Oskar Khayrullin

Variational Networks for MRI Reconstruction without Coil Sensitivities

MASTER'S THESIS

to achieve the university degree of
Diplom-Ingenieur

Master's degree programme
Computer Science

submitted to

Graz University of Technology

Supervisor

Univ.-Prof. Dr. Thomas Pock
Institute for Computer Graphics and Vision

Graz, Austria, Dec. 2020

Abstract

Magnetic resonance imaging is an important medical diagnostic tool. Unfortunately, due to physical limitations, this imaging technique is rather slow. In order to mitigate the disadvantage associated with slow scan times, accelerated data acquisition methods are utilized, resulting in a significant reduction in image quality. Therefore, over the years, much research has been aimed at improving the quality of MRI reconstruction, including a number of approaches that require the calculation of coil sensitivities. This thesis explores the potential of Fields-of-Experts-regularized variational networks for reconstruction of accelerated MRI data without explicit usage of coil sensitivities. First, multi-coil image-domain regularization approaches are discussed. Then, VN with fixed acquired data is derived, and methods of additional k-space regularization are proposed. We have trained a number of networks, and the results indicate that our models without coil sensitivities and with fewer parameters than the original VarNet by Hammernik et al. demonstrate comparable or even superior performance. In addition, the experiments show that the use of k-space regularization leads to an improvement in the quality of reconstruction.

Keywords. parallel MRI, image reconstruction, comressed sensing, variational network, deep learning

Kurzfassung

Die Magnetresonanztomographie ist ein wichtiges medizinisches Diagnoseinstrument. Leider ist dieses bildgebende Verfahren aufgrund physikalischer Einschränkungen ziemlich langsam. Um den mit langsamen Scanzeiten verbundenen Nachteil zu mildern, werden beschleunigte Datenerfassungsmethoden eingesetzt, was zu einer erheblichen Verringerung der Bildqualität führt. Daher wurde im Laufe der Jahre viel geforscht, um die Qualität der MRT-Rekonstruktion zu verbessern, einschließlich einer Reihe von Ansätzen, die die Berechnung von Spulenempfindlichkeiten erfordern. In dieser Arbeit wird das Potenzial von Fields-of-Experts-regulierten Variationsnetzwerken für die Rekonstruktion beschleunigter MRT-Daten ohne explizite Verwendung von Spulenempfindlichkeiten untersucht. Zunächst werden Multi-Coil-Bildbereichs-Regularisierungsansätze diskutiert. Dann wird Variationsnetzwerk mit fest erfassten Daten abgeleitet und es werden Methoden zur zusätzlichen k-Raum-Regularisierung vorgeschlagen. Eine Reihe von Netzwerken wurde trainiert und die Ergebnisse zeigen, dass unsere Modelle ohne Spulenempfindlichkeiten und mit weniger Parametern als das ursprüngliche VarNet von Hammernik u.a. eine vergleichbare oder sogar überlegene Leistung zeigen. Darüber hinaus zeigen die Experimente, dass die Verwendung der k-Raum-Regulierung zu einer Verbesserung der Qualität der Rekonstruktion führt.

Affidavit

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

The text document uploaded to TUGRAZonline is identical to the present master's thesis dissertation.

Date	Signature

Acknowledgments

I'm extremely grateful to Prof. Thomas Pock, who gave me the opportunity to work on this interesting topic, for his valuable advice, patient guidance, and for inspiring my interest in the field of deep-learning-based MRI reconstruction.

Special thanks to Kerstin Hammernik for helpful discussion and suggestions.

I very much appreciate the assistance and technical support provided by Erich Kobler. Besides, I would like to thank Patrick Knöbelreiter for introduction to PyTorch and supplying me with the implementation of the Adam algorithm, and Alexander Effland for explaining me the concept of TDV and his sincere willingness to help.

Also, thanks to my friend Leonid Margulis for the encouragement and translation of the abstract into German.

Finally, I must express my deepest gratitude to my parents for supporting me throughout my years of study.

Contents

List of Figures

List of Tables

1 Introduction

Contents

1.1 Parallel imaging

Magnetic Resonance Tomography (MRT/MRI) is increasingly popular, noninvasive technique of medical imaging. It works by detecting radio waves emitted by hydrogen atoms under excitation in a magnetic field. The frequency of emitted signal as well as resonance frequency depend on external field strength. This underlies the main technique of signal localization in MRI: computing the frequency components of the magnetic flux originating from tissue placed in a varying magnetic field. For this reason MRI scanners have a number of builtin coils: superconducting coil magnet for creating strong magnetic field, gradient coils for frequency and phase encodings, receiver coils for signal registration, etc.

Surface coils were introduced for the benefit of increasing Signal-to-Noise Ratio (SNR). Obviously, placing additional coils in the immediate vicinity of the inspected object improves detection of electromagnetic waves, which allows recording a signal of better quality. Early coil array designs were impaired by the issue of coil interactions. The solution to this problem was proposed in the fundamental work by Roemer et al., thereby allowing simultaneous multi-coil acquisition [1].

However, in contrast to full-body coil, due to the physical phenomenon of propagation of electromagnetic waves, surface coil sensitivities are highly non-uniform: detected signal strength depends on spatial location of the point emitting the signal. Therefore, individual coil images have different gradually changing intensity. Nevertheless, this property of surface coils has turned to be a valuable source of additional information in recovering the spatial location of the signal origin, besides standard frequency and phase encodings.

The main disadvantage of MRI compared to other modern medical imaging methods (in addition to high maintenance costs, e.g., expensive liquid helium needs to be topped up periodically) is long scan time. The possibility of multiple coil acquisition for accelerating MRI was theoretically proved by Kelton et al. [2] In general terms, acceleration is performed by reducing the number of phase encoding steps. The first practically applicable reconstruction methods for accelerated MRI were SMASH [3] and later SENSE [4]. Besides these two, other methods has since emerged.

In the following sections several reconstruction methods are presented, of which SENSE and GRAPPA [5] are the most common and commercially available.

1.2 SENSE

One of the main techniques in accelerated MRI reconstruction is SENSE (the name comes from SENSitivity Encoding), proposed by Pruessman et al. [4] In SENSE, data collected from a single coil is modeled by the equation

$$b_j = \mathrm{MFS}_j u + n_j, \tag{1.1}$$

where M is sampling operator, S_j is coil sensitivity, u is true image, and n_j is noise. Every voxel from the slice is encoded with coil sensitivity weight. Thus, SENSE poses MRI reconstruction as an inverse problem to the system

$$\mathbf{A}u = b \tag{1.2}$$

Here $\mathbf{A} \overset{\text{def}}{=} \mathrm{MFS}$ includes sensitivity weighting, Discrete Fourier Transform (DFT), and sampling operator with the reduction factor R. For the existence of unique solution, we additionally require that R cannot be greater than the number of coils. SENSE reconstruction then means finding the solution to the least squares problem

$$\min_u \frac{1}{2} \left\| \mathbf{A}u - b \right\|_2^2 \tag{1.3}$$

In the most simple form, i.e., with only quadratic terms, when the matrix $\mathbf{A}$ is relatively small (Cartesian sampling with 1D acceleration), and without taking noise into consideration, the solution can be found in a closed form:

$$u^* = \left(\mathbf{A}^{\mathrm{H}}\mathbf{A}\right)^{-1} \mathbf{A}^{\mathrm{H}} b, \tag{1.4}$$

where $\mathbf{A}^{\mathrm{H}}$ denotes conjugate transpose of the matrix $\mathbf{A}$. Fig. 1.1 depicts sample reconstruction using this scheme along with sensitivity-weighted reference combination.
In general, iterative schemes are preferred, e.g., variants of gradient descent, namely, Landweber iteration,

$$u^{t+1} = u^t - \alpha \mathbf{A}^{\mathrm{H}} \left(\mathbf{A} u^t - b \right), \tag{1.5}$$

or conjugate gradient algorithm (CG SENSE [6]).

SENSE is very flexible, works with arbitrary coil configurations, different sampling trajectories, but requires knowledge of the magnetic field of each coil. Thus, one must first solve the sensitivity estimation problem with adequate accuracy, because even small errors in coil sensitivity estimates affect reconstruction quality leaving artifacts. The original paper describes the method for coil sensitivity estimation. Other algorithms have since been developed, for example ESPIRiT [7].

Figure 1.1: An example of SENSE reconstruction with acceleration factor $R = 4$: a) exact solution, Eq. (1.4), b) CG-SENSE, c) fully sampled reference.

1.3 GRAPPA

Another widely accepted method of parallel imaging techniques is GRAPPA. It is similar to AUTO-SMASH [8], VD-AUTO-SMASH [9], but in contrast to these methods, GRAPPA recovers nonacquired data for all coils.

Unlike SENSE, GRAPPA is a direct coil-by-coil k-space domain reconstruction method. In essence, GRAPPA is based on the assumption that the following relation holds:

$$f_j = \sum_{c \in Coils} f_c * g_{jc}, \tag{1.6}$$

where f is a k-space, and g is a shift-invariant convolutional kernel. In other words, each point in k-space can be approximated by a weighted sum of neighboring points from all

coils. This means that the reconstruction of missing data x of coil j at position r is merely a linear combination:

$$x_{r,j} = \sum_{c \in Coils} \mathrm{M}_r\big(f_c * g_{r,jc}\big), \tag{1.7}$$

where M_r is a sampling operator, which selects data at position r, and g_r is a set of weights for approximating data at position r.

As stated above, it is assumed that g is shift-invariant. Therefore, kernel g_r can be approximated from the set of k-space patches where x_r and its neighboring points are known, Auto-Calibration Signal area (ACS):

$$\tilde{g}_{r,j} = \arg\min_{g_{r,j}} \frac{1}{2} \sum_{p \in ACS} \left| f_{p,j} - \sum_{c \in Coils} \mathrm{M}_p\big(f_c * g_{r,jc}\big) \right|^2 \tag{1.8}$$

In matrix form, where A_r is a matrix containing neighboring points in rows, b_r is a vector of $f_{p,j}$, and $g_{r,j}$ rearranged into a vector, we have

$$\tilde{g}_{r,j} = \arg\min_{g_{r,j}} \frac{1}{2} \left\| \mathrm{A}_r g_{r,j} - b_{r,j} \right\|_2^2 \tag{1.9}$$

Additionally, in order to reduce noise in the reconstruction, the Tykhonov regularization term can be added. The problem has analytical solution:

$$\tilde{g}_{r,j} = \left(\mathrm{A}_r^{\mathrm{H}} \mathrm{A}_r + \lambda \mathrm{Id}\right)^{-1} \mathrm{A}_r^{\mathrm{H}} b_{r,j} \tag{1.10}$$

The parameter λ acts as a trade-off between noise and aliasing artifacts.

Of course, reconstruction in frequency domain by the linear interpolation of neighboring frequency values might look counter-intuitive. Despite this, GRAPPA is stable and produces reconstructions of acceptable quality. However, images are noisy at high acceleration factors, which becomes an obstacle to practical use (see Fig. 1.2).

1.4 SPIRiT

In the SPIRiT algorithm developed by Lustig and Pauly the solution to the MRI reconstruction problem is constraint with calibration consistency [10]. The algorithm utilizes the basic GRAPPA assumption, Eq. (1.6) in matrix form,

$$\mathrm{G} f = f,$$

but posing the reconstruction as an inverse problem. This makes SPIRiT algorithm more general comparing to GRAPPA as it effectively allows reconstruction from arbitrary sampling patterns. Reconstruction of the entire k-space is obtained as a solution to the

optimization problem:

$$\min_x \frac{1}{2}\left\|\mathrm{G}x - x\right\|_2^2 + \frac{\lambda}{2}\left\|\mathrm{M}x - b\right\|_2^2 \tag{1.11}$$

Alternatively, the acquired data can be made immutable:

$$\min_x \frac{1}{2}\left\|(\mathrm{G}x - x)(\mathrm{M}_C x + \mathrm{M}b)\right\|_2^2, \tag{1.12}$$
$$\text{then } u = \mathcal{F}^{-1}[\mathrm{M}_C x + \mathrm{M}b]$$

Both problems, Eq. (1.11) and Eq. (1.12), can be solved directly with a linear solver (the resulting matrices are sparse) or using an iterative approach, e.g., conjugate gradient method. As with GRAPPA, SPIRiT requires a fully sampled ACS area to compute a calibration kernel, which is done in a similar manner.

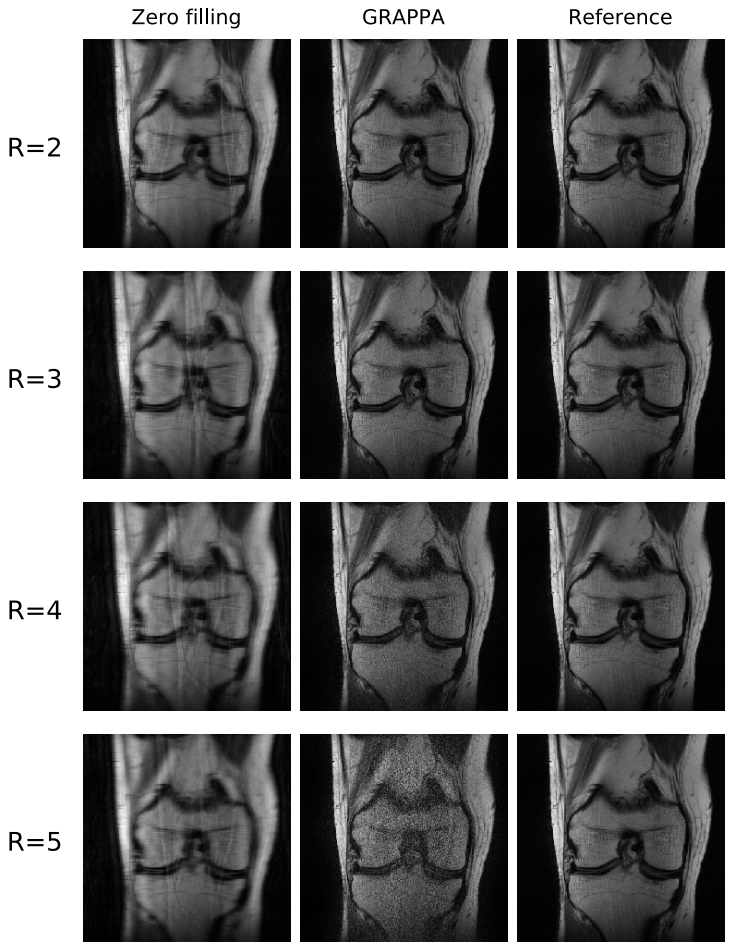


Figure 1.2: Noise amplification in GRAPPA reconstructions at different acceleration factors R.

1.5 Noise

According to Prussman et al., the upper bound of Signal-to-Noise Ratio (SNR) in reconstructed images from accelerated data is given by

$$\mathrm{SNR}_{reduced} = \frac{\mathrm{SNR}_{full}}{g_\rho \sqrt{R}}, \tag{1.13}$$

where $g_\rho \geq 0$ is the coil configuration geometry factor [4].

Roemer et al. compare few methods for combining coil images. Optimal SNR requires knowledge of absolute coil sensitivities [1].

It should also be noted, that spacial distribution of noise is non-uniform, which is clearly seen on the reconstructed images (see Fig. 1.3).

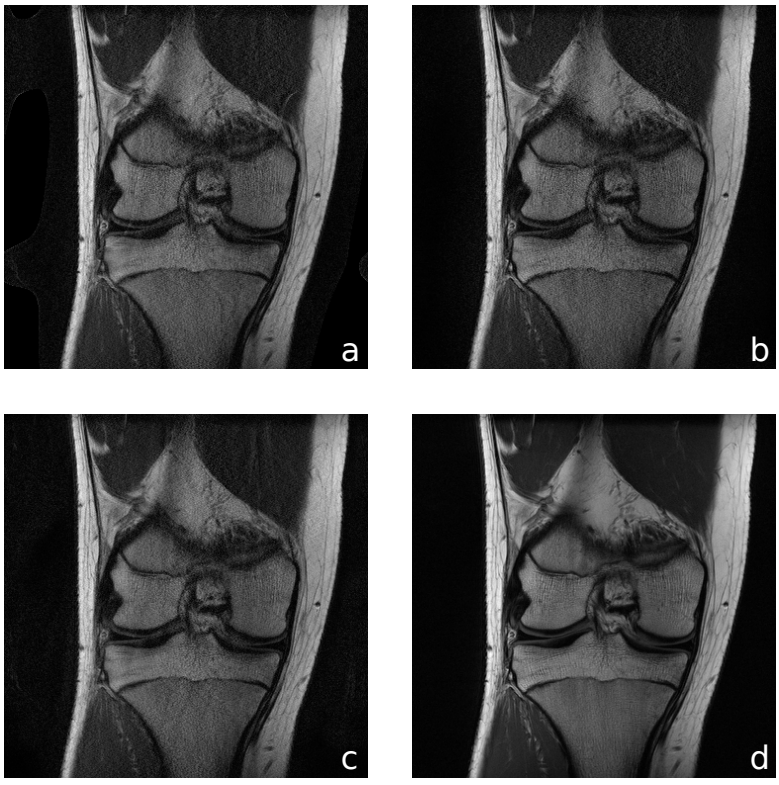


Figure 1.3: Comparison of reconstruction methods ($R = 4$): a) CG-SENSE, b) GRAPPA, c) SPIRiT, d) fully sampled reference.

1.6 More advanced methods

Following the success of GRAPPA, a number of methods have been proposed in order to improve the reconstruction quality of the algorithm. One of these is Nonlinear GRAPPA by Chang et al. [11] This approach utilizes the kernel method, i.e., the identity Eq. (1.6) is modified as follows:

$$b \;=\; \Phi(\mathrm{A})\, g,$$

where $\Phi : \mathbb{C}^{M\times K} \mapsto \mathbb{C}^{M\times N_K}$, $N_K \gg K$, is a nonlinear function that maps k-space data (here in the form of matrix A) into higher dimensional feature space.
The coefficients g can be computed similarly to Eq. (1.10):

$$\tilde{g} = \Big(\Phi(\mathrm{A})^{\mathrm{H}}\Phi(\mathrm{A})\Big)^{-1}\Phi(\mathrm{A})^{\mathrm{H}} b$$

For the kernel $\Phi(\cdot)$ the authors used polynomial function. Additionally, the size of feature space N_K should be chosen with care.

Complex nonlinear relations between neighboring points in k-space suggests using Neural Networks (NN) for this task. Akçakaya et al. proposed Robust Artificial-neural-networks for K-space Interpolation reconstruction (RAKI) [12]. It can be considered an advanced nonlinear NN-based variant of GRAPPA, where missing data is synthesized from acquired data using deep neural network:

$$f_j \;=\; \Phi(f;\, \theta_j) \tag{1.14}$$

The parameters θ_j of the network Φ are learned from the ACS area independly for every coil:

$$\tilde{\theta}_j \;=\; \arg\min_{\theta_j} \; \big\| f_{ACS} - \Phi(f_{ACS};\, \theta_j) \big\|_2^2$$

Then, reconstruction in k-space is computed using Eq. (1.14).

RAKI does not require large training datasets for learning, or, in other words, it is scan-specific. Moreover, noise amplification in reconstructions is greatly reduced compared to original GRAPPA especially at high acceleration rates. These nice properties make RAKI very attractive.

Another interesting method is AUTOMAP by Zhu et al. [13], a unified framework for image reconstruction, which models direct mapping from sensor representation (k-space) into image domain. Theoretically this method is interpreted as learning a robust low-dimensional joint manifold $\mathcal{M} = \mathcal{X} \times \mathcal{Y}$, where $\mathcal{X}$ is a manifold of k-space data, $\mathcal{Y}$ is a manifold of output images, and mapping functions ϕ_x and ϕ_y, as well as a between-manifold projection $g : \mathcal{X} \mapsto \mathcal{Y}$. The reconstruction process can be written as a function $f(x) \;=\; \phi_y \circ g \circ \phi_x^{-1}(x)$. This mapping is implemented with deep neural network.

AUTOMAP is robust against noise and other input perturbations, such as calibration errors [13]. However, this "brute force" approach to learned MRI reconstruction is inefficient due to a very large number of network parameters and, as a result, high memory usage and computing resources, which significantly limits image resolution [14].

Few recently emerged reconstruction methods based on iterative/cascading schemes are described at the end of Chapter 2.

2

MRI Reconstruction as an Optimization Problem

Contents

Image reconstruction tasks are often represented as problems of minimization of the energy functional in the following form:

$$\mathcal{E}(x, y) = \mathcal{D}(x, y) + \mathcal{R}(x)$$

Here $\mathcal{D}(x, y)$ is called *data fidelity*, and $\mathcal{R}(x)$ is *regularization*. This functional is closely related to the famous Bayes formula

$$p(x|y) = \frac{\overbrace{p(y|x)}^{likelyhood}\overbrace{p(x)}^{prior}}{p(y)},$$

where $p(x|y)$ denotes the probability of getting true image x given measurement y.

In situations of uncertainty, which in the context of accelerated MRI is mostly produced by undersampling and noise, we want to maximize the probability of obtaining the true image x, hence, we need to solve

$$\min_x \mathcal{E}(x, y)$$

A common choice for the data fidelity term is the sum of squared errors,

$$\mathcal{D}(x, y) = \left\|\mathbf{A}x - y\right\|_2^2$$

with **A** being a task-specific operator transforming x into data measurement domain. The things are more complicated with the regularizer, or prior, and much of research work has been done mainly in this direction. The theory of compressed sensing is of particular interest.

2.1 Compressed sensing

Compressed Sensing (CS) is a technique for sampling and reconstructing compressible signal. The main goal of the theory is to reduce the number of measurements (sampling rate) while maintaining the reconstruction accuracy. For instance, this directly means accelerated acquisition in MRI, or reduced exposure to ionizing radiation in CT.

CS can only be applied in the context of compressible data, for example, when signal is sparse. By exploiting the a-proiri knowledge of sparsity, it becomes possible to reconstruct highly undersampled signal with sufficient accuracy, in some cases exactly. If signal is not sparse, very often the amount of useful information, the Kolmogorov complexity, is much less than the amount of data collected at a measurement rate that satisfies the Nyquist-Shannon sampling theorem. This implies that at least in theory there must be a way to compress the signal, or, alternatively, sparsify it. Such principle of *transform sparsity* underlies regularization models used in CS framework.

A prominent work by Candès, Romberg, and Tao [15] demonstrates exact reconstruction of a signal from small set of frequencies. In the example application, they use total variation as complexity measure. Another paper with important theoretical results for the theory of CS was published by Donoho [16].

2.2 Regularization

One of the most important regularization models is total variation. In image processing Total Variation (TV) was first introduced in 1992 by Rudin, Osher, and Fatemi originally for image denoising (ROF model) [17]. Since then it has been adopted into many image processing applications including MRI reconstruction.

TV arises from the assumption that natural images are piecewise constant, that is, image gradients are sparse. TV can also be viewed as a special case of the regularizer

$$\mathcal{R}(u) = \int_\Omega \|\nabla u\|^p dx \quad \text{for } p = 1. \tag{2.1}$$

In discrete setting, TV is defined as

$$\mathrm{TV}(x) \overset{\text{def}}{=} \sum_{i=1}^{N} \sqrt{|(\mathrm{D}x)_{i,1}|^2 + |(\mathrm{D}x)_{i,2}|^2}, \tag{2.2}$$

where the linear operator D maps to discrete gradients.

By augmenting the SENSE model, Eq. (1.3), with the TV regularization, we obtain the following optimization problem:

$$u^* = \arg\min_u \ \mathrm{TV}(u) + \frac{\lambda}{2}\|\mathbf{A}u - b\|_2^2 \tag{2.3}$$

Due to the historical significance, I would like to pay special attention to the TV-regularized SENSE reconstruction, though without going into details.

The functional in Eq. (2.3) poses issues for optimization because of non-smoothness of the regularization term. Nevertheless, a number of algorithms have been proposed. The most efficient first-order methods to date work by solving the problem in primal-dual form. In the algorithm developed by Zhu and Chan [18], the ℓ_2-norm is replaced with the dual norm, thereby introducing the dual variable x:

$$\begin{gathered}\min_u \max_{x \in X} \langle \mathrm{D}u,\, x \rangle + \frac{\lambda}{2}\|\mathbf{A}u - b\|_2^2 \\ \text{where } X = \{\, x : \|x_i\|_2 \le 1,\ \ 1 \le i \le N \,\}\end{gathered} \tag{2.4}$$

The algorithm builds a sequence of both variables simultaneously, i.e., performing projected gradient ascent for the dual variable x and calculates the minimum for the primal variable u at every stage. The solution lies at the saddle point. But since original primal update is inefficient in our case, one of the simpler alternatives is to replace it with gradient descent. Below is a summary of an accelerated variant of the primal-dual algorithm by Chambolle and Pock [19] with an explicit gradient step on the primal variable u.

Choose $\tau > 0,\ \sigma > 0$

Initialize u^0(*e.g., zero filling*), $x^0 \leftarrow 0,\ \tilde{u}^0 \leftarrow u^0,\ k \leftarrow 0$

Repeat:

$x^{k+1} \leftarrow \mathcal{P}_X(x^k + \sigma \mathrm{D}\tilde{u}^k)$

$u^{k+1} \leftarrow u^k - \tau(\mathrm{D}^{\mathrm{T}} x^{k+1} + \lambda \mathbf{A}^{\mathrm{H}}(\mathbf{A}u^k - b))$

$\tilde{u}^{k+1} \leftarrow 2u^{k+1} - u^k$

$k \leftarrow k + 1$

The operator $\mathcal{P}_X$ denotes the projection of subvectors $x_i \in \mathbb{C}^2$ onto unit Euclidean disks, which can be easily calculated:

$$\mathcal{P}_X(x) \Leftrightarrow \frac{x_i}{\max\{1, \|x_i\|_2\}},\ 1 \le i \le N$$

Fig. 2.1 shows three reconstructions of the same slice with different data fidelity parameters. The solutions were obtained by calculating 1000 primal-dual iterations according to the scheme indicated above.

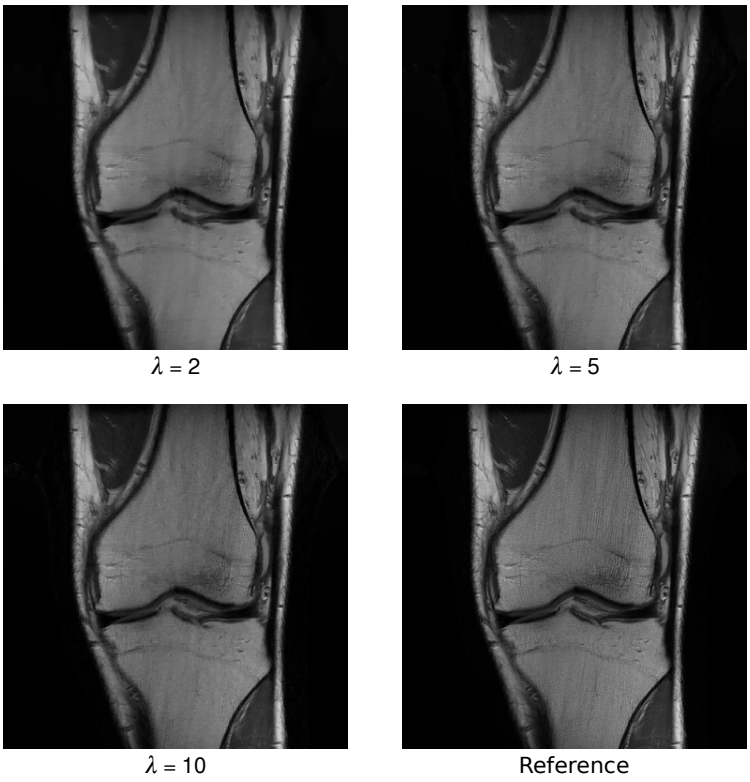


Figure 2.1: TV-regularized SENSE reconstruction with three different data fidelity parameters $\lambda = \{2, 5, 10\}$.

The main disadvantage of the TV regularization is the so-called staircasing effect. More advanced second order Total Generalized Variation (TGV) overcomes the issue of staircasing by including higher-order derivatives [20]. Knoll et al. have used TGV for MRI denoising and radial sampling SENSE reconstruction [21]. In their work, the reconstruction model is defined as follows:

$$\min_{u,\,v} \max_{x\in X,\,y\in Y,\,z} \langle \nabla u - v,\, x\rangle + \langle \nabla v,\, y\rangle + \langle \mathbf{A}u - b,\, z\rangle - \frac{1}{2\lambda}\|z\|_2^2 \tag{2.5}$$

Note that $\max_z \langle \mathbf{A}u - b, z\rangle - \frac{1}{2\lambda}\|z\|_2^2 = \frac{\lambda}{2}\|\mathbf{A}u - b\|_2^2$, i.e., the data fidelity term is also dualized, a trick that can be applied to Eq. (2.4) too. An example of TGV-regularized SENSE reconstruction is shown in Fig. 2.2.

However, the structural complexity of natural images has triggered the development of learned priors with probability distribution taken directly from the statistics.

Fields of Experts (FoE) is a framework for modeling image priors, which has drawn attention in computer vision research community due to its elegant representation and effectiveness. It is based on powerful sparse coding methods combined with Markov Random Field (MRF) models [22].
FoE models the image probability density as

$$p(x) \;=\; \frac{1}{Z(\Theta)} \prod_{k} \prod_{j=1}^{N} \phi_j(\mathbf{J}_j^{\mathrm{T}} x_{(k)}; \alpha_j)$$

Hence, the FoE regularizer in the energy functional has the following form:

$$\mathcal{R}(x) \;=\; -\sum_{k} \sum_{j=1}^{N} \log \phi_j(\mathbf{J}_j^{\mathrm{T}} x_{(k)}; \alpha_j),$$

$$\text{where experts } \; \phi_j(\mathbf{J}_j^{\mathrm{T}} x_{(k)}; \alpha_j) \;=\; \left(1 \,+\, \frac{1}{2}\big(\mathbf{J}_j^{\mathrm{T}} x_{(k)}\big)^2\right)^{-\alpha_j}$$

The linear filters $\mathbf{J}_j$ as well as the parameters $\alpha_j > 0$ are learned from data. Once trained, the regularizer is included into the energy functional, and solution is computed as regular.

More advanced learned priors have been developed. For example, one of the recent proposals is Total Deep Variation (TDV) [23], which is inspired by deep multiscale Convolutional Neural Networks (CNN) with residual connections. TDV has relatively low number of parameters, however, the models regularized with TDV achieve state-of-the-art performance. Furthermore, there is an ongoing research work on applying TDV to MRI reconstruction with promising results.

Figure 2.2: TV and TGV regularization in SENSE reconstruction with data fidelity parameter $\lambda = 5$: a) TV-SENSE, b) TGV-SENSE, c) fully sampled reference.

2.3 Variational networks

As shown by Barbu in [24], instead of learning the prior independently of the energy functional, training the FoE regularizer together with the inference algorithm yields considerable gains in performance. The inference algorithm there was obtained by unrolling first $N = \{1, 2, 3, 4\}$ iterations of the gradient descent scheme and trained through optimization of the loss function. Due to some reasons, the author used slow coordinate ascent method for optimization. To overcome this issue, "backpropagating" versions of such truncated iterative schemes were proposed by Domke in [25], where the gradients are computed according to the chain rule. Later, Chen et al. have shown that using different parameters at each stage further increases performance [26]. A similar concept, Variational Networks (VN) inspired by the proximal gradient method, were introduced in [27]. Effland et al. follow optimal control approach and learn optimal stopping time [28].

Of course, such learned iterative reconstruction schemes can be successfully used in different tasks involving functional minimization, e.g., denoising, deblurring, superresolution, inpainting, CT/MRI reconstruction, etc.

2.4 Variational networks for MRI reconstruction

The model by Hammernik et al. for MRI reconstruction originates from FoE-regularized SENSE method by unrolling first few steps of gradient descent [29].
Briefly, given the optimization problem

$$\min_u \ \mathcal{R}(u) \ + \ \frac{\lambda}{2}\big\|\mathbf{A}u - b\big\|_2^2, \tag{2.6}$$

where

$$\mathcal{R}(x) \ = \ \sum_k \sum_{j=1}^{N_K} \phi_j(\mathbf{K}_j u_{(k)}), \tag{2.7}$$

and $\mathbf{A} \overset{\text{def}}{=}$ MFS (see Section 1.2 for details).
Deriving the iteration step yields

$$u^{t+1} \ = \ u^t \ - \ \alpha^t \left(\sum_{j=1}^{N_K} \mathbf{K}_j^{\mathrm{T}} \phi_j'(\mathbf{K}_j u^t) \ + \ \lambda \mathbf{A}^{\mathrm{H}}\big(\mathbf{A}u^t - b\big) \right) \tag{2.8}$$

Then, the authors follow the ideas from [26] by using iteration-dependent parameters and more parametrized activation functions. Finally, they obtain

$$u^{t+1} \ = \ u^t \ - \ \sum_{j=1}^{N_K} \big[\mathbf{K}_j^t\big]^{\mathrm{T}} \phi_j^{t\prime}(\mathbf{K}_j^t u^t) \ - \ \lambda^t \mathbf{A}^{\mathrm{H}}\big(\mathbf{A}u^t - b\big), \quad 0 \le t \le T-1, \tag{2.9}$$

with activation functions

$$\phi_j^{t\prime}(x) \;=\; \sum_{k=1}^{N_W} w_{jk}^t \exp\left(-\frac{(x-\mu_k)^2}{\sigma^2}\right), \tag{2.10}$$

a weighed sum of Gaussian Radial Basis Functions (RBFs).

Let $f(u^0, s, b, m;\, \theta)$ denote the variational network with parameters θ fed with initial reconstruction u^0, coil sensitivity maps s, undersampled k-space data b, and sampling mask m. Parameters of the VN include filter kernels of convolutional operators $\mathbf{K}^t$, activation function weights w^t, and data fidelity coefficients λ^t. Training the VN is done by minimizing Mean-Squared-Error (MSE) loss function:

$$\min_{\theta} \left\{ \mathcal{L}(\theta) \stackrel{\text{def}}{=} \frac{1}{2|\mathcal{Z}|} \sum_{(u^0,s,b,m,u_{ref})\,\in\,\mathcal{Z}} \Big\| |f(u^0, s, b, m;\, \theta)|_\epsilon \;-\; |u_{ref}| \Big\|_2^2 \right\} \tag{2.11}$$

To ensure differentiability of the loss $\mathcal{L}(\theta)$, the absolute value of reconstruction is computed using the ϵ-smoothed norm, that is

$$|x|_\epsilon \stackrel{\text{def}}{=} \sqrt{\mathrm{Re}^2(x) + \mathrm{Im}^2(x) + \epsilon} \tag{2.12}$$

Additionally, the kernels are required to be zero-mean unit vectors, and the data fidelity coefficients are non-negative.

In general, trained VN demonstrates good reconstruction performance. Authors admit presence of aliasing artifacts on some slices, but these are challenging to remove, particularly, when SNR is low. In their experiments, the network was fed with zero-filling reconstructions.

2.5 Recent developments

More advanced forms of learnable regularization can be used instead of FoE, i.e., arbitrarily deep neural networks. For example, Aggarwal et al. in their work introduce the MoDL framework (MOdel-based reconstruction using Deep Learned priors) [30]. This approach is similar to the variational network described above but with some key differences. Particularly, it employs CNN as an artifact estimator:

$$\begin{aligned} &\min_u \; \frac{1}{2}\big\|\mathcal{N}(u)\big\|_2^2 \;+\; \frac{\lambda}{2}\big\|\mathbf{A}u - b\big\|_2^2, \\ &\text{where} \quad \mathcal{N}(u) \;=\; u \;-\; \mathrm{CNN}(u;\theta) \end{aligned} \tag{2.13}$$

Another major difference is the use of blocks of conjugate gradient algorithm within the network. These CG blocks have no trainable parameters, while the parameters of the CNN as well as the trade-off λ are shared across iteration steps.

The algorithm is obtained as follows. First, the nonlinear functional Eq. (2.13) is linearized at point u^t, which leads to the following optimization problem:

$$u^{t+1} \;=\; \arg\min_u \; \frac{1}{2}\big\|u - \mathrm{CNN}(u^t;\theta)\big\|_2^2 \;+\; \frac{\lambda}{2}\big\|\mathbf{A}u - b\big\|_2^2 \tag{2.14}$$

Then, the analytical solution yields the iteration scheme of the MoDL framework:

$$u^{t+1} \;=\; \big(\mathbf{I} + \lambda\mathbf{A}^{\mathrm{H}}\mathbf{A}\big)^{-1}\big(\mathrm{CNN}(u^t;\theta) + \lambda\mathbf{A}^{\mathrm{H}}b\big) \tag{2.15}$$

However, u^{t+1} is approximated in the CG sub-blocks because the exact solution above is impossible in general case.

The network was trained using MSE loss and demonstrates state-of-the-art performance.

Sriram et al. further develop the idea of variational networks [31]. Similarly, the iteration step originates from the gradient descent scheme, where intermediate values are expressed in k-space, and FoE-based regularization is replaced with CNN:

$$x^{t+1} \;=\; x^t \;-\; \underbrace{\mathrm{FS}\Big[\mathrm{CNN}(\mathrm{S}^{\mathrm{H}}\mathrm{F}^{-1}x^t;\,\theta^t)\Big]}_{\nabla\mathcal{R}} - \lambda^t\mathrm{M}(x^t - b), \tag{2.16}$$

Here the parameters are not shared across iterations. As before, this network requires sensitivity maps at every stage (cascade), and the most innovative feature is that they are obtained by evaluating additional Sensitivity Map Estimation (SME) module, which precedes image reconstruction cascades. The core of SME module is the CNN with the same structure as in cascades but with fewer parameters. Therefore, the algorithm first estimates sensitivity maps from k-space center, then the data is fed to the main part, Eq. (2.16). The whole network including SME module is trained end-to-end by maximizing Structured Similarity Index Measure (SSIM [32]).

The network with 12 cascades has relatively large number of parameters (about 30M). Nevertheless, the model implements stand-alone algorithm and produces reconstructions of very high quality.

At present, iterative schemes remain popular, and MRI researchers try different network architectures sometimes incorporating few networks into one framework, for example, Σ-net [33]. Other model may benefit from including additional k-space regularization for data consistency [34]. By the way, these models were developed for fastMRI challenge[1], which thus promotes MRI research.

[1] `https://fastmri.org`

Additionally, many modern MRI reconstruction models use integrated CNNs stemming from successful U-Net architecture [35]. The network was originally designed for image segmentation and won the ISBI cell tracking challenge[2] 2015. Another feature of U-Net is speed due to its relative simplicity. All these nice properties determined its popularity among the researchers. For instance, the recently proposed TDV regularization briefly mentioned in Section 2.2 is also inspired by U-Net [23]. It even got to the point that sometimes U-Net is applied directly to the undersampled k-space data [36].

[2] http://celltrackingchallenge.net

3

Variational Networks for MRI Reconstruction without Coil Sensitivities

Contents

The method described in Section 2.4 requires precomputation of coil sensitivity estimates, which are fed to the VN as a part of input. My work on this thesis was to investigate the performance of similar variational networks but without explicit usage of coil sensitivities.

First step is to remove the sensitivity weighting operator from the data fidelity term, i.e., $\mathbf{A} \overset{\text{def}}{=} \mathrm{MF}$, hence, $\mathbf{A}^{\mathrm{H}} = \mathrm{F}^{-1}\mathrm{M}$. Note that now the latter does not combine coil images. This raises the first question: how to apply the regularization? The following approaches can be tried:

- applying the regularizer on a combined image;
- applying the multi-channel regularizer.

The first option is very attractive due to its computational effectiveness, whereas the multi-channel model can be used to learn correlations between coil images.

Before proceeding further, we define the ϵ-smoothed sum-of-squares-combined image similar to the absolute value, Eq. (2.12):

$$\|x\|_\epsilon \overset{\text{def}}{=} \sqrt{\sum_{k=1}^{N_C} x_k \cdot \operatorname{conj}(x_k) + \epsilon}\,, \qquad \begin{pmatrix} x_1 \\ \vdots \\ x_{N_C} \end{pmatrix} \in \mathbb{C}^{N_C HW}, \quad \|\cdot\|_\epsilon : \mathbb{C}^{N_C HW} \mapsto \mathbb{R}^{HW}$$

3.1 Regularization in image domain

3.1.1 Approach 1: sum-of-squares-combined image

Here we consider the following optimization problem:

$$\min_u \; \mathcal{R}\big(\|u\|_\epsilon\big) \; + \; \frac{\lambda}{2}\big\|\mathbf{A}u - b\big\|_2^2 \tag{3.1}$$

The iteration step for the VN is

$$u^{t+1} \; = \; u^t \; - \; \hat{U}^t \operatorname{diag}^{-1}\big(\|u^t\|_\epsilon\big) \sum_{j=1}^{N_K} \big[\mathbf{K}_j^t\big]^{\mathrm{T}} \phi_j^{t\prime}\big(\mathbf{K}_j^t \|u^t\|_\epsilon\big) \; - \; \lambda^t \mathbf{A}^{\mathrm{H}}\big(\mathbf{A}u^t - b\big), \tag{3.2}$$

$$\text{where} \quad \hat{U}^t \; = \; \big[\operatorname{diag}(u_1^t), \, \ldots \, , \operatorname{diag}(u_{N_C}^t)\big]^{\mathrm{T}} \in \mathbb{C}^{N_C HW}, \quad 0 \le t \le T-1$$

This network, besides the performance advantage, has only a real set of filter kernels. Again, the kernels are constraint to be zero-mean unit vectors, and $\lambda^t \ge 0$.

3.1.2 Approach 2: multi-channel filters

This approach has more options for handling complex numbers. For example, one way is applying convolutional operator similar to the model from Hammernik et al. [29], i.e., taking the sum of real and imaginary components of the convolutions:

$$\mathbf{K}^t u \; \Leftrightarrow \; \operatorname{Re}(u) * \mathbf{k}_{\text{re}}^t \; + \; \operatorname{Im}(u) * \mathbf{k}_{\text{im}}^t$$

Another choice is using true complex convolutional operators, then applying activation functions on components:

$$\nabla_u \mathcal{R}(u^t; \theta^t) \; = \; \sum_{j=1}^{N_K} \big[\mathbf{K}_j^t\big]^{\mathrm{H}} \phi_j^{t\prime}(\mathbf{K}_j^t u^t),$$

where

$$\mathbf{K}^t u \; \Leftrightarrow \; \big(\operatorname{Re}(u) * \mathbf{k}_{\text{re}}^t \; - \; \operatorname{Im}(u) * \mathbf{k}_{\text{im}}^t\big) \; + \; i\,\big(\operatorname{Re}(u) * \mathbf{k}_{\text{im}}^t \; + \; \operatorname{Im}(u) * \mathbf{k}_{\text{re}}^t\big) \tag{3.3}$$

Additionally, instead of computing real-valued activations, we can use holomorphic functions to make the regularizer complex-differentiable. Other combinations are also possible. Certainly, these details significantly increase computational costs. Training is expected to be several times slower, while reconstruction accuracy improvement is not guaranteed. All these strategies of processing complex numbers in the regularizer are theoretically unclear, thus, the optimal configuration, if exists, can only be inferred empirically.

In general, phase information is irrelevant for image-domain regularization. Therefore, it can be applied on magnitude images, which eliminates the issue of complex numbers. Then in multi-channel regularization context the model is defined as

$$\min_u \ \mathcal{R}\big(|u|_\epsilon\big) \ + \ \frac{\lambda}{2}\big\|\mathbf{A}u - b\big\|_2^2 \tag{3.4}$$

For the VN we get

$$u^{t+1} \ = \ u^t \ - \ \operatorname{diag}\big(u^t\big)\operatorname{diag}^{-1}\big(|u^t|_\epsilon\big)\sum_{j=1}^{N_K}\big[\mathbf{K}_j^t\big]^{\mathrm{T}}\phi_j^{t\prime}\big(\mathbf{K}_j^t|u^t|_\epsilon\big) \ - \ \lambda^t\mathbf{A}^{\mathrm{H}}\big(\mathbf{A}u^t - b\big) \tag{3.5}$$

To sum up, these networks are fed with initial reconstruction u^0, which is a set of coil images, undersampled k-space, and sampling mask. The network outputs uncombined reconstruction. Final reconstruction in obtained by computing sum-of-squares combination (see Fig. 3.1).

Additionally, taking into account that data has different intensity per coil, we can distribute data fidelity trade-off parameter over coils:

$$\mathcal{D}(u) \ \stackrel{\text{def}}{=} \ \sum_{k=1}^{N_C}\frac{\lambda_k}{2}\big\|\mathbf{A}u_k - b_k\big\|_2^2$$

However, in all these models the data fidelity term does not encode gradient as in SENSE and rather acts as a soft constraint.

3.1.3 Recent work

One of the newest methods that does not require coil sensitivities, DeepcomplexMRI [37], uses CNN with multi-channel complex convolutional operators as in Eq. (3.3) and Rectified Linear Unit (ReLU) for the activation function. The results indicate that utilization of complex convolution improves visual quality of the reconstructions. Also, the authors state that DeepcomplexMRI with even half of the parameters of the real-valued network demonstrates comparable performance.

Bian et al. solve the problem of multiple coil images in a different way. They propose to learn the nonlinear operator $\mathcal{J} : \mathbb{C}^{N_C HW} \mapsto \mathbb{C}^{HW}$ that combines a set of coil images into a single complex image with homogeneous contrast across the FoV [38]. The operator $\mathcal{J}$ is implemented as two identical CNNs, one for real and the other for imaginary parts. Thus, the components are handled independently. The authors have obtained excellent results and claim that their method is the first *combine-then-regularize* approach to deep-learning-based MRI reconstruction.

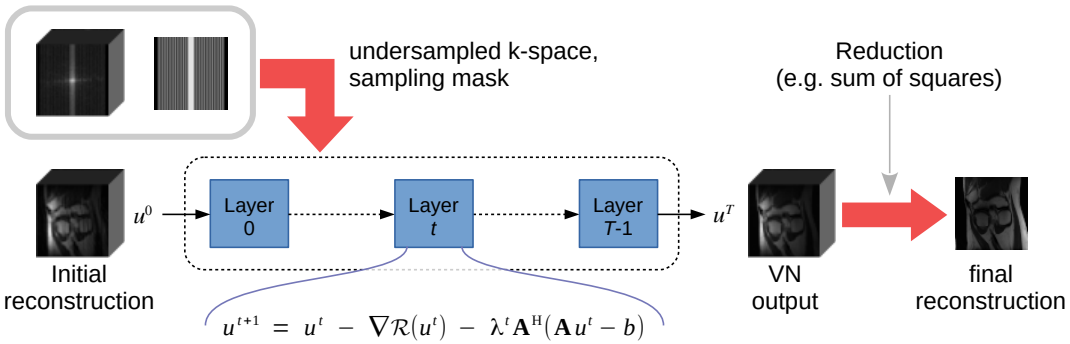


Figure 3.1: Reconstruction algorithm

3.2 Reconstruction in k-space

The reconstruction problem can also be formulated in Fourier domain:

$$\min_x \mathcal{R}\big(\underbrace{\mathrm{F}^{-1}x}_{u}\big) + \frac{\lambda}{2}\|\mathrm{M}x - b\|_2^2 \tag{3.6}$$

Although this representation does not give us any advantage over image-domain reconstruction, we can pose another problem by enforcing the data fidelity constraint. The following identity holds:

$$\min_x \left\{ \lim_{\lambda\to\infty} \mathcal{R}\big(\mathrm{F}^{-1}x\big) + \frac{\lambda}{2}\|\mathrm{M}x - b\|_2^2 \right\} = \min_x \mathcal{R}\big(\underbrace{\mathrm{F}^{-1}\left[\mathrm{M}b + \mathrm{M}_C x\right]}_{u}\big) \tag{3.7}$$

Then, the iteration step simplifies to

$$x^{t+1} = x^t - \mathrm{M}_C\mathrm{F}\nabla\mathcal{R}\big(\mathrm{F}^{-1}\left[\mathrm{M}b + \mathrm{M}_C x^t\right]\big), \quad 0 \le t \le T-1 \tag{3.8}$$

Now acquired data is fixed, and this model tries to recover missing data, an approach similar to inpainting.

3.3 Regularization in k-space

As discussed in Section 1.4, the SPIRiT algorithm [10] computes a solution to MRI reconstruction problem so that it agrees with the calibration, that is $\mathrm{G}x \approx x$. We may think that the term $(\mathrm{G}x - x)$ in the functional acts as a regularization of the k-space:

$$\min_x \underbrace{\|\mathrm{G}x - x\|_2^2}_{regularization} + \frac{\lambda}{2}\|\mathrm{M}x - b\|_2^2$$

Our models should benefit from such regularization. Actually, we then obtain the VNs derived from the FoE-regularized SPIRiT algorithm:

$$\min_x \mathcal{R}\big(\mathrm{F}^{-1}x\big) + \frac{\lambda}{2}\|\mathrm{M}x - b\|_2^2 + \frac{\mu}{2}\|\mathrm{G}x - x\|_2^2$$

$$x^{t+1} = x^t - \mathrm{M}_C\mathrm{F}\nabla\mathcal{R}\big(\mathrm{F}^{-1}x^t\big) - \lambda^t\mathrm{M}\big(\mathrm{M}x^t - b\big) - \mu^t\big(\mathrm{G}^{\mathrm{H}} - \mathrm{I}\big)\big(\mathrm{G} - \mathrm{I}\big)x^t$$

However, for reconstruction in k-space the variant with fixed acquired data is preferred:

$$\min_x \mathcal{R}\big(\mathrm{F}^{-1}\big[\mathrm{M}b + \mathrm{M}_C x\big]\big) + \frac{\mu}{2}\big\|\big(\mathrm{G}x - x\big)\big(\mathrm{M}b + \mathrm{M}_C x\big)\big\|_2^2$$

$$x^{t+1} = x^t - \mathrm{M}_C\mathrm{F}\nabla\mathcal{R}\big(\mathrm{F}^{-1}\big[\mathrm{M}b + \mathrm{M}_C x^t\big]\big) - \mu^t\mathrm{M}_C\big(\mathrm{G}^{\mathrm{H}} - \mathrm{I}\big)\big(\mathrm{G} - \mathrm{I}\big)\big(\mathrm{M}b + \mathrm{M}_C x^t\big) \quad (3.9)$$

The PRUNO algorithm [39] further elaborates the idea of k-space consistency by upgrading the SPIRiT constraint $(\mathrm{G} - \mathrm{I})$ with a set of kernels that approximate null space of calibration matrix. The estimation of null space can be obtained from singular value decomposition of calibration matrix: $\mathrm{A} = \mathrm{U}\Sigma\mathrm{V}^{\mathrm{H}}$. Due to the fundamental assumption of shift-invariance of correlation coefficients (which also suggests linear dependence), calibration matrix with sufficient number of rows must have null space. Theoretically, null space of the matrix A is spanned by the columns of V that correspond to zero singular values. In practice, a number of columns matching smallest singular values are chosen.

It should also be noted that one of the successful models in the 2019 fastMRI contest[1] includes null space operator for data consistency [34]. The operator is applied at every reconstruction stage.

Thus, with null space operator we obtain the following optimization problem:

$$\min_x \mathcal{R}\big(\mathrm{F}^{-1}\big[\mathrm{M}b + \mathrm{M}_C x\big]\big) + \frac{\mu}{2}\big\|\mathbf{N}\big(\mathrm{M}b + \mathrm{M}_C x\big)\big\|_2^2,$$

and derive the iteration scheme as

$$x^{t+1} = x^t - \mathrm{M}_C\mathrm{F}\nabla\mathcal{R}\big(\mathrm{F}^{-1}\big[\mathrm{M}b + \mathrm{M}_C x^t\big]\big) - \mu^t\mathrm{M}_C\mathbf{N}^{\mathrm{H}}\mathbf{N}\big(\mathrm{M}b + \mathrm{M}_C x^t\big) \quad (3.10)$$

[1] `https://fastmri.org`

The authors of PRUNO point that null space estimate should be sufficiently large. This means that straightforward implementation of the algorithm would be rather slow because computational costs depend on the number of vectors in null space. By taking into account the associative property of convolution, the sequence of linear operators $\mathbf{N}^{\mathrm{H}}\mathbf{N}$ may be combined into a single convolutional kernel: $\tilde{n}_{ij} = \sum_k n^{\mathrm{H}}_{ki} * n_{kj}$. However, the resulting kernel almost doubles in both dimensions, and under some conditions the performance gains may become negligible. In such cases convolution theorem may help. We rewrite the reconstruction problem in image domain:

$$\min_u \mathcal{R}(u) + \frac{\lambda}{2}\left\|\mathbf{A}u - b\right\|_2^2 + \frac{\mu}{2}\left\|\mathbf{N}\mathrm{F}u\right\|_2^2$$

$$u^{t+1} = u^t - \nabla_u \mathcal{R}(u^t) - \lambda^t \mathbf{A}^{\mathrm{H}}\left(\mathbf{A}u^t - b\right) - \mu^t \mathrm{F}^{-1}\mathbf{N}^{\mathrm{H}}\mathbf{N}\mathrm{F}u^t \tag{3.11}$$

By applying the convolution theorem, i.e.,

$$\mathrm{F}^{-1}\mathbf{N}^{\mathrm{H}}\mathbf{N}\mathrm{F}u^t \Leftrightarrow \left(\sum_{j=1}^{N_C} \mathcal{F}^{-1}\left(\tilde{n}_{ij} * \mathcal{F}[u_j^t]\right)\right)_{i=1}^{N_C} = \left(\sum_{j=1}^{N_C} \mathcal{F}^{-1}\left[\tilde{n}_{ij}\right] \cdot u_j^t\right)_{i=1}^{N_C} \Leftrightarrow \mathcal{F}^{-1}[\tilde{n}] \odot u^t,$$

we obtain

$$u^{t+1} = u^t - \nabla_u \mathcal{R}(u^t) - \lambda^t \mathbf{A}^{\mathrm{H}}\left(\mathbf{A}u^t - b\right) - \mu^t \mathcal{F}^{-1}[\tilde{n}] \odot u^t \tag{3.12}$$

This iteration scheme is advantageous since training and reconstruction speed is virtually independent from the number of vectors in null space as well as kernel sizes. At the same time, storing precomputed IFFT($\tilde{n}$) would pose a technical problem for training due to very high memory consumption.

The same approach may be applied to SPIRiT-regularized models, which would allow usage of larger kernels without performance penalty ($\tilde{g}$ is calculated similarly to $\tilde{n}$):

$$\min_u \mathcal{R}(u) + \frac{\lambda}{2}\left\|\mathbf{A}u - b\right\|_2^2 + \frac{\mu}{2}\left\|\left(\mathrm{G} - \mathrm{I}\right)\mathrm{F}u\right\|_2^2$$

$$u^{t+1} = u^t - \nabla_u \mathcal{R}(u^t) - \lambda^t \mathbf{A}^{\mathrm{H}}\left(\mathbf{A}u^t - b\right) - \mu^t \mathcal{F}^{-1}[\tilde{g}] \odot u^t \tag{3.13}$$

Finally, it is also possible to combine SENSE and SPIRiT (or PRUNO) into one model:

$$\min_u \mathcal{R}(u) + \frac{\lambda}{2}\left\|\mathbf{A}u - b\right\|_2^2 + \frac{\mu}{2}\left\|\left(\mathrm{G} - \mathrm{I}\right)\mathrm{FS}u\right\|_2^2,$$

where $\mathbf{A} \overset{\text{def}}{=} \mathrm{MFS}$, and $(\mathrm{G} - \mathrm{I})$ can be replaced with null space estimation $\mathbf{N}$. The resulting iteration scheme looks very similar:

$$u^{t+1} = u^t - \nabla_u \mathcal{R}(u^t) - \lambda^t \mathbf{A}^{\mathrm{H}}\left(\mathbf{A}u^t - b\right) - \mu^t \mathrm{S}^{\mathrm{H}}\left(\mathcal{F}^{-1}[\tilde{g}] \odot \mathrm{S}u^t\right) \tag{3.14}$$

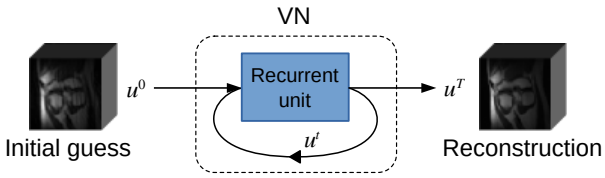
3.4 Feeding the VN

The most straightforward input to variational network is zero-filling reconstruction. In 1D acceleration these reconstructions exhibit inherent repeating aliasing artifacts. The images look superimposed multiple times and blurred in the direction of acceleration. Therefore, the algorithm's main objective is removing such aliasing artifacts, which sometimes may be difficult to achieve. Feeding zero images may also be tried, but this makes useless first layer's regularizer in our VNs. Therefore, this is not a good option.

One alternative is using GRAPPA reconstructions. For example, this strategy is used in [39] and it contributes to faster convergence, which is crucial for our algorithms based on truncated iterative schemes. The GRAPPA algorithm effectively removes strong artifacts, but in exchange for high noise levels. Hence, the VN's main objective is switched toward denoising, which seems to be an easier task. Moreover, image denoising problem is very well studied, and variational networks are proven effective in this task. However, computational costs of GRAPPA depend on sampling patterns. Regular patterns require computing only a few shift-invariant calibration kernels. Otherwise, GRAPPA becomes computationally inefficient.

3.5 VN with constant parameters

The models described above use iteration-dependent parameters. That is, each layer in a network has its own set of filter kernels, activation functions, data fidelity, and consistency trade-offs. This strategy is favored due to good results in [26], [29]. In contrast, the same set of parameters can be used in all layers, or, in other words, the VN has only one recurrent layer:



These networks are in active research, and some interesting results were obtained by Effland et al. [28] In the paper, the authors use the term "static" to refer to these types of networks. Here we also call them "static" or "recurrent" as opposed to "layered", which indicates iteration-dependent parameters.

The internal behavior of such VNs for MRI reconstruction is expected to be different from layered networks. Another feature is that they have much fewer parameters.

3.6 Training the VN

As in [29], we learn the VN parameters by minimizing Mean-Squared-Error (MSE) loss function:

$$\theta^* \;=\; \arg\min_{\theta} \left\{ \mathcal{L}(\theta) \;\stackrel{\text{def}}{=}\; \frac{1}{2|\mathcal{Z}|} \sum_{(u^0, b, m, u_{ref}) \,\in\, \mathcal{Z}} \Big\| \|f(u^0, b, m;\, \theta)\|_\epsilon \;-\; \|u_{ref}\| \Big\|_2^2 \right\} \tag{3.15}$$

MSE loss is a pixel-wise error metric, which is not the best option to train variational networks (see [40]). Nevertheless, this is still a good start.

Our loss function computes error between sum-of-squares-combined reconstructions and references. Another option is evaluating the error between coil images, e.g.,

$$\mathcal{L}(\theta) \;\stackrel{\text{def}}{=}\; \frac{1}{2|\mathcal{Z}|} \sum_{(u^0, b, m, u_{ref})} \sum_{k=1}^{N_C} \Big\| |f(u^0, b, m;\, \theta)_{(k)}|_\epsilon \;-\; |u_{ref(k)}| \Big\|_2^2 \tag{3.16}$$

Thus, we do supervised learning, an approach where every sample in training dataset comes with corresponding desired outcome, i.e., u_{ref}. This is a common strategy for training image restoration algorithms. As a disadvantage, it requires a large enough labeled dataset, which is not a big problem today[2].

In a nutshell, the model consumes input and generates output for each sample. Accuracy is measured by the loss function (MSE in our case), and the error is then back-propagated through the model, thereby fitting it to the training dataset. The trained model can then be evaluated for its performance on unseen data.

I do not include derivations of the gradients because modern specialized machine learning tools, such as PyTorch [41] and TensorFlow [42], compute them automatically.

[2] `https://fastmri.med.nyu.edu/`

4 Experiments and Results

Contents

4.1 Experimental setup

All models were implemented using PyTorch [41] with activation functions direcly in CUDA® [43] for performance reasons. All trainings and tests, as well as other related intensive computations were done on a workstation with four-core Intel® Core™ i7-960 @ 3.2GHz CPU, 24GB RAM, and two video cards, NVIDIA® GeForce® GTX TITAN X 12GB and NVIDIA® GeForce® 980 Ti 6GB, but using only one of them at a time.

The data for training and testing the variational networks was the same as in [29], but only coronal proton density sequence scans were used in my experiments. The data set consist of 20 patient scans with 20 slices each, making in total 400 slices. Half of these images, i.e., slices of patients 1–10, were used for training sessions and the other 200 slices for the tests. The slices are cropped into 368×368 pixels (along frequency encoding direction). The data set is normalized in a simple manner as in [29] and is processed into an optimizer with randomized batches of size 10. In order to simulate accelerated acquisition, the sampling masks were applied. In addition, the validation set of first 40 images from the testing data set was evaluated every 10^{th} epoch. The ground truth images were computed as root-sum-of-squares combination of fully sampled reconstructions.

For the optimizer base, Stochastic Gradient Descent (SGD) and the implementation of the block version of the Adam algorithm [44] were used. The latter was kindly provided to me by Patrick Knöbelreiter. After each optimization step, the parameter constraints are applied. These constraints include non-negativity of data fidelity and, where applicable, data consistency trade-off parameters, as well as normalization of filter kernels, which are required to be zero-mean unit vectors:

$$\lambda \geq 0, \quad \mu \geq 0, \quad \langle \mathbf{k}_j, \mathbf{1} \rangle = 0, \quad \|\mathbf{k}_j\|_2 = 1$$

The optimizations were usually performed in several sessions by manually adjusting the learning rate until convergence deterioration.

In order to assess the reconstructed image quality, I have used the following measures: Mean Squared Error (MSE), Normalized Root Mean Square Error (NRMSE), and Structural Similarity Index (SSIM [32]). The first two are calculated as follows:

$$\mathrm{MSE}(u, u_{ref}) = \frac{1}{HW}\|u - u_{ref}\|_2^2, \quad \mathrm{NRMSE}(u, u_{ref}) = \frac{\|u - u_{ref}\|}{\|u_{ref}\|}, \quad u, u_{ref} \in \mathbb{R}^{H \times W}$$

As a starting point, in order to reproduce the results from [29], I have implemented and trained the model described in the paper. The coil sensitivity data was precomputed in advance from ACS area of k-spaces for each slice with ESPIRiT algorithm [7]. Regarding the test results, only few slices out of 200 contain visible artifacts (though not strong). In general, the results are very good. The slightly modified version of the aforementioned model, e.g., with complex convolutions in the regularizer, produced similarly looking results, although quantitatively some models demonstrated better performance.

For MRI reconstruction without coil sensitivities, the variational networks with iteration-dependent parameters were tested first. The networks are made up of 10 similar layers. Different options for the FoE regularizer were tested, i.e., sum-of-squares-combined image, multi-channel 3D filters, convex convolutions, etc. Initially, the networks were fed with zero filling reconstructions.

In the attempts of archiving better results, some modifications were introduced: trade-off parameters were split such that each coil has its own parameter λ and μ (where applicable), the networks were trained with loss computed between individual coil images. The latter option was canceled due to high noise levels in reference reconstructions.

Then the experiments switched toward static iteration scheme with constant parameters. The recursion depth of these networks was set to 15. The behavior of this scheme comparing to the layered version is different with respect to the intermediate reconstructions.

To exploit the benefits of k-space reconstruction, the GRAPPA algorithm was implemented as described in [5]. For regular sampling with acceleration factor $R = 4$ in 2D MRI, the algorithm requires only three convolutional kernels, which means that in our case GRAPPA reconstructions can be efficiently computed. The direct reconstructions

Table 4.1: Tested models

Id	Description	Number of layers / steps, T	Number of filters, N_K	Activation function nodes, N_W	Filter size	Number of learnable parameters
VN1	Iteration-dependent parameters. The regularization is applied on SoS-combined image.	10	64	45	11×11	106,390
VN2	Iteration-dependent parameters. Multi-channel regularization, real and imaginary components are processed independently.	10	2×48	2×31	$15 \times 11 \times 11$	1,772,310
VN3	Static parameters. The regularization is applied on SoS-combined image.	15	64	45	11×11	10,639
VN4	Iteration-dependent parameters. Reconstruction in k-space with fixed acquired data. The image-domain regularization is applied on SoS-combined image.	10	64	45	11×11	106,240
VN5	Iteration-dependent parameters. The regularization is applied on SoS-combined image. Includes k-space consistency regularization.	10	64	45	11×11	106,540
VN6	Iteration-dependent parameters. Reconstruction in k-space with fixed acquired data. The image-domain regularization is applied on SoS-combined image. Includes k-space consistency regularization.	10	64	45	11×11	106,390
VN7	Static parameters. The regularization is applied on SoS-combined image. Includes k-space consistency regularization.	15	64	45	11×11	10,654
VN8	Iteration-dependent parameters. The regularization is applied on SoS-combined image. Includes k-space consistency regularization based on evaluation of null space of calibration matrix.	10	64	45	11×11	106,540

for acceleration factor $R = 2$ are almost as good as for fully sampled k-space, whereas reconstructions for $R = 4$ are noisy.

Again, the variational networks without coil sensitivities were tested, but this time by feeding the networks with noisy GRAPPA reconstructions. The networks tried to denoise the initial solutions.

Inspired by [10], the models were modified by including additional k-space regularization term. The experiments were done with SPIRiT kernels with sizes 5×5, 7×7, and 9×9. These kernels as well as GRAPPA kernels for initial reconstructions were computed separately for every slice and stored on HDD in order to speed up training start up process.

Next, an improved model based on PRUNO [39] with null space regularization term was tested. Two sets of precomputed null space approximations were used in the experiments: 100 kernels of size 5×5 for acceleration factor $R = 4$ and 150 kernels of size 9×9 for $R = 6$. In the network's implementation, the $\mathrm{IFFT}(\tilde{n})$ is computed for a batch fed to the network. This solves the memory-related issue while increasing the network throughput.

Finally, VN5 and VN8, as well as the network not included into Table 4.1 and implementing the last iteration formula from theoretical part, Eq. (3.14), i.e., using both sensitivities and null space regularization, were trained on data accelerated at $R = 6$.

4.2 Results

In all experiments the VNs with iteration-dependent parameters produce better results. The quantitative evaluation of the networks is shown in Table 4.2. When fed with zero filling reconstructions, neither of the networks were capable of removing aliasing artifacts from all testing images. However, some networks produced much better reconstructions than the others.

In particular, comparison of the reconstructions between VN1 and VN2 reveals that sum-of-squares regularization results in better image quality, only few slices contain striking aliasing artifacts. The network with componentwise multi-channel regularization, VN2, performs poorly. Although this VN has times more parameters, many reconstructed images are corrupted with apparent aliasing artifacts. On the other images, where artifacts were preserved by both methods, the more stronger ones were produced by VN2. For comparison, few slices are shown in Fig. 4.1.

Another drawback of multi-channel regularization is training and reconstruction time. For VN2 it takes almost 9 hours to do 100 epochs and 570ms to process a single slice, whereas VN1 trains at rate 1h 15m / 100 epochs, and requires only 90ms for reconstruction. Considering the large number of parameters and slow run times, its reconstruction performance is obviously inadequate. Therefore, VN2 was excluded from further experiments.

As expected, recurrent variational network VN3 fed with zero filling outputs lower quality results than its layered counterpart VN1. All reconstructions contain visible aliasing artifacts (see Fig. 4.1).

The network VN1 also outperforms VN4, where reconstruction is done by filling missing k-space data. Mostly, the results of both are comparable in quality.

One interesting observation is that training without filter kernel zero-mean constraint improves reconstruction quality (see Table 4.2), although there is only a subtle visual difference for the reader. To sum up, none of these models were able to completely eliminate the artifacts. However, the layered variational network with sum-of-squares-computed regularization, VN1, produced good-looking reconstructions by removing the artifacts from most of the slices. At the same time, some fine details also disappear.

A detailed look into the learned parameters of VN1 reveals that filters are very noisy, and structures (if present) are scarcely visible on most of them. The learned filter kernels of all 10 layers of the network are shown in Fig. 4.2.

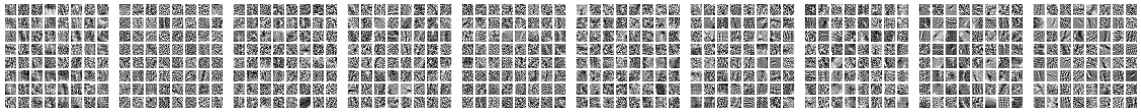


Figure 4.2: Learned filters of VN1 fed with zero filling
(The digital version of this document may be enlarged for detailed viewing.)

Additionally, Fig. 4.3 illustrates few filter kernels, corresponding activation functions and filter potentials (third row). The latter are calculated by integrating the activation functions, i.e.,

$$\phi_j^t(x) = \frac{\sqrt{2\pi}\,\sigma}{2} \sum_{k=1}^{N_W} w_{jk}^t \operatorname{erf}\left(\frac{x-\mu_k}{\sqrt{2}\,\sigma}\right) + C_j^t \quad \text{with } C_j^t \text{ s.t. } \min_x \phi_j^t(x) = 0.$$

As seen in the plot, the learned potentials come in different shapes, e.g., bell-shaped, V-shaped, sigmoid, and more complex functions, thereby reflecting variety of filter correlations.

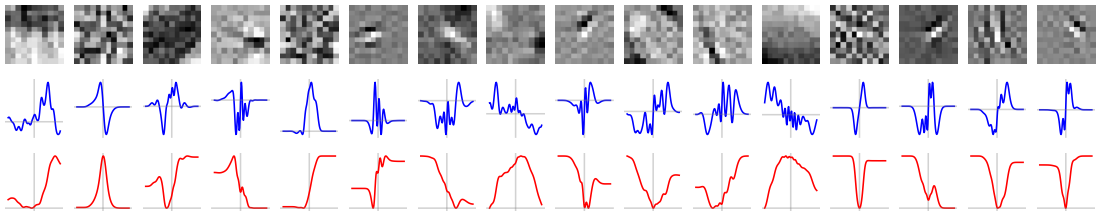


Figure 4.3: Selected filter kernels of VN1 (not on the same scale)

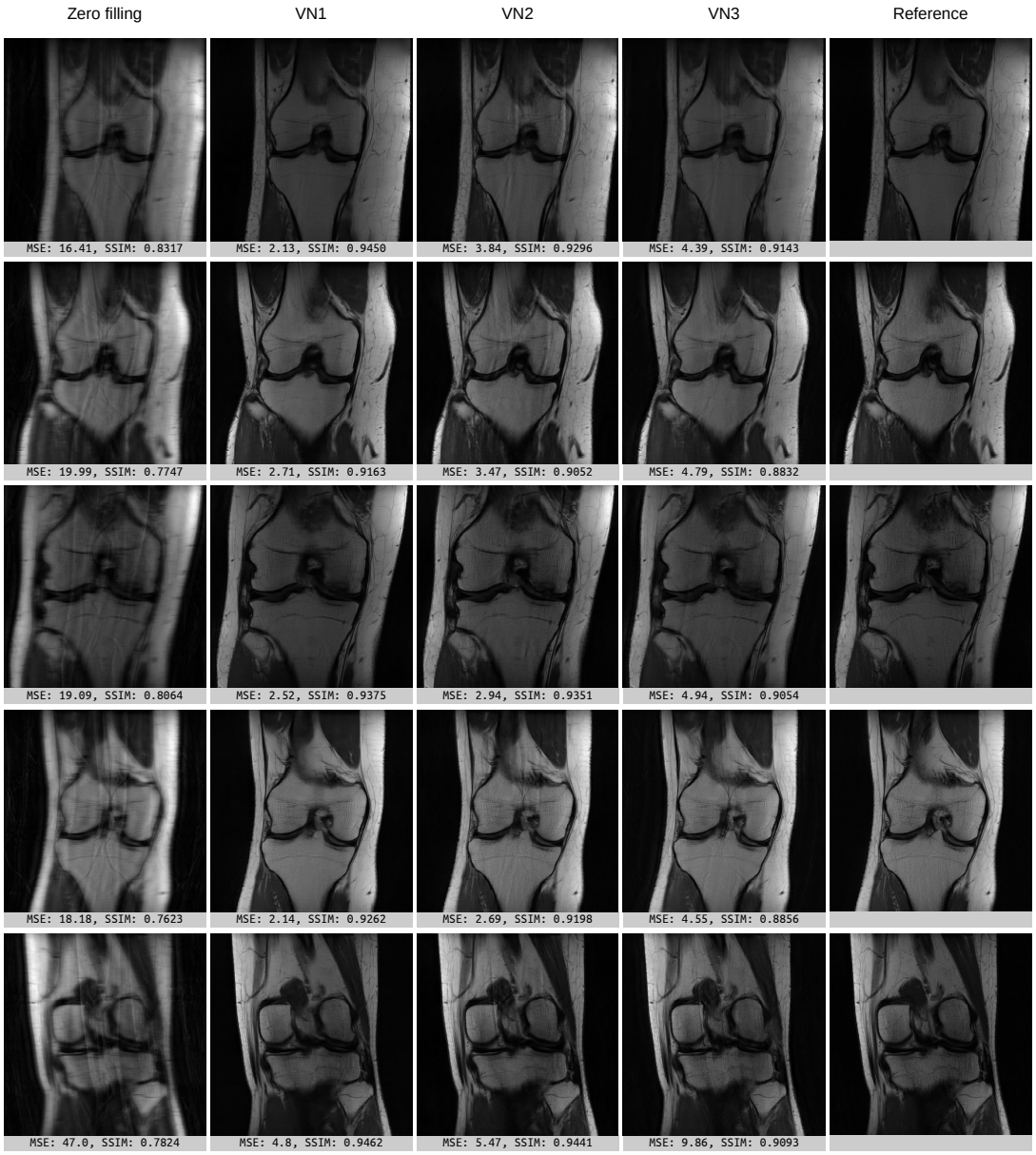


Figure 4.1: Comparison of reconstruction quality of variational networks fed with zero filling.

The data fidelity trade-off parameters have been learned for every coil channel and not equal but in general follow the same trend. In particular, there are few outliers in the layers $t \in \{1, 2, 4, 7\}$ that diverge from the main group, and in the layers $t \in \{5, 6\}$ the parameter values do not form a single cluster (see Fig. 4.4).

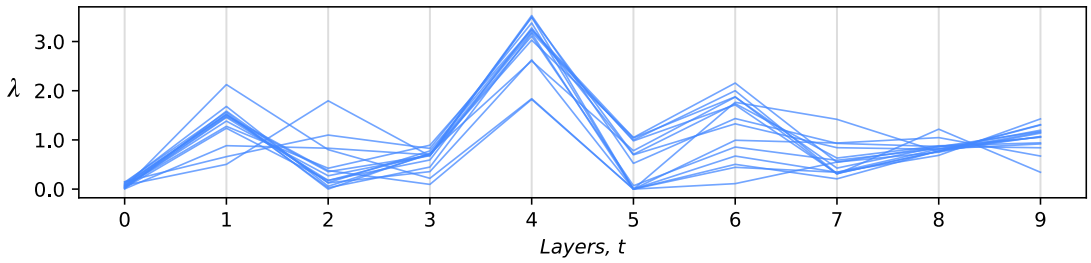


Figure 4.4: Learned data fidelity parameters of VN1

Looking at the intermediate steps, Fig. 4.5, it can be seen that the reconstruction process in the first 7–8 layers resembles a sequence of filtering operations. The most striking aliasing artifacts are removed in the layer $t = 4$ producing blurred image at $t = 5$. In the layers $t = 8$, $t = 9$ the image is visually refined, but very last layer suppresses some fine details. This can be attributed to the loss function used to train the network.

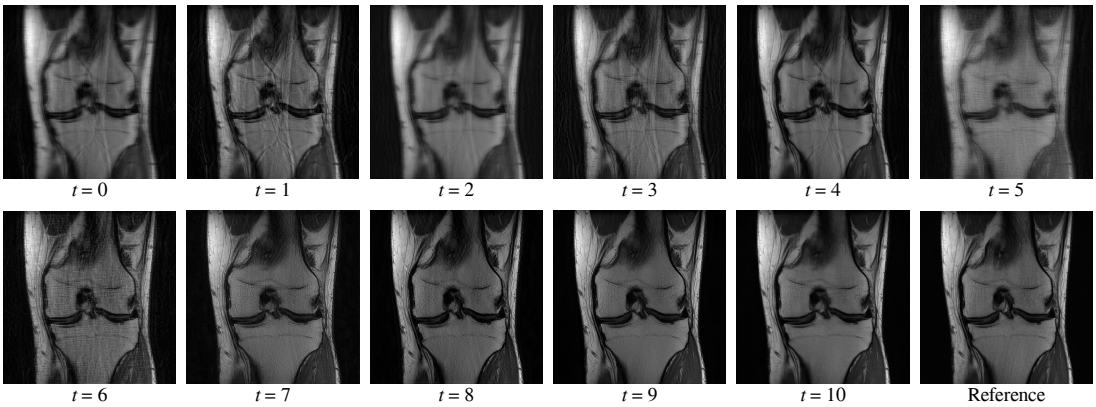


Figure 4.5: Intermediate reconstruction steps of VN1 fed with zero filling

For comparison, the intermediate steps of static variational network VN3 evolve with uniform changes gradually improving sharpness of the reconstruction. But still aliasing artifacts are not removed completely. Hence, besides producing reconstructions of lower quality, the network with 15 recurrent steps runs noticeably slower. At the same time, it has far less parameters, though the filters are very noisy. Fig. 4.6 depicts few of them.

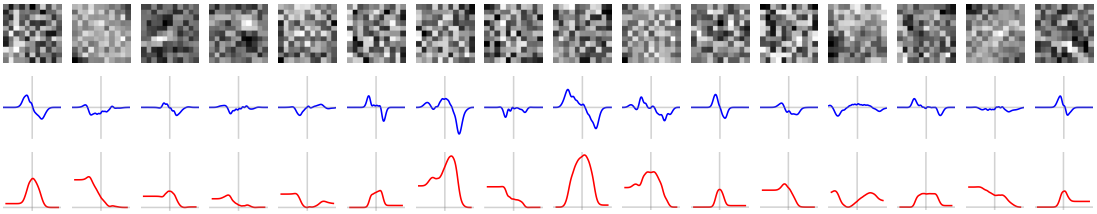


Figure 4.6: Selected filter kernels of VN3

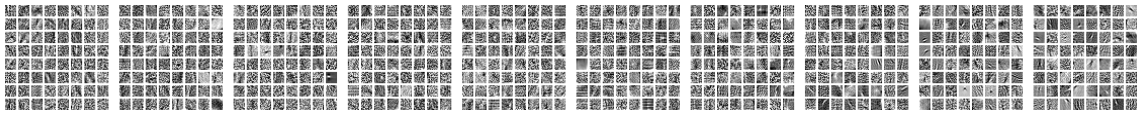


Figure 4.7: Learned filters of VN4 fed with zero filling
(The digital version of this document may be enlarged for detailed viewing.)

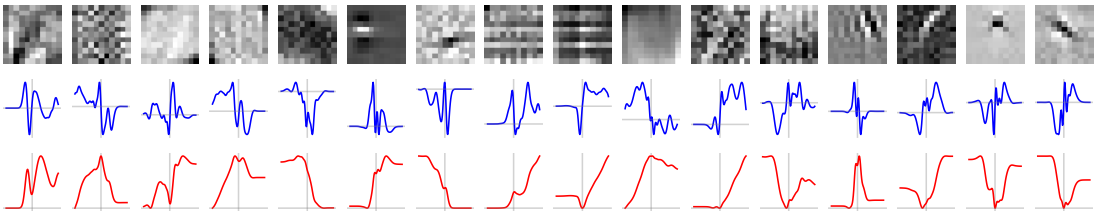


Figure 4.8: Selected filter kernels of VN4 (not on the same scale)

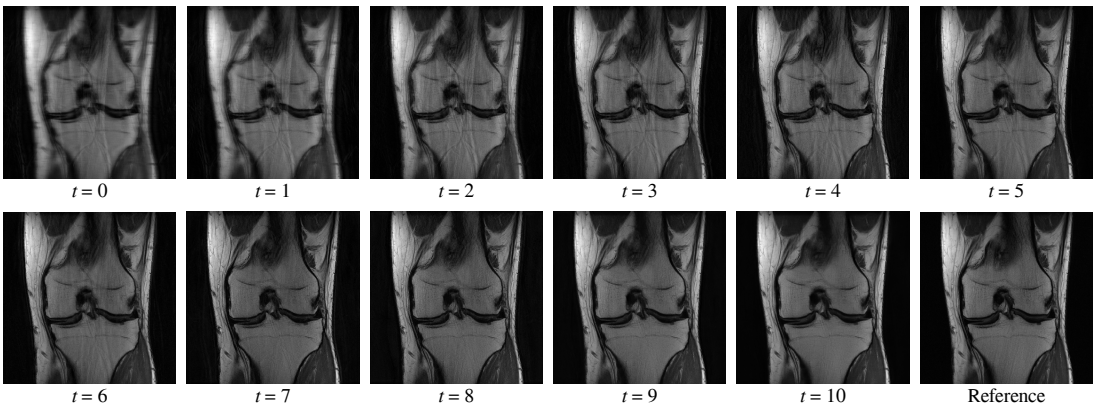


Figure 4.9: Intermediate reconstruction steps of VN4 fed with zero filling

Learned parameters of VN4 with fixed acquired k-space data are shown in Fig. 4.7 and Fig. 4.8. Likewise, we see a similar diversity of kernels and potential functions as of VN1, though the filter set seems to be of slightly better quality. Also, in the layer $t = 5$ there are a number of kernels with potentials that resemble ReLU (kernel in 11^{th} column in Fig. 4.8).

The internal reconstruction process of VN4 differs from that of VN1 (see Fig. 4.9). It more resembles iterative reconstruction with progressive improvement on every stage. That is, the aliasing artifacts are being gradually suppressed, producing pretty sharp image at the end.

4.2.1 GRAPPA initialization

Expectations from integrating GRAPPA algorithm into our reconstruction pipelines were met. Solely feeding noisy GRAPPA solutions without altering the networks' architectures substantially improves reconstruction quality. Specifically, the aliasing artifacts almost disappear. The reconstructions look much better, strong artifacts are not present at all. Likewise, training the networks without imposing zero-mean constraint on filter kernels results in lower MSE, NRMSE, as well as higher SSIM. Few reconstructions are shown in Fig. 4.10.

In these experiments, layered networks VN1 and VN4 produced similar results, whereas visual quality of the reconstructions generated by static variational network VN3 is slightly lower. It tends to smooth image, thereby suppressing fine details.

Computing GRAPPA reconstructions does not add significant run-time overhead to the whole algorithm. For example, PyTorch GPU implementation takes approximately 70 milliseconds per slice.

Fig. 4.11 shows learned filter kernels of VN1 and VN4. These filters look much better than the filters learned by feeding zero filling solutions. For example, in the layer $t = 6$ of VN1 most of the filters represent texture patterns. It is also interesting that few distinct checker patterns can be seen within the learned filter set of VN4 (see Fig. 4.14).

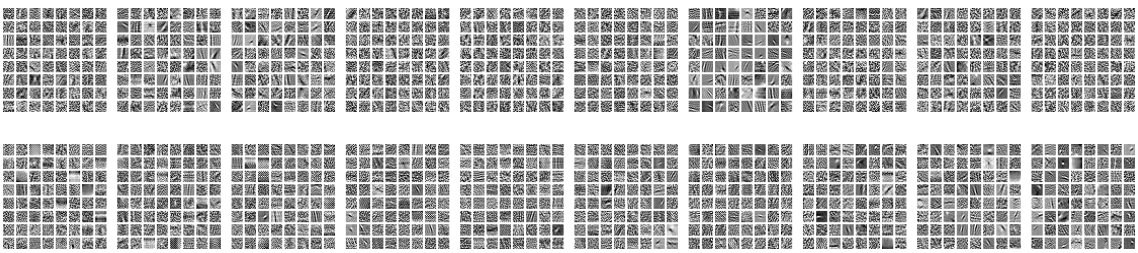


Figure 4.11: Learned filter kernels of VN1 (top) and VN4 (bottom) fed with GRAPPA (The digital version of this document may be enlarged for detailed viewing.)

Figure 4.10: Comparison of reconstruction quality of variational networks fed with GRAPPA.

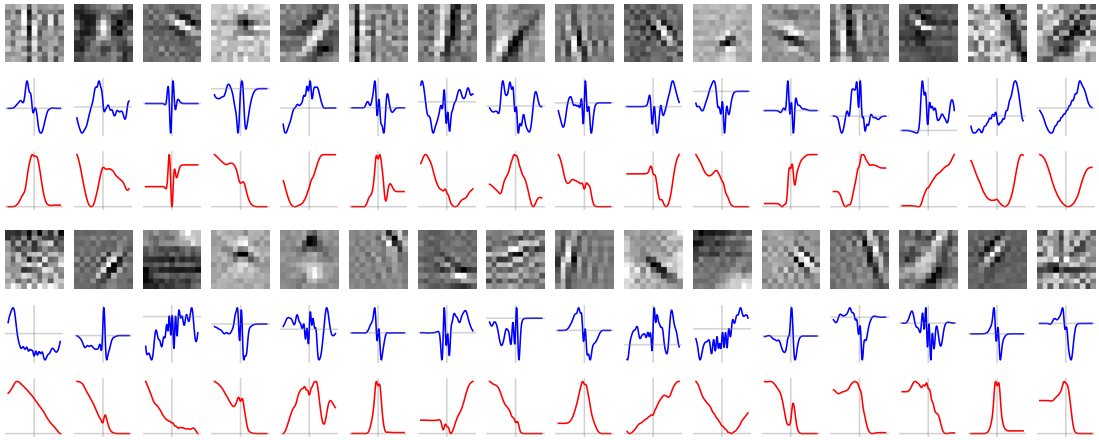


Figure 4.12: Selected filter kernels of VN1 (not on the same scale)

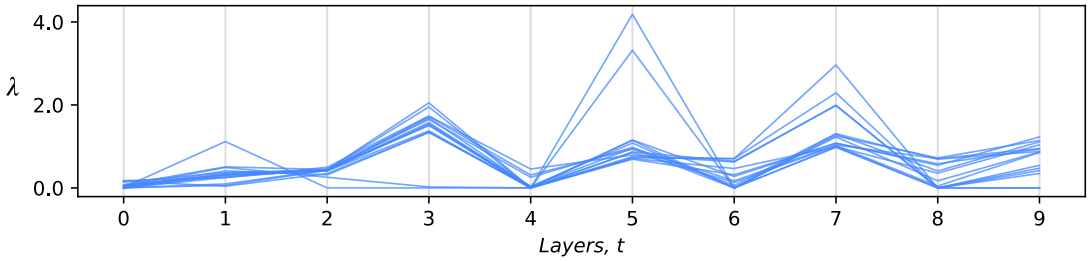


Figure 4.13: Learned data fidelity parameters of VN1 fed with GRAPPA

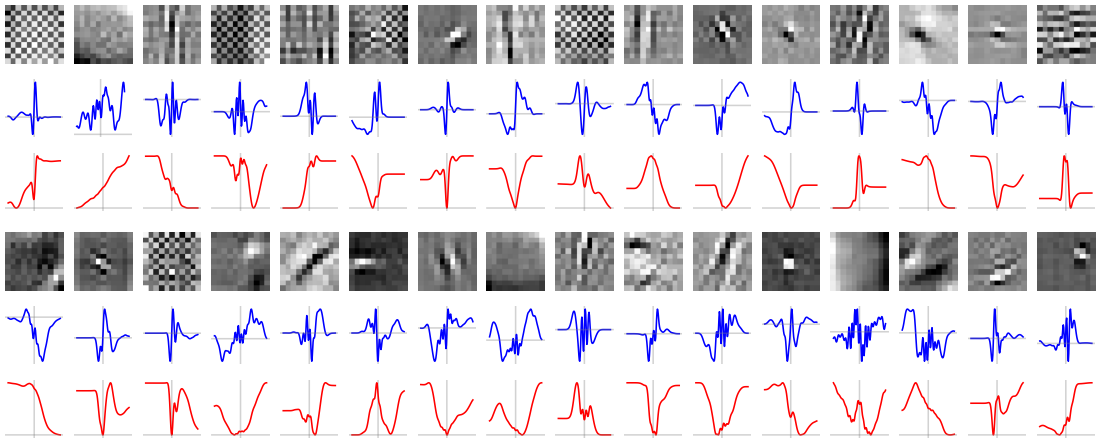


Figure 4.14: Selected filter kernels of VN4 (not on the same scale)

In this experiment the data fidelity parameters of VN1 have formed more compact distribution. Coil channels 3, 8 with low SNR are obvious outliers. Their parameter values have almost zero at $t = 3$ and then reach peak at $t = 5$ (see Fig. 4.13).

Surprisingly, inspecting intermediate reconstruction steps of VN1 fed with GRAPPA reveals that aliasing artifacts reappear again at $t = 4$ and are gradually removed toward the end (see Fig. 4.15). This behavior may be perceived as suboptimal, because the GRAPPA algorithm is used specially for the purpose of artifact removal. Nevertheless, this feeding strategy results in better reconstruction quality compared to zero filling and is supported by the quantitative evaluation (Table 4.2). Similarly, image sharpness is reduced a bit in the last layer.

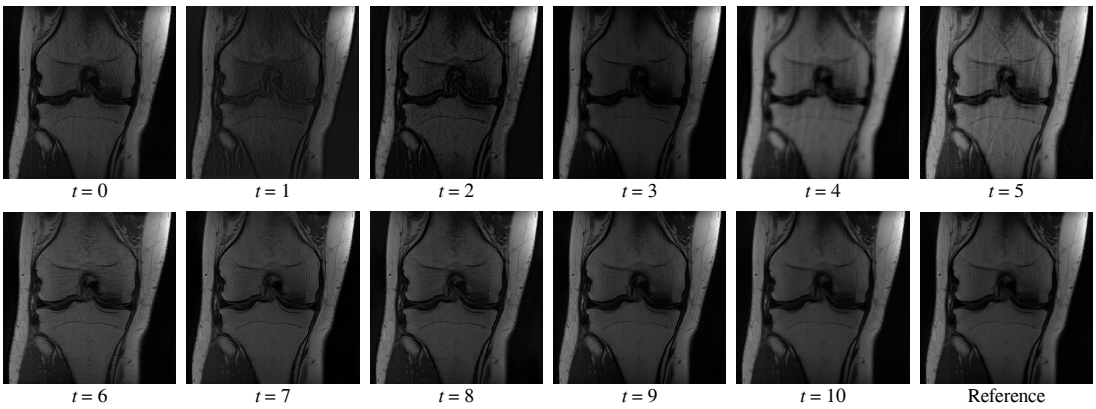


Figure 4.15: Intermediate reconstruction steps of VN1 fed with GRAPPA

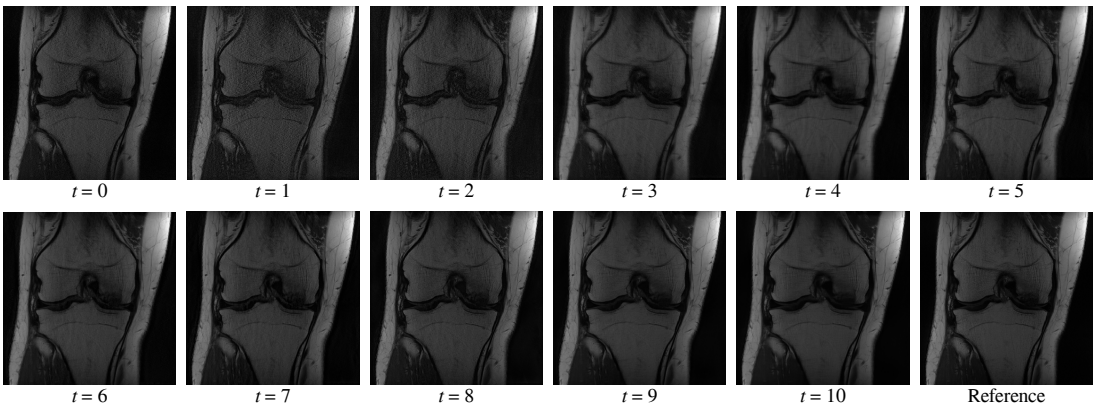


Figure 4.16: Intermediate reconstruction steps of VN4 fed with GRAPPA

Some minor artifacts can be seen on the reconstruction steps of VN4, although the process differs significantly (see Fig. 4.16). But still, reappearance of artifacts especially in VN1 suggests the possible benefits of using k-space regularization in the form of SPIRiT or null space operators.

For completeness, few filter kernels of static VN3 are shown in Fig. 4.17.

The networks trained with GRAPPA as input cannot be used to reconstruct from zero filling solution and vise-versa. More precisely, only in the first case the networks are able to slightly improve reconstruction, that is, make image less blurred. Therefore, this is not of practical interest.

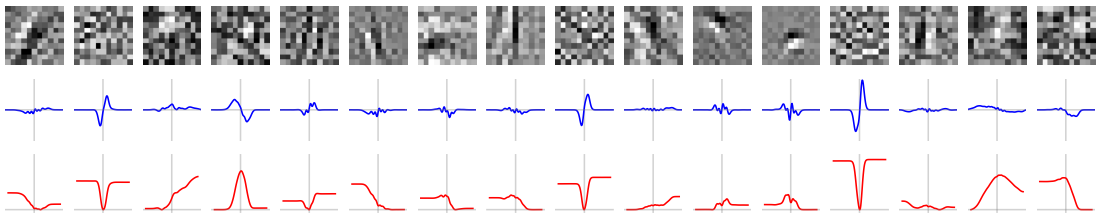


Figure 4.17: Selected filter kernels of VN3

4.2.2 Models with k-space regularization

Four networks with k-space consistency operators were trained in this experiment setting. There was no significant improvement in reconstruction quality. The networks VN5, VN6, and VN8 had similar performance (see Fig. 4.18).

All these networks have been trained to accept GRAPPA reconstructions as input. Additionally, the methods also require calculation of either SPIRiT kernels or null space estimates.

Comparison of all layered variational networks gives the impression that image-domain reconstruction models, i.e., VN1, VN5, and VN8, are more robust in terms of artifact suppression, while k-space reconstruction methods with fixed acquired data produce images with better detail. However, there are unusual noise-like artifacts in few slices produced by VN6. Such artifacts were not previously observed. They are especially noticeable in the background (for example, the slice in 2^{nd} row, 3^{rd} column, Fig. 4.18).

I will quickly go through the learned parameters. Fig. 4.19 shows filter kernels of layered networks implementing image-domain reconstruction, i.e., VN5 and VN8. Besides noise removal kernels, for example, minimizing at zero cross correlation, there are many filters in the form of structure patterns for image sharpening.

The learned trade-off parameters of VN5 are shown below in Fig. 4.20. Once more, coil channels 3, 8 diverge from the main group. For comparison, the trade-off parameters of VN8 look more coherent, no outliers (see Fig. 4.21).

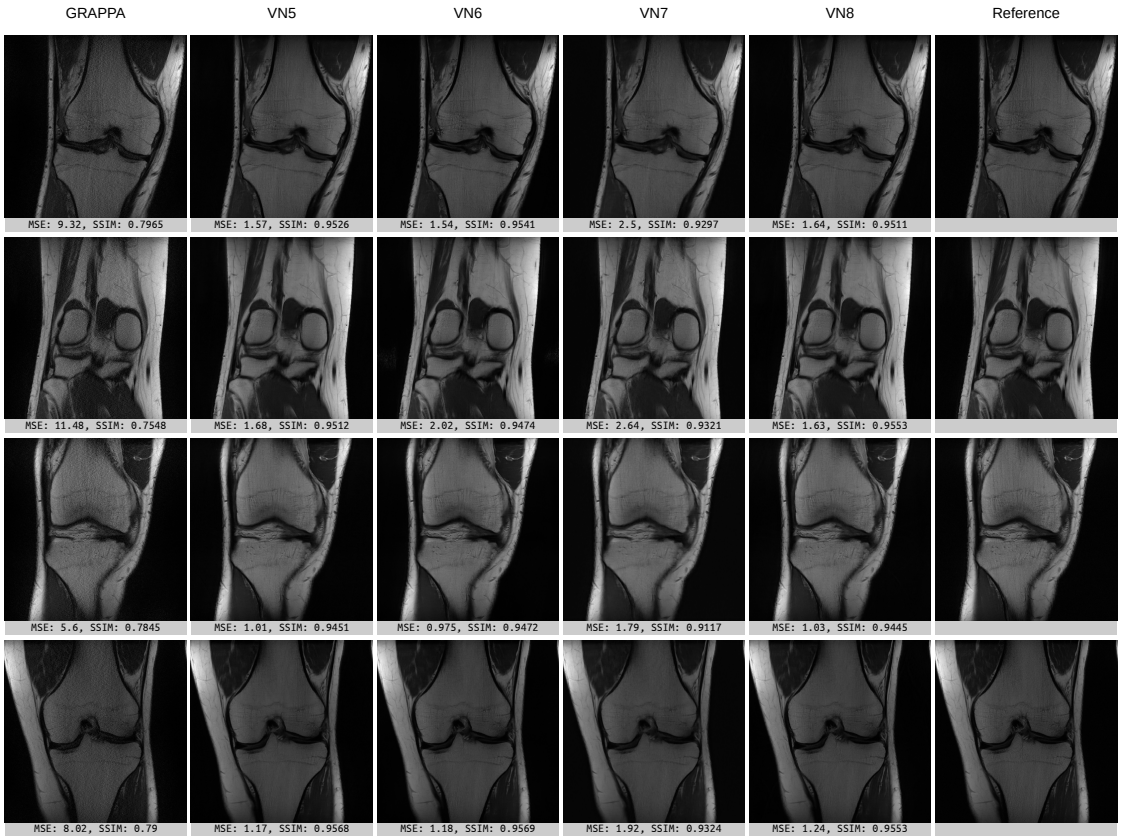


Figure 4.18: Comparison of reconstruction quality of variational networks with k-space regularization

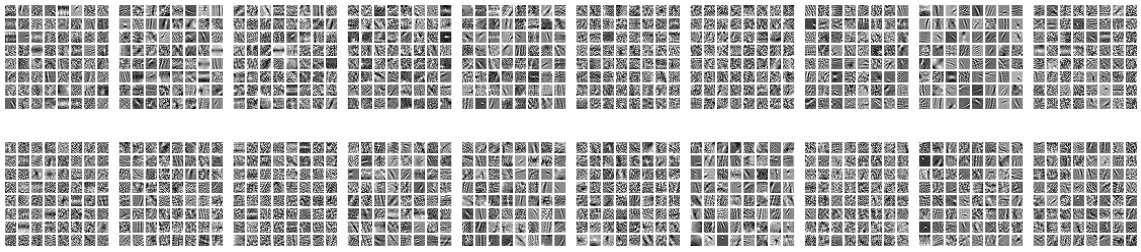


Figure 4.19: Learned filter kernels of VN5 (top) and VN8 (bottom)
(The digital version of this document may be enlarged for detailed viewing.)

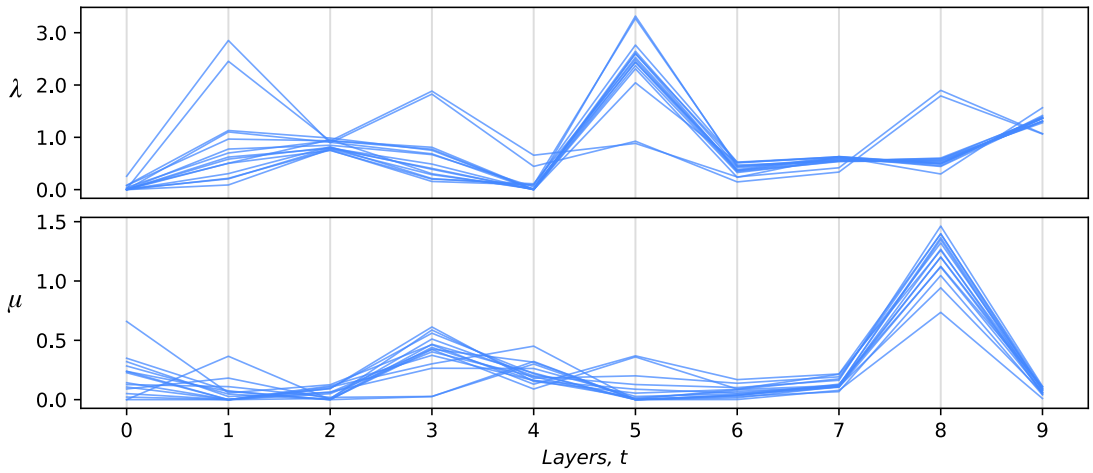


Figure 4.20: Learned data fidelity and k-space consistency trade-off parameters of VN5

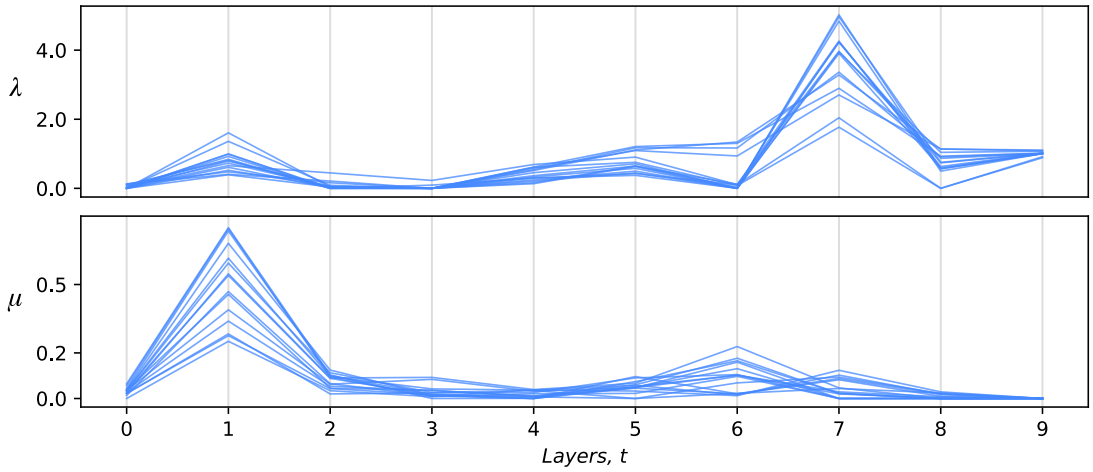


Figure 4.21: Learned data fidelity and k-space consistency trade-off parameters of VN8

Only intermediate reconstruction steps of VN8 are included. The process depicted in Fig. 4.22 is quite unique. The image seems to be ready at $t = 7$, but some artifacts show up at $t = 8$, possibly due to high data fidelity parameters and low k-space consistency trade-offs (see Fig. 4.21). Nevertheless, the image looks pretty fine at the end.

The learned filter kernels of variational network VN6 are very interesting (see Fig. 4.23). These are the most beautiful filters of all experiments. As with VN4, many checker patterns can be seen. The filters from 9^{th} layer are also rather unusual. Few filter kernels are shown in detail in Fig. 4.24.

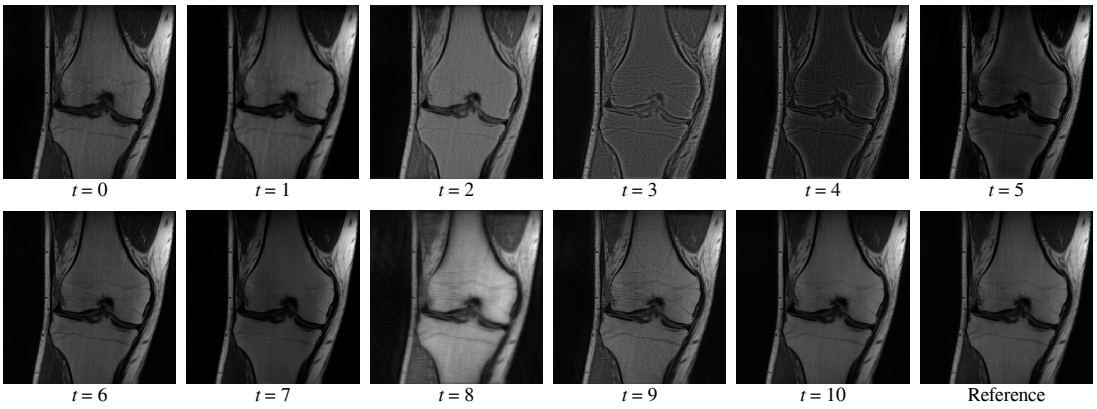


Figure 4.22: Intermediate reconstruction steps of VN8

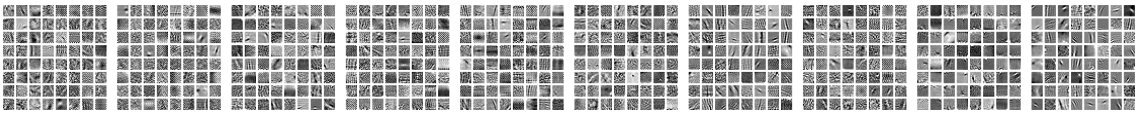


Figure 4.23: Learned filter kernels of VN6
(The digital version of this document may be enlarged for detailed viewing.)

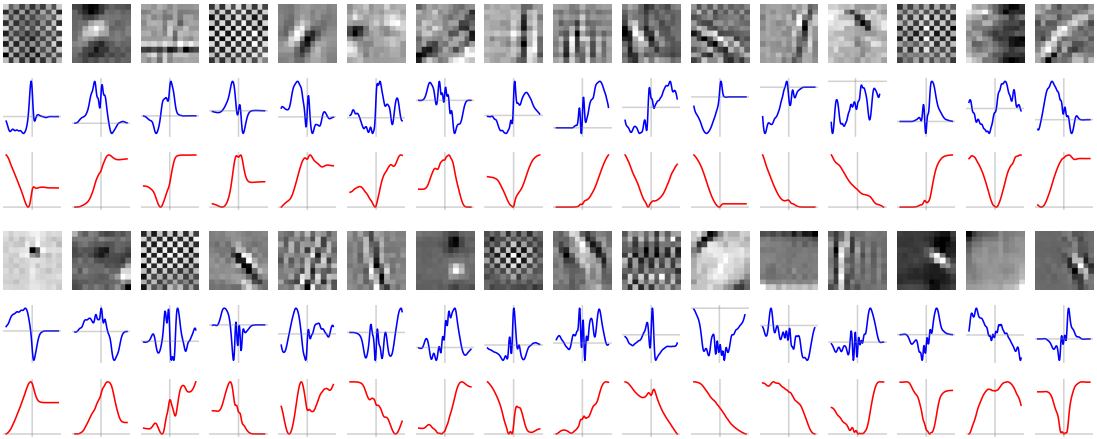


Figure 4.24: Learned filter kernels of VN6

4.2.3 GRAPPA vs zero filling

The experiment results demonstrate that feeding GRAPPA reconstructions to the networks is advantageous at regular acquisition pattern with acceleration factor $R = 4$. For comparison, few slices produced by VN1 are shown in Fig. 4.25. Besides, this is confirmed by the numerical data in Table 4.2. However, for layered networks, GRAPPA feeding does not lead to dramatic improvement of final reconstruction. In general, zero-filling-fed VNs produce quite good images, the striking artifacts only appear in a few slices. Additionally, the most advanced of our non-sensitivity models, VN8, was trained with zero filling input. Similarly, noticeable aliasing artifacts are only present in a small number of slices.

In contrast, our static variational networks benefit much more from GRAPPA feeding. The main issue with the static VNs is excessive smoothing of fine details, thereby significantly reducing the quality of the final reconstruction.

Finally, Fig. 4.26 depicts reconstructions obtained using four variational networks. This comparison also includes the original VarNet by Hammernik et al. [29] The results are comparable in quality, though GRAPPA-fed VNs produce slightly better images.

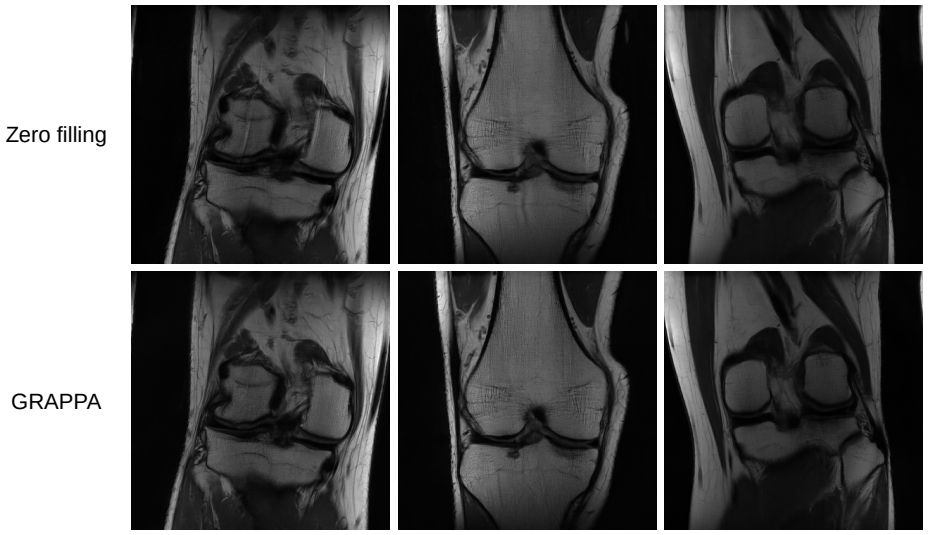


Figure 4.25: Comparison of feeding methods of VN1

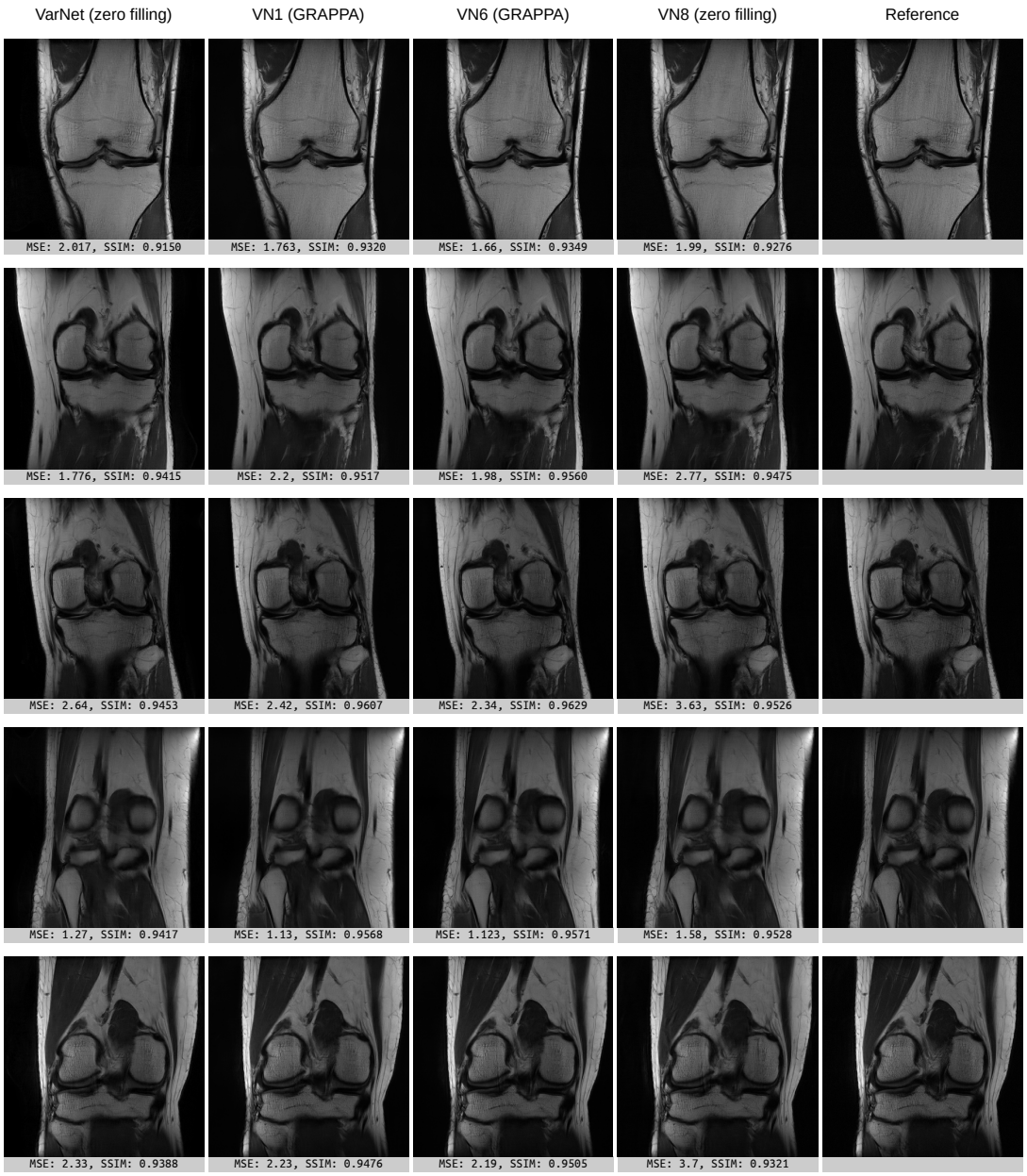


Figure 4.26: Comparison of reconstruction quality of VN1, VN6, VN8, and VarNet

Table 4.2: Quantitative comparison of variational networks for reconstruction of MRI data with acceleration factor $R = 4$ ($R_{eff} = 3.54$) without using coil sensitivity estimates.

Algorithm	Input	Number of parameters	Reconstruction time[a] (ms)	MSE on training set	MSE	NRMSE	SSIM
Zero filling				16.1554	16.7692	0.1466	0.8006
GRAPPA			70[b]	7.9671	7.4278	0.0992	0.7924
VarNet[c]	Zero filling	131,050	260	1.2534	1.4982	0.0452	0.9375
VN1	Zero filling	106,390	90	2.1056	2.6927	0.0592	0.9279
VN1, no zero-mean[d]				1.9058	2.4577	0.0562	0.9343
VN1	GRAPPA			1.8409	1.7980	0.0497	0.9325
VN1, no zero-mean				1.2954	1.3460	0.0429	0.9471
VN2	Zero filling	1,772,310	570	1.4053	2.4030	0.0562	0.9307
VN3 (static)	Zero filling	10,639	125	4.2777	4.8070	0.0813	0.8947
	GRAPPA			2.2697	2.1615	0.0551	0.9218
VN4 (k-space)	Zero filling	106,240	90	1.9875	2.4351	0.0569	0.9309
	GRAPPA			1.4574	1.5149	0.0458	0.9420
VN5	GRAPPA	106,540	195	1.2371	1.2996	0.0423	0.9471
VN6 (k-space)		106,390	180	1.2198	1.2881	0.0422	0.9487
VN7 (static)		10,654	295	2.1934	2.1042	0.0545	0.9225
VN8	Zero filling	106,540	150	1.5003	1.8681	0.0496	0.9409
	GRAPPA			1.2695	1.3277	0.0427	0.9472

[a] For variational networks, this timing excludes initial reconstruction costs.
[b] Including calculation of three 5×2 convolutional kernels.
[c] Model by Hammernik et al. with sensitivities [29]. Included for comparison.
[d] VN1 has been trained with and without zero-mean constraint, i.e., $\langle \mathbf{k}_j, \mathbf{1} \rangle = 0$. It has not been applied to other networks.

4.2.4 Reconstruction at higher acceleration rates

3D MRI, where undersampling can be done in two phase-encoded dimensions (2D acceleration), has greater potential to acquire data at a higher acceleration rates. Additionally, more data acquisition patterns are available in 3D MRI, e.g., spiral trajectories. In the case of 2D MRI with Cartesian acquisition, higher acceleration factors would inevitably result in substantial deterioration of reconstruction quality.

I have trained three layered networks, VN5, VN8, and VN9 (derived from the model with null space regularization and sensitivity estimates) for the acceleration factor $R = 6$. The quantitative results of the experiments are presented in Table 4.3. As it was expected, reconstruction quality of the networks at this acceleration factor is much lower. In general, the images look clear, however, fine detail are lost to a greater extend, and aliasing artifacts appear more often. Again, feeding GRAPPA solutions seems to be advantageous, although the GRAPPA reconstructions themselves have limited practical significance due to very high noise level.

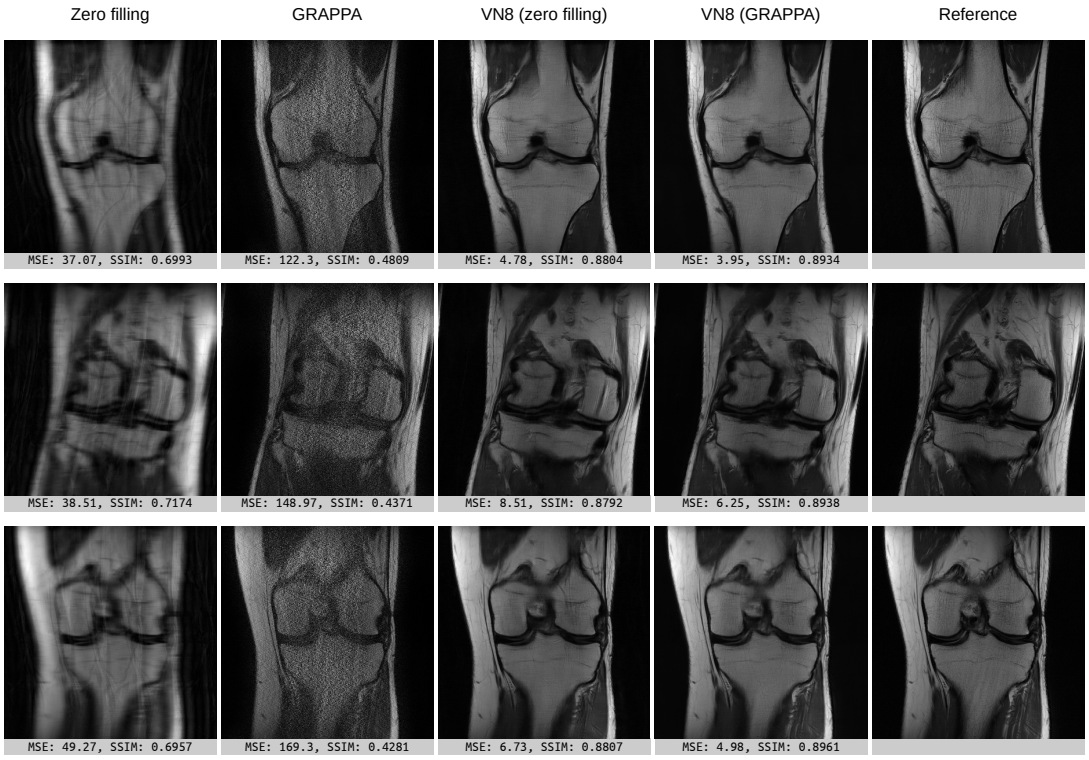


Figure 4.27: Reconstructions produced by VN8 ($R = 6$)

Fig. 4.27 depicts reconstruction of three slices for comparison. The GRAPPA solutions were computed without Tykhonov regularization term, hence, they are very noisy. The same GRAPPA reconstructions were fed to VN8 (4^{th} column). The network VN5 with SPIRiT regularization has only been trained for GRAPPA input and produced results similar to those of VN8.

As regards the network with sensitivities, VN9, the reconstructions are visibly better than those of GRAPPA-fed VN8. Few slices are shown in Fig. 4.28. Furthermore, the network runs slightly faster, though the reconstruction process involves multiple sensitivity weighting operations.

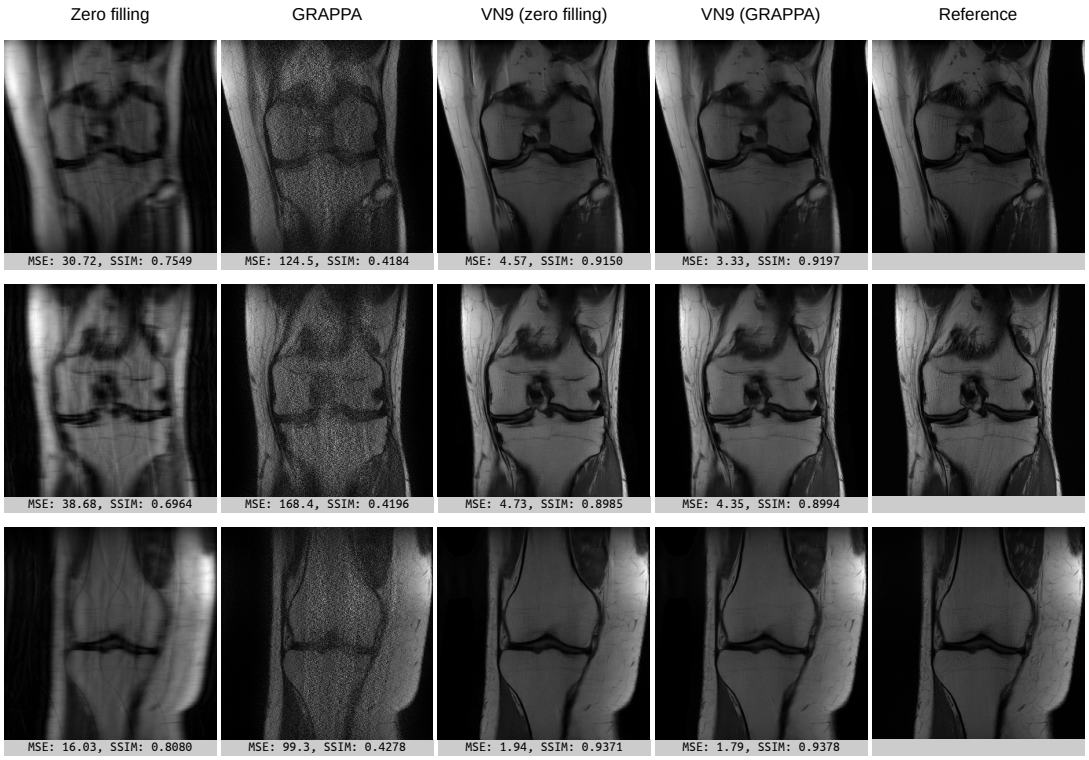


Figure 4.28: Reconstructions produced by VN9 ($R = 6$)

Table 4.3: Quantitative comparison of variational networks for reconstruction of MRI data with acceleration factor $R = 6$ ($R_{eff} = 5.04$).

Algorithm	Input	Number of parameters	Reconstruction time[a] (ms)	MSE on training set	MSE	NRMSE	SSIM
Zero filling				24.4909	25.8323	0.1831	0.7617
GRAPPA			135[b]	83.6917	89.5789	0.3260	0.5120
VN5	GRAPPA	106,540	150	2.7109	3.1278	0.0648	0.9057
VN8	Zero filling	106,540	150	3.4400	4.4000	0.0767	0.8944
	GRAPPA			2.7160	3.0994	0.0644	0.9109
VN9[c]	Zero filling	106,260	140	2.6672	3.3330	0.0672	0.9131
	GRAPPA			2.5872	2.8951	0.0626	0.9161

[a] For variational networks, this timing excludes initial reconstruction costs.
[b] Including calculation of three 7×2 convolutional kernels.
[c] The network with coil sensitivities, Eq. (3.14).

5 Discussion

In general terms, the experiments have confirmed that most of the tested variational networks are capable of producing high quality reconstructions from undersampled MRT Coronal PD data. Since coil sensitivity estimates are not fed as a part of the input to the networks[1], this removes some computationally intensive overhead, thereby shortening the run time of the whole reconstruction algorithm.

The optimal architecture with respect to reconstruction quality and computational costs still remains unknown. However, one of the main tasks of this work was to test different FoE-based regularization methods. A number of options have been tried, and it turned out that the models with the regularization applied on sum-of-squares[2] combination works best. This approach not only superior in reconstruction quality, but also more computationally efficient. Moreover, applying image-domain regularization, i.e., Fields of Experts, on real-valued data is theoretically justified. At the same time, using multi-channel filters on magnitude coil images does not lead to better reconstructions for a number of reasons, for example, highly non-uniform intensities and different SNR. Therefore, based on experimental outcomes, all other options in the same scenario are not recommended, which is also demonstrated by the unimpressive results of VN2 and its slow run time (see Fig. 4.1 and Table 4.2). In contrast, models with SoS-applied regularization generally produce good results. Nevertheless, I must conclude the following. Studying the latest advances in learned MRI reconstruction, it becomes evident that today Fields of Experts prior model for this task (and perhaps for all learning-based image processing in general) is mostly outdated.

Besides intensities, complex data in MRI delivers phase information. For illustrative purposes, Fig. 5.1 shows coil magnitude images and phase images of artifact-free slice reconstructed with VN8. It is visible that phase changes gradually, or, in other words, phase images should be smooth. Since aliasing artifacts affect phase information, this

[1] VN9 with sensitivities has been trained only for comparison.
[2] Actually root-sum-of-squares combination.

suggests of having additional regularization in phase domain. However, designing such phase regularizer is not a trivial task. Following the trend of machine learning, data-driven models may be considered a possible solution. For example, Lee et al. train two multi-scale neural networks with large receptive fields to estimate aliasing artifacts, one for magnitudes and the other for phase images. The reconstruction is then obtained by subtracting the estimated artifact map from corrupted input image [45].

More often complex MR image components are treated as channels and concatenated together to a form a single vector, followed by real-valued multi-channel convolution. Yet, we see that when used with a powerful prior, this leads to very competitive results. On the other hand, DeepcomplexMRI employs complex convolution, which the authors believe is beneficial due to the correlation between real and imaginary parts [37]. This, of course, almost doubles the computational load. Besides, the following dilemma arises when modeling MRI reconstruction without coil sensitivities: to combine or not to combine the coil images before evaluating the prior. Intuitively, I would prefer the combine-then-regularize paradigm supported by the impressive results of Bian et al. [38] as well as our own experiments. But it is too early to draw conclusions on this issue. Perhaps, another possible approach that could be explored is processing MR images as 2D vector fields. Fig. 5.2 depicts three coil images with interesting vector patterns.

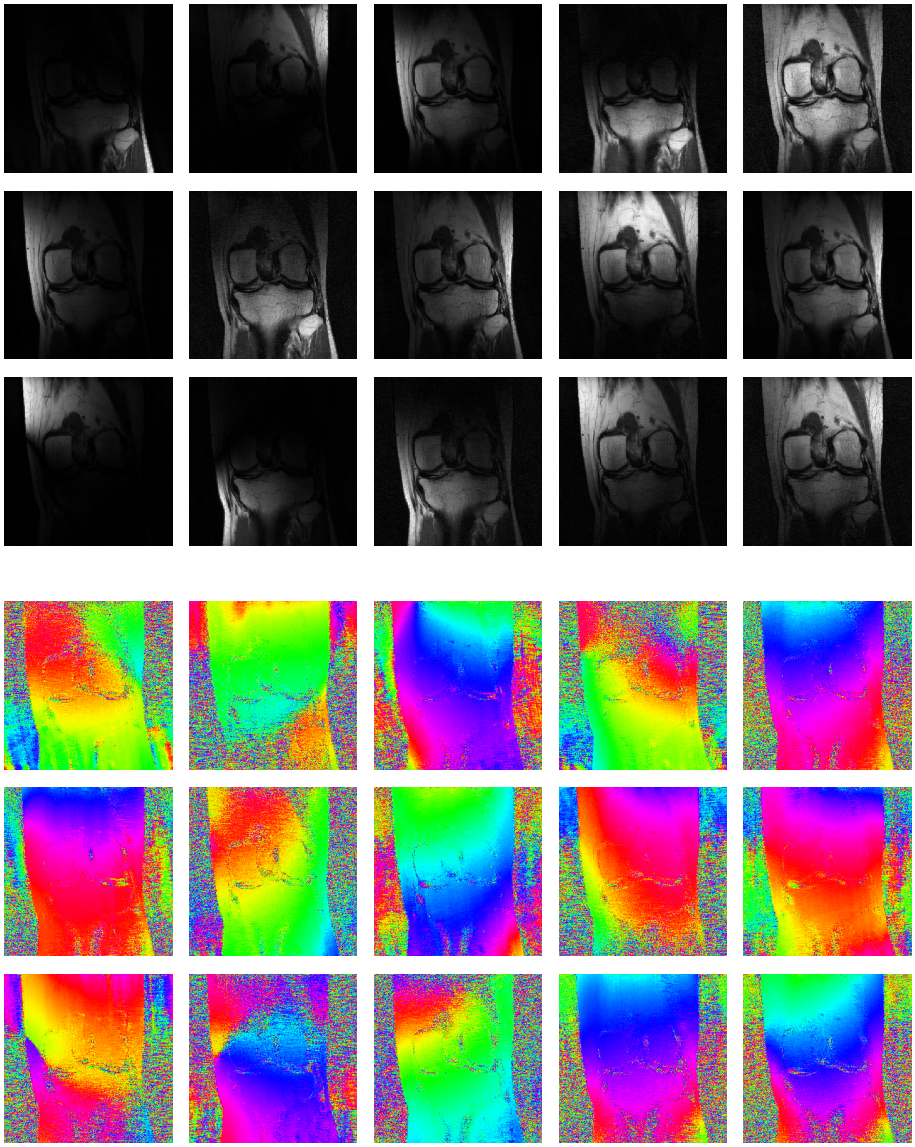
We have also observed that even our best reconstructions lose some fine details anyway. Although, this seems to be unavoidable when processing accelerated MRI data, still, it is believed that using better error metric could improve reconstruction quality. Specifically and in simple words, loss should depend not only on pixel values, but also on the local structure of an image. It was already shown that training a variational network using SSIM as loss function is beneficial [40]. It results in sharper images compared to squared ℓ_2 loss training.

Sriram et al. train their model, E2E-VarNet briefly described in Section 2.5, using $J(u, u_{ref}) = -\text{SSIM}(u, u_{ref})$, though they also admin excessive smoothness of reconstructions [31]. To make the images look more natural, they add a small amount of Gaussian noise.

Totay, advanced forms of loss functions are used to train some state-of-the-art models. The following composition of SSIM and ℓ_1-norm might be a good choice (for example, it was used to learn the parameters of Σ-net [33] and GrappaNet [36]):

$$J(u, u_{ref}) \;=\; -\text{SSIM}(|u|, |u_{ref}|) \;+\; \lambda \ell_1(|u|, |u_{ref}|)$$

Zhao et al. compare the performance of image restoration networks trained with different loss functions [46]. It turned out that not only the perceptually-motivated metrics, such as Multi-Scale Structural Similarity Index (MS-SSIM [47]), are superior to MSE, but even training with ℓ_1 leads to better results. However, the authors show that SSIM and MS-SSIM alone cannot be considered an ideal loss function. They finally propose the combination of MS-SSIM and multi-scale Gaussian weighted ℓ_1 (see [46] for details).



The phase images are rendered in hue colors:

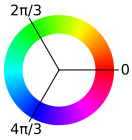
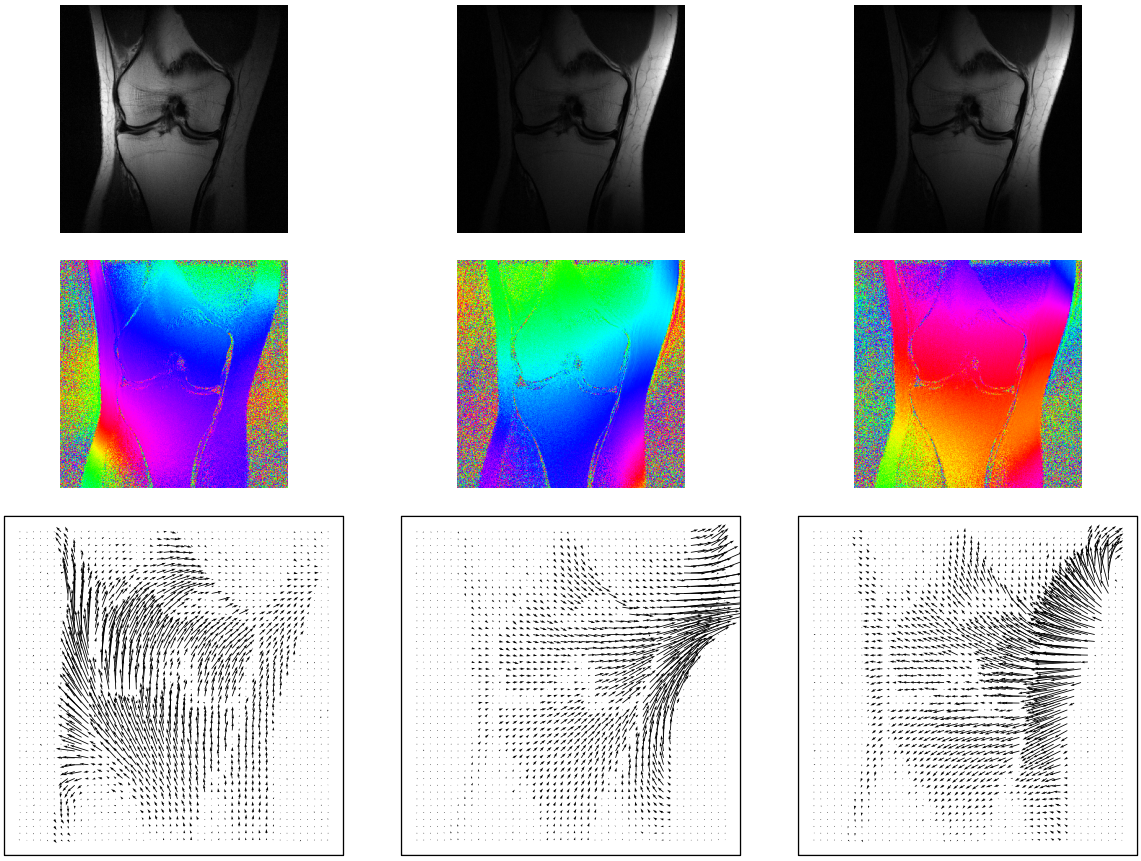


Figure 5.1: Magnitude and phase images of a slice reconstructed by VN8



The phase images are rendered in hue colors:

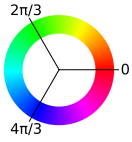


Figure 5.2: MRT complex images from three coils displayed as 2D vector fields (bottom) with corresponding magnitude and phase images (top and middle rows)

Despite this, it seems that the optimal loss function for training deep neural networks with applications to MRI reconstruction remains an open question. Moreover, an advanced k-space normalization is recommended to bring the data to the same magnitude scale. Solving these problems is one of the objectives of further research.

Additionally, our results indicate that feeding good initial guess to the tested models leads to an improvement in reconstruction quality, i.e., the advantage of GRAPPA over zero-filling. However, to the best of my knowledge, GRAPPA reconstructions cannot be efficiently computed in some cases. For comparison, regular Cartesian sampling with acceleration factor $R = 4$ requires computation of only three kernels, whereas non-regular non-uniform k-space acquisition patters, such as radial or spiral in 3D MRI, would require

computation of many more [10]. For higher acceleration rates, usage of GRAPPA solutions as input is even more questionable.

Regarding k-space consistency regularization, there might be room for improvement. That is, its relatively small contribution to improving the quality of MRI reconstruction suggests that in this simple schemes, derived from steepest gradient descent, it is not used efficiently. Hence, outcome of the reconstruction process strongly depends on the regularization in image domain, while the k-space consistency operator plays only an auxiliary role.

Finally, variational network with fixed acquired data, that is, VN4 and its derivatives, have one possible useful property. Namely, the MR image gradually improves as it passes through the network. This is different to the layered image-domain reconstruction VNs (see for comparison Fig. 4.5 and Fig. 4.9). It should also be noted that the scheme with fixed acquired data is limited to Cartesian sampling.

6 Conclusion

In this work I have tested a number of variational networks for reconstruction of accelerated MRI data without coil sensitivities. We have obtained good-looking results for a number of models, which illustrates that variational network with Field of Experts regularization is still a powerful reconstruction tool. However, I have to admit that these VNs can no longer compete with the latest deep learning MRI reconstruction techniques. But in comparison with them, our networks have a very low number of parameters (e.g., ∼106K of our VNs versus 30M of E2E-VarNet [31] or 480M of GrappaNet [36]). The use of advanced regularization is expected to result in a significant improvement in reconstruction quality.

The experiment outcomes clearly indicate that the layered networks outperform the static ones. Additional k-space regularization, either SPIRiT or PRUNO, improves reconstruction quality, and it has been shown that using large kernels does not increase computational costs.

Also, for me personally, it was a great opportunity to work on CUDA-enabled hardware, in addition to gaining in-depth knowledge of the subject.

A

List of Acronyms

ACS	Auto-Calibration Signal
CG	Conjugate Gradient
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CS	Compressed Sensing
CT	Computed Tomography
FoE	Fields of Experts
FoV	Field of View
GPU	Graphics Processing Unit
HDD	Hard Disk Drive
IFFT	Inverse Fast Fourier Transform
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
MRT	Magnetic Resonance Tomography
MS-SSIM	Multi-Scale Structural Similarity Index
MSE	Mean Squared Error
NRMSE	Normalized Root Mean Squared Error
ReLU	Rectified Linear Unit
SNR	Signal-to-Noise Ratio
SoS	Sum-of-Squares
SSIM	Structural Similarity Index
TDV	Total Deep Variation
TGV	Total Generalized Variation
TV	Total Variation
VN	Variational Network

Bibliography

[1] P. Roemer, W. Edelstein, C. Hayes, S. Souza, and O. Mueller, "The NMR phase array," *Magnetic Resonance in Medicine*, vol. 16, no. 2, pp. 192–225, 1990. (page 1, 6)

[2] J. Kelton, R. Magin, and S. Wright, "An algorithm for rapid image acquisition using multiple receiver coils," in *SMRM, 8th Annual Meeting, Amsterdam*, p. 1172, 1989. (page 2)

[3] D. Sodickson and W. Manning, "Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays," *Magnetic Resonance in Medicine*, vol. 38, no. 4, pp. 591–603, 1997. (page 2)

[4] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 952–962, 1999. (page 2, 6)

[5] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magnetic Resonance in Medicine*, vol. 47, no. 6, pp. 1202–1210, 2002. (page 2, 28)

[6] K. Pruessmann, M. Weiger, P. Boernert, and P. Boesiger, "Advances in sensitivity encoding with arbitrary k-space trajectories," *Magnetic Resonance in Medicine*, vol. 46, no. 4, pp. 638–651, 2001. (page 3)

[7] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig, "ESPIRiT – an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA," *Magnetic Resonance in Medicine*, vol. 71, no. 3, pp. 990–1001, 2014. (page 3, 28)

[8] P. Jakob, M. Griswold, R. Edelman, and D. Sodickson, "AUTO-SMASH: a self-calibrating technique for SMASH imaging," *MAGMA*, vol. 7, no. 1, pp. 42–54, 1998. (page 3)

[9] R. M. Heidemann, M. A. Griswold, A. Haase, and P. M. Jakob, "VD-AUTO-SMASH imaging," *Magnetic Resonance in Medicine*, vol. 45, no. 6, pp. 1066–1074, 2001. (page 3)

[10] M. Lustig and J. M. Pauly, "SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space," *Magnetic Resonance in Medicine*, vol. 64, no. 2, pp. 457–471, 2010. (page 4, 23, 30, 53)

[11] Y. Chang, D. Liang, and L. Ying, "Nonlinear GRAPPA: A kernel approach to parallel MRI reconstruction," *Magnetic Resonance in Medicine*, vol. 68, no. 3, pp. 730–740, 2012. (page 7)

[12] M. Akçakaya, S. Moeller, S. Weingärtner, and K. Uğurbil, "Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging," *Magnetic Resonance in Medicine*, vol. 81, no. 1, pp. 454–465, 2019. (page 7)

[13] 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, 2018. (page 7, 8)

[14] F. Knoll, K. Hammernik, C. Zhang, S. Moeller, T. Pock, D. K. Sodickson, and M. Akçakaya, "Deep learning methods for parallel magnetic resonance image reconstruction," *arXiv preprints*, 2019, arXiv:1904.01112 [eess.SP]. (page 8)

[15] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006. (page 10)

[16] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006. (page 10)

[17] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, pp. 259–268, 1992. (page 10)

[18] M. Zhu and T. Chan, "An efficient primal-dual hybrid gradient algorithm for total variation image restoration," *UCLA CAM Report*, 05 2008. (page 11)

[19] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011. (page 11)

[20] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM Journal on Imaging Sciences (SIIMS)*, vol. 3, no. 3, pp. 492–526, 2010. (page 12)

[21] F. Knoll, K. Bredies, T. Pock, and R. Stollberger, "Second order total generalized variation (TGV) for MRI," *Magnetic Resonance in Medicine*, vol. 65, no. 2, pp. 480–491, 2010. (page 12)

[22] S. Roth and M. J. Black, "Fields of experts," *International Journal of Computer Vision*, vol. 82, no. 2, pp. 205–229, 2009. (page 13)

[23] 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), Seattle, WA, USA*, pp. 7546–7555, June 2020. Available: `https://arxiv.org/abs/2001.05005`. (page 13, 17)

[24] A. Barbu, "Training an active random field for real-time image denoising," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2451–2462, 2009. (page 14)

[25] J. Domke, "Generic methods for optimization-based modeling," in *International Conference on Artificial Intelligence and Statistics, La Palma, Canary Islands*, pp. 318–326, 2012. (page 14)

[26] Y. Chen, W. Yu, and T. Pock, "On learning optimized reaction diffusion processes for effective image restoration," in *28th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA*, June 2015. (page 14, 25)

[27] E. Kobler, T. Klatzer, K. Hammernik, and T. Pock, "Variational networks: Connecting variational methods and deep learning," in *39th German Conference on Pattern Recognition (GCPR 2017), Basel, Switzerland*, pp. 281–293, September 2017. (page 14)

[28] A. Effland, E. Kobler, K. Kunisch, and T. Pock, "Variational networks: An optimal control approach to early stopping variational methods for image restoration," *Journal of Mathematical Imaging and Vision*, vol. 62, no. 3, pp. 396–416, 2020. (page 14, 25)

[29] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, "Learning a variational network for reconstruction of accelerated MRI data," *Magnetic Resonance in Medicine*, vol. 79, no. 6, pp. 3055–3071, 2018. (page 14, 20, 25, 26, 27, 28, 43, 45)

[30] 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, 2019. (page 15)

[31] A. Sriram, J. Zbontar, T. Murrell, A. Defazio, C. L. Zitnick, N. Yakubova, F. Knoll, and P. Johnson, "End-to-end variational networks for accelerated MRI reconstruction," *arXiv preprints*, 2020, arXiv:2004.06688 [eess.IV]. (page 16, 50, 55)

[32] 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, 2004. (page 16, 28)

[33] J. Schlemper, C. Qin, K. Hammernik, and J. Duan, "Σ-net: Ensembled iterative deep neural networks for accelerated parallel MR image reconstruction," in *33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada*, December 2019. (page 16, 50)

[34] Anonymous author(s), "An auto-calibrating deep learning algorithm for undersampled MRI reconstruction," in *33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada*, December 2019. (page 16, 23)

[35] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, pp. 234–241, 2015. Available: `https://arxiv.org/abs/1505.04597`. (page 17)

[36] A. Sriram, J. Zbontar, T. Murrell, C. L. Zitnick, A. Defazio, and D. K. Sodickson, "GrappaNet: Combining parallel imaging with deep learning for multi-coil MRI reconstruction," *arXiv preprints*, 2019, arXiv:1910.12325 [eess.IV]. (page 17, 50, 55)

[37] S. Wang, H. Cheng, L. Ying, T. Xiao, Z. Ke, H. Zheng, and D. Liang, "DeepcomplexMRI: Exploiting deep residual network for fast parallel MR imaging with complex convolution," *Magnetic Resonance Imaging*, vol. 68, pp. 136–147, 2020. Available: `https://arxiv.org/abs/1906.04359`. (page 21, 50)

[38] W. Bian, Y. Chen, and X. Ye, "Deep parallel MRI reconstruction network without coil sensitivities," *arXiv preprints*, 2020, arXiv:2008.01410 [eess.IV]. (page 22, 50)

[39] J. Zhang, C. Liu, and M. E. Moseley, "Parallel reconstruction using null operations," *Magnetic Resonance in Medicine*, vol. 66, no. 5, pp. 1241–1253, 2011. (page 23, 25, 30)

[40] K. Hammernik, F. Knoll, D. Sodickson, and T. Pock, "L2 or not L2: Impact of loss function design for deep learning MRI reconstruction," in *Proceedings of the International Society of Magnetic Resonance in Medicine (ISMRM)*, 2017. (page 26, 50)

[41] *PyTorch*. `https://pytorch.org/`. (page 26, 27)

[42] *TensorFlow*. `https://www.tensorflow.org/`. (page 26)

[43] *CUDA Toolkit Documentation*. `https://docs.nvidia.com/cuda/`. (page 27)

[44] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprints*, 2014, arXiv:1412.6980 [cs.LG]. (page 28)

[45] D. Lee, J. Yoo, and J. C. Ye, "Deep artifact learning for compressed sensing and parallel MRI," *arXiv preprints*, 2017, arXiv:1703.01120 [cs.CV]. (page 50)

[46] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2017. Available: `https://arxiv.org/abs/1511.08861`. (page 50)

[47] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1398–1402, 2003. (page 50)