions on the MRI system, the sampling procedure as described above is a linear function, mapping real-world information to k-space frequency data in $k \in \mathbb{C}^m, m \in \mathbb{N}$. We want to call this operator the MRI sampling operator, or MRI forward operator, and denote it by $A : \mathbb{R}^n \to \mathbb{C}^m$. The measurement process is then be represented by

$$Ax = k \tag{1}$$

where k denotes k-space data and x the reconstructed image. The task of MR image reconstruction can then be viewed simply as the task of finding pre-images under this operator of given, measured, k-space data. This means reconstruction is an inverse problem. If A was bijective, then for each measurement data k there would exist a unique reconstruction x, which could be determined simply by inverting the linear operator. However, in practice A is usually not invertible. For one, the measured data k is subject to noise. With A possibly ill-conditioned, this may lead to solutions that are not necessarily a good representation of the underlying real-world data, or no solution may exist at all. Also, if undersampling is used, A loses its injectivity, so the solution may not be unique. Therefore, we need a robust way of recovering a - in some sense plausible - reconstruction x, even if the recorded MRI data k is insufficient to fully determine the reconstruction.

To this end, we first note that if (1) has a unique solution x, then x is the unique minimizer of the optimization problem

$$\underset{x\in\mathbb{C}^n}{\operatorname{argmin}}\frac{1}{2}\|Ax-k\|_2^2. \tag{2}$$

If however if there is no unique solution to (1), for the reasons mentioned above, then a solution to (2) may not represent a reasonable reconstruction. To overcome this problem, one can add a regularizing term Ψ to the energy functional in (2), which enforces constraints on its solution to make it closer resemble the solution to (1) if k were fully sampled and noise-free k-space data. This can also be interpreted as using prior knowledge about the fully sampled reconstructions. We thus obtain the regularized optimization problem

$$\underset{x\in\mathbb{C}^n}{\operatorname{argmin}}\frac{1}{2}\|Ax-k\|_2^2+\Psi(x). \tag{3}$$

Assuming the regularization term Ψ to be differentiable, the gradient of the target functional reads

$$\nabla_x\left(\frac{1}{2}\|Ax-k\|_2^2+\Psi(x)\right)=A^*(Ax-k)+\Phi(x) \tag{4}$$

where A^* denotes the adjoint operator of A and $\Phi:=\nabla_x\Psi$ the derivative of the regularization term. Using this, we can write down the gradient descent scheme to approximate the solution to the problem (3):

$$x^{t+1}=x^t-\lambda(A^*(Ax^t-k)+\Phi(x^t)) \tag{5}$$

For $\Phi=0$ this is known as a *Landweber iteration* [54]. The choice of Ψ and Φ remains open for now.

The MRI forward operator for Cartesian MRI reconstruction

In the following, the MRI forward operator A is described in detail, first for fully-sampled and then undersampled Cartesian data, after which the case of non-Cartesian sampling is discussed.

As discussed before, k-space frequency data can be transformed to image data using the Fourier transform. Therefore, in the simplest case of single-coil fully sampled data, the MRI sampling operator A is equal to the Fourier transform $\mathcal{F}$. However, in practice one almost exclusively uses multi-coil receiver arrays. In this case, the sampling operator has to take into account the sensitivity profiles of the coils, which describe the sensitivity of each coil to signals received in space. They are given as *sensitivity maps* $S=(S_1,...,S_{n_c})\in\mathbb{C}^{n_c}$ in image space. For a given image $x\in\mathbb{C}^N$, the i^{th} coil observes the image $S_i\odot x$, where $\odot$ denotes pointwise matrix multiplication. This can be elegantly described by the *expand operator*:

$$\mathcal{E}:\mathbb{C}^n\to\mathbb{C}^{n_c\times n},\quad \mathcal{E}(S,x)=\mathcal{E}_S(x)=(S_1\odot x,...,S_{n_c}\odot x) \tag{6}$$

This is a linear map. The MRI forward operator can then be written compactly as

$$A = \mathcal{F} \circ \mathcal{E}_S \tag{7}$$

where the Fourier transform is understood to operate coil-wise, which is always the convention in the following.

For reconstructing measured data, one has to combine multi-coil images with the help of sensitivity maps. This can be achieved with the following function, called the *coil reduction operator*:

$$\mathcal{R} : \mathbb{C}^{n_c \times n} \to \mathbb{C}^n, \quad \mathcal{R}(S, x) = \mathcal{R}_S(x) = \sum_{l=1}^{n_c} \bar{S}_l \odot x \tag{8}$$

Assuming we have normalized sensitivity maps S, i.e. S satisfy $\sum_{l=1}^{n_c} |S_l|^2 = 1$ point-wise, then $\mathcal{E}_S(x)$ is actually is the left inverse to the coil expansion as defined above, meaning it holds $\mathcal{R}_s(\mathcal{E}_S(x)) = x$ for all $x \in \mathbb{C}^n$.

Because of its importance in the variational approach described above, the adjoint operator A^* should also be characterized explicitly. It is well-known that the Fourier transform is a unitary operator, in the discrete case with respect to the standard euclidean scalar product on $\mathbb{C}^n$. Therefore we have $\mathcal{F}^* = \mathcal{F}^{-1}$. Furthermore, the expand and reduce operators as defined above are mutually adjoint, as one can easily see. Therefore, the adjoint MRI forward operator is given by

$$A^* = (\mathcal{F} \circ \mathcal{E}_S)^* = \mathcal{E}_S^* \circ \mathcal{F}^* = \mathcal{R}_S \circ \mathcal{F}^{-1}. \tag{9}$$

So far, we described the most basic case of fully sampled Cartesian multi-coil data, where k-space is sampled sufficiently dense according to the Nyquist for the selected field of view and spatial resolution. If undersampling is included additionally, this has to be incorporated into the forward operator. If we denote the undersampling simply as a map $\mathcal{M} : \mathbb{C}^n \to \mathbb{C}^n$ which zeros specific locations at which no data was sampled, then the forward operator reads [45]

$$A = \mathcal{M} \circ \mathcal{F} \circ \mathcal{E}_S, \tag{10}$$

where we understand $\mathcal{M}$ to operate coil-wise. One can easily see that the masking operator $\mathcal{M}$ is linear and self-adjoint. Thus, in this case the adjoint forward operator is given by

$$A^* = \mathcal{R}_S \circ \mathcal{F}^{-1} \circ \mathcal{M} \tag{11}$$

However, $\mathcal{M}$ is obviously not invertible, and thus the forward operator A is not unitary in this case.

The MRI forward operator for non-Cartesian data

For non-Cartesian acquisitions, A additionally has to include the projection onto the k-space trajectory. In the case of 2D imaging, let us denote the sampling trajectory as a vector $\tau \in \mathbb{R}^{m\times 2}$, which specifies a sampling location in $\mathbb{R}^2$ for each of m off-grid sampling points.

By convention, the forward operation which maps from a Cartesian grid to off-grid locations is called *de-gridding* or sometimes *adjoint gridding*, while the operation mapping from off-grid locations to a Cartesian grid is called *gridding*. We denote these maps by $\mathcal{G}_\tau : \mathbb{C}^m \to \mathbb{C}^n$ and $\mathcal{G}_\tau^* : \mathbb{C}^n \to \mathbb{C}^m$. Note that whether these two operations are inverse to each other depends on the trajectory and the gridding method, but in general they are not. The Fourier transform together with a gridding operator is commonly called the *non-uniform Fourier transform* (nuFFT). Then the entire acquisition process can be expressed by the forward operator

$$A = \mathcal{G}_\tau^* \circ \mathcal{F} \circ \mathcal{E}_S \tag{12}$$

Possible undersampling does not have to be explicitly included, since the sampling trajectory τ, already contains the full information about the sample locations.

For the adjoint operator, depending on the gridding method and the trajectory, one has to additionally include a density compensation term $\mathcal{D}_\tau$ before gridding, to account for possible non-uniform sampling of k-space. This is discussed in the next section in more detail. In this case, the adjoint operator is given by

$$A^* = \mathcal{R}_S \circ \mathcal{F} \circ \mathcal{G}_\tau \circ \mathcal{D}_\tau. \tag{13}$$

An illustration of the adjoint operator can be seen in Figure 2.

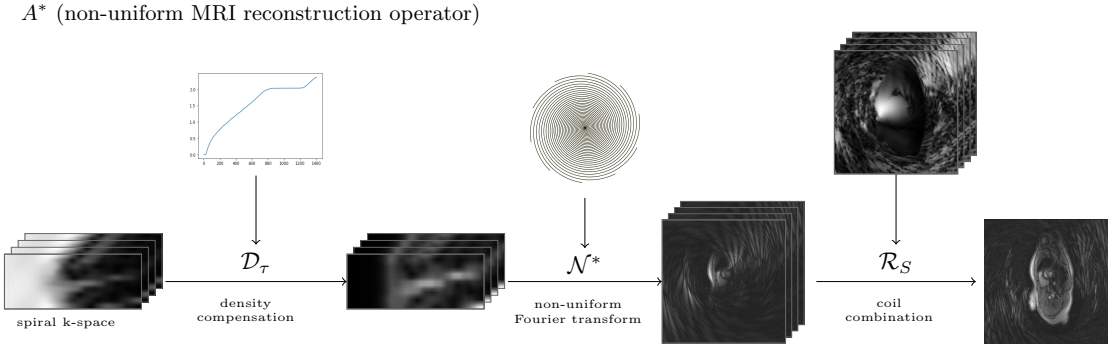


Figure 2: Illustration of the adjoint MRI forward operator A^* for non-Cartesian multi-coil data.

Density Compensation Functions

For the spiral trajectory used here (see section 4.3.1), the sampling density of the non-Cartesian data is higher near the center of k-space than in the periphery. Therefore,

if one were to directly apply gridding on this data, the resulting k-space would misrepresent the measured data, as the signal levels are overestimated in regions with higher sampling density and underestimated in regions with lower sampling density. This leads to degraded image quality in the reconstruction. Also, it prevents the inverse operation, which we want to call de-gridding, from being actually inverse to the gridding operator. Therefore, one usually compensates for these effects by modulating the magnitude of the non-Cartesian k-space data by a function representing the inverse local sampling density, called a density compensation function (DCF). Various methods for computing DCFs exist, e.g. analytically [55], by estimating density through a Voroni tessellation [56] or through iterative methods [57]. As non-Cartesian sampling patterns become more widely used due to their efficiency in k-space sampling, the interest in DCFs continues to rise, with new methods for their determination being published to this date [58]. The DCF is specific to the k-space sampling pattern τ only. Therefore it needs to be computed just once an can then be used for gridding k-space data from this trajectory.

Usually one is only interested in gridding non-Cartesian data once before applying the Fourier transform to obtain an image. In the VN model however, we actually repeatedly transform between non-Cartesian k-space and image space, respectively after executing a gradient descent step. This makes it important to avoid compounding of gridding errors, which would otherwise lead to entirely unusable results. This leads to the idea of approaching the computation of DCFs through the requirement that the gridding and de-gridding operations are mutually inverse. If there is a unique solution to this problem, this will also result in the DCF with the best image quality, as it is reasonable to assume image quality to be best when the off-grid data and on-grid representation are maximally consistent in the sense of the gridding operations.

Specifically, we obtain the DCF $\mathcal{D}_\tau$ as the scalar function that minimizes the squared L_2-norm between some k-space data multiplied by is to the result after nuFFT-transforming to image space and back several times, that is

$$\mathcal{D}_\tau = \underset{D \in \mathbb{R}^{\mathbb{M}}}{\operatorname{argmin}} \sum_{i=1}^{I} \| k_i - \mathcal{N}^{-1}(\cdots \mathcal{N}(D \cdot k_i)) \|_2^2 \tag{14}$$

where $\{k_1, ...k_I\}$ is a set of k-space data. We transform to image space and back several times, as this amplifies the effects of improperly represented k-space by compounding. This could increase the effectiveness of the norm difference in the optimization. The exact appearance of the sample k-spaces $\{k_1, ...k_I\}$ matters only moderately. The corresponding sample images should cover a reasonable range of features which translate into common k-space features being represented, to equally distribute the influence of the DCF on all important regions of k-space.

We compute the DCF for a spiral trajectory of 40 spiral arms as described in section 4.3.1. For this trajectory, it is a priori clear that the sampling density is decreasing

monotonically along each spiral arm. Therefore we minimize only on the subspace of monotonically increasing functions, which is enforced by projecting on this space after each optimization step. As k-space data, we selected 500 images of segmented cine frames and transformed them into synthetic off-grid k-space data. The minimization (14) was implemented in PyTorch [59] and carried out by the ADAM optimizer [60] with a learning rate of 10^{-2} for 104 epochs. We use three nuFFTs to image space and back. The mean of resulting DCF along the 40 spiral arms is shown in Figure 3.

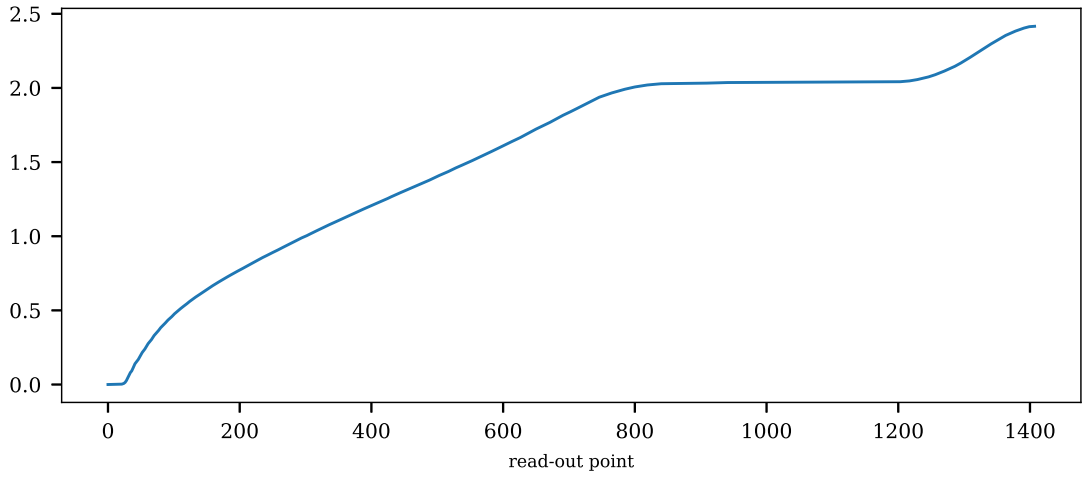


Figure 3: Shape of the computed density compensation function for a spiral trajectory (mean of all 40 spiral arms)

The variational network model

Let us now go back to equation (5). We established in section 4.3.3 that, with a proper regularization term Φ, this gradient descent scheme approximates a solution to the MRI reconstruction problem (1). In compressed sensing, the regularization term Φ is chosen carefully by exploiting sparsity of the signal in some domain. However, this choice is not at all straight-forward, and it is unclear which one is best suited.

Following an entirely different approach, the idea of variational networks is, to use machine learning to obtain the regularization operator Φ from data. To this end, Φ is realized by some artificial neural network, to which the task of enforcing a proper solution then falls to. Many variants of this idea have been published [2, 14, 15, 23, 26, 45, 47, 61–63].

Usually, in a gradient descent scheme, the number of iterations is either dynamically determined by a termination criterion, or, less ideally, fixed at a high number, after which convergence is likely. The latter is the case for most CS models. However, both of these approaches are problematic for a VN, in particular for its training. Since a dynamic number of iterations is not practical, as this would change the size of the network upon

inference, one has to use a fixed number of steps. But if the number of iterations is high, the network is very deep, as the regularization networks are applied subsequently. This leads to vanishing gradient problem, as well as high computational complexity. In particular for non-Cartesian data, where gridding and de-gridding has to be performed, this would result in very long runtimes. Thus, one uses a small, fixed number of steps, typically around 10, which are called cascaces. An illustration can be seen in Figure 4.

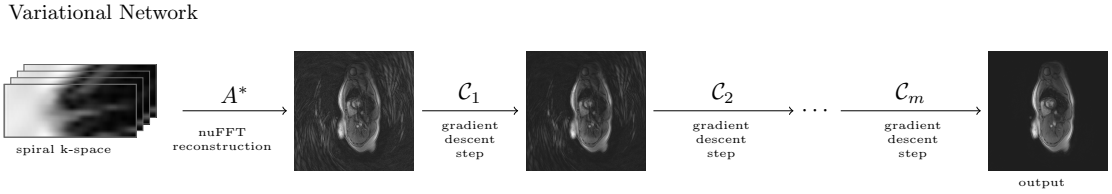


Figure 4: Illustration of a variational network. After transforming to image space using the adjoint MRI operator A^*, a fixed number of cascades is applied subsequently.

In each cascade, a step

$$x^{t+1} = x^t - \lambda(A^*(Ax^t - k) + \Phi(x^t)) \tag{15}$$

is performed, with Φ a neural network. The term $\lambda(A^*(Ax^t - k)$, the gradient of $\frac{\lambda}{2}\|Ax - k\|_2^2$, enforces agreement of the network output with measured data. It is called the *data consistency term*. For the architecture of the regularization term Φ, a large number of options have been used, usually CNNs. Many of the later publications use U-Nets [45], but other options include plain CNNs [23], residual neural networks [63] or a fields of experts model [2]. Just like in other networks, it is trained by using gradient backpropagation from the network output. An illustration of one cascade can be seen in Figure 5.

In some models, a separate network for each cascade is used [2, 14, 45], while in others the same network is applied repeatedly [15, 63]. The latter has the advantage of dramatically reducing the number of trainable parameters of the entire network by a factor of the number of cascades (and thus likely also the required training data). However, using separate networks allows more degrees freedom in the gradient descent operation.

Variational network model for non-Cartesian k-space data

For reconstructing the data in this study, we use a VN model with appropriate forward operator for non-Cartesian data as described previously. Gridding is achieved by using the Kaiser-Bessel gridding operator implemented in [64], after applying the DCF described in section 4.3.3.

As in other recent publications [18, 45], we chose a U-Net model for realizing the regularization terms in the VN.

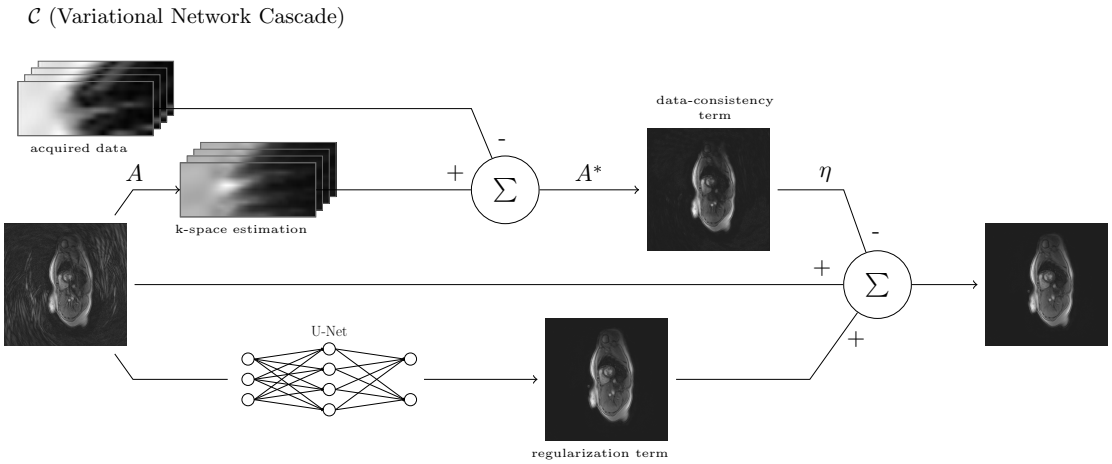


Figure 5: Illustration of one cascade of a variational network. It performs a regularized gradient descent step, consisting of a data consistency term (upper part) and a regularization term (lower part) realized by a neural network. A variety of architectures have been proposed.

Features are extracted by four subsequent convolutional blocks with downsampling by average pooling. Subsequently, up-sampling is achieved by four transpose convolutions with stride and kernel size 2 (followed by instance normalization and a leaky ReLU activation function), and further convolutional blocks. The network features skip connections, i.e. the intermediate outputs in the downsampling path are concatenated before the convolutional blocks in the upsampling path.

Each convolutional block consist of a multi-channel convolution with kernel size 3×3, followed by instance normalization and a leaky ReLU activation function (with slope 0.2), all of which is repeated two times. In the first convolution, the number of channels is doubled (in the downsampling path) or halved (in the upsampling path). Due to the skip connections, the transpose convolutions for up-sampling also halve the number of channels.

All network parameters are real numbers, and the complex image data is treated as a real image with two channels. The first convolutional block expands these two channels to 18. No bias is applied in any of the convolutions except the last one before the output. This architecture results in around $2.5 \cdot 10^6$ trainable parameters per U-Net. The entire VN with 10 cascades has a total of $24.5 \cdot 10^6$ trainable parameters.

4.3.4 Network training

We used the structural similarity index measure (SSIM) as loss function. It is perceptually motivated [65] and was preferred over the mean-squared error (MSE), since the latter often introduces spatial blurring [66].

For computing the gradients during training, we used the common ADAM optimizer

[60] with a learning rate of $5 \cdot 10^{-3}$. Due to GPU memory constraints, the batch size was set to 1. As we found it to introduce slight spatial blurring, no data augmentation was applied. We executed the training procedure for 12 epochs while monitoring the SSIM loss curves. On the hardware specified in 4.3.6, this took a total of approximately 105 hours.

After epoch 7 the train loss continued to decrease, while the validation loss started to increase. We interpreted this as a sign that overfitting was present after this point. Therefore, the weights of the network after training until epoch 7 were subsequently used. This corresponds to $93.34 \cdot 10^3$ training steps, which took approximately 56 hours.

The training and validation loss curves can be seen in Figure 6.

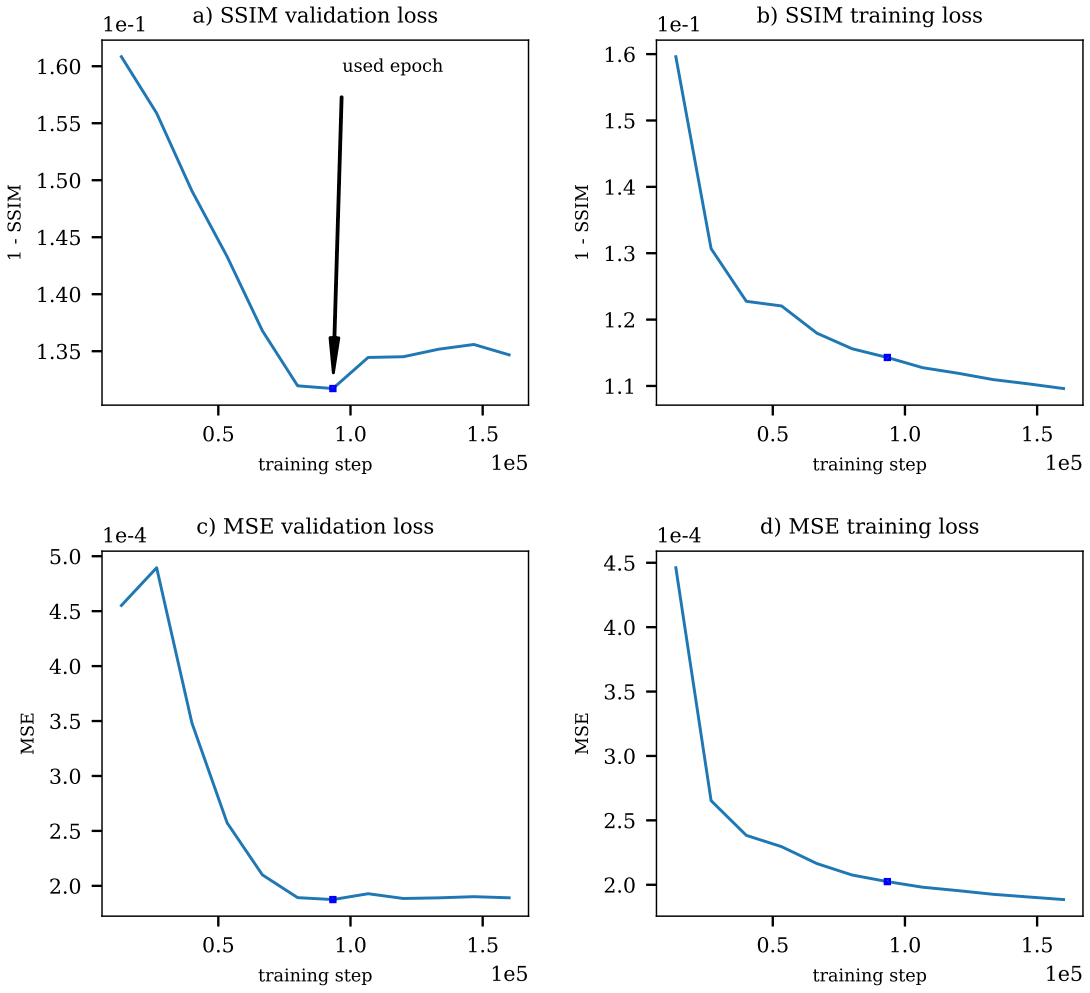


Figure 6: Training and validation loss curves of the variational network. The training was carried out with SSIM as loss function. MSE was computed for additional monitoring only.

4.3.5 Comparison methods

To evaluate the performance of the VN model in reconstructing undersampled real-time data, presented above, we chose two alternative state-of-the-art methods to compare it against: A U-Net model and a low-rank plus sparse model. The latter was originally introduced in [67] and previously used for a similar application in [21].

The U-Net model is designed as follows: It first applies an adjoint nuFFT to transform the undersampled spiral data into image space, which it then passes through a single U-Net for correction of undersampling artifacts. The U-Net is the same network as was used for regularization in the VN, but with 5 pooling layers and 30 channels in the first layer, resulting in a number of $2.7 \cdot 10^7$ trainable parameters in total. This size was chosen to enable a fair comparison by approximately matching the number of parameters in the U-Net model to the number of parameters in the VN model, which contains 10 U-Nets of smaller size (in total $2.5 \cdot 10^7$ parameters). As loss function, we also use SSIM. The training and validation datasets were the same as for the VN, but we applied additional data augmentation in the form of random rotations during training. We trained for 16 epochs ($2.1 \cdot 10^5$ steps) which, due to much faster inference than the VN model, took approximately 35h.

4.3.6 Implementation

The neural network models were implemented in the open-source machine learning framework PyTorch [59]. Parts of the VN and the U-Net implementation were modified from the freely available implementation in the fastMRI project [66]. The nuFFT and Kaiser-Bessel gridding operators were provided by the open-source library TorchKBnuFFT [64]. For computing coil sensitivity maps we used the method described in [53]. The source code is publicly available on GitHub [68].

Both neural network models were implemented to run on an NVIDIA Titan RTX GPU. The LRS model was implemented in MATLAB® (The MathWorks, Inc., Natick, Massachusetts, United States) and supported CPU execution only. It was described in [21] where it was used for reconstruction of data acquired with the CRISPI method.

4.3.7 Reconstruction time

With the implementations as described above, we investigated the runtimes for the various reconstruction methods by reconstructing a representative real-time acquisition from the evaluation dataset, consisting of 12 slices and 60 real-time frames per slice. Both the VN and the U-Net model operate on each time frame separately, while the LRS model applies a temporal model and therefore reconstructs the entire series slice-wise. Accumulating the overall reconstruction time per slice for VN and U-Net, we tested for differences in the runtimes per slice, by comparing the runtime of the methods first all

together with a Friedman test, and subsequently pair-wise with Wilcoxon signed-rank tests.

4.3.8 Quantitative evaluation

We evaluated the reconstruction performance of the VN in several ways, and compared it to that of the U-Net and LRS model.

For a comparison by means of quantitative parameters, we calculated SSIM and root mean-squared error (RMSE) between the respective reconstructions and reference images. However, as discussed before, since it is impossible to obtain fully sampled real-time reference data, we generated a synthetic dataset from segmented spiral images: For one of the volunteers in the evaluation dataset (see section 4.3.2), we generated segmented spiral cine images as described in 4.3.1 from breath-hold acquisitions of 9 slices with 20 frames per slice. These were de-gridded back to their acquisition pattern of 40 spiral arms. By discarding 30 of these, we obtained pairs of undersampled spiral data (10 arms, as the real-time acquisitions) and reference images.

The undersampled spiral data can then be reconstructed by various methods, and the result compared to the reference images by means of quantitative measures. We are mostly interested in an accurate depiction of the heart, and not so much the background or other anatomy. Therefore, we applied a circular mask that tightly covers the entire heart prior to computing SSIM and RMSE. This results in scores on 180 images for VN, LRS, U-Net and also the naive reconstruction by simple nuFFT.

For interpreting the results, we first conducted a Friedman test to test for whether the reconstruction method has a significant impact on the performance. Post-hoc, we applied Wilcoxon signed-rank tests for each pairing of two of these methods, to investigate differences in SSIM or RMSE scores between the methods in detail.

4.3.9 Expert reader study

Quantitative parameters like SSIM and RMSE are well-known to have number of shortcomings for judging the visual quality of images, which is why an evaluation by human experts is imperative [18].

To this end, we reconstructed all data in the evaluation dataset with VN, U-Net and LRS, and submitted these together with the Cartesian reference acquisitions for evaluation by two expert radiologists with four and ten years of experience in cardiac imaging. The evaluation was carried out in the categories delineation/sharpness, artifacts, myocardium-blood contrast, temporal resolution and noise.

We used a four-point Likert scale, with the following grades: 1: poor, 2: moderate, 3: good, 4: very good. For each of the 14 subjects in the evaluation dataset, each rater provided one such score per category. The two raters were blinded to each other as well as to the reconstruction type.

The performance of the reconstruction methods was compared by first conducting a Friedman test, to investigate a significant impact of the reconstruction method on the scores. Post-hoc, Wilcoxon signed-rank tests were conducted on the scores of each pairing of methods, separately for each category. For the statistical analysis, the average grade of the two readers was used, to alleviate the effects of inter-rater variability.

4.3.10 Quantification of functional parameters

For evaluation of clinical performance, the proposed method was applied to functional imaging of the heart. On the entire evaluation dataset, along with the gold standard Cartesian reference images, we determined cardiac functional parameters, specifically end-diastolic volume (EDV), end-systolic volume (ESV), stroke volume (SV) and ejection fraction (EF). To enable the analysis of inter-rater variability, the evaluation was done repeatedly, by the same two expert radiologists as in 4.3.9. Both evaluated the two sets of images independently from each other using a dedicated medical imaging software (cvi42, Circle Cardiovascular Imaging Inc., Calgary, Canada).

Subsequently, the functional parameters derived from the two methods were compared through Bland-Altman analysis [69, 70]. We used the method variant as described in [70], which specifically takes into account the repeated evaluation.

Additionally, to provide context to the deviations between the methods, we estimated the inter-rater variability, also through Bland-Altman analyses. This was done for each parameter separately on the real-time VN images and the segmented Cartesian acquisitions.

4.3.11 Investigation of alternative loss function

In the scientific literature, many potential loss terms for training deep learning models are presented [4], each with their own advantages and disadvantages. The loss function is one of many choices, that a scientist or engineer working with machine learning has to make when designing a new artificial neural network model. Often, it is not clear, which term is best for a specific application, and the investigation of different loss functions and their differences remains an active field of research. We attempt to contribute to this discussion, by investigating the impact of choosing SSIM or mean-squared error (MSE) as loss function for our model. This was done by repeating the training procedure with the same parameters, but changing the loss term from SSIM, which we used in the main model, to MSE, and then comparing the two resulting models. The two models will be referred to as MSE-VN and SSIM-VN. If no loss term is specified, SSIM-VN is meant.

The training and validation loss curves for MSE-VN can be seen in Figure 7. As shown there, we trained the MSE-VN model for 14 epochs, but used the weights after training for 10 epochs, or $133.3 \cdot 10^3$ steps, because the MSE validation loss stopped decreasing. Both training loss curves are very similar for MSE-VN and SSIM-VN (see

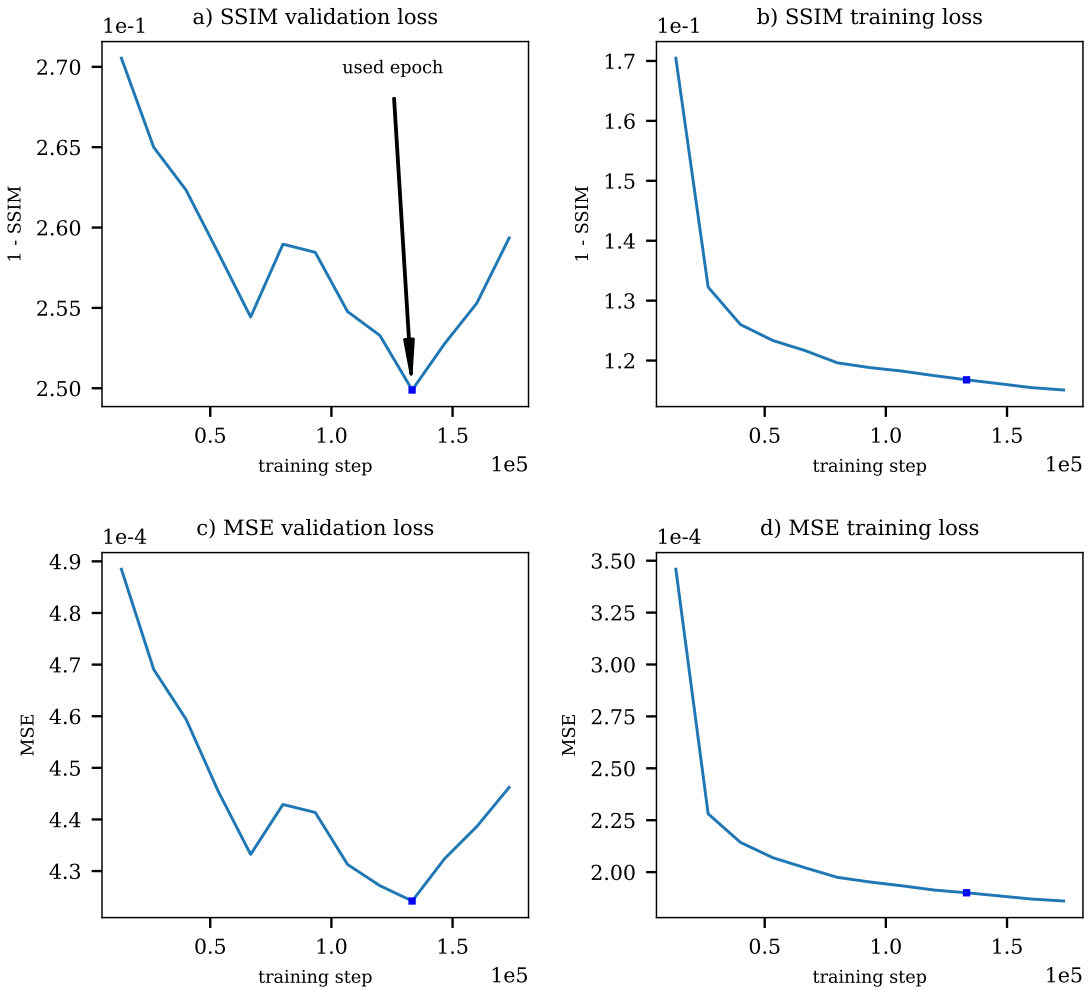


Figure 7: Training and validation loss for training of the VN model trained with MSE as loss function. For MSE-VN, the SSIM loss curves were computed for additional monitoring, but not used in the training procedure.

Figure 6). The validation loss curves however indicate substantially lower MSE and SSIM validation loss values for MSE-VN than for SSIM-VN.

For computing quantitative scores, we used the same method and synthetic dataset as in 4.3.8. The scores between the two models were compared with a Wilcoxon signed-rank test for each parameter. Furthermore, we selected one slice of the dataset, and compare reconstructions of several frames by MSE-VN and SSIM-VN visually.

4.4 Results

4.4.1 Effects of variational network reconstruction in k-space

Fig. 8 shows the k-space data of an exemplary undersampled real-time frame together with its VN and direct nuFFT reconstruction. For an exemplary coil, the off-grid data

on the 10 spiral arms (see Fig. 1) can be seen in Fig 8 a). Kaiser-Bessel gridding transforms this data to k-space on a Cartesian grid, shown in Fig. 8 b). An inverse Fourier transform then results in a nuFFT reconstruction of the undersampled data (also in b). The artifacts arising from the undersampling are clearly noticeable.

In Fig. 8 c), one can see the reconstruction by the VN method. The undersampling artifacts are substantially attenuated. In the corresponding k-space, the spiral pattern is no longer apparent.

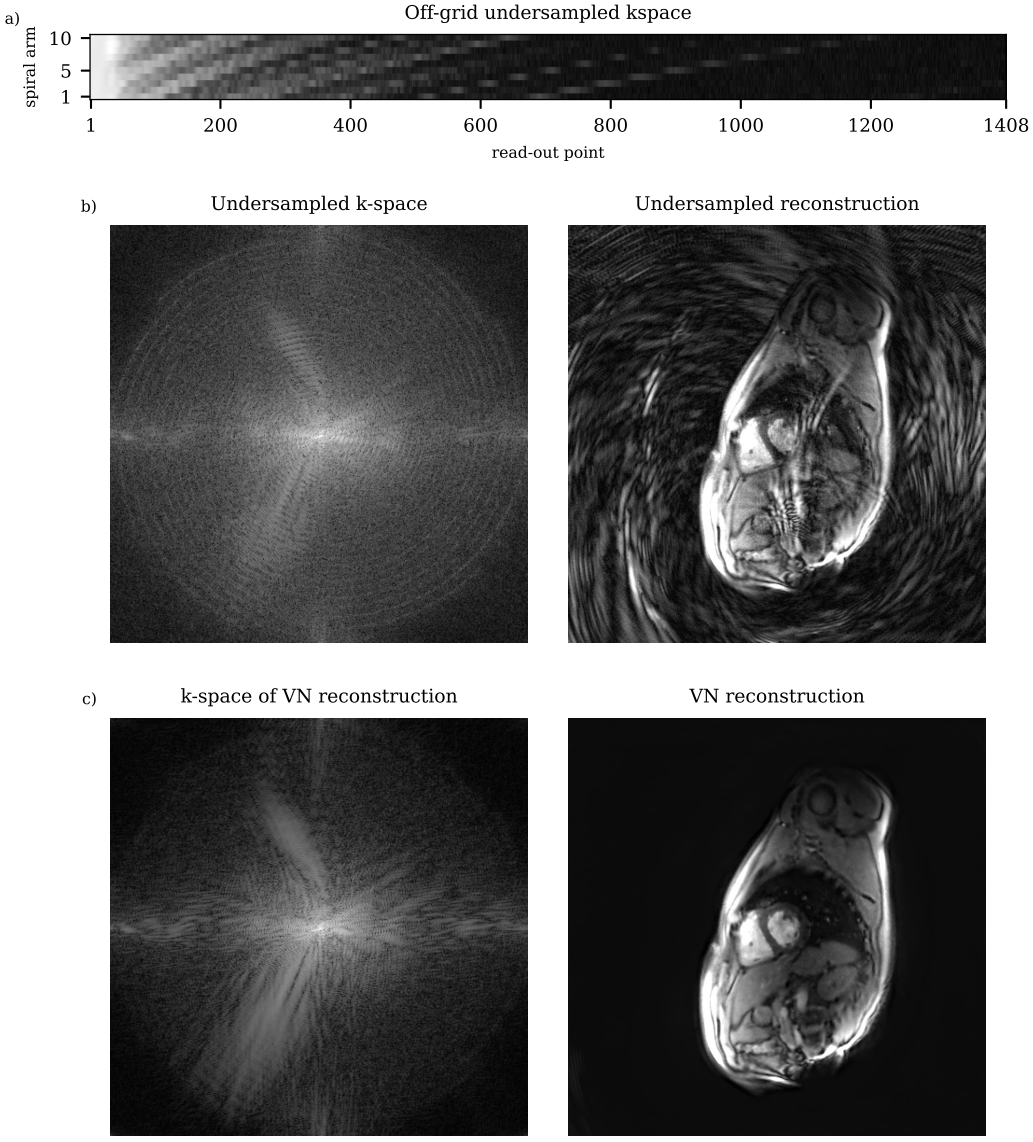


Figure 8: Effects of the VN reconstruction for a representative real-time frame. a) shows off-grid spiral k-space data (log of magnitude) b) shows the same dataset, after performing a nuFFT reconstruction, together with the corresponding on-grid kspace (log of magnitude). c) shows the reconstruction by the variational network, together with the corresponding k-space (log of magnitude).

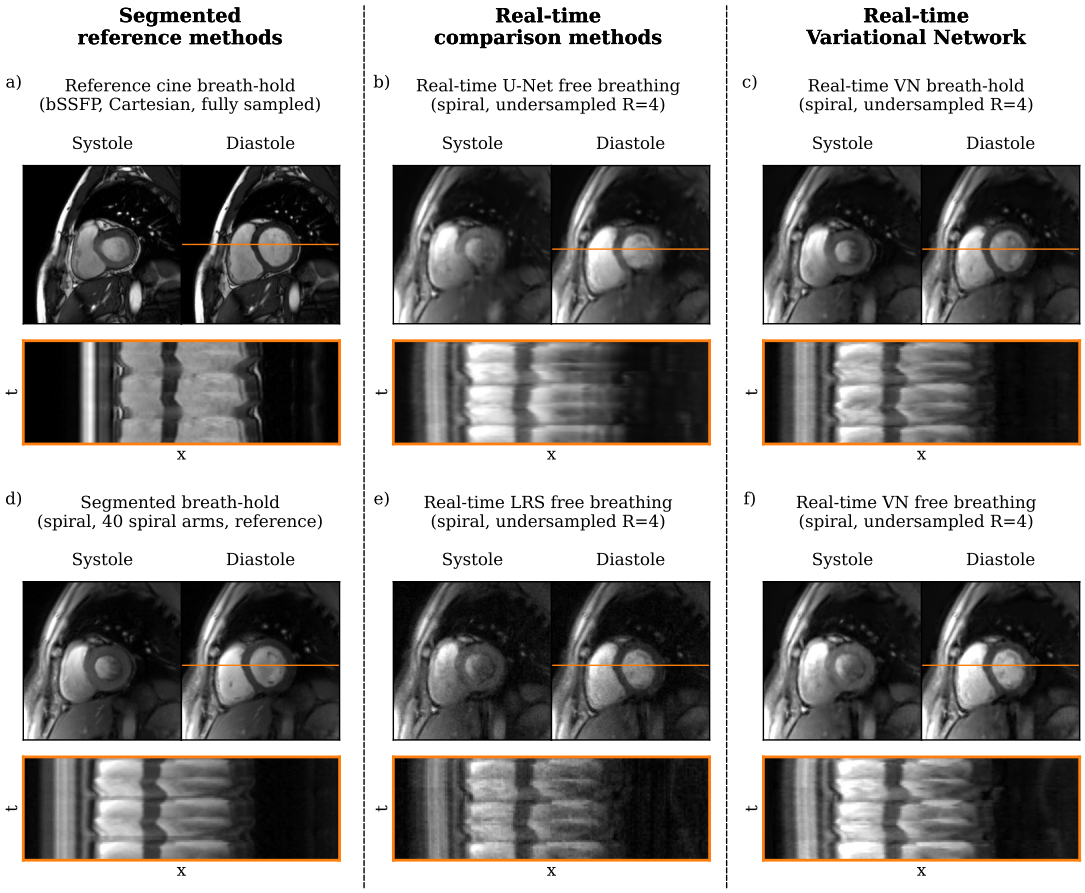


Figure 9: Systolic and diastolic frames obtained by different imaging approaches in a healthy participant. The temporal cross-sections (x-t plot) show the signal intensity of the indicated line across all reconstructed cardiac phases reconstructed. For the real-time acquisition, this corresponds to 50 frames. For the segmented approaches, 25 frames were depicted repeatedly a) and in 20 frames d). The image in a) was obtained in breath-hold with a fully sampled Cartesian balanced SSFP (bSSFP) protocol, as typically used in clinical practice. The patient position is slightly different from the other acquisitions, most likely due to motion between the acquisitions. Images in c) and d) are obtained from the same spiral acquisition in breath-hold. In d), data were binned into segmented frames, whereas in c) the data were reconstructed as real-time frames using VN. The images in b), e), and f) are reconstructed from the same acquisition in free breathing using a U-Net model, a low-rank plus sparse model, and the VN model, respectively. In b)–f), signal voids due to flow can be observed in the left ventricle close to the lateral wall. A dynamic view of this comparison is also available (see Supplementary Material Information at the end of chapter 4).

4.4.2 Qualitative comparison of reconstruction methods

For comparing the image quality of VN reconstruction to the alternative methods described in 4.3.5 as well as the clinical reference technique, Fig. 9 shows systolic and diastolic frames for six different reconstructions as well as x-t plots. The latter show

the indicated orange line across 50 frames, where the temporal dimension is represented by the y-axis. A more detailed qualitative comparison can be seen in Fig. 10, where we used the synthetic data as described in section 4.3.2 to compare the undersampled reconstructions to spiral segmented fully sampled references with error maps. As supplementary material (see Supplementary Material Information at the end of chapter 4), one can see cardiac dynamics in 9 slices of a representative patient from the evaluation dataset simultaneously.

The first column of Fig. 9 displays segmented reference methods: a) shows the clinical reference Cartesian bSSFP acquisitions described in 4.3.2 while d) shows segmented reconstructions from spiral acquisitions (see 4.3.1). Because of the bSSFP acquisition, the contrast on a) is noticeably different from the other subplots, where the acquisitions are based on spoiled gradient echo.

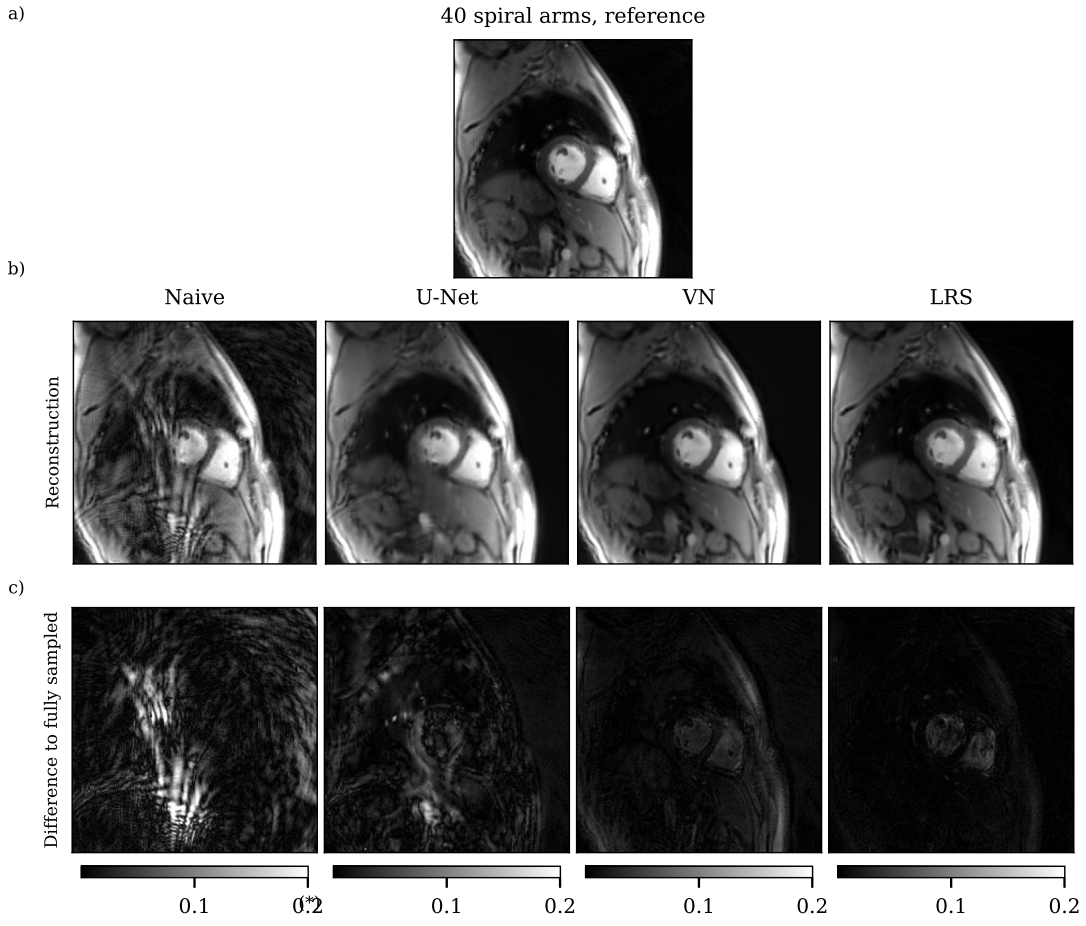


Figure 10: Qualitative comparison of reconstructions by methods for reconstruction of undersampled real-time data. To obtain a fully sampled spiral reference reconstruction for generation of error maps, the undersampled data was generated by retrospectively undersampling a segmented spiral acquisition. *The error map of the naive reconstruction is clipped, i.e. values above the indicated range are shown as white.

VN	U-Net	LRS
0.89 ± 0.03 s	0.07 ± 0.01 s	40.50 ± 0.69 s

Table 1: Runtimes of the different methods for reconstructing a single real-time frame. Since the LRS model applies a temporal model, and thus reconstructs slice wise, the number given here was obtained by dividing the runtime for an entire slice by the number of real-time frames per slice (60). Note that VN and U-Net were implemented on GPU, while LRS ran on CPU only.

In the center column, one can see reconstructions by the comparison reconstruction methods for real-time data, U-Net and LRS, described in 4.3.5. In the x-t-plot in b), streaks in t-direction can be seen, which are presumably residual artifacts that remain stationary over time.

Finally, the last column shows reconstructions by the presented VN. c) shows a frame from an acquisition in breath-hold, while the frame in f) was acquired in free breathing. As can be expected due to the real-time acquisition, no difference in quality between the two can be observed.

In all real-time reconstructions, one can observe dark signal voids in the cardiac chambers and in the x-t plots. These are due to blood flow. Since turbulence resulting from flow is aperiodic, and the segmented methods average over several cardiac cycles, these regions appear smoother in both segmented reconstructions.

In Figure 10, large reconstruction errors can be observed in the U-Net output. These may be residual artifacts or reconstruction errors. When comparing the U-Net error map the that of the naive reconstruction, some similarities can be seen, but also some errors that are not due to undersampling artifacts. In the lung, as well as the abdomen, hallucinated blood vessels can be observed.

The VN and LRS reconstructions can barely be differentiated from the fully sampled reconstruction. In the error maps, one can see that for VN, the largest differences to fully sampled appear in the bright fat, and some moderate differences in the heart. For LRS, the differences to fully sampled in the heart are slightly higher than for VN. As the heart is the area of the image with the most dynamic, a potential reason for this observation may be that the LRS method applies a temporal model.

4.4.3 Speed of reconstructions

The results of the investigation and comparison of runtimes of the tested reconstruction methods (see section 4.3.7) can be seen in Table 1.

Additionally, computing sensitivity maps took on average 17.95 ± 0.03 s per slice. For VN, reconstructing the representative measurement with 12 slices and 60 real-time

frames per slice took 14.39 min in total. The statistical tests showed the runtimes to be significantly different between all methods together as well as for each pair-wise comparison.

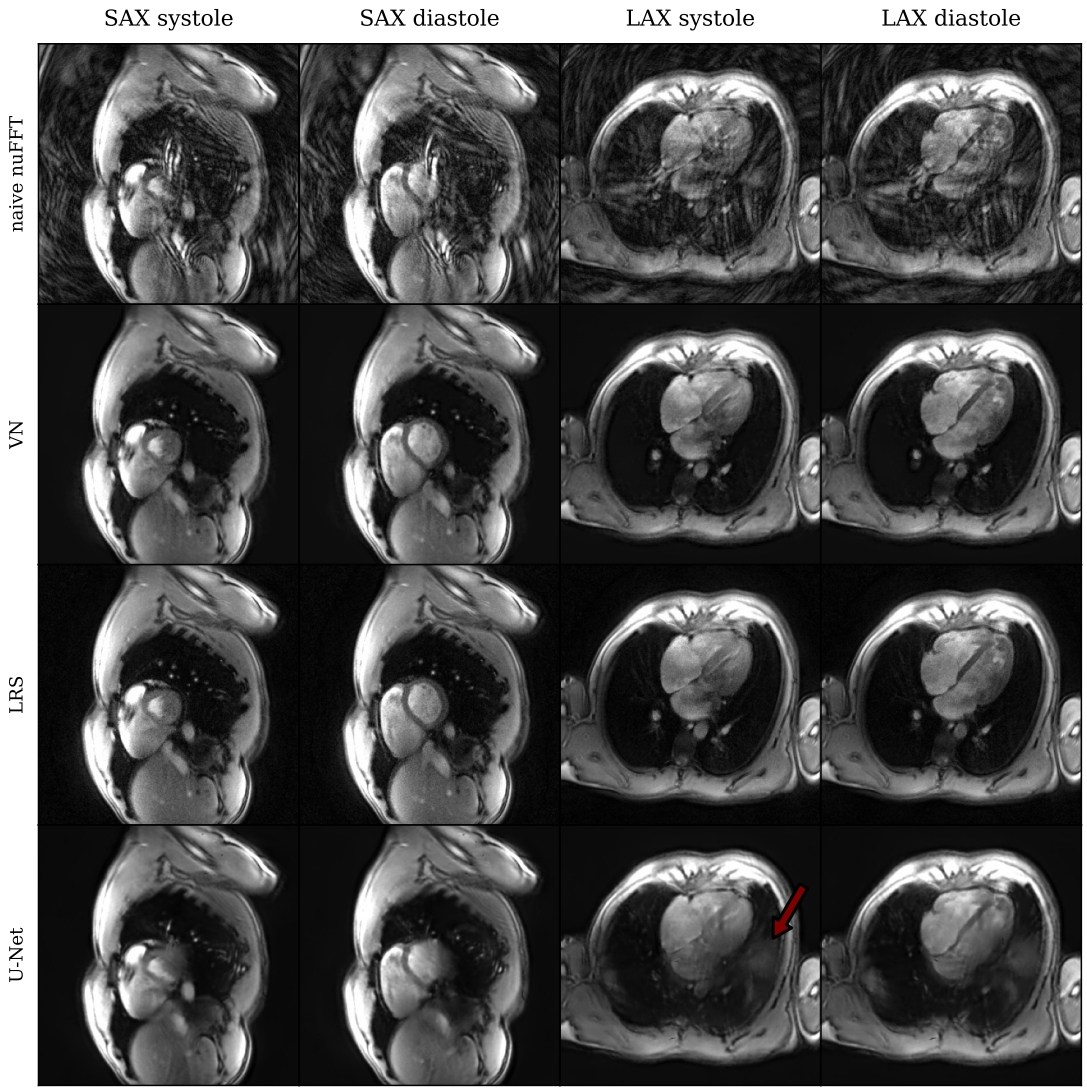


Figure 11: Comparison of reconstructions by the examined methods of short-axis (SAX) and long-axis (LAX) data in systole and diastole of a healthy volunteer, which is not part of the training or evaluation dataset. The VN and U-Net models were trained on a dataset containing only short-axis training images. The red arrow indicates substantial reconstruction errors in the U-Net reconstructions of long-axis images. Note that the quality of the long-axis images is below the clinical standard, presumably due to flow artifacts resulting from the real-time acquisition.

4.4.4 Investigation of generalization properties on long-axis images

To assess the generalization properties of the compared reconstruction methods, we applied them to short axis (SAX) and long axis (LAX) data of an additional (not part

	Naive	U-Net	VN	LRS
SSIM	0.67 ± 0.04	0.83 ± 0.04	0.94 ± 0.01	0.95 ± 0.03
RMSE $\times 10^2$	4.40 ± 0.65	2.36 ± 0.24	0.92 ± 0.25	1.05 ± 0.30

Table 2: Results of the quantitative analysis, comparing structural similarity index measure (SSIM) and root-mean squared error (RMSE) between different reconstruction methods. The scores were computed from reconstructing a synthetic dataset of 180 images from 9 slices and comparing the respective reconstructions to fully sampled reference reconstructions with the respective quantitative measure. A circular mask around the heart was applied before computing scores.

of the training or evaluation set) healthy volunteer, shown on Figure 11. The reconstruction quality of the SAX images is in line with the observations from the previous chapter, where errors can be observed for U-Net, but VN and LRS show similar, high image quality. On the LAX images, this holds true as well, but the quality of the U-Net reconstruction is substantially inferior. At the side of the thorax (see red arrow), strong hallucinations are present.

This is remarkable, because the U-Net and VN models were trained on only SAX images, yet the VN is able to reconstruct LAX images without apparent loss of quality, while the U-Net is not.

4.4.5 Quantitative comparison of reconstruction methods

The quantitative scores that were computed as described in section 4.3.8 can be seen in Table 2. As expected, the naive reconstruction without correction of undersampling artifacts performs poorly. The U-Net achieves somewhat better, yet by no means good scores. VN and LRS are scored highly on both measures, where SSIM score is almost identical and RMSE scores are slightly lower for VN. From these data, the Friedman test showed there to be a significant difference ($p = 0.05$) in scores between the reconstruction methods, separately for SSIM and RMSE. The post-hoc Wilcoxon signed-rank tests concluded that each pair-wise comparison exhibits a significant difference at $p = 0.05$. Thus, excluding the naive reconstruction, U-Net performs worst, and VN best in terms of RMSE while LRS performs best in SSIM.

4.4.6 Expert reader study

The mean scores with standard deviations from the expert reader study as described in section 4.3.9 are listed in Table 3.

The Cartesian reference cine achieves the highest scores. As in the evaluation by quantitative parameters, in most categories, U-Net performs mediocre to poor, while VN

and LRS show similarly high scores.

The Friedman test showed significant differences between the methods in all categories ($p = 0.05$). The subsequent pair-wise comparison through Wilcoxon signed-rank tests confirmed that Cartesian cine is rated significantly ($p = 0.05$) better than all other methods in all categories, except that no significant difference in temporal resolution compared to LRS could be found. On the other hand, U-Net is scored worst in all except the category noise. Between LRS and VN, no significant difference in scores could be determined at $p = 0.05$, except that LRS exhibits a significantly higher level of noise.

	Cartesian cine	U-Net	VN	LRS
Delineation/Sharpness	4.0 ± 0.2	1.8 ± 0.6	3.0 ± 0.7	2.9 ± 0.5
Artifacts	3.6 ± 0.6	1.3 ± 0.5	2.6 ± 0.6	2.8 ± 0.5
Myocardium-Blood contrast	3.9 ± 0.3	2.1 ± 0.8	3.0 ± 0.7	2.9 ± 0.7
Temporal resolution	3.7 ± 0.6	2.0 ± 0.7	3.0 ± 0.8	3.1 ± 0.7
Noise	3.9 ± 0.3	2.3 ± 0.6	2.5 ± 0.5	1.6 ± 0.5

Table 3: Results of the expert reader study, comparing Cartesian segmented cine and three reconstruction methods for spiral real-time data through expert scores in five categories. The scores were given by two experienced radiologists on a four-point Likert scale, rating scans of fourteen subjects.

4.4.7 Investigation of alternative loss function

The mean values and standard deviations of the quantitative analysis for MSE-VN and SSIM-VN are shown in Table 4. Both the SSIM and RMSE scores are slightly inferior for MSE-VN compared to SSIM-VN. The Wilcoxon signed-rank test showed the differences in both SSIM and RMSE to be significant at $p = 0.05$.

For a visual comparison of MSE-VN and SSIM-VN, Figure 12 shows reconstructions by the two models, together with error maps. No difference is immediately noticeable in the reconstructions, only in the error map can differences be observed. Concerning the background, SSIM-VN seems to have higher error rates in bright regions, in particular fat, while MSE-VN has a higher error level in the dark background. In the heart, slight reconstruction errors in the form of small dots can be observed in both models, but on a similar overall level and for the most part in the same areas.

	trained with SSIM (SSIM-VN)	trained with MSE (MSE-VN)
SSIM score	0.94 ± 0.01	0.92 ± 0.01
RMSE score $\times 10^2$	0.92 ± 0.25	0.96 ± 0.10

Table 4: Comparison of structural similarity index measure (SSIM) and root mean-squared error (RMSE) scores between a variational network (VN) model trained with SSIM (SSIM-VN) and one trained with mean-squared error (MSE-VN). The scores are computed on the same dataset of 180 images from 9 slices with references from a spiral cine sequence as in section 4.3.9. They quantify the difference to the fully sampled ground truth cine images, after applying a circular mask around the heart.

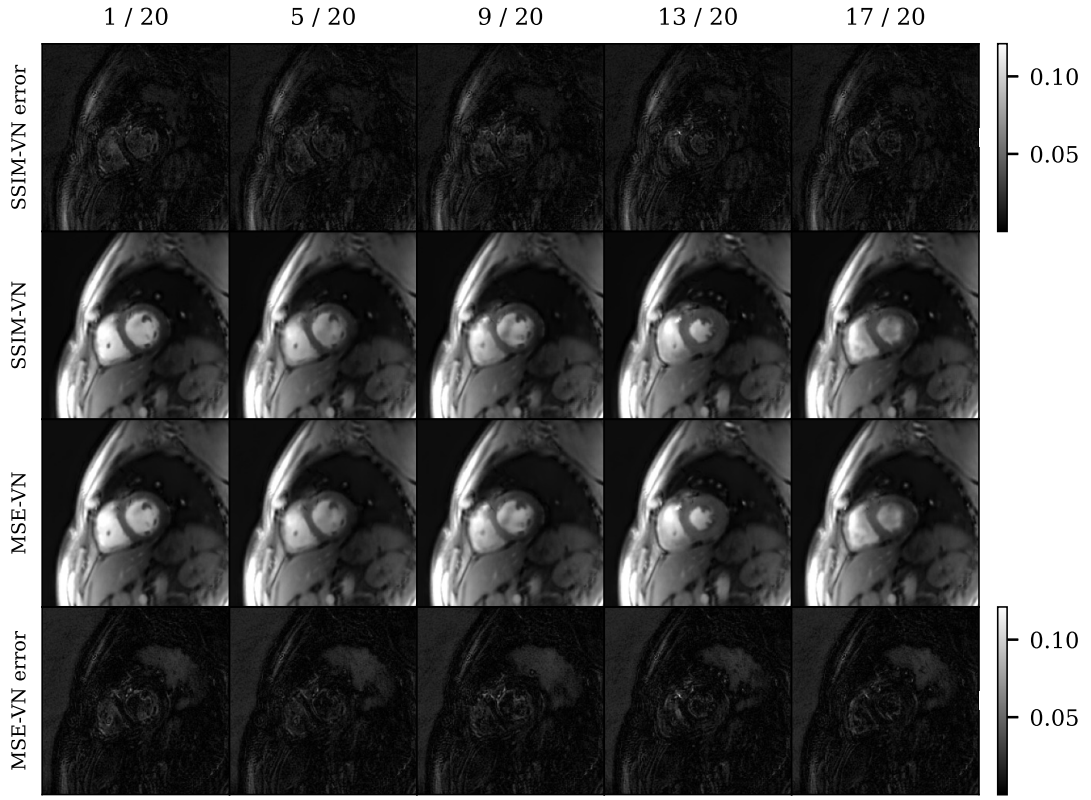


Figure 12: Comparison between the variational network trained with SSIM as training loss (SSIM-VN) and the one trained with MSE as training loss (MSE-VN). To obtain a dataset with ground truth reconstructions for computing error maps, segmented spiral cine reconstructions of an exemplary slice were de-gridded to obtain synthetic data, which was used as input for both networks.

4.4.8 Derived functional parameters

The cardiac functional parameter values that were derived from the Cartesian cine reference and the real-time VN reconstruction are shown in Figure 13. Their corresponding comparisons between the methods and the observers through Bland-Altman analyses can

be seen in Figure 14.

For the comparison between the reference cine and the real-time VN method, we observe moderate deviations on average for EDV, ESV and SV, and very good agreement for EF. The samples do not appear to show any proportional bias. However, for all parameters, the limits of agreement around the mean deviation are quite large, between 10 - 30 % of the reference mean value. This indicates substantial variation between the methods.

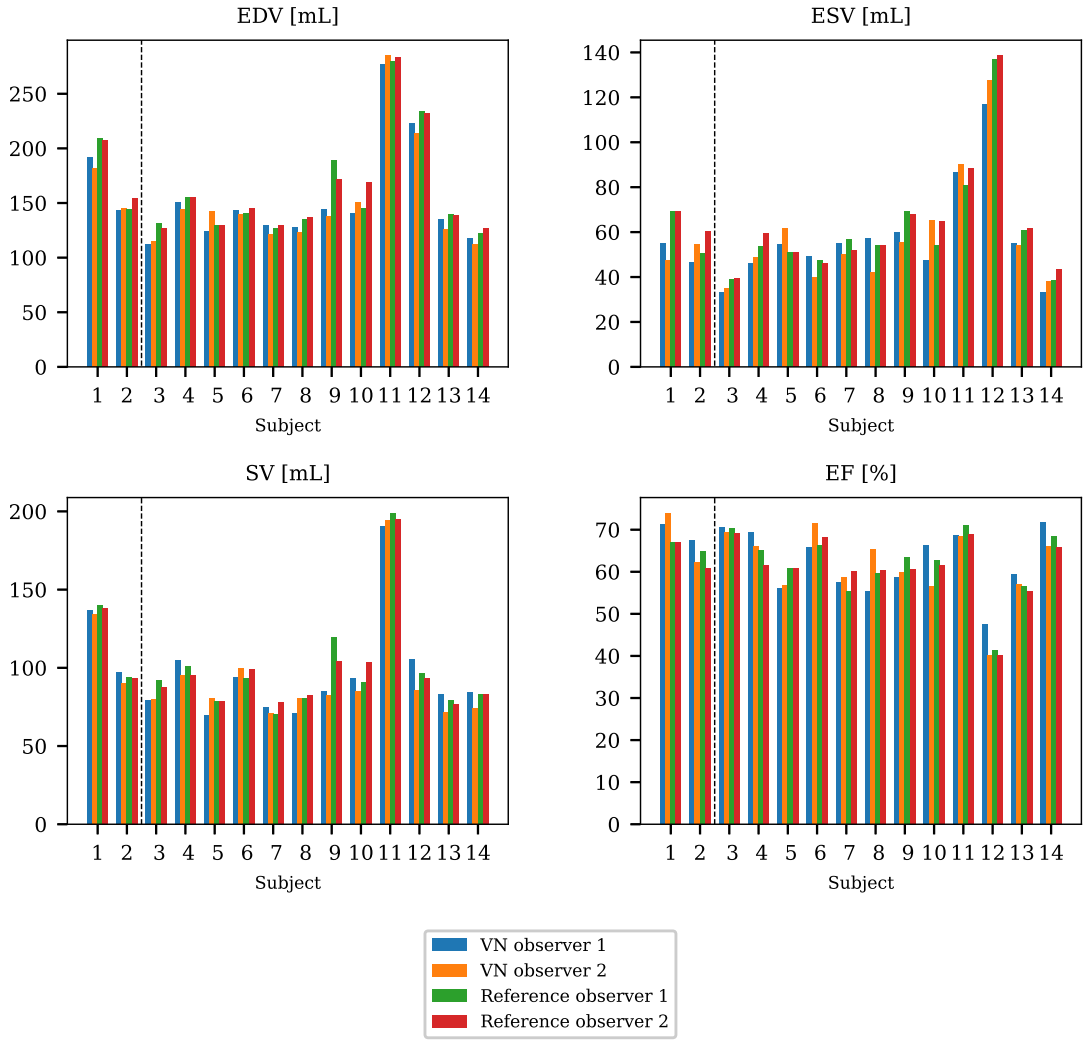


Figure 13: Functional cardiac parameters resulting from the analysis by two observers on real-time reconstructions by VN and the corresponding clinical Cartesian bSSFP cine images. The dashed line separates healthy volunteers (2) from clinical patients (12).

However, the limits of agreement for the inter-rater variation is, at least when comparing the observers on the VN images, similarly high. On the reference images, the raters agreement is better, yet the limits of agreement are still around 10% of the mean of the reference value. The average of all subjects is similar between the raters for all parameters.

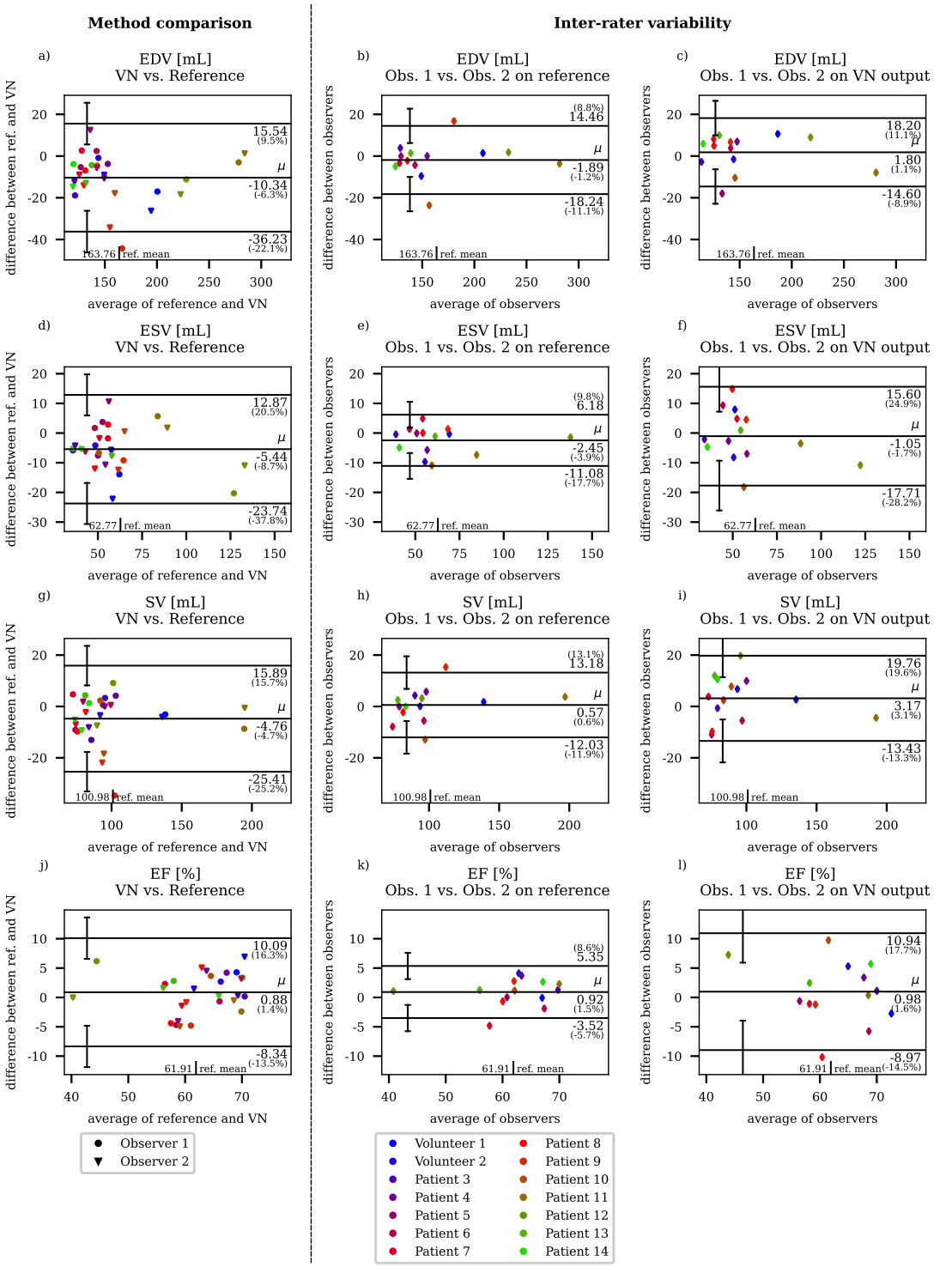


Figure 14: Bland-Altman plots of the functional cardiac parameters, comparing the real-time acquisitions with VN reconstruction to the clinical Cartesian bSSFP cine (first column) and the corresponding inter-rater variability between the two raters. The inter-rater variability is computed separately on the VN and reference images (second and third column). For providing context to the absolute values, each Bland-Altman plot shows the mean value of the reference method, and gives all absolute numbers in relation to this value (%-values in brackets).

4.5 Discussion

In this chapter, we presented a variational network model for reconstruction of undersampled spiral real-time cardiac frames. Subsequently, we performed extensive evaluation and investigated its suitability for functional imaging of the heart.

For benchmarking the performance of the presented VN method in reconstructing undersampled MR data, we chose to compare it to a previously-published LRS method [21, 67] and a general U-Net approach, which has been used in numerous publications [17, 18, 51] before. Together with the VN method, we can interpret these as part of a spectrum: On one end, the LRS model is purely based on a specific modeling of the MR reconstruction process, where insights into the structure of the acquired signal (i.e. sparsity) are used for reducing artifacts. The U-Net model on the other end is a general model for image processing, which was originally introduced for segmentation [71]. It does not use any additional information about the acquisition process or k-space data, only the images themselves. The MRI-specific adaptation that enables artifact reduction is completely learned from training images. The VN model can be interpreted as an attempt to incorporate elements from both: Its structure is modeled after the same gradient descent iteration as in compressed sensing, but the regularization is learned from image data. This is motivated by the aim, to achieve good reconstruction performance due to superior modeling capabilities of the U-Nets and the gradient descent structure, while also enforcing data consistency.

In terms of reconstruction performance and image quality, we observed that the U-Net is the fastest, but also delivers by far the worst image quality, on almost all measures. Its output is often blurry, and artifact reduction is moderate at best. Furthermore, it is very specific to the anatomy and image properties it was trained on. When applying it to long-axis images, which were not represented in the training set, its performance degraded even further, with hallucinations clearly visible.

The LRS model on the other hand is completely agnostic to anatomy. Its image quality was generally high, with only minor to no residual artifacts, and sharp contrast. However, it suffered from increased noise. Note, that the parameters of the algorithm have to be set by hand, which is a time-consuming task for most compressed sensing methods. Also it is essential to mention that the LRS model used here additionally applies a model in the temporal dimension. This allows it to exploit sparsity in this domain, which gives it a major advantage over the other models that do not incorporate any temporal information.

The Cartesian cine references clearly achieve the highest scores of all images and in all categories. However, it is unclear if these superior scores are due to these images being acquired with a segmented, and therefore fully sampled method, or simply a difference in contrast, since most of the reference images had bSSFP contrast. Also, another difference

was that a Cartesian trajectory was used.

Finally, the VN model achieved good image quality. According to quantitative measures (SSIM and RMSE) and expert ratings, the obtained image quality was comparable to the LRS method, even though no temporal model was applied in the VN. Remarkably, in its application to long-axis images, which were not represented in the training set at all, the image quality and artifact reduction appeared to work equally well. We attribute this superior generalization to the data consistency mechanisms inherent in the VN model.

As with any supervised deep learning method, one possible way for improving performance is through improving the training dataset. In our case, the training dataset is relatively small with data from only 11 subjects. This is due to the fact that our training data is expensive, as it was acquired in vivo specifically for this purpose. Thus, performance gains could be expected from increasing the amount of training data. An efficient way to do so, would be to use synthetic data. Though a lot of care has to be dedicated towards making the simulated data realistic, the amount of data that could potentially be obtained in this way is essentially unlimited, as large archives of clinical cardiac measurements exist in clinical environments.

We chose SSIM as loss function in accordance with many prominent contributions proposed recently [17, 18]. Because mean-squared error represents another popular choice, we retrained our VN with mean-squared error and compared the resulting image quality. In summary, only minor differences were found for our application.

In our study, we used a spiral sequence based on spoiled gradient echo, because it was available in out institution from previous development [21]. However, with respect to functional CMR, balanced SSFP sequences have several advantages in comparison to our spiral spoiled gradient-echo prototype sequence, like higher signal-to-noise ratio (SNR) and superior contrast between myocardium and blood [72]. Accordingly, the difference of the contrast between clinical routine acquisition (bSSFP) and the accelerated image series (spoiled gradient echo) represents a considerable limitation of the comparison shown in Figure 9. Furthermore, in the real-time images, e.g. Fig. 9 and prominently also 11, dark patches can be observed in the cardiac chambers. This might have been caused by turbulent blood flow, since spoiled gradient-echo sequences are generally sensitive to inflow effects. The reference acquisitions based on bSSFP did not suffer from signal voids in a comparable manner. Furthermore, since the flow is aperiodic, it is averaged out to some degree for segmented cine, but not for the real-time acquisitions. While the focus in this work has clearly been on the reconstruction method, implementing a spiral pulse sequence based on bSSFP represents the next logical step toward clinical application.

In comparison with other techniques that exploit extreme dimension reductions by reconstructing a maximum of domains at once [73, 74] our approach is comparably limited with respect to possible undersampling rates, as we chose to process each 2D real-time

image of the dynamic series individually. Conversely, however, the potential of overfitting in these domains is also prevented, which excludes the notorious risk of temporal blurring. We feel that the 2D approach based on an efficient k-space trajectory represents a sufficient and at the same time highly robust strategy for real-time cardiac MRI. Nevertheless, the VN method could be extended by using a more sophisticated regularizer, such as the recently-introduced total deep variation [19, 75].

It has been a recent subject of discussion [76–78] whether the term "real-time" MRI should continue to be used for methods which produce dynamic images where each time frame depicts a succession of events in the same order that they occurred during the data acquisition. Alternatively, it is proposed to reserve this term for methods that additionally provide these images within a short time, so-called latency, after the data acquisition. Various latencies are cited as potential requirements for this term, but all are in the range of up to a few seconds. We share the opinion of Nayak [77], that the current use of the term "real-time" as only constituting requirements on the data acquisition is intuitive and expressive, and therefore have used the term throughout this work in this way. Additional low latency should be expressed by the addendum of the adjective "interactive". However, we want to stress that latency is an issue of supreme importance, as has already been remarked [76]. It may well be the pivotal requirement hindering novel imaging techniques from use in clinical routine.

One major improvement of the proposed method with respect to many previous publications of fast cardiac MRI was to decrease latency. On our hardware described above, we achieved reconstruction times of 0.9 s per frame, 14.39 min for an entire series (including the computation of sensitivity maps), with our current implementation. Thus, it can not and should not be considered "interactive real-time MRI" by any standard. However, we want to point out that even if the method is not fully interactive, decreasing latency from hours to minutes is a key development, in particular for the determination of functional parameters, where a latency of this length appears acceptable in clinical practice. The model-based LRS approach, demanded considerably longer reconstruction times. It is noteworthy that LRS was implemented and ran on a CPU only. Nevertheless, the model structure including a singular value decomposition is - in general - more difficult to parallelize, and the iterations needed for optimization (70) exceed the number of cascades of the VN (10) by far. Thus, in our opinion, the VN represents the best trade-off between image quality and run-times for the acceleration techniques tested. Although the reconstruction time for an entire exam is still too long (14.39 min) for applying VN in clinical routine, moderately upscaling the hardware can shorten runtimes toward acceptable delays for most applications. Furthermore, we want to point out the fact that our VN method reconstructs each frame individually, which provides high flexibility with respect to how entire series are processed. Therefore, overall reconstruction times could be extremely shortened by processing individual images in parallel (e.g. on multiple

GPUs).

In the presented Bland-Altman analysis on the agreement of cardiac functional parameters, we found only small fluctuations on average, but the associated limits of agreement showed rather large deviation between the methods. Several factors may contribute to this. First, the clinical protocol for the segmented reference measurements was rather inhomogeneous, with in particular (due to different heart rates of the subjects) the temporal resolution varying between 26 and 56 ms per frame. Also, not all but most of the clinical references were acquired with a bSSFP sequence, so a difference in contrast may also have an impact. Furthermore, the inter-rater variability also contributes to the difference between the methods. The magnitude of the inter-rater variability according to the limits of agreement in the Bland-Altman plots on the VN images is even approximately as large as the deviations between the methods. Interestingly, the inter-rater variability on the reference images was only approximately half as high. This may be explained by the difference in contrast, as the bSSFP images generally have better myocardium-blood contrast, making the segmentations more consistent. Also, measurements with bSSFP contrast are the clinical standard, to which the raters are more accustomed. Still, it is very likely that variability in the observers contributed to the large limits of agreement between the methods. Furthermore, for each of the limits of agreement, the associated 95% confidence intervals were rather large, which limits the conclusions that can be drawn from the limits of agreement. Note that in small sample sizes, the common choice of 1.96σ we followed here overestimates the limits of agreement [79]. In general, the analysis presented here is of qualitative nature, as no reliable quantitative assertions can be made from such a small sample size. We conclude that the presented data does not in any way suggest the methods to disagree, but is insufficient to show agreement with statistical significance. This would require a larger sample size and would be substantially supported by a more consistent environment.

Recent work, mainly in the field of machine learning based post processing of medical images, makes an entirely automatic quantification of cardiac function conceivable. Various publications [7, 8] already demonstrated semantic segmentation of the heart with human level accuracy. The latter in conjunction with the presented acquisition and reconstruction method would reduce the workload for a functional MR exam of the heart to just one minute of scan-time in free breathing without the need for any further manual intervention. Undoubtedly, this would represent a substantial advancement with respect to the current state-of-the-art. Similar techniques could be applied for other MRI applications, paving the way to improved accessibility, increased patient comfort, reduced costs and even to entirely new kinds of investigations.

4.6 Conclusion

We presented a variational network architecture that is capable of reconstructing cardiac real-time data acquired with an undersampled spiral trajectory in free breathing. The method was extensively evaluated and compared to other state-of-the-art methods. We found that the VN is able to reconstruct images with high quality, equivalent to current model-based methods. A comparison to the current clinical standard for cardiac functional imaging remained inconclusive. In comparison to a purely image based U-Net trained on the same task, we found that the data consistency mechanisms inherent in the VN architecture constitute a considerable advantage, leading to better image quality, higher robustness and better generalization properties.

Supplementary Material Information

As supplementary material, there are two videos available, which have been referenced in the text above. These can be accessed online on the website of the original publication. A short description and the download links are given in the following:

- Video V1: Depiction of cardiac dynamics by the VN reconstruction in an exemplary patient for nine slices simultaneously.
  Link: *https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fmrm.29357&file=mrm29357-sup-0002-VideoS1.mp4*

- Video V2: A comparison of the following dynamic acquisitions: a fully sampled Cartesian reference cine (25 heart phases) acquired in a healthy participant in breath-hold a); spiral real-time frames acquired in the same participant in free breathing, reconstructed with a U-Net, low-rank plus sparse, and the VN model, respectively b), e), f); and a spiral acquisition in the same participant in breath-hold, binned into 20 segmented frames d) and reconstructed by VN as real-time frames c).
  Link: *https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fmrm.29357&file=mrm29357-sup-0003-VideoS2.mp4*

Funding Information

This work was supported by the German Ministry for Education and Research under Research Grant 05M20WKA.

5 Assessment of resolution and noise in magnetic resonance images reconstructed by data driven approaches

The contents of this chapter are adapted from the paper [80]

"Assessment of resolution and noise in magnetic resonance images reconstructed by data driven approaches"

by

Jonas Kleineisel[1,2], Katja Lauer[1,3], Alfio Borzì[3], Thorsten A. Bley[1], Herbert Köstler[1], Tobias Wech[1,2]

who are affiliated with the following institutions:

[1] Department of Diagnostic and Interventional Radiology, University Hospital Würzburg

[2] Comprehensive Heart Failure Center, University Hospital Würzburg

[3] Chair of Mathematics IX: Scientific Computing, Institute of Mathematics

at the University of Würzburg in Würzburg, Germany.

The original article is currently in press at *Zeitschrift für Medizinische Physik* and is available online (*https://doi.org/10.1016/j.zemedi.2023.08.007*) under the Creative Commons BY 4.0 license (*https://creativecommons.org/licenses/by/4.0/*). The following chapter has been slightly modified from the original text, though the majority of the content was reproduced without modifications. This or similar adaption of this work is permitted under the Creative Commons BY license.

Previous results have been published as a conference abstract [81].

5.1 Abstract

Introduction: Deep learning-based approaches are increasingly being used for the reconstruction of accelerated MRI scans. However, presented analyses are frequently lacking in-detail evaluation of basal measures like resolution or signal-to-noise ratio. To help closing this gap, spatially resolved maps of image resolution and noise enhancement (g-factor) are determined and assessed for typical model- and data-driven MR reconstruction methods in this paper.

Methods: MR data from a routine brain scan of a patient were undersampled in retrospect at R = 4 and reconstructed using two data-driven (variational network (VN), U-Net) and two model based reconstructions methods (GRAPPA, TV-constrained compressed sensing). Local resolution was estimated by the width of the main-lobe of a local point-spread function, which was determined for every single pixel by reconstructing images with an additional small perturbation. g-factor maps were determined using a multiple replica method.

Results: GRAPPA showed good spatial resolution, but increased g-factors (1.43 - 1.84, 75% quartile) over all other methods. The images delivered from compressed sensing suffered most from low local resolution, in particular in homogeneous areas of the image. VN and U-Net show similar resolution with mostly moderate local blurring, slightly better for U-Net. For all methods except GRAPPA the resolution as well as the g-factors depend on the anatomy and the direction of undersampling.

Conclusions: Objective image quality parameters, local resolution and g-factors have been determined. The examined data driven methods show less local blurring than compressed sensing. The noise enhancement for reconstructions using CS, VN and U-Net is elevated at anatomical contours but is drastically reduced with respect to GRAPPA. Overall, the applied framework provides the possibility for more detailed analysis of novel reconstruction approaches incorporating non-linear and non-stationary transformations.

5.2 Purpose

Together with recent advances in machine learning and neural network-based computer vision, data-driven methods have seen increasing interest in the MRI academic literature. Several distinct tasks are suited for the application of machine learning, including reconstruction of undersampled data [2, 14]. The data-driven modelling capabilities of neural networks have shown enormous potential for improving and accelerating reconstruction tasks, and are capable of coping with undersampled acquisitions.

However, unlike in classical MRI, the transformation properties and quality of advanced reconstruction methods, data-driven as well as model-driven, depend on local structure and are not uniform across the imaged field of view. Therefore signal abnor-

malities, e.g. pathological structures, should be judged in view of the local quality and reliability of the imaging method. To this end, local and objective measures of the quality of reconstructions obtained from data-driven methods are needed. These results could moreover prove valuable in the development of new reconstruction methods: Knowing where and how the quality of an image is compromised is certainly crucial for precisely analyzing and systematically improving a reconstruction technique.

While "image quality" actually describes a number of several aspects, the performance of new approaches is frequently assessed by determining only a single global score like RMSE or SSIM. This undoubtedly leads to straightforward assessments, however, it also over-simplifies the evaluation in many cases. Especially image resolution is not considered precisely by these scores. Since classical MRI constitutes a linear and shift-invariant (also called stationary) transform in good approximation, the resolution is uniform and can be analyzed by determining a single point-spread function [82]. Model- and data-driven reconstruction methods, however, in most cases introduce non-linear and non-stationary aspects.

To overcome this problem, resolution has to be determined locally, at every spatial location. Suitable techniques have been developed many years ago. Comprehensive theoretical considerations have first been presented by Fessler and Rogers [83]. These were geared towards applications in emission and transmission tomography, where more research was done subsequently [84]. Later, Wech et al. [85] applied these concepts to cardiac MRI, by determining resolution maps for images reconstructed with iterative thresholding compressed sensing algorithms. However, the recent rise of deep learning-based, highly non-linear image reconstruction methods makes this topic more current than ever and leaves their assessment with these tools to be desired. Chan et al. [86] lately assessed resolution for more recent reconstruction methods. They showed perturbation responses in single pixels and checkerboard tests for the whole image. This provided an improved assessment with respect to mere visual inspection, however, the proposed maps are not easily interpretable as local resolution was not quantified. Additionally, the only deep learning method they evaluated was a direct U-Net reconstruction, while unrolled gradient schemes like the variational network architecture proved to provide superior performance in the meantime.

Besides resolution, the signal-to-noise ratio (SNR) of an MR image is a decisive quality measure. For determining the noise attenuation and amplification properties of reconstruction methods, g-factor maps are well suited. These can be computed by using a multiple-replica method [87]. By combining the evaluation of resolution with that of noise behavior, we aim to obtain a more comprehensive picture of the properties of current MRI reconstruction methods than previous studies.

By investigating resolution and noise of model- and data-driven reconstruction methods, this work has the following aims: First, it serves to better understand methods cur-

rently being developed in academic research, thereby closing the gap toward widespread application. Furthermore, it demonstrates how quality of advanced reconstruction methods can be evaluated systematically and how these evaluations can be interpreted. Ultimately, this could increase diagnostic quality by aiding the analysis of MR images.

5.3 Methods

5.3.1 Measuring resolution using local point-spread functions

Local point-spread functions for general image processing

In the following let us understand a greyscale image with n pixels simply as a vector in $\mathbb{R}^n$. Then any kind of imaging method can be modeled as a function $T : \mathbb{R}^n \to \mathbb{R}^n$, wich maps some real-world quantity at n spatial locations onto a representation by an image. Note that the complex case is also covered by choosing n and m appropriately and using the isomorphy $\mathbb{R}^2 \simeq \mathbb{C}$. In the case of MRI, T encompasses the multi-coil sampling procedure together with all subsequent reconstruction operations.

The reconstructed image representes a discretization of proton density or another quantity to which the specific sequence is sensitive. If $a \in \{1, ..., n\}$ denotes an index corresponding to a spatial location, and $e_a = (\delta_{ka})_{k=1,...,n}$ the standard basis vector at that location, then be_a represents a signal which consists only of a single point with intensity b as location a.

For common MRI sampling and reconstruction methods, it is reasonable to assume that T is continuously differentiable. Then, by definition of the differential, there exists a gradient $\nabla T : \mathbb{R}^n \to C^0(\mathbb{R}^n, \mathbb{R}^n)$ and a residual $r : \mathbb{R}^n \to \mathbb{R}^n$, such that

$$T(O + be_a) = T(O) + b\nabla T(O)e_a + r(b) \tag{16}$$

where it holds $\lim_{b\to 0} \frac{r(b)}{b} \to 0$. That is, the effect that a small change be_a in the input quantity has on the reconstruction $T(O)$ can be completely described by the partial derivative $\nabla T(O)e_a$. Thus, from the definition it follows

$$\lim_{b\to 0} \frac{(T(O + be_a) - T(O))}{b} = \nabla T(O)e_a. \tag{17}$$

Therefore, we define the *local point-spread function* (LPSF) [83, 85] as

$$LPSF(T, O, a) := \frac{(T(O + be_a) - T(O))}{b} \tag{18}$$

for some small $b > 0$. If $b > 0$ is small enough we have

$$LPSF(T, O, a) \approx \nabla T(O)e_a. \tag{19}$$

The LPSF can be thought of as the change of the reconstruction $T(O)$ at location a, given a small change in the reconstructed object O at location a. If T is linear, one

can immediately see from (18) that $LPSF(T, O, a) = T(e_a)$, i.e. the LPSF does not depend on the baseline perturbed image. If the reconstruction method T is additionally shift-invariant, it further does not depend on a.

The key idea for determining resolution from the shape of an LPSF is that details in the input can be distinguished, if they are represented by distinguishable features in the reconstruction [83, 85]. According to the Rayleigh criterion [85, 88], an average human observer perceives two points in an image as belonging to different objects, if the signal intensity on the line between these locations drops to below 81% of the maximum of the brightness of the two points.

Thus, whether two points from the input can be distinguished in the output, depends crucially on how distinct the points are represented in the output: If the two points are both spread out too much by the reconstruction method, they may become indistiguishable. Since the LPSF characterizes the output representation of a point perturbation in an input object, the width of the main lobe of an LPSF can provide a measure of resolution [83, 85].

In the case of classical MRI, i.e. fully sampled single-coil data, the reconstruction operator T is a Fourier transform. In this case, due to the linearity and shift-invariance, the LPSF is sinc-shaped everywhere. One finds that for two identical sinc-shaped LPSF, the Rayleigh-criterion is met if the distance between their maxima is no less than the width of the main lobe at $\frac{2}{\pi} \approx 64\%$ of the maximal height. Motivated by this observation, for an arbitrary reconstruction method, we take the width of the main lobe at 64% of the maximum height as a measure of resolution. Any value larger than 1 corresponds to a situation where point inputs are spread out, such that they may not be distinguished from neighbouring points, which corresponds to a loss of resolution.

Application to undersampled MRI

Now, assume that we have some fully sampled MRI k-space data, which represents some real-world object O, together with a reconstruction method denoted by T. Then we apply the concepts described in the previous section as follows:

First, we transform O coil-wise to image space. For a spatial location a, the image is perturbed by adding a small perturbation to the pixel in a. Empirically, we found that $T(O + be_a) - T(O) \propto b$ holds true approximately if we choose the amplitude of the perturbation no larger than 0.1% of the maximum of the root-sum-of-squares reconstruction of O for all the tested reconstruction methods and for both of the images that we applied them to. The amplitude of the perturbation in each coil is chosen proportional to the signal level of the coil in a. Furthermore, we found that for the examples studied in this paper, the complex phase of the perturbation had no substantial effect, and thus we always keep the phase of perturbed pixels constant. For the perturbed coil-wise images we then transformed to k-space and applied undersampling. From these "raw

data", we then computed a perturbed reconstruction. An unperturbed reconstruction was obtained simply by applying the model reconstruction to the undersampled original data. As detailed above, the difference between the perturbed and unperturbed reconstruction then yields the LPSF. The entire process is illustrated in Figure 15. From the two-dimensional LPSF, the resolution in the two spatial dimensions was then determined from the shape of the LPSF. For this, we extracted the row (respectively, the column), and applied 5-fold Fourier interpolation before measuring the width of the main lobe at 64% of its maximum. This procedure yielded two resolution maps, for the resolution in horizontal and vertical direction.

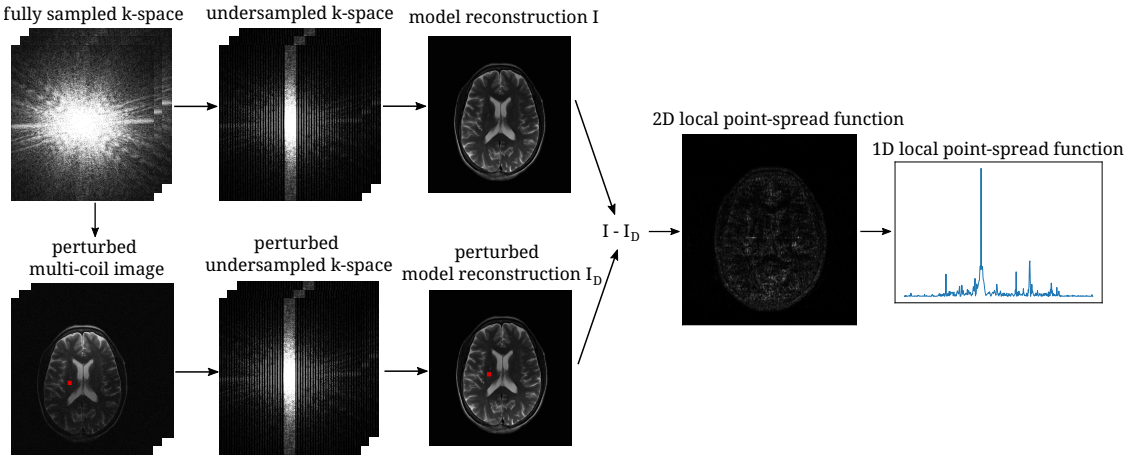


Figure 15: Illustration of the approach for determining local point-spread functions. Fully sampled k-space data is retrospectively undersampled with and without previous perturbation in a single pixel (marked red). By subsequently applying some reconstruction method on the two k-space data, we obtain two model reconstructions. The difference of these represents the 2D local point-spread function (LPSF) for the location of the perturbed pixel. A horizontal or vertical profile is then used for measuring the width of the main lobe.

5.3.2 Determining g-factor maps with a pseudo multiple replica method

Using the notation from the previous section, our goal in the following is to establish a method for measuring the noise properties of a MRI reconstruction method.

Previously, we denoted the entire imaging method by an operator $T : \mathbb{R}^n \to \mathbb{R}^n$, going from real-world quantities to an image representation, both on a discrete grid. By distinguishing between the acquisition and the reconstruction, the imaging operator can be split into two parts:

$$T = E \circ S \tag{20}$$

Here, $S : \mathbb{R}^n \to \mathbb{C}^{c \times d}$ denotes the sampling procedure, which maps from a real-world quantity of interest to complex multi-coil MR k-space data. The dimension of size c

denotes the coils. The reconstruction operator $E : \mathbb{C}^{c\times d} \to \mathbb{R}^n$ takes the data and reconstructs a greyscale image.

The sampling procedure is inevitably corrupted by additive noise, such that for a given object O, the reconstruction can be expressed by

$$T(O) = E(S(O) + N) \tag{21}$$

where $N \in \mathbb{R}^{c\times d}$ contains complex normally distributed noise. Thus, the measured k-space data that is to be reconstructed has the form $k = S(O) + N$. But neither the ground-truth data $S(O)$ or the noise N can be extracted from the data, in order to investigate how the reconstruction operator E responds to the noise. To get around this, we can add additional synthetic noise $\tilde{N}$ to the data, and infer from the response $E(S(O) + N + \tilde{N})$ how E reacts to noise in the input data. This is called the *pseudo multiple replica method* and has been described previously [89].

However, due to physical coupling of the receiver coils between each other, the noise N is correlated between the coils. To accurately simulate noise $\tilde{N}$, this correlation has to be taken into account. Since it can not be accurately measured from the k-space data k, we use a separate acquisition without an RF-pulse, where consequently only noise is acquired. If we denote the measured noise data by $M \in \mathbb{C}^{c\times d}$, the covariance matrix is given by

$$\Psi := \frac{1}{n} M\, M^* = \left(\frac{1}{n} \sum_{k=1}^{n} M_{i,k} \overline{M_{j,k}} \right)_{i,j=1,\dots,n} \in \mathbb{C}^{c\times c}, \tag{22}$$

where we used the fact that the expected value for the noise is 0. The matrix entry $\Psi_{i,j}$ gives the covariance between the i^{th} and j^{th} coil. Once these correlations are known, synthetic noise with the same correlation as the measured noise can be generated via

$$\tilde{N} = \sqrt{\Psi}\, N = \left(\sum_{j=1}^{n} \Psi_{i,j}^{\frac{1}{2}} N_{j,k} \right)_{\substack{i=1,\dots,c \\ j=1,\dots,m}} \tag{23}$$

where $N \in \mathbb{C}^{c\times m}, N \sim \mathcal{N}(0,1)$ and $\sqrt{\Psi}$ is the positive root of Ψ. It can be obtained from Ψ as follows: From the definition, it is immediately obvious that Ψ is Hermitian and positive definite, and therefore diagonizable. Thus, with a matrix V of orthonormal eigenvectors, Ψ can be written as

$$\Psi = V\, \mathrm{diag}(\lambda_1, ..., \lambda_c)\, V^{-1} \tag{24}$$

where $\lambda_1, ..., \lambda_c \geq 0$ are the non-negative eigenvalues of Ψ. Then the positive root $\sqrt{\Psi}$ can be obtained via

$$\sqrt{\Psi} = V \mathrm{diag}(\sqrt{\lambda_1}, ..., \sqrt{\lambda_c}) V^{-1}. \tag{25}$$

With the method as described above we can generate r noise samples $\tilde{N}_1, ..., \tilde{N}_r$, which can be added to the measured k-space data k to obtain pseudo-repetitions $k + \tilde{N}_1, ..., k + \tilde{N}_r$ of the measurement. By applying undersampling, and subsequently the reconstruction operator E, we thus get r different accelerated reconstructions. Furthermore, since the pseudo-repetitions are fully sampled, we can also reconstruct them classically, i.e. by Fourier transform and root-sum-of-squares coil combination. This yields r non-accelerated reconstructions. By computing the standard deviations σ_{acc} and σ_{normal} of the accelerated and non-accelerated reconstructions, the g-factor map is obtained as

$$g := \frac{\sigma_{acc}}{\sigma_{normal}\sqrt{R}} \tag{26}$$

where R denotes the acceleration factor in the undersampling. It gives a spatially resolved ratio of how much the reconstruction method for accelerated data increases noise over a non-accelerated acquisition. Since a reduction in acquisition time by a factor of R inevitably results in an increase in noise by a factor of $\sqrt{R}$, this effect is compensated for by dividing by $\sqrt{R}$ in equation (26). If $g(a) \geq 1$ at some spatial location a, then noise is increased by more than what is expected from the shortened acquisition, i.e. the method amplifies noise. Conversely, if $g(a) \leq 1$ then noise is suppressed at location a.

5.3.3 Reconstruction methods

For gaining insights into their reconstruction properties, we chose a selection of popular model-based and data-driven methods:

GRAPPA: As classical parallel imaging baseline, we used a GRAPPA [90] implementation [91].

Compressed sensing: A TV-regularized compressed sensing (CS) model, which is implemented in the Berkley Advanced Reconstruction Toolbox (BART) [92]. The optimization problem is solved with the ADMM optimizer [93]. The parameters (number of iterations and regularization weight) were individually adjusted for the T_1- and T_2-weighted image.

Variational network: For data-driven reconstructions, we used the publicly available challenge dataset and the baseline models of the 2020 fastMRI challenge [18]. The pre-trained models were provided by fastMRI. They are trained on the fastMRI 2020 multicoil brain challenge dataset, consisting of 4469 T_1, T_1 post-contrast, T_2, and FLAIR acquisitions with additional 1378 for validation. The dataset and the models are described in detail in the accompanying paper [18] and a previous publication [66]. The variational network (VN) model consists of 12 cascades, and the regularization terms are realized by U-Nets with 4 up- and downsampling steps and 18 channels in the first layer. The entire VN model has $2.99 \cdot 10^7$ trainable parameters.

U-Net: Like the VN model, the U-Net model was one of the baseline models of the 2020

fastMRI challenge [18], and was trained on the same dataset. It has the same structure as the regularizers in the VN model, but uses 256 channels in the first layer. It has a total of $4.96 \cdot 10^8$ trainable parameters.

All methods were applied to data, which were retrospectively undersampled with an acceleration factor of $R = 4$. Undersampling was applied in phase-encoding direction, which corresponds to horizontal direction in image space. The undersampling masks, which differ between the methods due to their specific requirements, can be seen in Figure 16. For all methods, a fully sampled region in the center of k-space with size 8% of phase encoding lines, called the autocalibration signal (ACS) region, was used. For GRAPPA, the calibration was performed on this region and the reconstruction then applied to k-space where every fourth line was sampled, without ACS region, to ensure $R = 4$. For CS, a random undersampling pattern with ACS region was used. For both VN and U-Net, an undersampling pattern with ACS region was used, where the width of the gaps between the sampled lines alternates between 4 and 5 lines. This matches the undersampling in the training data. Due to a different matrix size, the random and equispaced masks for the T_1-weighted scans differ, but follow the same principles.

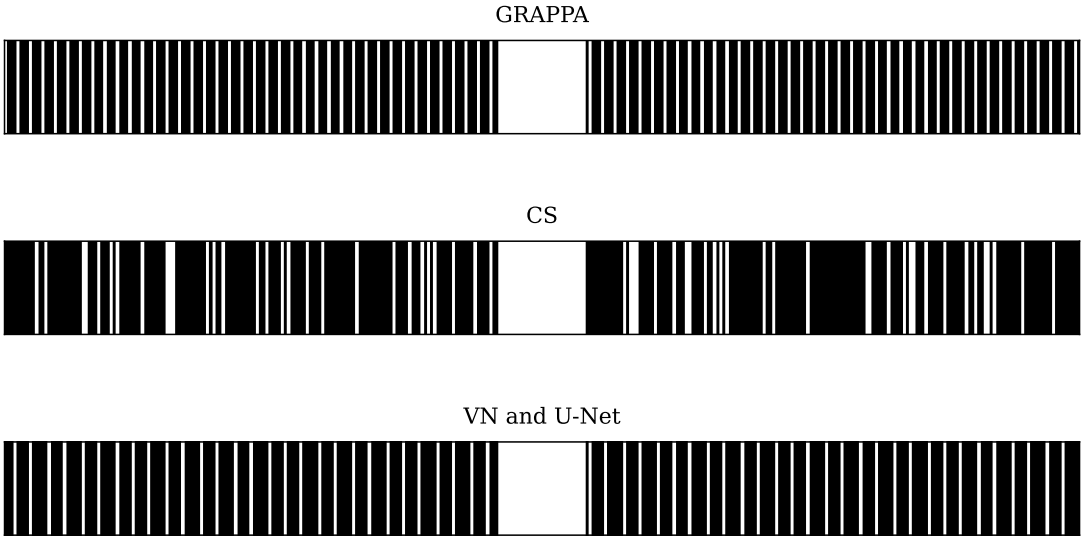


Figure 16: Undersampling patterns with acceleration $R = 4$ used for masking the T_1-weighted scans. Only a section of the vertical size is shown. Due to a different matrix size, the random and equispaced masks for the T_2-weighted scans differ, but follow the same characteristics. GRAPPA uses a regular undersampling pattern where every fourth line is sampled and uses the fully sampled center region for kernel calibration (ACS region), but not for reconstruction, such that $R = 4$ is achieved. The other methods include the ACS region. CS uses a random undersampling pattern, while VN and U-Net alternate between gaps of 4 and 5 missing lines between each sampled lines, to ensure $R = 4$ and consistency with the masks in the training data.

5.3.4 Acquisitions

Raw data of a fully sampled MR-investigation of the brain on a clinical 3T scanner (MAGNETOM Prisma fit, Siemens Healthcare GmbH, Erlangen, Germany) was exported as a representative dataset. The subject was a 57-year old woman. Images were acquired to rule out cerebral manifestation of a B-Non-Hodgkin Lymphoma. Apart from a few unspecific white matter lesions no intracranial pathology was found. Transversal T_1-weighted images were acquired using a spoiled gradient echo sequence (TE = 4.28 ms, TR = 321 ms, flip angle 90°, FOV = 235 × 235 mm^2, in-plane spatial resolution = 0.61 × 0.68 mm^2, matrix size 384 × 346) and T_2-weighted images were acquired using a TSE-sequence (TE = 91 ms, TR = 6610 ms, flip angle 150°, FOV = 235 × 235 mm^2, in-plane spatial resolution = 0.46 × 0.51 mm^2, matrix size 512 × 461), well within the distribution of the NYU training dataset. A head coil with 20 channels was used. As described in section 5.3.2, we additionally acquired a noise measurement for computing g-factors.

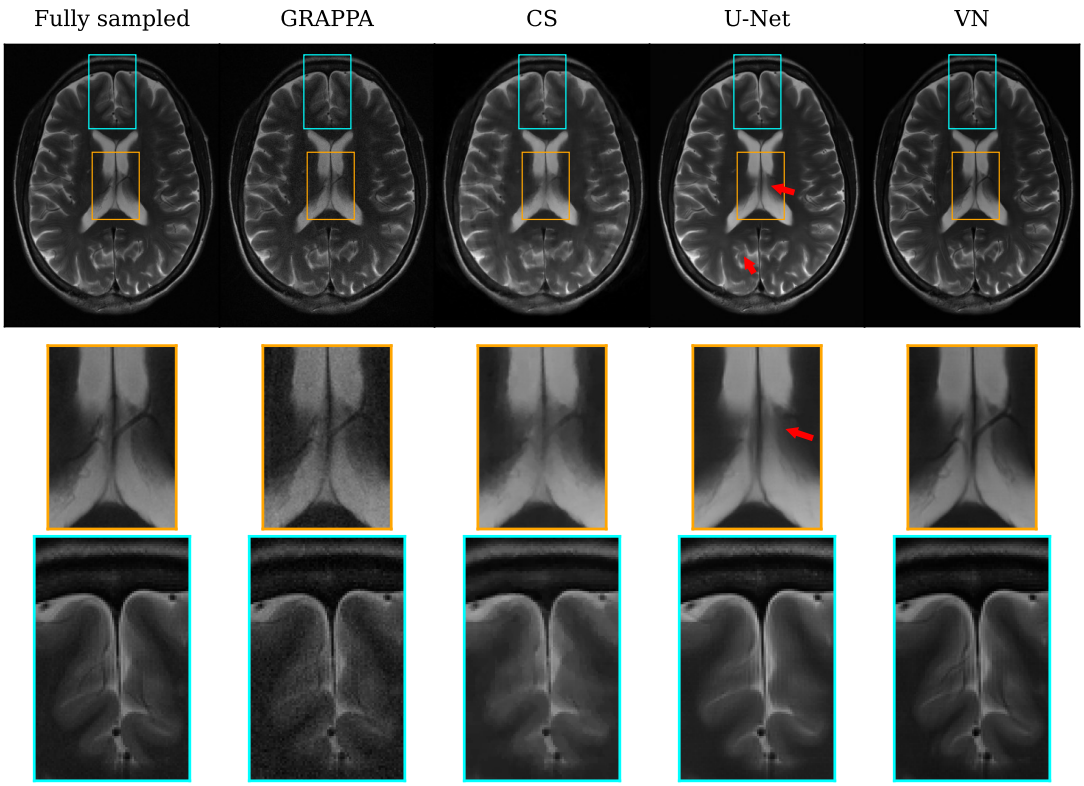


Figure 17: Reconstructions of a central slice of the T_2-weighted acquisition by the tested methods. The red arrows indicate substantial reconstruction errors in the U-Net output where anatomy is misrepresented due to hallucinations. The yellow and blue boxes show anatomical structure in detail in the lower two rows. One can see substantial noise in the GRAPPA reconstructions, while the other model reconstructions appear smoothed and alter the appearance of small details in some cases.

Our study was approved by the institutional review board of the University of Würzburg and the requirement of written informed consent by the study participant was waived due to the retrospective study design (decision number 20220805 01 following the request on 05.08.2022). The data was fully anonymized for data analysis.

5.4 Results

5.4.1 Qualitative comparison of reconstructions

Figures 17 and 18 show reconstructions of a central slice of the T_2- and T_1-weighted acquisitions by the reconstruction methods as detailed above. In general, all methods seem to correct for undersampling artifacts at $R = 4$ effectively. However, especially in the enlarged ROIs (orange and blue boxes), some important differences can be noticed: With respect to detail and sharpness, GRAPPA seems to preserve all information present in the reference, but adds some noise. CS exhibits some clearly noticeable spatial blurring, such that finer anatomical details are lost. The U-Net reconstruction appears to exhibit

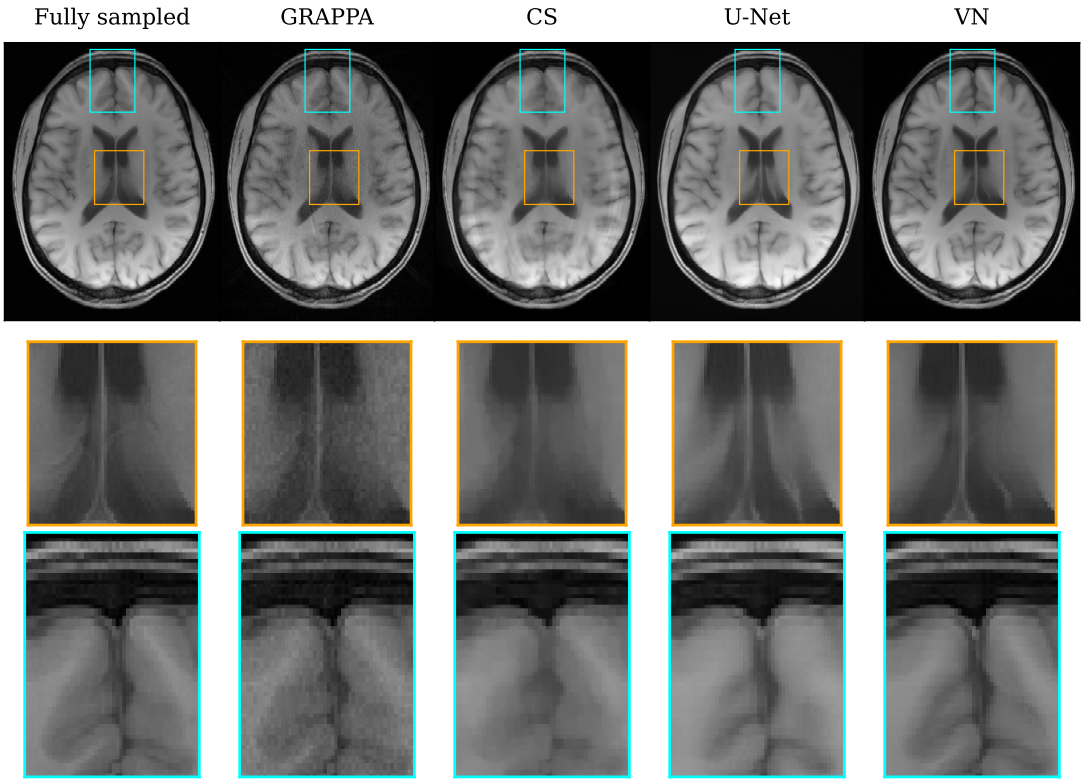


Figure 18: Reconstructions of a central slice of the T_1-weighted acquisition by the tested reconstruction methods. The yellow and blue boxes show anatomical structure in detail in the lower two rows. Like in the T_2-weighted images, one can see substantial noise in the GRAPPA reconstructions, while the other model reconstructions appear smoothed and alter the appearance of small details in some cases.

low apparent noise and only minimal blurring. But, critically, in the locations indicated by red arrows in the T_2-weighted images, anatomy is misrepresented by hallucinations which hide specific features. The VN on the other hand does not show similar systematic errors for this example, and exhibits an apparently high image quality similar to the reference.

5.4.2 Resolution maps

Representative LPSFs, one with a narrow and one with a wider main lobe can be seen in Figure 19. In the 2D LPSFs, one can see that in particular CS and U-Net also exhibit changes in pixels sometimes spatially quite far away from the location where the perturbation was set.

Figures 20 and 21 show the resolution maps in both spatial dimensions obtained from the proposed method for the T_1- and T_2-weighted images. Cumulative frequency analysis (CFA) of the resolution maps and g-factor maps is shown in Figure 22, where a mask was applied to only take pixels in the anatomy into account.

For the GRAPPA method, one can see from the maps and the CFA, that except for some outliers, the maps are constant at a value around 1, indicating no substantial loss in resolution with respect to nominal k_{max}, homogeneously across the image.

Compressed sensing clearly exhibits the largest broadenings of the LPSF's main lobes w, i.e. the lowest resolution. The resolution maps appear non-continuous, with even neighboring pixels being assigned clearly different resolutions. In pixels near edges, the width w of the LPSF's main lobe is lower than in homogeneous areas, where values between $w = 1.5 - 2.5$ are common, indicating blurring there. The behavior is similar in the resolution maps for both directions, with more blurring in the undersampled direction.

Similarly, U-Net and VN show blurring in the undersampled horizontal direction in most of the pixels. Their general level of blurring is less pronounced compared to CS and lowest for U-Net. A higher resolution close to 1 is preserved close to edges. In the fully sampled vertical direction, the behavior is reversed, as U-Net and VN show moderate blurring at sharp anatomical edges, while the width of the main lobe is reduced to values smaller than 1 for homogeneous regions in the brain and the ventricles. The maps and in particular the CFA present the resolution of U-Net and VN to be similar, while VN shows slightly higher overall blurring and U-Net exhibits more outliers.

As clearly visible from the resolution maps and CFA, resolution in all methods is noticeably lower in the horizontal direction, where undersampling was applied.

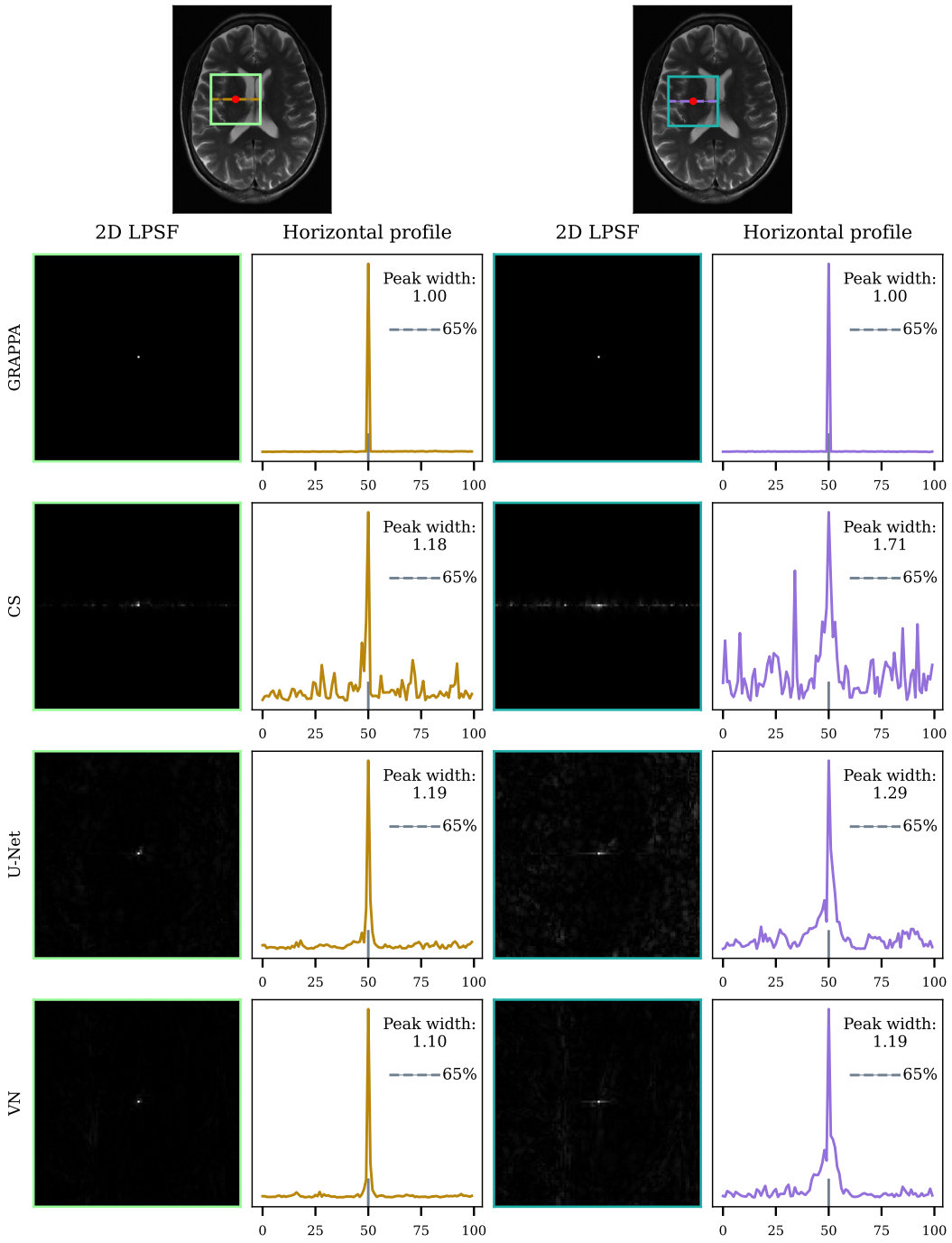


Figure 19: Representative local point-spread functions (LPSF) for two pixels of the T_2-weighted image for all tested reconstruction methods. The first row shows the respective locations of the pixels, and indicates a region of interest (ROI) to be investigated. Below, the 2D LPSFs in this ROI as well as their profile in horizontal direction are shown. The corresponding widths that were computed from these profiles at 65% of the peak height after interpolation are noted. The locations were selected to show an example of a pixel where the methods exhibit narrower (in the two left columns) and wider (in the two right columns) widths of the LPSF's main lobe, corresponding to lower and higher level of local blurring. Especially in CS and U-Net, one can observe substantial signal in the 1D and 2D LPSF outside the main lobe.

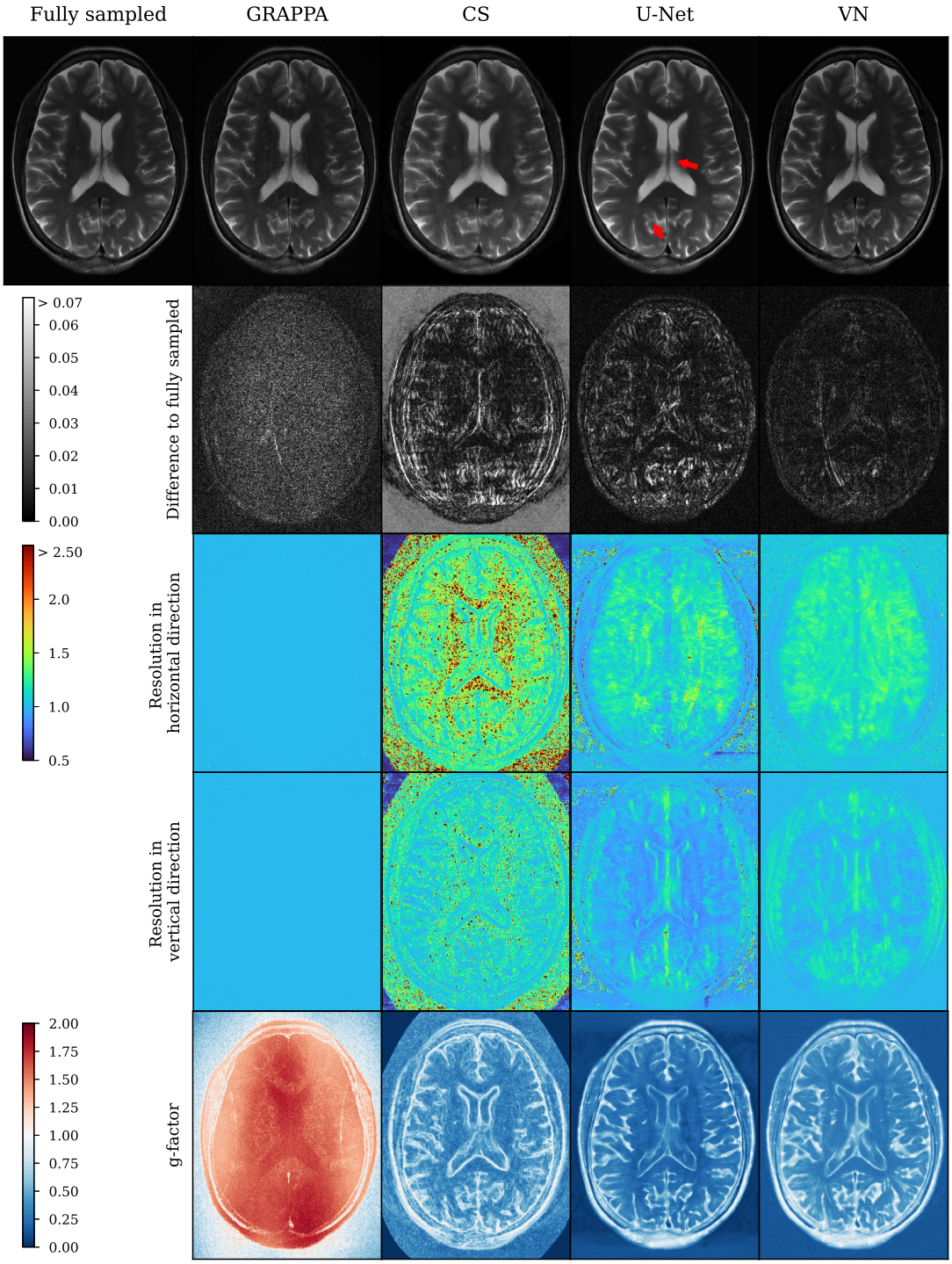


Figure 20: (Caption on the next page.)

Figure 20: (Figure on the previous page.) Reconstructions of a T_2-weighted slice by GRAPPA, compressed sensing (CS), U-Net and variational network (VN), together with error maps, resolution maps in both spatial directions and g-factor maps. The red arrows indicate substantial reconstruction errors in the U-Net output where anatomy is misrepresented due to hallucinations (see also Fig. 17). Since the output of the reconstruction models is scaled arbitrarily, for computing the error maps the output was rescaled with a factor minimizing the l_2-difference to the reference. The background of the CS reconstruction appears bright in the error map due to the root-sum-of-squares coil combination, which leads to noise with mean > 0 in the background of the reference, while the CS model removes the noise prior to coil combination. The width of the main lobe of the local point-spread function (LPSF) w is interpreted as a measure of resolution (3rd and 4th row), where values larger than 1 correspond to a loss of resolution, i.e. local blurring. GRAPPA shows no blurring throughout all locations. In CS and deep learning-based methods, w correlates with anatomical structure. Undersampling was applied in horizontal direction (see Fig. 16), where generally more blurring is indicated by the resolution maps. Values of the g-factor below 1 correspond to a suppression of noise, while values above 1 show noise amplification. All methods except GRAPPA suppress noise in most areas, while GRAPPA shows the known characteristics of noise amplification.

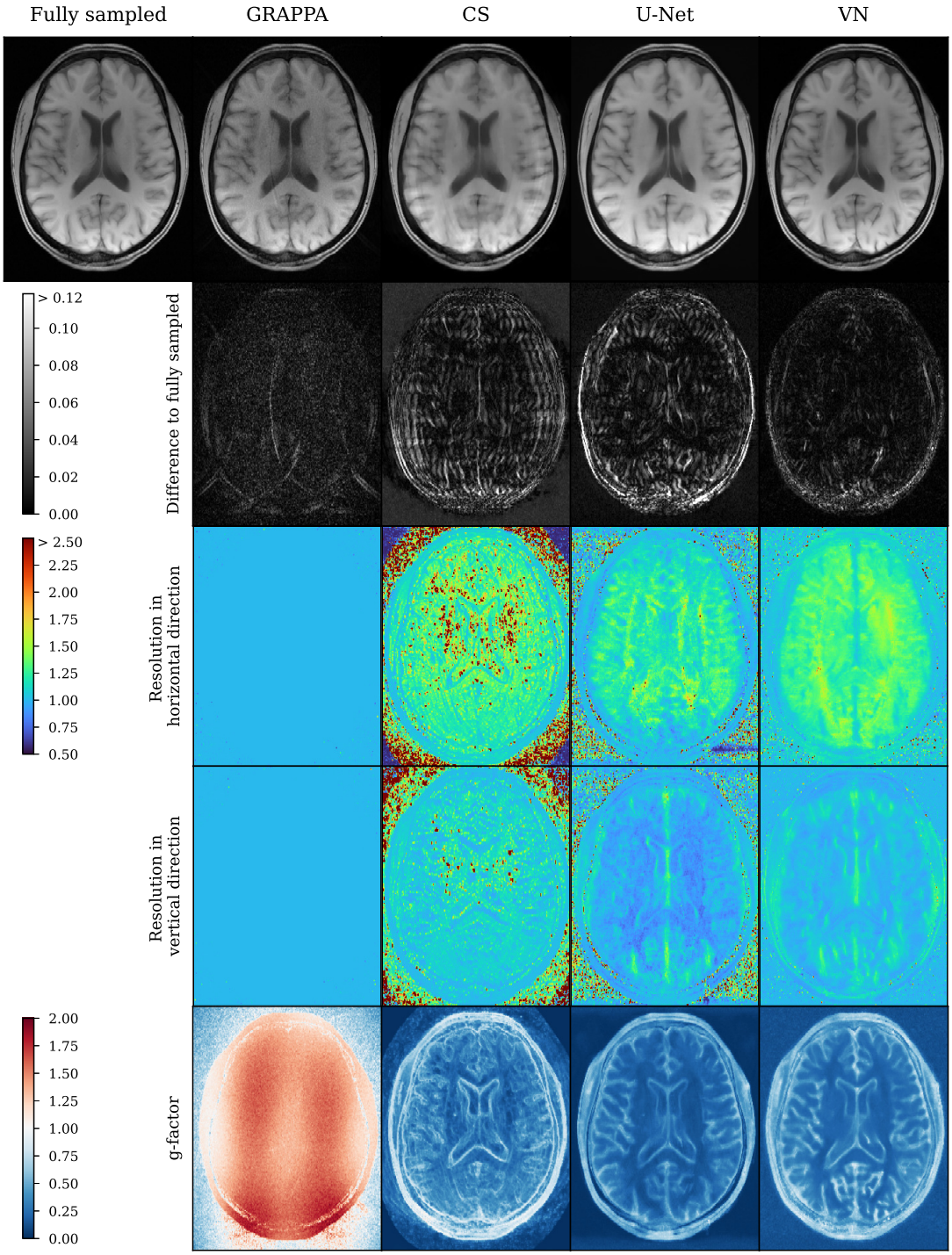


Figure 21: Reconstructions of a T_1-weighted slice by the tested methods together with resolution maps in both spatial directions and g-factor maps. The results appear similar to the ones obtained in the T_2-weighted image (Fig. 20), with ideal resolution for GRAPPA and blurring correlated with anatomical structure for the remaining model- and data-driven reconstruction techniques. With respect to noise, GRAPPA again features slightly elevated g-factors while the application of CS, U-Net and VN resulted in denoised images.

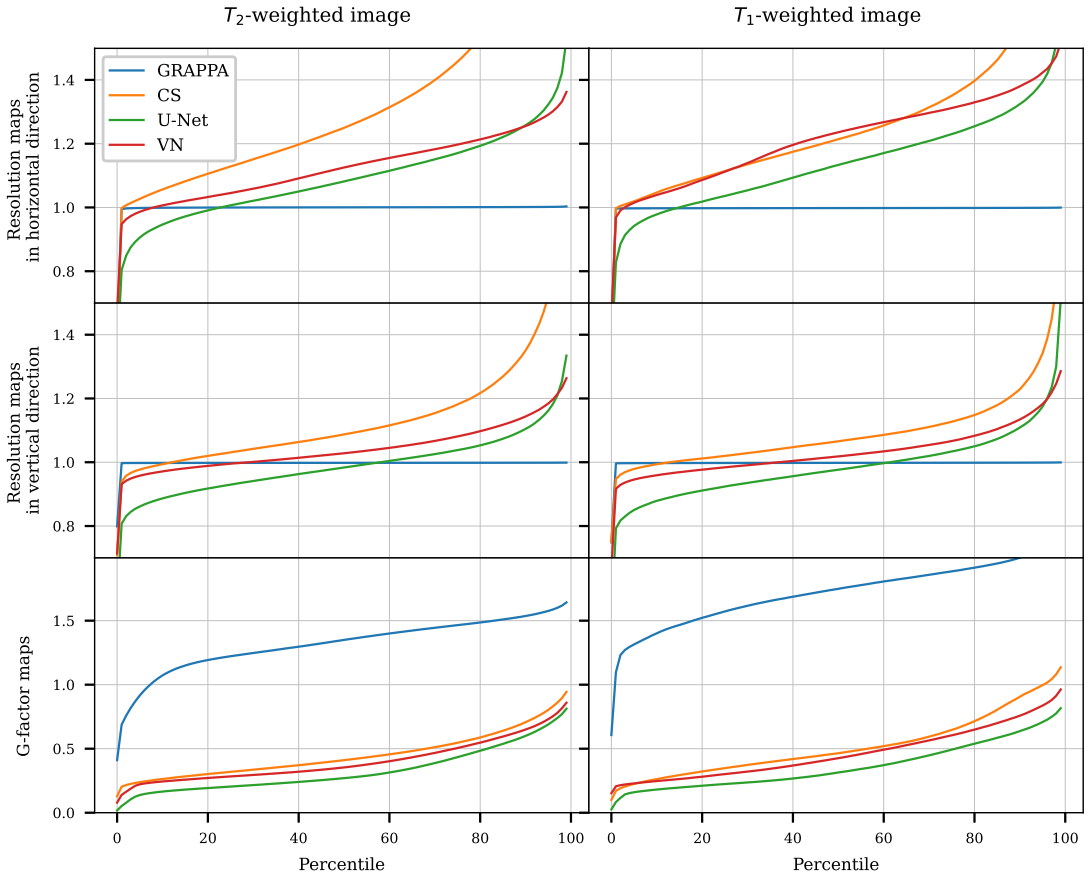


Figure 22: Cumulative frequency analysis (percentile plots) of the resolution maps in horizontal and vertical direction and the g-factor maps that are shown in Fig. 20 and 21 for the T_2 and T_1-weighted image. A mask was applied to only take values inside the anatomy into account. The x-axis specifies percentages, while the y-axis gives the corresponding percentile of the number of pixels in the image. Resolution scores above 1 indicate local blurring in those pixels. The g-factor measures noise amplification (if > 1) or suppression (if < 0). From the curves, it is obvious that GRAPPA shows no blurring basically everywhere, and high g-factors above 1 in most pixels. The other methods exhibit g-factors below 1 almost everywhere. For the resolution scores, CS shows values substantially above 1 in most pixels, and the highest values of all methods. VN and U-Net appear similar, with the widths of the LPSFs for VN somewhat larger than for U-Net in most pixels. However, from the crossings of the green and red line one can see that U-Net has more outliers.

5.4.3 g-factor maps

Figures 20 and 21 show the g-factor maps that were computed by the multiple-replica method as detailed above, while the corresponding cumulative frequency analysis (CFA) can be seen in Figure 22.

GRAPPA exhibits the classical g-factor maps as shown in various studies [87]. Their geometry is driven by the arrangement of receiver coils.

CS, U-Net and VN, all show a drastic denoising. However, this denoising is not present at edges of the images where the g-factor are close to 1. For CS, U-Net and VN, the g-factor maps appear rather similar.

5.5 Discussion

In this work, local spatial resolution and local noise enhancement is determined for data-driven and model-based reconstructions of undersampled MRI data. Local point-spread functions for each pixel and a multiple-replica approach were applied for quantifying resolution as well as noise amplification in four different reconstruction methods. We found that GRAPPA exhibits a homogeneously uncompromised resolution, but shows considerable noise amplification, as it has been described previously [87]. For all other investigated methods, the resolution depended on image structure. TV-regularized CS suffers from a substantial decrease in resolution in homogeneous areas of the image, but not at edges. To a lesser extent, this holds true for the resolution of U-Net and VN in horizontal direction though also some scattered blurring independent of structure can be seen in the resolution maps. In U-Net and VN in vertical direction, the resolution also depends on anatomical structure and is close to 1 around edges. The blurring for data-driven methods is substantially decreased compared to CS, and slightly higher for VN compared to U-Net. g-factor maps are similar between VN and U-Net, showing noise suppression almost everywhere except around hard edges.

While we are confident that the methods for investigating resolution and noise properties used here provide information on the general behavior of tested reconstruction methods, all results in this paper are initially specific to the anatomy, contrast, acquisition hardware etc. of the images we examined. Besides general similarity, some differences between the T_1- and the T_2-weighted image remain, like lower resolution in the T_1-weighted image in the horizontal direction than in the T_2-weighted one in all methods, especially the VN (see Fig. 20, 21 and 22) and visibly lower reconstruction quality in the CS reconstruction of the T_1-weighted image (see Fig. 3 and 4) than of the T_2-weighted one. Since these issues are consistent with the fact that the T_1-weighted image was acquired (see section 5.3.4) with a lower in-plane spatial resolution (0.61×0.68 mm^2) than the T_2-weighted one (0.46×0.51 mm^2), we think that the acquired resolution and matrix size of the input data does have some effect on the performance of the reconstruction methods and the methods for examining resolution and noise. For data-driven methods, a good match between the training dataset and test images is also important. Investigating these relationships in detail in a more comprehensive study including a variety of anatomies, contrasts, resolutions and hardware would be a worthwhile goal for future work.

We found that there is a substantial difference in the width of the LPSF's main lobe

for the vertical and horizontal directions. The values for w in the horizontal direction, where undersampling was performed, were clearly higher than in the vertical direction, see Fig.20, 21 and 22. This might be taken into account when choosing the direction in which to perform undersampling.

The general idea of introducing some kind of perturbation to evaluate how a reconstruction method responds has already been used in medical imaging [83–86]. Nevertheless, the approach of measuring the LPSF's main lobe has profound advantages over some previously presented methods. In particular, through the process of measuring the width of local point-spread functions, we obtain a *quantitative* measure of resolution, that can easily be interpreted as the ability to discern information by means of a single map. However, this comes at the cost of losing some of the information present in the 2D LPSFs, as only the width of its main lobe is measured and the rest of its shape is not considered. Furthermore, the spatially resolved g-factor also provides clear information on the amplification or suppression of noise for the individual imaging and reconstruction process.

In our results, we saw a clear difference in the maps of resolution and noise between parallel imaging, represented by GRAPPA, and the other reconstruction methods CS, U-Net and VN. GRAPPA interpolation is achieved by a convolution in k-space, separately on the data of each receiver coil. The uniform operation in k-space allows GRAPPA to retain ideal resolution, but at the cost of increased noise. Due to their similarity with GRAPPA in terms of mechanisms and qualitative output, we believe that an assessment of resolution of other parallel imaging techniques like SENSE would have similar results, though a detailed investigation would be of interest. The other examined methods (CS, U-Net and VN) are applied in image space. U-Net operates only on a coil-combined reconstruction. CS and VN use the spatial information by the coil sensitivity maps in their data consistency terms, but not in their regularization terms, which operate on coil-combined images. We consider these mechanisms as a trade-off between resolution and noise, as it allows them to achieve noise suppression and good artifact correction, but results in the introduction of local blurring.

As the underlying goal of MRI is answering clinical questions, it is important how our results translate to the diagnostic value of images reconstructed with the examined methods. For GRAPPA, we did not find any impairment by blurring or otherwise augmented anatomy, in accordance with the theory of the method. The risk here is that the noise is increased to a level where subtle anatomical or pathological details would get lost. The other methods, on the contrary, consistently reduced noise. However, all, and CS the most, showed varying degrees of local blurring, which can hide abnormalities. For the data-driven methods U-Net and VN, pathologies that were not represented in the training set pose a further important challenge. Unlike the purely model-based CS method, these data-driven methods have the ability to systematically suppress some

image components, or fabricate an output which appears realistic to human observers, while not being representative of the measured data. In general, the reconstruction defects in deep learning-based methods are diverse and often subtle [18]. For example, even though the U-Net reconstructions appear sharp and no residual aliasing artifacts are visible, we observed hallucinations. We want to stress the fact that the dataset was by no means specifically selected for this reason. Other studies have reported similar errors [18, 30]. Though this is certainly a rare occurrence for well-designed methods, the mere possibility raises the still open question of how these kind of reconstruction faults could be discovered, should they appear in potential clinical application.

The methods we presented here can give a spatially resolved indication of the ability to depict small details and the level of noise. High resolution and low noise are certainly important prerequisites for proper reconstructions, and thus the maps we presented can identify areas in which lower quality is possible. However, locally quantifying fidelity of reconstructions to the measured data in a sufficient manner is a further step, and the results presented here do not allow conclusions in this regard. For example, in Fig. 20 and 21 we saw overall higher values of w in VN than in U-Net, and yet the VN does not exhibit obvious reconstruction defects like U-Net. Some recent publications attempt to measure fidelity by repeated reconstructions of various kind, to identify areas in which the reconstructions are associated with uncertainty [19, 94]. Still, we believe that further development of functional and reliable methods that can provide quantifications of local reconstruction fidelity is desirable, and can have a positive impact on the trustworthiness of machine learning driven MRI.

A central assumption in the local approach, was to use sufficiently small perturbations, in order to be able to use local linearity. This is justified, since the reconstruction operator is usually smooth, and thus locally linear. On the other hand, the perturbation should not be chosen too small, to avoid the perturbation being lost in the noise. By reconstructing and checking the linearity condition, we found that, when we choose the amplitude as 0.1% of the maximal pixel, all the tested reconstruction method behave almost linear for both the T_1 and the T_2-weighted image. However, this is highly specific to the reconstruction methods and should be checked for each new method and each new image.

Note that, as described in the methods section, we used different retrospective sampling patterns for the methods. This may introduce slightly different representation of information simply due to the undersampling pattern and not directly the reconstruction method. However, using a single pattern for all methods would have introduced a much more substantial bias, since the methods have different requirements for undersampling mask structure, e.g. GRAPPA requires regular undersampling while compressed sensing benefits from random sampling. Furthermore, due to these patterns being so specific to the reconstruction methods, one could actually regard the sampling as a part of the

entire method, which consequently has to be included when comparing to other methods.

An obstacle to overcome for widespread application of local point-spread functions may be computational time. For computing a resolution map, a reconstruction with the method to be tested needs to be computed for each pixel of the input image. Since the computation of the LPSF is not computationally demanding, almost all of the time for computing a resolution map is spent on reconstructions. Therefore the computation times for a resolution map depends heavily on the speed of the reconstruction method. This is manageable for fast reconstruction methods, but problematic for slower ones. Thus, to overcome this issue, the reconstruction method itself can be accelerated through efficient implementations or appropriate hardware, like GPUs. Still, the determination of local resolution may not be suited for routine computation in many images.

A clear advantage of the resolution measurement on the other hand is its simplicity. The algorithm requires in essence only the computation of two reconstructions together with some basic signal processing for determining the width of the main lobe of the LPSF. We provide an implementation on GitHub (https://github.com/expRad/dlassessment).

5.6 Conclusion

We demonstrated methods for objective quality assessment of non-linear and non-stationary MRI reconstruction methods, namely a multiple-replica method for examining noise and a method for measuring local spatial resolution. These were applied to GRAPPA, compressed sensing, variational network and U-Net reconstructions. This study may serve as an example for the kind of critical and objective analysis of novel reconstruction methods that is necessary for advancing these techniques toward clinical application.

Funding Information

This work was supported by the German Ministry for Education and Research under Research Grant 05M20WKA.

6 Summarising Discussion

With the overarching goal of improving clinical MRI, the intention in this work was to advance MR image reconstruction. In the MR imaging procedure, reconstruction is a key component. Not only does it have a direct impact on image quality, but it can also make undersampling possible, which has been considered as a viable option for achieving acceleration in the scientific literature [4]. Acceleration can increase efficiency, accessibility, patient comfort and reduce cost. In some applications, like dynamic imaging, it is even required for applying MRI in the first place. Therefore, much effort has been dedicated to improving signal processing and reconstruction algorithms in MRI. Previously, CS algorithms have been established to this end. More recently, great interest has been taken in the rise of DL in image processing. Therefore, these algorithms have been applied to MR reconstruction in the past years. With a large amount of active research currently being conducted, they are constantly improving, and in some areas can already be seen as the state-of-the-art.

For this task, a promising network architecture is a variational network (VN). It incorporates data consistency mechanisms that promote fidelity of the reconstructions with measured data, leading to more robustness. In the first chapter, a VN model for reconstruction of undersampled spiral data for functional imaging of the heart was presented. There are a number of alternative approaches and possible improvements of the VN architecture that was used in this thesis, some of which are discussed in the following.

Uniquely, the VN model explicitly incorporated the non-Cartesian data in the data consistency term by transforming between on-grid and off-grid k-space in each cascade of the network. This approach is rather puristic in the sense that the entire MRI sampling procedure is modeled in the forward and adjoint operator of the VN. However, it has the disadvantage the a gridding and de-gridding operation has to be performed for each data consistency step, i.e. in each cascade. Even with computations on GPU, these operations are comparatively slow, which lead to reconstruction times of about 0.9 s per image where most of the time is spend on gridding. An alternative approach here would be to initially transform to a Cartesian grid, and then perform the entire VN computation there, which is much faster. The challenge then is to perform data consistency on the Cartesian k-space, while the data are measured with spiral waveforms. Since data consistency

should only be performed on k-space positions where data was actually measured, using convolution gridding like it was done here is problematic, as it spreads out the measured data in k-space. While it may be possible to come up with an approach to perform data consistency on such a gridded k-space, another approach would be to use an alternative method for transforming to a Cartesian grid. An option is grappa operator gridding (GROG) [95], where the sampled off-grid positions are shifted and transformed to the closest on-grid point, using the information gained from comparing different receiver coil channels. This results in Cartesian k-space data with defined sampling positions, and thus a VN that operates entirely on a Cartesian grid can easily be implemented. For data similar to the spiral cardiac data presented in chapter 4, such an approach was attempted. However, performance was inferior to a VN that performs data consistency off-grid. This may be due to errors that were introduced in the GROG operation, which were then enforced together with data consistency, thus degrading image quality. Nevertheless, the potential for substantial acceleration warrants further investigation in this direction.

In a recently-presented paper [45], a VN model was used that performed the gradient descent operation in k-space rather than image space. While the authors performed experiments and found that in their application this improved performance, most current publications use VNs operating in image space [2, 15, 17, 18]. This is problematic when using Kaiser-Bessel gridding, as the undersampling mask can not be modeled correctly. But using such a VN operating in k-space could be possible for non-Cartesian data, by using GROG to initially transform the undersampled data to a Cartesian grid. This could increase performance. Furthermore, the authors also included a learnable component that estimates the coil sensitivity maps, which they found increased performance over using pre-computed sensitivity maps. It would be worthwhile to explore whether these ideas can be beneficial for in the applications presented in this thesis.

Considering for a moment the motivation for the VN model (described in section 4.3.3), the idea for the regularization term Φ in the equation

$$x^{t+1} = x^t - \lambda(A^*(Ax^t - k) + \Phi(x^t)) \tag{27}$$

is that is it uses prior knowledge to enforce high-quality network output and suppress undersampling artifacts. In compressed sensing, Φ is realized by a hand-crafted model term, while in the VN model the number of cascades is fixed and in each cascade an individual regularization term is learned from data. However, in view of how regularized gradient descent usually operate (e.g. in compressed sensing), it seems somewhat unnatural that different regularization terms enforce different image features in each cascade. One could think that it may just as well be possible to learn a single regularizer from data, that enforces desirable image features, regardless of how many gradient descent steps have taken done before. Apart from reducing the number of trainable parameters in a VN model by a factor equal to the number of cascades (usually around 10), this

would have the benefit that the number of iterations would not have to be fixed, and could be adjusted to the specific input, as is done in compressed sensing. To investigate this approach, the VN model from chapter 4 was re-trained, as well as several other VN models for other applications. The number of cascades was kept constant, but only used a single U-Net as regularizer, which was used to compute the regularization terms in all cascades. This approach did result in a network model that was in principle able to reduce undersampling artifacts. However, performance both in terms of the loss function as well as visual image quality consistently stayed behind the classical model, even though the difference was only small in some tested applications. The reasons for this are not clear. It may be that using different regularizers in each cascade leads to more flexibility in the operation of the model or the higher parameter count could increase the modeling capability. Nevertheless, even though no clear benefits could yet be found in this approach, by potentially employing different training approaches, or tuning the hyperparameters (e.g. the size of the regularizing U-Net) it may well be possible to achieve an innovative, and more flexible VN model.

The variational network model in chapter 4 was extensively evaluated, by comparing it to segmented techniques including the clinical gold standard, a CS and a U-Net model in a variety of categories. These included quantitative scores like SSIM and RMSE, an expert reader study and performance in cardiac functional imaging. The goal of this evaluation was to compare the various methods based on the quality of their output. Image quality encompasses a large number of aspects. In medical imaging, the first priority is clearly diagnostic value, i.e. the ability of an image to convey information in a form that a trained radiologist can read to answer a specific clinical questions. This is important, since it differentiates the concept of image quality in medical imaging from that in general image processing. For example, up to a certain level, radiologists are able to read through noise quite well, meaning that in spite of noise they are able to discern structures that a normal, untrained observer may not. Also, not all parts of an image are of interest, e.g. in functional imaging of the heart the main feature of interest is the location of the myocardium-blood border, to enable an accurate segmentation for computation of quantitative scores. On the other hand, image artifacts that look similar to structures that are of interest may be more problematic. In a perfect world, assessment of image quality in MRI would therefore be conducted on diagnostic value directly. Unfortunately, this world is not perfect, and diagnostic value is quite hard and usually expensive to measure. As a positive example, the 2020 fastMRI challenge [18] should be mentioned. Some of the submitted reconstruction models were evaluated for depiction of pathological structures, by reconstructing a dataset that included a range of common pathologies, the depiction of which was then judged by a panel of radiologists. However, this is unfortunately more of an exception than the norm, and even here, due to the high effort, was done only to evaluate the finalists, which were selected on the

basis of quantitative scores. In chapter 4, diagnostic value was assessed by computing cardiac functional parameters for the clinical gold standard and the proposed VN model. The other reconstruction methods (U-Net, LRS) were omitted from this comparison due to the high effort that is necessary for manually segmenting cardiac images, though these results would certainly have been of interest. Instead, like in many studies, these were evaluated based on quantitative scores, both model-based with respect to a reference (SSIM and RMSE) and through ratings by expert radiologists. While ratings by experts are always possible, the effort is also relatively high due to the manual process. Also, access to expert ratings can be limited. Depending on the rated categories and the clinical query, it may allow some limited conclusions regarding diagnostic value. For the study presented in chapter 4, the expert ratings could provide some information on the ability of the reconstructions to accurately locate the myocardium-blood border. Furthermore, these serve to generally evaluate the image quality of the examined models, which may be of interest for potential other applications of these models. Lastly, assessment of image quality based on quantitative metrics like RMSE or SSIM is the most common form of evaluation. The required effort is relatively low, in particular for supervised DL applications where some sort of trainings dataset is already present. After all, the foundation of DL builds on the idea that good image quality in the output of a network can be achieved by minimizing such a loss function. While this is true to some extent, the information that model-based metrics of image quality provide is generally only very coarse. These loss terms simply can not capture all detailed aspects of image quality with respect to diagnostics, as these also depend on the clinical question. This is also true for perceptually motivated terms like SSIM. Also, while evaluation on the basis of quantitative scores is feasible in most cases, it can also be problematic in others. The evaluation of the presented model had to rely on synthetic data that was generated from cine reconstructions using a special spiral sequence (see Table 2), since no directly fully sampled acquisitions are possible in cardiac imaging.

By this point, it seems clear that evaluation of reconstruction methods and their output image quality in medical imaging is challenging, with considerable downsides to all common approaches. As a potential step towards improving this situation, chapter 5 was dedicated to demonstrating advanced, and not yet widely used, approaches for assessing image quality in MRI reconstruction. It showed methods for assessment of resolution and noise in MR reconstructions by non-linear, in particular data-driven, algorithms. Unlike the purely image-based evaluations discussed above, these explicitly utilize the acquired k-space data and the reconstruction method, to - essentially - 'locally probe' the reconstruction procedure of the examined image with either a small perturbation for assessing local blurring or with repeated, noisy reconstructions for assessing noise amplification or suppression. In a sense, both of these methods analyze the image by locally testing how the reconstruction would have changed in the presence of either a local perturbation

or input noise. Extracting additional information by repeated application of the reconstruction method for assessment increases complexity, in particular computational effort, but it also allows some conclusions that would not have been possible based only on the reconstructed image alone. It should be noted that these results are specific to the examined image and reconstruction method, and care has to be taken before generalizing. Although rather similar behavior in the reconstruction methods between the two tested brain images was observed, the small sample size clearly limits the assertions that are possible with respect to a reconstruction method in general. Still, even for analyzing just single images, these methods may be valuable tools. They allow precise local analysis of resolution and noise, which can be combined together to get a spatially resolved assessment of image quality. However, directly infering conclusions about diagnostic value is not possible. For example, the U-Net reconstructions were clearly flawed, as was apparent from the error maps and the observed hallucinations, but this was not reflected in the quantitative maps of resolution and noise. In the VN reconstructions, with clearly less hallucinations, the quantitative maps actually indicated lower resolution in most of the image (see Fig. 20, 21, 22). Therefore, these approaches do not aim to replace any of the methods for analyzing MRI reconstructions discussed above. In particular, it should be stressed again that for clinical applications, there currently does not exist any alternative for reliably assessing diagnostic value other than analysis of clinical results by experts. Instead, the presented methods should be seen as additional tools, with which some important aspects of image quality can be judged. As such, the results above show promising potential for enriching the analysis of MRI reconstructions.

Next to cardiac imaging, MRI is also particularly suited for abdominal imaging, due to its superior contrast in soft tissue. For visualizing the biliary and pancreatic duct, magnetic resonance cholangiopancreatography (MRCP) is currently considered the gold standard, and is commonly used in clinical routine. MRCP consists of a heavily T_2-weighted, three-dimensional acquisition, covering the entire biliary tract. Its structure, and in particular pathologies like stenoses in the biliary ducts are of principal interest to clinicians. This presents challenges to the imaging system, as high spatial resolution, and thus acquisition time, is necessary to detect small abnormalities. As it is therefore currently not possible to record all necessary data in a single breath-hold, one uses respiratory gating, which increases acquisition time even more. This is problematic, not only from an economic point of view, but also since avoiding patient motion becomes increasingly difficult with longer scan times. Together with residual breathing motion due to imperfect gating, this results in scans commonly being corrupted by motion artifacts. To overcome these difficulties, an acceleration of MRCP acquisitions beyond classical techniques is imperative. This may be possible by reconstructing undersampled acquisitions robustly with a VN. However, a hurdle to overcome is collecting an appropriate dataset for fitting the model. This is particularly problematic in the case of MRCP, since fully

sampled acquisitions are frequently corrupted by motion artifacts. One possible solution may be to use an appropriate synthetic dataset. Another approach to even entirely overcome the need for training data are self-supervised training schemes, which have recently been applied to MRI reconstruction [63]. To explore these approaches, a 3D VN model for Cartesian data, where all operations like Fourier transforms or convolutions in the U-Net model are three-dimensional, was implemented. It was trained on one hand in the classical supervised manner on a synthetic dataset generated from 807 MRCP volumes, which were exported from the University Hospital Würzburg's clinical archive. On the other hand the self-supervised technique recently presented by Yaman et al [63], with which the network weights can be fitted on only the data of a single acquisition was also implemented. The resulting models were evaluated on prospectively undersampled MRCP acquisitions at $R = 12$, and compared to a compressed sensing model. Preliminary results were published as a conference abstract [96] and some representative reconstructions can be seen in Fig. 23. One can observe that both VN models as well as the CS model are in prinicple able to effectively reduce the undersampling artifacts that are clearly visible in the naive reconstruction without any artifact correction. In the output of the self-supervised VN and the CS model, slight blurring can be seen, while the supervised VN shows some noise. However, for a robust and detailed assessment of these models and training schemes, a more comprehensive study is necessary. To this end, a larger patient study is currently being conducted, where undersampled MRCP data are acquired at different acceleration factors. While this is a challenging undertaking due to frequent corruption by gating errors, the aim is to develop a reliable, undersampled MRCP protocol with a reconstruction based on a 3D VN. In summary, while these results are certainly still preliminary, they point towards another promising application of VNs in clinical MRI.

For application to this or similar 3D reconstruction methods, generalizing the approaches for assessment of resolution and noise that were presented in chapter 5 could be a worthwhile aim for future work. Most steps could be done just the same, and for the pseudo-multiple replica method for computing g-factor maps no changes are necessary beyond using 3D volumes instead of 2D images. The method for assessing resolution would require measuring the width of the main lobe of a 3D instead of a 2D LPSF. Straight-forward, this can be done by adding a third direction into which the 3D LPSF is sliced. However, this raises the question, which is also valid for 2D images, whether there is not an easier, more interpretable approach then examining two or three maps for the spatial directions separately. For example, one could slice the LPSF's main lobe at some pre-determined height, and then measure the area (or, the volume respectively) of the resulting shape. Regardless of dimensionality, this would result in a single map, combining the information of the two maps for the spatial directions that can be seen in Figures 20 and 21. However, while increasing simplicity, some information would be

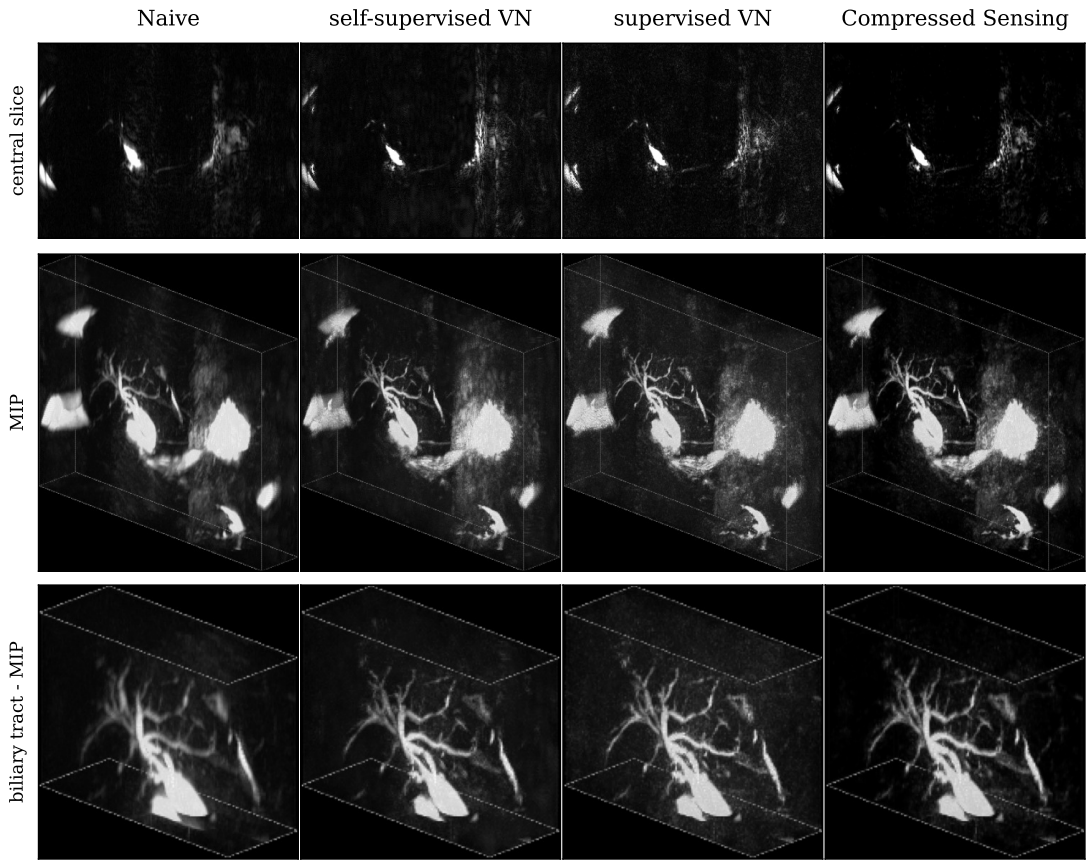


Figure 23: Comparison of reconstructions of prospectively undersampled (R = 12) magnetic resonance cholangiopancreatography (MRCP) using a 3D variational network trained with a supervised and a self-supervised method, and a compressed sensing model. The first row shows a central slice, while the lower rows show maximum intensity projections. Compared to the naive reconstruction without any artifact correction, both VN models as well as the CS model are able to effectively reduce undersampling artifacts.

lost, like the difference in behavior that was observed for the two spatial dimensions. Going even beyond this, it is worth pointing out that the concept of analyzing MRI reconstruction methods through local point-spread functions, or more general by any kind of perturbations, has a large potential for making more detailed assessments possible. For example, one could perturb an image not only in single pixels, but multiple pixels at once, and potentially even use carefully designed structures as perturbations. Developing ideas on how to best take into account the structure and shape of the 2D LPSF beyond just measuring the width of the main lobe could also be beneficial, as with the approach presented above, much information is disregarded (see Fig. 19). In a recent paper [86], the authors attempted an approach where all pixels are perturbed simultaneously, and further presented a copious list of other possible ideas. Finally, expanding the range of applications of the methods from chapter 5 by generalizing to non-Cartesian

reconstruction methods would be useful. This could prove challenging due to the added complexity of gridding, and cases like in chapter 4 where fully sampled data is difficult or impossible to obtain pose additional obstacles. However, investigating the effects of non-Cartesian reconstruction on resolution and noise would certainly be of interest and a valuable extension to the results presented above.

With the motivation of improving magnetic resonance imaging in a clinical context, this thesis identified reconstruction as a key component to facilitate progress. In the scientific literature, the current advances in signal and image processing through novel deep learning algorithms have recently demonstrated considerable potential for applications in MRI reconstruction. In particular variational networks, where the architecture is informed by the nature of the task as an inverse problem by unrolling a gradient descent iteration have been established as suitable [2, 14, 15, 17, 18]. Therefore, examining VNs took a central place in this thesis. The following paragraph reflects on these results. First, in all presented applications, it was found that VNs are well suited for reconstructing undersampling artifacts in MRI. They are highly versatile, and successful applications to 2D Cartesian, 2D spiral and 3D Cartesian reconstruction problems were presented. Perceived image quality was generally good. Slight blurring around edges was observed on occasion, both visually (Fig. 10) and on the basis of local point-spread functions (Fig. 20, 21), but to an extend acceptable for clinical use (Tab. 3). Noise is generally low in the output images (Fig. 20, 21, Tab. 3), as it is actively suppressed by the learned regularizers, in particular in homogenous areas. Like other advanced methods, the produced resolution and noise in the output is not homogenous, but adaptive to the image structure (Fig. 20, 21). Compared to purely image-based data-driven artifact correction based on a U-Net, clearly superior performance was observed in the VN on all occasions. While computational complexity and runtimes are lower for U-Nets (Tab. 1), these showed more residual artifacts and blurring (Fig. 9, 10), substantially more hallucinated image structures (Fig. 17, 18), and worse generalization beyond the trained data domain (Fig. 11). Comparisons to model-based compressed sensing methods were more balanced, with performance overall similar, but details depending on the specific CS model. In different cases, lower levels of noise (Fig. 10, Tab. 3), blurring (Fig. 23) and residual artifacts (Fig. 18) were observed in VN over CS. In dynamic imaging, performance was overall on par, even though the latter was able to use additional information by applying a temporal model (Fig. 9, Tab. 3). Unlike VNs, CS models do not need to be trained and are agnostic to anatomy (Fig. 11), but often require longer runtimes (Tab. 1, Section 4.5). For applying VN models to data with non-Cartesian sampling trajectories, a novel approach, where the entire sampling process including gridding and de-gridding was modeled in the forward and adjoint operator, was found to perform well (Chapter 4).

Overall, VNs showed excellent performance in clinical MRI. They seem to combine

the good modeling capabilities of deep learning with the data fidelity and robustness of compressed sensing.

With a large host of publications on manifold applications in the past years, it may be conceivable - and is certainly up to debate - to call VNs, or more general networks of unrolled gradient descent structure, the current gold standard for DL-based MRI reconstruction. Meanwhile, the field of image processing and with it MRI reconstruction is continuing to evolve. In the very recent past, two new kind of new DL models, namely Transformer and Diffusion Models have been gaining traction. While Transformers can be used as a building block inside the regularizers of a VN or other neural networks, Diffusion Models represent an entirely new approach to extracting and modeling prior knowledge from a training dataset. Some first, promising applications to MRI reconstruction have been presented [97–99], though it is certainly still too early to evaluate the benefits of these techniques. It does however turn attention to the remakable speed and energy with which innovation in the field of MRI reconstruction is driven today.

7 References

[1] M. Hardy and H. Harvey, "Artificial intelligence in diagnostic imaging: Impact on the radiography profession," *The British Journal of Radiology*, vol. 93, no. 1108, p. 20 190 840, Apr. 2020. DOI: 10.1259/bjr.20190840.

[2] K. Hammernik *et al.*, "Learning a variational network for reconstruction of accelerated MRI data," *Magnetic Resonance in Medicine*, vol. 79, no. 6, pp. 3055–3071, 2018. DOI: 10.1002/mrm.26977.

[3] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487–492, Mar. 2018. DOI: 10.1038/nature25988.

[4] A. Bustin, N. Fuin, R. M. Botnar, and C. Prieto, "From compressed-sensing to artificial intelligence-based cardiac MRI reconstruction," *Frontiers in Cardiovascular Medicine*, vol. 7, 2020. DOI: 10.3389/fcvm.2020.00017.

[5] E. Ahishakiye, M. B. Van Gijzen, J. Tumwiine, R. Wario, and J. Obungoloch, "A survey on deep learning in medical image reconstruction," *Intelligent Medicine*, vol. 1, no. 3, pp. 118–127, Sep. 1, 2021. DOI: 10.1016/j.imed.2021.03.003.

[6] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, no. 1, pp. 221–248, 2017. DOI: 10.1146/annurev-bioeng-071516-044442.

[7] M. J. Ankenbrand, D. Lohr, W. Schlötelburg, T. Reiter, T. Wech, and L. M. Schreiber, "Deep learning-based cardiac cine segmentation: Transfer learning application to 7T ultrahigh-field MRI," *Magnetic Resonance in Medicine*, vol. 86, no. 4, pp. 2179–2191, 2021. DOI: 10.1002/mrm.28822.

[8] W. Bai *et al.*, "Automated cardiovascular magnetic resonance image analysis with fully convolutional networks," *Journal of Cardiovascular Magnetic Resonance*, vol. 20, no. 1, p. 65, Sep. 14, 2018. DOI: 10.1186/s12968-018-0471-x.

[9] C. Tinauer *et al.*, "Interpretable brain disease classification and relevance-guided deep learning," *Scientific Reports*, vol. 12, no. 1, p. 20 254, Nov. 24, 2022. DOI: 10.1038/s41598-022-24541-7.

[10] A. Veeramuthu *et al.*, "MRI brain tumor image classification using a combined feature and image-based classifier," *Frontiers in Psychology*, vol. 13, 2022.

[11] C. Floyd, "An artificial neural network for SPECT image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 10, no. 3, pp. 485–487, Sep. 1991. DOI: 10.1109/42.97600.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, vol. 25, Curran Associates, Inc., 2012.

[13] S. Wang *et al.*, "Accelerating magnetic resonance imaging via deep learning," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2016, pp. 514–517. DOI: 10.1109/ISBI.2016.7493320.

[14] 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*, vol. 37, no. 2, pp. 491–503, Feb. 2018. DOI: 10.1109/TMI.2017.2760978.

[15] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-based deep learning architecture for inverse problems," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 394–405, Feb. 2019. DOI: 10.1109/TMI.2018.2865356.

[16] G. Zeng *et al.*, "A review on deep learning MRI reconstruction without fully sampled k-space," *BMC Medical Imaging*, vol. 21, no. 1, p. 195, Dec. 24, 2021. DOI: 10.1186/s12880-021-00727-9.

[17] F. Knoll *et al.*, "Advancing machine learning for MR image reconstruction with an open competition: Overview of the 2019 fastMRI challenge," *Magnetic Resonance in Medicine*, vol. 84, no. 6, pp. 3054–3070, 2020. DOI: 10.1002/mrm.28338.

[18] M. J. Muckley *et al.*, "Results of the 2020 fastMRI challenge for machine learning MR image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 40, no. 9, pp. 2306–2317, Sep. 2021. DOI: 10.1109/TMI.2021.3075856.

[19] D. Narnhofer, A. Effland, E. Kobler, K. Hammernik, F. Knoll, and T. Pock, "Bayesian uncertainty estimation of learned variational MRI reconstruction," *IEEE Transactions on Medical Imaging*, vol. 41, no. 2, pp. 279–291, Feb. 2022. DOI: 10.1109/TMI.2021.3112040.

[20] M. Stich *et al.*, "Gradient waveform pre-emphasis based on the gradient system transfer function," *Magnetic Resonance in Medicine*, vol. 80, no. 4, pp. 1521–1532, 2018. DOI: 10.1002/mrm.27147.

[21] P. Eirich *et al.*, "Cardiac real-time MRI using a pre-emphasized spiral acquisition based on the gradient system transfer function," *Magnetic Resonance in Medicine*, vol. 85, no. 5, pp. 2747–2760, 2021. DOI: 10.1002/mrm.28621.

[22] H. Scholten, D. Lohr, T. Wech, and H. Köstler, "Fast measurement of the gradient system transfer function at 7 T," *Magnetic Resonance in Medicine*, vol. 89, no. 4, pp. 1644–1659, 2023. DOI: 10.1002/mrm.29523.

[23] J. Schlemper *et al.*, "Nonuniform variational network: Deep learning for accelerated nonuniform MR image reconstruction," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, D. Shen *et al.*, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2019, pp. 57–64. DOI: 10.1007/978-3-030-32248-9_7.

[24] D. Shen *et al.*, "Rapid reconstruction of highly undersampled, non-cartesian real-time cine k-space data using a perceptual complex neural network (PCNN)," *NMR in Biomedicine*, vol. 34, no. 1, e4405, 2021. DOI: 10.1002/nbm.4405.

[25] Y. Zhang, H. She, and Y. P. Du, "Dynamic MRI of the abdomen using parallel non-cartesian convolutional recurrent neural networks," *Magnetic Resonance in Medicine*, vol. 86, no. 2, pp. 964–973, 2021. DOI: 10.1002/mrm.28774.

[26] A. Kofler, M. Haltmeier, T. Schaeffter, and C. Kolbitsch, "An end-to-end-trainable iterative network architecture for accelerated radial multi-coil 2d cine MR image reconstruction," *Medical Physics*, vol. 48, no. 5, pp. 2412–2425, 2021. DOI: 10.1002/mp.14809.

[27] M. O. Malavé *et al.*, "Reconstruction of undersampled 3d non-cartesian image-based navigators for coronary MRA using an unrolled deep learning model," *Magnetic Resonance in Medicine*, vol. 84, no. 2, pp. 800–812, 2020. DOI: 10.1002/mrm.28177.

[28] L. Fan *et al.*, "Rapid dealiasing of undersampled, non-cartesian cardiac perfusion images using u-net," *NMR in Biomedicine*, vol. 33, no. 5, e4239, 2020. DOI: 10.1002/nbm.4239.

[29] Z. Ramzi, J.-L. Starck, and P. Ciuciu, "Density compensated unrolled networks for non-cartesian MRI reconstruction," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2021, pp. 1443–1447. DOI: 10.1109/ISBI48211.2021.9433912.

[30] J. Kleineisel *et al.*, "Real-time cardiac MRI using an undersampled spiral k-space trajectory and a reconstruction based on a variational network," *Magnetic Resonance in Medicine*, vol. 88, no. 5, pp. 2167–2178, 2022. DOI: 10.1002/mrm.29357.

[31] J. Kleineisel, P. Eirich, J. F. Heidenreich, H. Köstler, T. A. Bley, and T. Wech, "Real-time cardiac MRI using spiral read-outs and a Variational Network for data-driven reconstruction," in *Proc. Intl. Soc. Mag. Reson. Med. 29*, 2021.

[32] K. S. Nayak, Y. Lim, A. E. Campbell-Washburn, and J. Steeden, "Real-time magnetic resonance imaging," *Journal of Magnetic Resonance Imaging*, vol. 55, no. 1, pp. 81–99, 2022. DOI: 10.1002/jmri.27411.

[33] P. Kellman, F. H. Epstein, and E. R. McVeigh, "Adaptive sensitivity encoding incorporating temporal filtering (TSENSE)†," *Magnetic Resonance in Medicine*, vol. 45, no. 5, pp. 846–852, 2001. DOI: 10.1002/mrm.1113.

[34] V. Muthurangu, P. Lurz, J. D. Critchely, J. E. Deanfield, A. M. Taylor, and M. S. Hansen, "Real-time assessment of right and left ventricular volumes and function in patients with congenital heart disease by using high spatiotemporal resolution radial k-t SENSE," *Radiology*, vol. 248, no. 3, pp. 782–791, Sep. 1, 2008. DOI: 10.1148/radiol.2482071717.

[35] M. Beer *et al.*, "Free breathing cardiac real-time cine MR without ECG triggering," *International Journal of Cardiology*, vol. 145, no. 2, pp. 380–382, Nov. 19, 2010. DOI: 10.1016/j.ijcard.2010.02.052.

[36] J. A. Steeden, G. T. Kowalik, O. Tann, M. Hughes, K. H. Mortensen, and V. Muthurangu, "Real-time assessment of right and left ventricular volumes and function in children using high spatiotemporal resolution spiral bSSFP with compressed sensing," *Journal of Cardiovascular Magnetic Resonance*, vol. 20, no. 1, p. 79, Dec. 6, 2018. DOI: 10.1186/s12968-018-0500-9.

[37] H. Haji-Valizadeh *et al.*, "Validation of highly accelerated real-time cardiac cine MRI with radial k-space sampling and compressed sensing in patients at 1.5t and 3t," *Magnetic Resonance in Medicine*, vol. 79, no. 5, pp. 2745–2751, 2018. DOI: 10.1002/mrm.26918.

[38] C. Unterberg-Buchwald *et al.*, "Real time cardiac MRI and its clinical usefulness in arrhythmias and wall motion abnormalities," *Journal of Cardiovascular Magnetic Resonance*, vol. 16, no. 1, P34, Jan. 16, 2014. DOI: 10.1186/1532-429X-16-S1-P34.

[39] T. Wech *et al.*, "High-resolution functional cardiac MR imaging using density-weighted real-time acquisition and a combination of compressed sensing and parallel imaging for image reconstruction," *RöFo - Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*, vol. 182, no. 8, pp. 676–681, Aug. 2010. DOI: 10.1055/s-0029-1245504.

[40] D. Voit, S. Zhang, C. Unterberg-Buchwald, J. M. Sohns, J. Lotz, and J. Frahm, "Real-time cardiovascular magnetic resonance at 1.5 t using balanced SSFP and 40 ms resolution," *Journal of Cardiovascular Magnetic Resonance*, vol. 15, no. 1, p. 79, Sep. 12, 2013. DOI: 10.1186/1532-429X-15-79.

[41] T. Kido *et al.*, "Compressed sensing real-time cine cardiovascular magnetic resonance: Accurate assessment of left ventricular function in a single-breath-hold," *Journal of Cardiovascular Magnetic Resonance*, vol. 18, no. 1, p. 50, Aug. 24, 2016. DOI: 10.1186/s12968-016-0271-0.

[42] K. S. Nayak, C. H. Cunningham, J. M. Santos, and J. M. Pauly, "Real-time cardiac MRI at 3 tesla," *Magnetic Resonance in Medicine*, vol. 51, no. 4, pp. 655–660, 2004. DOI: 10.1002/mrm.20053.

[43] 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*, vol. 57, no. 6, pp. 1086–1098, 2007. DOI: 10.1002/mrm.21236.

[44] K. Block, M. Uecker, and J. Frahm, "Model-based iterative reconstruction for radial fast spin-echo MRI," *IEEE Transactions on Medical Imaging*, vol. 28, no. 11, pp. 1759–1769, Nov. 2009. DOI: 10.1109/TMI.2009.2023119.

[45] A. Sriram *et al.*, "End-to-end variational networks for accelerated MRI reconstruction," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, A. L. Martel *et al.*, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2020, pp. 64–73. DOI: 10.1007/978-3-030-59713-9_7.

[46] T. Küstner *et al.*, "CINENet: Deep learning-based 3d cardiac CINE MRI reconstruction with multi-coil complex-valued 4d spatio-temporal convolutions," *Scientific Reports*, vol. 10, no. 1, p. 13 710, Aug. 13, 2020. DOI: 10.1038/s41598-020-70551-8.

[47] N. Fuin *et al.*, "A multi-scale variational neural network for accelerating motion-compensated whole-heart 3d coronary MR angiography," *Magnetic Resonance Imaging*, vol. 70, pp. 155–167, Jul. 1, 2020. DOI: 10.1016/j.mri.2020.04.007.

[48] C. Oh, D. Kim, J.-Y. Chung, Y. Han, and H. Park, "ETER-net: End to end MR image reconstruction using recurrent neural network," in *Machine Learning for Medical Image Reconstruction*, F. Knoll, A. Maier, and D. Rueckert, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2018, pp. 12–20. DOI: 10.1007/978-3-030-00129-2_2.

[49] 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*, vol. 38, no. 1, pp. 280–290, Jan. 2019. DOI: 10.1109/TMI.2018.2863670.

[50] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-net for compressive sensing MRI," in *Advances in Neural Information Processing Systems*, vol. 29, Curran Associates, Inc., 2016.

[51] 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*, vol. 81, no. 2, pp. 1143–1156, 2019. DOI: 10.1002/mrm.27480.

[52] B. A. Hargreaves, *Variable-density spiral design functions*, https://mrsrl.stanford.edu/~brian/vdspiral/. Online, accessed 7 June 2019.

[53] D. O. Walsh, A. F. Gmitro, and M. W. Marcellin, "Adaptive reconstruction of phased array MR imagery," *Magnetic Resonance in Medicine*, vol. 43, no. 5, pp. 682–690, May 2000. DOI: 10.1002/(sici)1522-2594(200005)43:5⟨682::aid-mrm10⟩3.0.co;2-g.

[54] L. Landweber, "An iteration formula for fredholm integral equations of the first kind," *American Journal of Mathematics*, vol. 73, no. 3, pp. 615–624, 1951. DOI: 10.2307/2372313.

[55] R. D. Hoge, R. K. Kwan, and G. B. Pike, "Density compensation functions for spiral MRI," *Magnetic Resonance in Medicine*, vol. 38, no. 1, pp. 117–128, Jul. 1997. DOI: 10.1002/mrm.1910380117.

[56] V. Rasche, R. Proksa, R. Sinkus, P. Bornert, and H. Eggers, "Resampling of data between arbitrary grids using convolution interpolation," *IEEE Transactions on Medical Imaging*, vol. 18, no. 5, pp. 385–392, May 1999. DOI: 10.1109/42.774166.

[57] J. G. Pipe and P. Menon, "Sampling density compensation in MRI: Rationale and an iterative numerical solution," *Magnetic Resonance in Medicine*, vol. 41, no. 1, pp. 179–186, 1999. DOI: 10.1002/(SICI)1522-2594(199901)41:1⟨179::AID-MRM25⟩3.0.CO;2-V.

[58] N. Dwork *et al.*, "Least squares optimal density compensation for the gridding non-uniform discrete fourier transform," *arXiv:2106.06660*, Jun. 16, 2021.

[59] A. Paszke *et al.*, "PyTorch: An imperative style, high-performance deep learning library," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 721, Red Hook, NY, USA: Curran Associates Inc., Dec. 8, 2019, pp. 8026–8037.

[60] D. P. Kingma and J. L. Ba, "ADAM: A method for stochastic optimization," in *International Conference on Learning Representations*, 2015.

[61] K. Hammernik, J. Schlemper, C. Qin, J. Duan, R. M. Summers, and D. Rueckert, "Systematic evaluation of iterative deep neural networks for fast parallel MRI reconstruction with sensitivity-weighted coil combination," *Magnetic Resonance in Medicine*, vol. 86, no. 4, pp. 1859–1872, 2021. DOI: 10.1002/mrm.28827.

[62] C. M. Sandino, P. Lai, S. S. Vasanawala, and J. Y. Cheng, "Accelerating cardiac cine MRI using a deep learning-based ESPIRiT reconstruction," *Magnetic Resonance in Medicine*, vol. 85, no. 1, pp. 152–167, 2021. DOI: 10.1002/mrm.28420.

[63] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya, "Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data," *Magnetic Resonance in Medicine*, vol. 84, no. 6, pp. 3172–3191, 2020. DOI: 10.1002/mrm.28378.

[64] M. Muckley, *Torchkbnufft*, https://github.com/mmuckley/torchkbnufft. Online, accessed 1 Aug. 2022.

[65] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004. DOI: 10.1109/TIP.2003.819861.

[66] J. Zbontar *et al.*, *fastMRI: An open dataset and benchmarks for accelerated MRI*, Dec. 11, 2019. DOI: 10.48550/arXiv.1811.08839.

[67] 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*, vol. 73, no. 3, pp. 1125–1136, 2015. DOI: 10.1002/mrm.25240.

[68] J. Kleineisel *et al.*, *GitHub - expRad/spiralvarnet*, https://github.com/expRad/spiralvarnet. Online, accessed 30 June 2019.

[69] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *The Lancet*, vol. 327, no. 8476, pp. 307–310, Feb. 8, 1986. DOI: 10.1016/S0140-6736(86)90837-8.

[70] ——, "Measuring agreement in method comparison studies," *Statistical Methods in Medical Research*, vol. 8, no. 2, pp. 135–160, Apr. 1, 1999. DOI: 10.1177/096228029900800204.

[71] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

[72] N. Kawel-Boehm *et al.*, "Reference ranges ("normal values") for cardiovascular magnetic resonance (CMR) in adults and children: 2020 update," *Journal of Cardiovascular Magnetic Resonance*, vol. 22, no. 1, p. 87, Dec. 14, 2020. DOI: 10.1186/s12968-020-00683-3.

[73] L. Feng, L. Axel, H. Chandarana, K. T. Block, D. K. Sodickson, and R. Otazo, "XD-GRASP: Golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing," *Magnetic Resonance in Medicine*, vol. 75, no. 2, pp. 775–788, 2016. DOI: 10.1002/mrm.25665.

[74] A. G. Christodoulou *et al.*, "Magnetic resonance multitasking for motion-resolved quantitative cardiovascular imaging," *Nature Biomedical Engineering*, vol. 2, no. 4, pp. 215–226, Apr. 2018. DOI: 10.1038/s41551-018-0217-y.

[75] E. Kobler, A. Effland, K. Kunisch, and T. Pock, "Total deep variation for linear inverse problems," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020, pp. 7546–7555. DOI: 10.1109/CVPR42600.2020.00757.

[76] B. Dietz, B. G. Fallone, and K. Wachowicz, "Nomenclature for real-time magnetic resonance imaging," *Magnetic Resonance in Medicine*, vol. 81, no. 3, pp. 1483–1484, 2019. DOI: 10.1002/mrm.27487.

[77] K. S. Nayak, "Response to letter to the editor: Nomenclature for real-time magnetic resonance imaging," *Magnetic Resonance in Medicine*, vol. 82, no. 2, pp. 525–526, 2019. DOI: 10.1002/mrm.27770.

[78] M. Bock, L. Traser, A. C. Özen, M. Burdumy, B. Richter, and M. Echternach, "Reply to letter to the editor: "nomenclature for real-time magnetic resonance imaging"," *Magnetic Resonance in Medicine*, vol. 81, no. 3, pp. 1485–1485, 2019. DOI: https://doi.org/10.1002/mrm.27599.

[79] A. Carkeet, "Exact parametric confidence intervals for bland-altman limits of agreement," *Optometry and Vision Science*, vol. 92, no. 3, Mar. 2015. DOI: 10.1097/OPX.0000000000000513.

[80] J. Kleineisel, K. Lauer, A. Borzì, T. A. Bley, H. Köstler, and T. Wech, "Assessment of resolution and noise in magnetic resonance images reconstructed by data driven approaches," *Zeitschrift für Medizinische Physik*, 2023, in press. DOI: 10.1016/j.zemedi.2023.08.007.

[81] K. Lauer, J. Kleineisel, A. Borzi, T. A. Bley, and T. Wech, "Assessment of resolution and noise in MR images reconstructed by data driven approaches," in *Proc. Intl. Soc. Mag. Reson. Med. 30*, 2022.

[82] A. Maier, S. Steidl, V. Christlein, and J. Hornegger, Eds., *Medical Imaging Systems: An Introductory Guide*, vol. 11111, Lecture Notes in Computer Science, Cham: Springer International Publishing, 2018. DOI: 10.1007/978-3-319-96520-8.

[83] J. Fessler and W. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction: Space-invariant tomographs," *IEEE Transactions on Image Processing*, vol. 5, no. 9, pp. 1346–1358, Sep. 1996. DOI: 10.1109/83.535846.

[84] S. Ahn and R. M. Leahy, "Analysis of resolution and noise properties of nonquadratically regularized image reconstruction methods for PET," *IEEE Transactions on Medical Imaging*, vol. 27, no. 3, pp. 413–424, Mar. 2008. DOI: 10.1109/TMI.2007.911549.

[85] T. Wech *et al.*, "Resolution evaluation of MR images reconstructed by iterative thresholding algorithms for compressed sensing," *Medical Physics*, vol. 39, no. 7, pp. 4328–4338, 2012. DOI: 10.1118/1.4728223.

[86] C.-C. Chan and J. P. Haldar, "Local perturbation responses and checkerboard tests: Characterization tools for nonlinear MRI methods," *Magnetic Resonance in Medicine*, vol. 86, no. 4, pp. 1873–1887, 2021. DOI: 10.1002/mrm.28828.

[87] 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*, vol. 62, no. 3, pp. 739–746, 2009. DOI: 10.1002/mrm.22066.

[88] P. Blümler, B. Blümich, R. E. Botto, and E. Fukushima, *Spatially Resolved Magnetic Resonance: Methods, Materials, Medicine, Biology, Rheology, Geology, Ecology, Hardware: Methods, Medicine, Biology, Rheology, Geology, Ecology, Hardware.* Weinheim: Wiley-VCH, Dec. 1, 1998, 760 pp.

[89] P. M. Robson, A. K. Grant, A. J. Madhuranthakam, R. Lattanzi, D. K. Sodickson, and C. A. McKenzie, "Comprehensive quantification of signal-to-noise ratio and g-factor for image-based and k-space-based parallel imaging reconstructions," *Magnetic Resonance in Medicine*, vol. 60, no. 4, pp. 895–907, 2008. DOI: 10.1002/mrm.21728.

[90] M. A. Griswold *et al.*, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magnetic Resonance in Medicine*, vol. 47, no. 6, pp. 1202–1210, 2002. DOI: 10.1002/mrm.10171.

[91] N. McKibben, *PyGRAPPA*, https://github.com/mckib2/pygrappa. Online, accessed 28 Nov. 2022.

[92] *BART Toolbox for Computational Magnetic Resonance Imaging*, https://mrirecon.github.io/bart/. Online, accessed 31 Oct. 2022.

[93] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, Jul. 25, 2011. DOI: 10.1561/2200000016.

[94] G. Luo, M. Blumenthal, M. Heide, and M. Uecker, "Bayesian MRI reconstruction with joint uncertainty estimation using diffusion models," *Magnetic Resonance in Medicine*, vol. 90, no. 1, pp. 295–311, 2023. DOI: 10.1002/mrm.29624.

[95] N. Seiberlich, F. A. Breuer, M. Blaimer, K. Barkauskas, P. M. Jakob, and M. A. Griswold, "Non-cartesian data reconstruction using GRAPPA operator gridding (GROG)," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1257–1265, Dec. 2007. DOI: 10.1002/mrm.21435.

[96] J. Kleineisel, B. Petritsch, T. A. Bley, H. Köstler, and T. Wech, "Reconstruction of accelerated MR cholangiopancreatography using supervised and self-supervised 3D Variational Networks," in *Proc. Intl. Soc. Mag. Reson. Med. 31*, 2023.

[97] T. Wech, O. Schad, and J. Kleineisel, "Physics-informed reconstruction of undersampled MR data using a reverse diffusion model trained with magnitude-only images," in *Proc. Intl. Soc. Mag. Reson. Med. 31*, 2023.

[98] D. Hu, Y. Zhang, J. Zhu, Q. Liu, and Y. Chen, "TRANS-net: Transformer-enhanced residual-error AlterNative suppression network for MRI reconstruction," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–13, 2022. DOI: 10.1109/TIM.2022.3205684.

[99] Y. Korkmaz, T. Cukur, and V. Patel, *Self-supervised MRI reconstruction with unrolled diffusion models*, Jun. 28, 2023. DOI: 10.48550/arXiv.2306.16654.

8 Curriculum vitae

9 Affidavit

Affidavit

I hereby confirm that my thesis entitled "Variational networks in magnetic resonance imaging - Application to spiral cardiac MRI and investigations on image quality" is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Würzburg, _____ _____

Jonas Kleineisel

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation "Variational Networks in der Magnetresonanztomographie - Anwendung auf spirale Herzbildgebung und Untersuchungen zur Bildqualität" eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Würzburg, _____ _____

Jonas Kleineisel

A Publications

Papers

1. J. Kleineisel, J. F. Heidenreich, P. Eirich, N. Petri, H. Köstler, B. Petritsch, T. A. Bley, and T. Wech, "Real-time cardiac MRI using an undersampled spiral k-space trajectory and a reconstruction based on a variational network", *Magnetic Resonance in Medicine*, vol. 88, no. 5, pp. 2167–2178, 2022, DOI: 10.1002/mrm.29357

2. J. Kleineisel, K. Lauer, A. Borzì, T. A. Bley, H. Köstler, and T. Wech, "Assessment of resolution and noise in magnetic resonance images reconstructed by data driven approaches", *Zeitschrift für Medizinische Physik*, in press, 2023, DOI: 10.1016/j.zemedi.2023.08.007

Conference Proceedings

1. J. Kleineisel, P. Eirich, J. F. Heidenreich, H. Köstler, T. A. Bley, and T. Wech, "Real-time cardiac MRI using spiral read-outs and a variational network for data-driven reconstruction," in *Proc. Intl. Soc. Mag. Reson. Med. 29*, 2021.

2. J. Kleineisel, H. Köstler, and T. Wech, "Reconstructing accelerated MR data using neural networks," presented at the *EUREKA! Symposium Würzburg*, 2022.

3. K. Lauer, J. Kleineisel, A. Borzì, T. A. Bley, and T. Wech, "Assessment of resolution and noise in MR images reconstructed by data driven approaches," in *Proc. Intl. Soc. Mag. Reson. Med. 30*, 2022.

4. J. Kleineisel, B. Petritsch, T. A. Bley, H. Köstler, and T. Wech, "Reconstruction of accelerated MR cholangiopancreatography using supervised and self-supervised 3d variational networks," in *Proc. Intl. Soc. Mag. Reson. Med. 31*, 2023.

5. T. Wech, O. Schad, and J. Kleineisel, "Physics-informed reconstruction of undersampled MR data using a reverse diffusion model trained with magnitude-only images," in *Proc. Intl. Soc. Mag. Reson. Med. 31*, 2023.

B Acronyms

CT computed tomography

MRI magnetic resonance imaging

MR magnetic resonance

PI parallel imaging

CS compressed sensing

DL deep learning

VN variational network

CNN convolutional neural network

GPU graphics processing unit

IRB institutional review board

FOV field of view

bSSFP balanced steady-state free precession

LRS low-rank plus sparse

ECG electrocardiogram

TR repetition time

TE echo time

DCF density compensation function

nuFFT non-uniform Fast Fourier Transform

SSIM structural similarity index measure

MSE mean-squared error

RMSE root mean-squared error

EDV end-diastolic volume

ESV end-systolic volume

SV stroke volume

EF ejection fraction

GRAPPA GeneRalized Autocalibrating Partial Parallel Acquisition

SNR signal-to-noise ratio

LPSF local point-spread function

ADMM alternating direction method of multipliers

SAX short axis

LAX long axis

MRCP magnetic resonance cholangiopancreatography

GROG grappa operator gridding

C Statement of contributions

Statement of individual author contributions and of legal second publication rights to manuscripts included in the dissertation

Manuscript 1: J. Kleineisel, J. F. Heidenreich, P. Eirich, N. Petri, H. Köstler, B. Petritsch, T. A. Bley, T. Wech, "Real-time cardiac MRI using an undersampled spiral k-space trajectory and a reconstruction based on a variational network", Magnetic Resonance in Medicine, vol. 88, no. 5, pp. 2167–2178, 2022								
Participated in	**Author Initials**, Responsibility decreasing from left to right							
Study Design, Methods Development	J.K.	T.W.	H.K.	P.E.				
Data Collection	T.W.	J.K.	J.H.	B.P.	N.P.	P.E.	T.B.	
Data Analysis and Interpretation	J.K.	J.H.	B.P.	T.W.	H.K.			
Manuscript Writing:	J.K.	T.W.	H.K.	J.H.	B.P.			
Introduction	J.K	T.W.	H.K.	J.H.	B.P.			
Materials & Methods	J.K	T.W.	H.K.	J.H.	B.P.			
Discussion	J.K	T.W.	H.K.	J.H.	B.P.			
First Draft	J.K.	T.W.						

Explanations (if applicable): -

Manuscript 2: J. Kleineisel, K. Lauer, A. Borzì, T. A. Bley, H. Köstler, T. Wech, "Assessment of resolution and noise in magnetic resonance images reconstructed by data driven approaches", Zeitschrift für Medizinische Physik, in press, 2023						
Participated in	**Author Initials**, Responsibility decreasing from left to right					
Study Design, Methods Development	J.K.	K.L.	T.W.	H.K.		
Data Collection	T.W.	J.K.	K.L.	H.K.	T.B.	
Data Analysis and Interpretation	J.K.	K.L.	T.W.	H.K.	A.B.	
Manuscript Writing:	J.K.	T.W.	H.K.	A.B.		
Introduction	J.K.	T.W.	H.K.	A.B.		
Materials & Methods	J.K.	T.W.	H.K.	A.B.		
Discussion	J.K.	T.W.	H.K.	A.B.		
First Draft	J.K.	T.W.				

Explanations (if applicable): -

If applicable, the doctoral researcher confirms that she/he has obtained permission from both the publishers (copyright) and the co-authors for legal second publication.

The doctoral researcher and the primary supervisor confirm the correctness of the above mentioned assessment.

Doctoral Researcher's Name Date Place Signature

Primary Supervisor's Name Date Place Signature

Statement of individual author contributions to figures/tables of manuscripts included in the dissertation

Manuscript 1: J. Kleineisel, J. F. Heidenreich, P. Eirich, N. Petri, H. Köstler, B. Petritsch, T. A. Bley, T. Wech, "Real-time cardiac MRI using an undersampled spiral k-space trajectory and a reconstruction based on a variational network", Magnetic Resonance in Medicine, vol. 88, no. 5, pp. 2167–2178, 2022								
Figure	**Author Initials**, Responsibility decreasing from left to right							
1	J.K.	P.E.	T.W.	H.K.				
2	J.K.							
3	J.K.							
4	J.K.							
5	J.K.							
6	J.K.							
7	J.K.							
8	J.K.							
9	J.K.	T.W.	J.H.	B.P.	H.K.			
10	J.K.	T.W.						
11	J.K.							
12	J.K.							
13	J.H.	B.P.	J.K.	T.W.	N.P.	T.B.	H.K.	
14	J.K.	T.W.	J.H.	B.P.	H.K.			
Table	**Author Initials**, Responsibility decreasing from left to right							
1	J.K.							
2	J.K.							
3	J.H.	B.P.	J.K.	T.W.				
4	J.K.							

Explanations (if applicable): -

Manuscript 2: J. Kleineisel, K. Lauer, A. Borzì, T. A. Bley, H. Köstler, T. Wech, "Assessment of resolution and noise in magnetic resonance images reconstructed by data driven approaches", Zeitschrift für Medizinische Physik, in press, 2023					
Figure	**Author Initials**, Responsibility decreasing from left to right				
15	J.K.				
16	J.K.				
17	J.K.	K.L.	T.W.		
18	J.K.	K.L.	T.W.		
19	J.K.	T.W.	H.K.		
20	J.K.	T.W.	H.K.	K.L.	
21	J.K.	T.W.	H.K.	K.L.	
22	J.K.				

Explanations (if applicable): -

I also confirm my primary supervisor's acceptance.

Doctoral Researcher's Name Date Place Signature

D Acknowledgments

First and foremost, I want to thank my dedicated and incredibly kind supervisor **Prof. Dr. Tobias Wech** without whom this thesis or any of its results would never have been possible. He generously invested his time and energy whenever I needed help of any kind, and taught me the great value of friendliness and respect through his example. I feel truly very lucky to have enjoyed such outstanding supervision and support during the entirety of my doctoral project.

This extends to **Prof. Dr. Herbert Köstler**, who was equally always willing to help whenever needed. He has been leading our research group for many years with experience and foresight, building a welcoming and creative environment along the way. For this environment, I am also grateful to all my past and present colleagues in the Department of Experimental Radiology, including **Dr. Andreas Weng, Dr. Anne Slawig, Katharina Rath, Dr. Johannes Portmann, Hannah Scholten, Philipp Nunn, Dr. Philipp Eirich,** and **Oliver Schad**, working with whom was a great joy every day. Furthermore, I am thankful to my clinical colleagues **Dr. Julius Heidenreich** and **PD Dr. Bernhard Petritsch**, without whom this thesis would probably only contain data from approximately two patients instead of several dozens.

I want to thank **Prof. Dr. Alfio Borzì** for his continued support over the years, without him I would have missed the opportunity that became this thesis in the first place.

Also, I want to thank the **German Ministry for Education and Research**, for funding me and the other doctoral researchers in the iDeLIVER project. To these, namely **Nadja Vater, Zijin Yang** and **Mathias Feinler**, I am grateful for the friendly discussions during our cooperation.

Additionally, I wish to thank the **Graduate School of Life Sciences**, for contributing to the successful completion of my thesis through a well-designed framework. Also, for all the pleasant annual meetings I am grateful to the members my thesis committee, of which **Prof. Dr. Michael Laßmann** and **Dr. Johannes Tran-Gia** are yet to be mentioned.

Finally, I want to express my gratitude to my parents **Petra** and **Ralf Kleineisel**, and my beautiful girlfriend **Daniela Schaub** for their continuing support.