

The influence of hybrid modelling on deep learning-based MRI reconstruction performance

Mathijs Vanhaverbeke

Thesis submitted for the degree of
Master of Science in
Biomedical Engineering, option
Biomedical Data Analytics

Supervisors:

Dr. ir. D. Christiaens
Dr. T. Dresselaers

Academic year 2023 – 2024

The influence of hybrid modelling on deep learning-based MRI reconstruction performance

Mathijs Vanhaverbeke

Thesis submitted for the degree of
Master of Science in
Biomedical Engineering, option
Biomedical Data Analytics

Supervisors:

Dr. ir. D. Christiaens
Dr. T. Dresselaers

Assessors:

Prof. dr. M. Blaschko
Dr. W. Gsell

© Copyright KU Leuven

Without written permission of the supervisors and the author it is forbidden to reproduce or adapt in any form or by any means any part of this publication. Requests for obtaining the right to reproduce or utilize parts of this publication should be addressed to Faculteit Ingenieurswetenschappen, Kasteelpark Arenberg 1 bus 2200, B-3001 Leuven, +32-16-321350.

A written permission of the supervisors is also required to use the methods, products, schematics and programmes described in this work for industrial or commercial use, and for submitting this publication in scientific contests.

Preface

I want to express my gratitude to the Medical Imaging Research Center for enabling this thesis through their provision of opportunity, infrastructure, and expertise. Special thanks are due to Daan and Tom for their guidance throughout this research process. They consistently offered assistance whenever I encountered challenges or had inquiries. Tom, I thoroughly enjoyed our data collection process and your more physics-based way of looking at scientific problems was really enriching for me. Additionally, I am deeply thankful to Dominique Delaere for imparting knowledge on navigating remote server environments, demonstrating incredible patience in troubleshooting and providing assistance whenever needed. Last but not least, I extend heartfelt appreciation to my family and friends for their unwavering support throughout this journey. It has been a fulfilling and enlightening experience, and I am grateful to have shared it with each of you.

Mathijs Vanhaverbeke

Contents

Preface	i
Abstract	iii
List of Figures	iv
List of Tables	vi
List of Abbreviations	viii
1 Introduction	1
2 Literature Review	5
2.1 Principles of Magnetic Resonance Imaging	6
2.2 Classical reconstruction techniques for undersampled k-space data	8
2.3 Background on Artificial Neural Networks	14
2.4 DL-based reconstruction techniques for undersampled k-space data	19
2.5 Conclusion	28
3 Methods: Data Analysis	29
3.1 Used dataset	30
3.2 Preprocessing	34
3.3 Reconstruction evaluation metrics	38
3.4 Conclusion	40
4 Methods: The Hybrid Models	41
4.1 Hybrid Neural Network pipeline architecture	41
4.2 Training paradigm	43
4.3 Conclusion	44
5 Results & Discussion: Model Training And Evaluation	45
5.1 Model training	46
5.2 Model evaluation process	48
5.3 Model evaluation outcomes	51
5.4 Conclusion	58
6 Conclusion	59
A Reconstruction examples	61
Bibliography	69

Abstract

The influence of hybrid modelling on deep learning-based MRI reconstruction performance.

Magnetic Resonance imaging is a crucial diagnostic tool, particularly valued for its superior soft tissue contrast and safety compared to other modalities like Computed Tomography imaging. However, long acquisition times pose challenges for patient comfort and healthcare efficiency. Unfortunately, naive acceleration of the Magnetic Resonance imaging process often results in image artifacts, compromising the diagnostic quality of the images.

Traditional image reconstruction methods such as GRAPPA, SENSE, and CS have been the standard in clinical practice to mitigate this, but are limited by moderate acceleration factors and long iteration times. Recent advancements in deep learning have shown promise in addressing these limitations. Nonetheless, such models remain inherently data-hungry. In an attempt to deal with this, the novel approach of hybrid modelling has emerged, integrating classical reconstruction techniques with deep learning models to potentially improve reconstruction performance for the same amount of training data available.

This work employed the fastMRI brain dataset to train and evaluate three different hybrid neural network models: GRAPPA U-Net, SENSE U-Net, and CS U-Net. Additionally, two conventional deep learning baseline models were created: ZF (ACS) U-Net and ZF (no ACS) U-Net, trained with and without the presence of an autocalibration signal region in k-space respectively. These models were evaluated on test scans accelerated with acceleration factors of $R=4$ and $R=8$. Used evaluation metrics include traditional measures like the NMSE, PSNR, MSSIM index, and VGG loss, as well as a newly proposed Singular Value Decomposition Metric to assess error structure.

Experimental results reveal that hybrid models significantly outperform conventional deep learning and classical reconstruction methods, especially in the absence of an autocalibration signal region in k-space. The CS U-Net model demonstrated the highest resilience, effectively handling an acceleration factor of at least $R=4$ without noticeable artifacts in inspected test scans. These findings underscore the potential of hybrid modelling to enhance Magnetic Resonance image reconstruction, suggesting that further optimization and exploration of hybrid architectures could yield even greater improvements.

List of Figures

2.1	Illustration of equation 2.11. Both its inverse nature and its different components become clear. Magnitudes are shown. Image taken from Hou <i>et al.</i> [27].	9
2.2	Illustration of the GRAPPA workflow. Images taken from [32].	12
2.3	Illustration of the SENSE reconstruction process for an acceleration factor of R=3, utilizing 4 receiver coils. In this scenario, the signals originating from three equidistant pixels within the object (depicted on the left) converge or 'fold' onto a single pixel in each of the downsized (for illustrative purposes) coil images. Image taken from Hamilton <i>et al.</i> [25].	13
2.4	Illustration of the full GrappaNet model. The first network, referred to as f_1 , processes the multi-coil k-space data that underwent R-fold undersampling and transforms it into k-space data mimicking R'-fold undersampling with the same number of coils ($R' < R$). Subsequently, the GRAPPA kernel \mathbf{K} , separately estimated via the ACS region, is applied to this result. This process then fills in the last missing k-space data points. The value of R' is deliberately kept small enough to ensure accurate image reconstruction using conventional parallel imaging methods like GRAPPA. In the original work, R'=2. Image taken from Sriram <i>et al.</i> [24].	25
2.5	Illustration of DeepMRIRec. Image taken from Alam <i>et al.</i> [58].	25
3.1	Exemplary raw data of an MR scan volume slice, part of the fastMRI brain dataset. The displayed data was acquired through a FLAIR imaging sequence and measured with 16 coils and an MR image slice matrix size of 640x320. Note that no preprocessing was applied here yet, and that the slice has a FOV of 320x320.	32
3.2	Illustration of the different k-space sampling mask types used to train the three different hybrid model pipelines from Chapter 4. Here, R=4.	35
3.3	Illustration of the magnitude of the coil sensitivities \mathcal{S}_i calculated by the BART toolbox for the data shown in Figure 3.1.	37
4.1	The general architecture of the investigated HNN pipelines.	43

5.1	Learning curves displaying the evolution of model performance on the validation dataset during training, once for ZF (ACS) U-Net trained on R=4 training data (A) and once for ZF (ACS) U-Net trained on a mixture of R=4 and R=8 training data (B). The x-axis displays the training epochs converted into training steps (here, equal to the number of image slices in the training dataset times the number of epochs) and the learning curve itself is slightly smoothed for visual purposes, though the original curve is still shown with a lower opacity. The y-axis displays the loss values, calculated and averaged across the validation dataset.	47
5.2	Results of the SVDM verification experiments described in section 5.2. These graphs display the influence of Gaussian noise (A), Gaussian blurring (B), the acceleration factor R (C), and the ACS fraction size (D) on the SVDM. Std = standard deviation.	49
5.3	Figures facilitating the desired intra-paradigm comparisons based on the data displayed in Table 5.2. Each column of figures visualizes an intra-paradigm comparison, with each column-figure analyzing one evaluation metric. In each figure, data is displayed for test scans with R=4 (orange), R=8 (green), or all test scans together (blue). The left column (A) focuses on the SENSE reconstruction paradigm, the middle column (B) focuses on the CS reconstruction paradigm, and the right column (C) focuses on the GRAPPA reconstruction paradigm.	53
A.1	Magnitude of the ground truth of the test scan volume slice of the reconstructions presented in Figures A.2, A.3, and A.4.	61
A.2	Reconstruction outputs of the different intra-SENSE paradigm reconstruction methods: ZF (SENSE mask) (A), SENSE (B), ZF (no ACS) U-Net (C), and SENSE U-Net (D). This concerns the same test scan volume slice as the one shown in Figure A.1. Here, R=4. For each reconstruction, the magnitude of the slice of interest is shown, together with its residual error map and evaluation metric values.	62
A.3	Reconstruction outputs of the different intra-CS paradigm reconstruction methods: ZF (CS mask) (A), CS (B), ZF (no ACS) U-Net (C), and CS U-Net (D). This concerns the same test scan volume slice as the one shown in Figure A.1. Here, R=4. For each reconstruction, the magnitude of the slice of interest is shown, together with its residual error map and evaluation metric values.	63
A.4	Reconstruction outputs of the different intra-GRAPPA paradigm reconstruction methods: ZF (GRAPPA mask) (A), GRAPPA (B), ZF (ACS) U-Net (C), and GRAPPA U-Net (D). This concerns the same test scan volume slice as the one shown in Figure A.1. Here, R=4. For each reconstruction, the magnitude of the slice of interest is shown, together with its residual error map and evaluation metric values.	64
A.5	Magnitude of the ground truth of the test scan volume slice of the reconstructions presented in Figures A.6, A.7, and A.8.	65

LIST OF FIGURES

A.6 Reconstruction outputs of the different intra-SENSE paradigm reconstruction methods: ZF (SENSE mask) (A), SENSE (B), ZF (no ACS) U-Net (C), and SENSE U-Net (D). This concerns the same test scan volume slice as the one shown in Figure A.5. Here, R=8. For each reconstruction, the magnitude of the slice of interest is shown, together with its residual error map and evaluation metric values.	66
A.7 Reconstruction outputs of the different intra-CS paradigm reconstruction methods: ZF (CS mask) (A), CS (B), ZF (no ACS) U-Net (C), and CS U-Net (D). This concerns the same test scan volume slice as the one shown in Figure A.5. Here, R=8. For each reconstruction, the magnitude of the slice of interest is shown, together with its residual error map and evaluation metric values.	67
A.8 Reconstruction outputs of the different intra-GRAPPA paradigm reconstruction methods: ZF (GRAPPA mask) (A), GRAPPA (B), ZF (ACS) U-Net (C), and GRAPPA U-Net (D). This concerns the same test scan volume slice as the one shown in Figure A.5. Here, R=8. For each reconstruction, the magnitude of the slice of interest is shown, together with its residual error map and evaluation metric values.	68

List of Tables

2.1	Overview of the most important MR image reconstruction methods found in the current literature. ACS = autocalibration signal, \mathcal{S}_i = coil sensitivity map estimations.	27
3.1	The different sequences part of the fastMRI brain dataset [7].	31
3.2	The official fastMRI brain dataset split.	33
5.1	Results of the comparative experiment described in subsection 3.1.3 investigating the influence of R_{training} on model performance. The model performances were evaluated on test scans with $R=4$, $R=8$, or on all test scans together. Average metric values are displayed. Performance differences for $R_{\text{evaluation}}=4+8$ are statistically significant for every metric (Wilcoxon signed-rank test: $p_{\text{corrected}}<.0001$).	46
5.2	The results of all performed reconstruction method evaluations. The model performances were evaluated on test scans with $R=4$, $R=8$, or on all test scans together. Average metric values are displayed. Performance differences for $R_{\text{evaluation}}=4+8$ are statistically significant for every metric and HNN model (Wilcoxon signed-rank test: $p_{\text{corrected}}<.0001$). Values are rounded to have three significant figures.	52

List of Abbreviations

Abbreviations

ACS	Autocalibration Signal
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
CS	Compressed Sensing
CT	Computed Tomography
DL	Deep Learning
FOV	Field of View
FT	Fourier Transform
GAN	Generative Adversarial Network
GRAPPA	Generalized Autocalibrating Partial Parallel Acquisition
HNN	Hybrid Neural Network
IFT	Inverse Fourier Transform
MR	Magnetic Resonance
MRI	Magnetic Resonance Image
MSSIM	Mean Structural Similarity
NMSE	Normalized Mean Squared Error
PReLU	Parametric Rectified Linear Unit
PSNR	Peak Signal to Noise Ratio
RF	Radiofrequency
ReLU	Rectified Linear Unit
SENSE	Sensitivity Encoding
SVDM	Singular Value Decomposition Metric
SNR	Signal-to-noise ratio
SSIM	Structural Similarity
TE	Echo Time
TR	Repetition Time
ZF	Zero-Filling

Chapter 1

Introduction

Magnetic Resonance (MR) imaging is an imaging modality which provides unparalleled insights into the human body's intricate structures, particularly in the context of neurological and soft tissue related diseases. Naturally, MR imaging has become indispensable in modern medical diagnostics. Advantages of MR images over images from other medical imaging techniques are related to both the resulting image quality and the potential health hazards coupled with the imaging process.

To briefly elucidate, MR images exhibit a superior soft tissue contrast compared to images from other imaging modalities, such as e.g. Computed Tomography (CT) images [1][2]. The consequence of this is two fold; water, muscle, fat and other soft tissues can be differentiated more clearly, and contrast agents may be used less frequently for the MR imaging process than for other imaging processes. The latter advantage also implies an improvement in the safety of the image acquisition process, since the use of contrast agents carries some risk and may lead to kidney failure for patients who require numerous contrast scans throughout their lives [1][3]. Additionally, magnetic resonance, the physical principle on which MR imaging relies, doesn't require the use of ionizing radiation [2][4]. This makes MR imaging a much safer process, since ionizing radiation dose is a limiting factor in multiple other imaging modalities such as e.g. x-ray or CT imaging [5].

However, despite these upsides to the MR imaging process, an MR scan still needs to be artifact-free and have an appropriately high enough resolution to have any diagnostic value. This means that a lot of data needs to be collected during an MR scan, as the spatial resolution and the quality of an MR image is proportional to the amount of data collected [2]. Due to this, clinical MR image acquisition can be a lengthy process, sometimes exceeding 30 minutes depending on the body part and sequences being imaged [6][7]. This is problematic both from the patient's point of view and the healthcare provider's point of view for multiple reasons.

From the patient's point of view, it can be said that a combination of the long acquisition time with several other properties of the MR imaging process results in some degree of patient discomfort. First and foremost, patients have to stay completely still during an MR scan in order to avoid motion artifacts, often

1. INTRODUCTION

including ghosting and blurring in the image [8]. This necessity for immobility can be particularly challenging for long scan times, especially for pediatric patients, who are sometimes even sedated because of this despite the potential risks associated with anesthesia [1]. Next to this, MR scanners can, depending on the image acquisition sequence, generate substantial acoustic noise. This is because MR scanners use strong magnetic fields with rapidly switching magnetic field gradients (see Chapter 2) which generate large Lorentz forces acting on the scanner's gradient coils themselves, thus creating loud knocking noises [1][9]. This acoustic noise further diminishes patient comfort during longer MR scans [9]. A last source of patient discomfort linked to the long MR image acquisition time is related to the deposition of radiofrequency (RF) power in the patient's body, leading to tissue heating, and since its intensity is proportional to the on-time of the scanner, patient discomfort is again related to the duration of the MR imaging process [1][10].

From the healthcare provider's point of view, it can be said that the long acquisition time of the MR imaging process may negatively affect the efficacy of the clinical pathology screening and evaluation process. For example, as previously mentioned, long MR evaluations are often paired with the presence of motion artifacts in the resulting image: studies even report the presence of motion artifacts in up to 29.4% of inpatient MR examinations at a single institution [11]. These motion artifacts may obscure critical anatomical details from the physician, putting the healthcare patient at risk of inappropriate treatment or misdiagnosis. Because of this, MR scan sequences may need to be repeated, leading to unnecessary economic costs for the healthcare provider which can range from anywhere between \$92.600 and \$139.000 per scanner per institution per year [11]. Next to this, the MR image acquisition duration also has a direct impact on the institution's finances; MR scanners are typically associated with a base-case operational cost estimate of \$530-\$590 per hour [11][12]. Lengthier MR image acquisitions are thus also leading to a higher cost per MR image. Additionally, long acquisition times may render certain time-sensitive MR imaging procedures like fetal or cardiac MR imaging, where target movement is unavoidable and temporal resolution is important, unfeasible without costlier hardware and decreasing the image's spatial resolution [4]. Lastly, long acquisition times also lead to long waiting times or access times for the healthcare provider's patients. A study which was performed at the University Hospital of Mannheim showed that the average access time, excluding emergencies, was 3 days (with an interquartile range of 5 days) for inpatients and 44 days (with an interquartile range of 47 days) for outpatients [13]. Timely access to MR scanners is critical for early disease detection, ensuring that treatment can commence at the earliest stages, significantly impacting patient outcomes and healthcare costs associated with advanced disease stages.

The benefits of shorter scan times are thus evident, and consequently, the pursuit of enhanced MR imaging speed has remained a prominent research objective ever since the inception of MR imaging in the 1970s [7]. However, simply reducing the MR image acquisition time by reducing the number of MR measurements taken without enhancing the image reconstruction process can lead to a new set of challenges since,

as previously stated, the quality of an MR image is proportional to the number of MR measurements. Decreasing the scan duration this way thus often results in a lower signal-to-noise ratio (SNR), decreased spatial resolution, and increased sensitivity to image artifacts such as aliasing artifacts. Because of this, the primary objective of numerous reconstruction techniques has been the reconstruction of high-resolution and artifact-free or artifact-suppressed MR images from undersampled or flawed MR measurements [2].

Many deterministic, classical reconstruction techniques concerning the above problem statement already exist, with GRAPPA reconstruction [14], SENSE reconstruction [15], and Compressed Sensing (CS) based reconstruction [16][17] perhaps being the most noteworthy classical reconstruction techniques [2][18]. However, although these reconstruction methods have made certain achievements, they still face the challenges of long iteration times and relatively low acceleration rates [19]. Therefore, in recent years, deep learning (DL) methods have also been put forward for the MR image reconstruction task. By training a DL model on fully sampled MR data, it is possible to reconstruct high-quality images from undersampled data. Since DL methods are very flexible and data-driven, these techniques offer the potential to significantly enhance the speed and quality of MR image reconstruction even further compared to the aforementioned classical reconstruction methods; comparative studies have already shown that DL methods can match and even exceed the ability of classical reconstruction methods to reduce the time required for data acquisition without sacrificing image quality [2][19].

Nevertheless, even with these initial encouraging findings, DL techniques still exhibit a statistical character, necessitating substantial amounts of data for intricate tasks like MR image reconstruction from undersampled measurements. Consequently, recent research is in its preliminary phases of exploring the potential fusion of established classical reconstruction methods with newly proposed DL models to alleviate some of this data hunger (see subsection 2.4.2). The author of this work proposes to term these new models as '**hybrid models**'.

Unfortunately, thus far, only a few of these hybrid models have been published, and they appear to concentrate solely on frameworks involving GRAPPA, which is just one of the existing classical reconstruction techniques available. Moreover, no comprehensive scientific performance comparison studies have been conducted yet to ascertain if these hybrid models truly represent a new, optimal approach.

In this work, a concise overview of MR image acquisition and the most pertinent classical and DL-based reconstruction methods will be presented, followed by a contribution to the consideration of DL-based methods for the MR image reconstruction task using undersampled MR data by comparing and investigating new hybrid neural network pipelines for the task at hand.

1. INTRODUCTION

Research questions This thesis will investigate the following research questions: **Does hybrid modelling have the capacity to increase the performance of the exemplary conventional deep learning-based MRI reconstruction pipeline investigated in this thesis? Which classical reconstruction method has a higher synergy with deep learning models? What is the largest acceleration factor that can reasonably be achieved with the proposed models?**

Thesis outline In Chapter 2, a review of the current literature on MR image reconstruction will be conducted, encompassing both classical reconstruction algorithms and neural network-based methods. The components utilized in the design of these neural networks will also be discussed. Furthermore, unexplored avenues within the existing literature, which hold potential significance for this study or future research, such as hybrid modelling, will be highlighted.

Deep neural networks possess an inherent data requirement due to the extensive number of learnable parameters in their architecture. Consequently, Chapter 3 will discuss the dataset employed for network training. Additionally, the various preprocessing steps applied to that dataset will equally be elaborated upon in this chapter. Lastly, different evaluation metrics that will be used further on are discussed, with even a new evaluation metric being introduced.

Chapter 4 will then describe the proposed network architecture and the training paradigm utilized for the experiments presented and discussed in Chapter 5. In this subsequent chapter, these experiments are given further explanations where necessary and their results are discussed in search for an answer to the proposed research questions of this work.

In a last chapter, Chapter 6, these research questions will finally be answered.

Chapter 2

Literature Review

This literature review provides a background on neural networks for MR image reconstruction using undersampled MR data. Its sections also aim to position this topic in the broader scope of MR image reconstruction in general.

First, a concise background on magnetic resonance imaging is provided. This section will provide fundamental concepts and a mathematical basis required to grasp the general problem statement as well as the content of the other sections.

This is followed by the discussion of the most noteworthy classical reconstruction techniques used for MR image reconstruction using undersampled MR data. These techniques are still regarded as the golden standard for MR image reconstruction, as these techniques are used by most MR scanners in clinical practice [2].

Next, a general background on Artificial Neural Networks (ANNs) in light of MR image reconstruction is given, followed by a last section which discusses significant state-of-the-art deep learning methods for MR image reconstruction. This section mainly focuses on trained neural networks, although small sections about untrained neural networks and other extensions to the MR data reconstruction problem are provided at the end as well. Explicit distinctions between conventional deep learning models and hybrid deep learning models are made through the use of subsections, as this is relevant for the research question of this thesis.

2. LITERATURE REVIEW

2.1 Principles of Magnetic Resonance Imaging

The fundamental principles of Magnetic Resonance (MR) imaging are well-explained in [20] and [21]. All of the information in this section is derived from these works.

MR imaging is based on the physical principle of MR; the interaction of elementary particles, like protons, with radiofrequency (RF) waves in the presence of a magnetic field. Protons, also referred to as hydrogen atoms and highly prevalent in living organisms [4], have the intrinsic property of spin, which is quantized according to quantum theory: for protons there are two states of spin, spin up ($s_z = \frac{1}{2}$) and spin down ($s_z = -\frac{1}{2}$). If no external magnetic field is present, then each spin state is equally likely to occur as both spin states have an identical spin state energy level. However, if an external magnetic field $\mathbf{B} = (0,0,B_0)$ is present, more protons will exhibit a spin state which matches with the direction of the magnetic field vector as this spin state has a lower associated energy level:

$$E = -\mu_z B_0 = -\gamma J_z B_0 = -\gamma \hbar s_z B_0 \quad (2.1)$$

with μ_z the z-component of the magnetic dipole moment of the proton, γ the gyromagnetic ratio of the proton, J_z the z-component of the spin angular momentum of the proton, \hbar the reduced Planck constant and s_z the spin state of the proton in the z-direction. This means that for a given volume element ΔV , a net magnetization vector $\mathbf{M} = \sum_{j \in \Delta V} \boldsymbol{\mu}_j$ will arise, aligned with and precessing around the external magnetic field \mathbf{B} . From equation 2.1, it also follows that the energy required for a transition from a spin up state to a spin down state equals $\Delta E = \hbar \gamma B_0 = \hbar \omega_0$, with $\omega_0 = \gamma B_0$ being called the resonance (angular) frequency or Larmor frequency. This corresponds to the absorption of a photon of an electromagnetic wave with a frequency of $f_0 = \frac{\omega_0}{2\pi}$, as $E_{photon} = hf$. From a classical mechanics point of view, by applying an RF pulse at the resonance frequency for a given amplitude and time interval, the net magnetization vector \mathbf{M} may thus be 'flipped' or 'rotated' with respect to the z-axis over an angle α , creating a transversal magnetization vector component which induces signals in the MR scanner's receiver coils before relaxing back to the equilibrium state. It's these signals which eventually allow the creation of a full MR image measurement, by repeatedly alternating RF pulse excitations and MR measurements with a repetition time TR and an echo time TE.

At each MR measurement, multiple MR signals $s_r^{TR,TE}$ from multiple voxels r are being measured at once, leading to a measured signal S :

$$S^{TR,TE} = \sum_r s_r^{TR,TE} \quad (2.2)$$

This creates an entanglement problem, where signals from different voxels can't be differentiated. In order to obtain spatial information, phase shifts are introduced between the precessing net magnetization vectors of different voxels:

$$S^{TR,TE} = \sum_r s_r^{TR,TE} e^{-j\phi_r} \quad (2.3)$$

2.1. Principles of Magnetic Resonance Imaging

This is done by making the external magnetic field \mathbf{B} spatially varying, as the precessing frequency, equal to the resonance (angular) frequency ω_0 , is related to \mathbf{B} via $\omega_0 = \gamma B_0(x, y, z)$. The external magnetic field \mathbf{B} then becomes:

$$\mathbf{B} = (0, 0, B_0(x, y, z)) = (0, 0, B_0 + \mathbf{G} \cdot \mathbf{r}) \quad (2.4)$$

with \mathbf{G} the linear magnetic field gradient, generated via magnetic gradient coils, and \mathbf{r} the spatial location of the voxel of interest. Assuming no motion of the protons, this results in a phase shift of:

$$\phi(\mathbf{r}) = \gamma(\mathbf{G} \cdot \mathbf{r})\Delta t = 2\pi\mathbf{k} \cdot \mathbf{r} \quad (2.5)$$

with \mathbf{k} an auxiliary variable completely determined by machine parameters, i.e. the sequence of applied gradients $\mathbf{G}(t)$:

$$2\pi\mathbf{k} = \gamma \int_0^{TE} \mathbf{G}(\tau) d\tau \quad (2.6)$$

Combination of equations 2.5 and 2.6 with the continuous form of equation 2.3 results in the now-famous reconstruction basis of MR images:

$$S(\mathbf{k}) = \int_{\mathbf{r}} s(\mathbf{r}) e^{-j\phi(\mathbf{r})} d\mathbf{r} = \int_{\mathbf{r}} s(\mathbf{r}) e^{-j2\pi\mathbf{k} \cdot \mathbf{r}} \quad (2.7)$$

From the mathematical form of equation 2.7, it follows that $S(\mathbf{k})$ is the Fourier Transform (FT) of $s(\mathbf{r})$. MR image measurements, $S(\mathbf{k})$, are thus actually performed in the frequency domain, which is termed the **k-space** in the special case of the MR imaging framework. In order to reconstruct the data or the image $s(\mathbf{r})$ in the image domain, many measurements for different values of \mathbf{k} are performed, ideally until the k-space is fully sampled (defined by the required image resolution), followed by the application of an Inverse Fourier Transform (IFT) to the k-space data.

Typically, every line in k-space (repeatedly sampled either along k_x or k_y) is collected following a single RF excitation pulse. This direction in k-space is termed the frequency encoding direction, while its perpendicular in-slice direction is named the phase encoding direction. The phase encoding direction is thus the direction which has the most significant impact on the MR image acquisition time. It is also important to note that, since the signals induced in the MR scanner's receiver coils have a phase as well as a magnitude, the values of $S(\mathbf{k})$ are complex.

An important implication following from the FT-relationship established by equation 2.7, is that the Nyquist criterion needs to be respected in the k-space in order to avoid aliasing in the image domain. From this, it follows that e.g. for the k_x direction in k-space:

$$\frac{1}{\Delta k_x} \geq 2x_{max} = FOV_x \quad (2.8)$$

This explains why plain undersampling without modifying the image reconstruction method, as mentioned in Chapter 1, causes aliasing artifacts and why acceleration of the MR acquisition via simplistic undersampling is not feasible.

2.2 Classical reconstruction techniques for undersampled k-space data

As previously derived, simply applying the IFT to undersampled k-space data results in aliasing artifacts. MR image reconstruction using undersampled k-space data, aiming to speed up the MR image acquisition, thus requires a new mathematical reconstruction method. For this, it is useful to pose the MR image reconstruction problem as an inverse problem [22]. Considering one 2D, single-coil, fully sampled k-space measurement, the inverse problem formulation becomes:

$$\mathbf{y}^{full} = \mathcal{F}\mathbf{x} + \boldsymbol{\eta} \quad (2.9)$$

where $\mathbf{y}^{full} \in \mathbb{C}^{N_1 \times N_2}$ is the fully sampled 2D k-space data, $\mathbf{x} \in \mathbb{C}^{N_1 \times N_2}$ is the MR image that needs to be reconstructed, \mathcal{F} is the FT operator and $\boldsymbol{\eta}$ is random noise, typically assumed to have a zero-mean Gaussian distribution [2].

Continuing the formulation of the reconstruction problem at hand in the same framework, one can represent the case where undersampled k-space data is considered as follows:

$$\mathbf{y} = \mathcal{M} \odot \mathcal{F}\mathbf{x} + \boldsymbol{\eta} \quad (2.10)$$

where \mathbf{y} is the undersampled k-space data, \mathcal{M} is a binary sampling mask and \odot is the element-wise multiplication operator [23]. Here, the binary sampling mask represents the physical undersampling process in k-space. This new inverse mathematical framework now serves as the basis of the classical reconstruction techniques that are discussed below.

2.2.1 Parallel Imaging

Parallel imaging stands as a pivotal advancement in the realm of MR scan time reduction, establishing its indispensability in clinical and research applications. It is unique in the sense that it is also a hardware development besides a software development; parallel imaging entails the deployment of multiple physical receiver coils in the MR scanner to capture multiple distinct perspectives of the imaged subject, which can then be intelligently combined to reduce the required number of k-space samples necessary for the obtention of an artifact-free MR image [18][24].

This leads to a system of N_c equations, where the individual equations are modified versions of equation 2.10:

$$\mathbf{y}_i = \mathcal{M} \odot \mathcal{F}(\mathcal{S}_i \odot \mathbf{x}) + \boldsymbol{\eta}_i \quad (2.11)$$

where $i = \{1, 2, \dots, N_c\}$ is the receiver coil index, N_c is the number of employed receiver coils that have a unique sensitivity variation along the direction that is accelerated and \mathcal{S}_i is the complex-valued coil sensitivity map of the i -th coil [25][26]. The different components of equation 2.11 are visualized in Figure 2.1.

2.2. Classical reconstruction techniques for undersampled k-space data

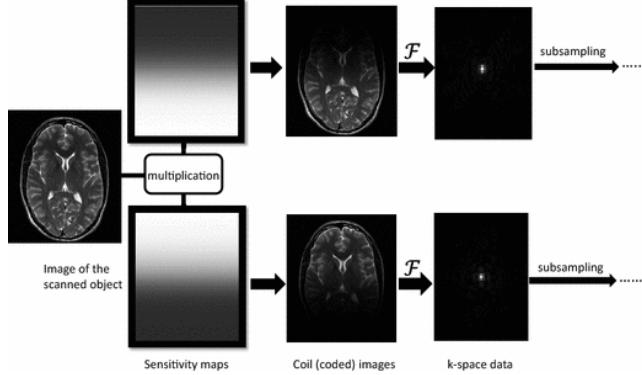


FIGURE 2.1: Illustration of equation 2.11. Both its inverse nature and its different components become clear. Magnitudes are shown. Image taken from Hou et al. [27].

Coil sensitivity maps can be estimated using different methods, either relying on complete or partial autocalibration measurements (methods like ESPIRiT, as detailed in [28]) or on mathematical relationships such as learned correlations, certain transformations or eigenvalue approaches [26][28][29]. Putting everything together, and collapsing the operation $\mathcal{M} \odot \mathcal{F}(S_i \odot \mathbf{x})$ into one operation $E\mathbf{x}$ with E called the hybrid encoding operator, the system of N_c equations can be written as:

$$\mathbf{y} = E\mathbf{x} + \boldsymbol{\eta} \quad (2.12)$$

where \mathbf{y} corresponds to the vectorized multi-coil k-space data of size $N_c N_1 N_2 \times 1$, E corresponds to the hybrid encoding operator matrix of size $N_c N_1 N_2 \times N_1 N_2$, \mathbf{x} corresponds to the vectorized MR image of size $N_1 N_2 \times 1$, and $\boldsymbol{\eta}$ corresponds to the vectorized additive white Gaussian noise of size $N_c N_1 N_2 \times 1$ [22]. If the k-space is fully sampled, then equation 2.12 is clearly overdetermined. This means that it may be solved via numerical methods such as the least squares method, which can be derived here from the maximum likelihood estimation of \mathbf{x} since the noise $\boldsymbol{\eta}$ is assumed to be Gaussian white noise [2][22]. The following can be shown for $N = N_c N_1 N_2$:

$$\begin{aligned}
 \hat{\mathbf{x}} &= \operatorname{argmax}_{\mathbf{x}}(p(\mathbf{y}|\mathbf{x})) \\
 &= \operatorname{argmax}_{\mathbf{x}}(\log(p(\mathbf{y}|\mathbf{x}))) \\
 &= \operatorname{argmax}_{\mathbf{x}}(\log(\prod_{n=1}^N \frac{1}{\sqrt{\pi 2\sigma^2}} e^{-\frac{(y_n - (E\mathbf{x})_n)^2}{2\sigma^2}})) \\
 &= \operatorname{argmax}_{\mathbf{x}}(N \log(\frac{1}{\sqrt{\pi 2\sigma^2}}) - \sum_{n=1}^N \frac{(y_n - (E\mathbf{x})_n)^2}{2\sigma^2}) \\
 &= \operatorname{argmin}_{\mathbf{x}}(\sum_{n=1}^N (y_n - (E\mathbf{x})_n)^2) \\
 &= \operatorname{argmin}_{\mathbf{x}} \|\mathbf{y} - E\mathbf{x}\|_2^2
 \end{aligned} \quad (2.13)$$

2. LITERATURE REVIEW

Note that the above least squares formulation arises from the Gaussian white noise model, and utilizing an alternative noise model would yield a different optimization formulation [22]. Equation 2.13 may continue to be solved in a general way according to the methods presented in [30]:

$$\begin{aligned}\hat{\mathbf{x}} &= \operatorname{argmin}_{\mathbf{x}} ((\mathbf{y} - \mathbf{E}\mathbf{x})^H(\mathbf{y} - \mathbf{E}\mathbf{x})) \\ &= \operatorname{argmin}_{\mathbf{x}} (\mathbf{y}^H\mathbf{y} - \mathbf{x}^H\mathbf{E}^H\mathbf{y} - \mathbf{y}^H\mathbf{E}\mathbf{x} + \mathbf{x}^H\mathbf{E}^H\mathbf{E}\mathbf{x}) \\ &= \operatorname{argmin}_{\mathbf{x}} (\mathbf{y}^H\mathbf{y} - 2\mathbf{y}^H\mathbf{E}\mathbf{x} + \mathbf{x}^H\mathbf{E}^H\mathbf{E}\mathbf{x}) \\ &= \operatorname{argmin}_{\mathbf{x}} (f(\mathbf{x}))\end{aligned}\tag{2.14}$$

$$\begin{aligned}\Rightarrow \frac{df(\mathbf{x})}{d\mathbf{x}} \Big|_{\mathbf{x}=\hat{\mathbf{x}}} &= 0 \\ \Leftrightarrow -2\mathbf{E}^H\mathbf{y} + 2\mathbf{E}^H\mathbf{E}\hat{\mathbf{x}} &= 0 \\ \Leftrightarrow \hat{\mathbf{x}} &= (\mathbf{E}^H\mathbf{E})^{-1}\mathbf{E}^H\mathbf{y}\end{aligned}\tag{2.15}$$

where H is the Hermitian transpose, as the hybrid encoding operator matrix \mathbf{E} is complex-valued [22]. The reconstructed MR image $\hat{\mathbf{x}}$ can thus be obtained by computing the Moore-Penrose generalized inverse of \mathbf{E} . This remains possible as long as the number of rows of \mathbf{E} is larger than or equal to the number of columns of \mathbf{E} , i.e. as long as the number of gradient encoding steps is reduced by no more than a maximum factor of N_c [22][31]. Concretely, this thus means that parallel imaging allows MR image acquisition with a maximal possible acceleration factor equal to the number of employed receiver coils N_c . In practical applications, however, the maximum attainable acceleration factor is typically constrained and falls short of the aforementioned theoretical limit due to the overlap and non-orthogonality of the coil sensitivities [25].

In this general, most simple parallel imaging case, computing equation 2.15 can be computationally intensive due to the typically large size of the hybrid encoding matrix \mathbf{E} . For instance, with eight coils, a 128x128 MR image matrix and 4-fold Cartesian undersampling, the matrix \mathbf{E} has 32,768 rows and 16,384 columns [25]. This makes direct inversion a cumbersome and impractical approach. However, diverse and more efficient methods to solve the inversion problem have been identified, and these approaches can generally be categorized into two primary classes (apart from some exceptions): SENSE-like methods, which operate within the image space, and GRAPPA-like methods, which operate within the k-space [24][25]. The three most notable of these methods will be discussed next.

2.2.2 GRAPPA

Generalized Autocalibrating Partial Parallel Acquisition (GRAPPA) is a k-space-based parallel imaging-based technique that makes predictions about the unrecorded k-space data points in the undersampled k-space by calculating them as a weighted sum of nearby sampled k-space data points from all the receiver coils. Since these weights are spatially invariant, their application can be thought of as a complex-valued 3D convolution in k-space, going from N_c input channels to N_c output channels [24]. Thus, in a formal sense, the unobserved k-space points, denoted as \mathbf{y}^u , are determined based on the observed k-space points, denoted as \mathbf{y} , through a convolution operation with the GRAPPA kernel, denoted as \mathbf{K} :

$$\mathbf{y}^u = \mathbf{K} \circledast \mathbf{y} \quad (2.16)$$

where \circledast represents the convolution operation [24]. Of course, GRAPPA's most critical step involves estimating the convolution kernel's weights. This estimation process follows a specific procedure. First, GRAPPA presupposes the existence of a fully sampled area within k-space, known as the **autocalibration signal** (ACS) region, typically positioned at the center of k-space [2]. Subsequently, a subset of k-space lines within this ACS region is deliberately obscured, generating a simulated dataset comprising acquired k-space points, denoted as \mathbf{y}^{ACS_1} , and missing k-space points, denoted as \mathbf{y}^{ACS_2} . The weights \mathbf{K} of the GRAPPA convolution kernel are then estimated by minimizing the least squares cost function [2][24]:

$$\hat{\mathbf{K}} = \operatorname{argmin}_{\mathbf{K}} \|\mathbf{y}^{ACS_2} - \mathbf{K} \circledast \mathbf{y}^{ACS_1}\|_2^2 \quad (2.17)$$

Once GRAPPA's kernel \mathbf{K} has been estimated, the undersampled k-space slices of the different coils can be interpolated. In theory, fully sampled multi-coil k-space data are then obtained. From this result, it is possible to generate an MR image for each receiver coil, by applying the IFT to their corresponding k-space data. This gives N_c MR images, which should each be equivalent to the term $(\mathcal{S}_i \odot \mathbf{x})$ from equation 2.11 for the i-th coil. Next, these need to be combined into one single MR image. This can be done in multiple different ways, but often the root-sum-of-squares combination method is chosen as this is both mathematically simple and near-SNR optimal [31]. The workflow of GRAPPA is summarized in Figure 2.2.

The GRAPPA technique has demonstrated excellent performance when dealing with uniform undersampling and it's the preferred method employed in product sequences of Siemens and GE scanners [2]. The self-calibrating feature of GRAPPA becomes especially intriguing in situations where obtaining coil sensitivity maps is challenging, like in cardiac and abdominal imaging or when dealing with single-shot applications [14]. Despite this, GRAPPA reconstructions can unfortunately still show aliasing or other artifacts for high acceleration factors. Generally, this happens for acceleration factors $R > 3$ [22][25].

2. LITERATURE REVIEW

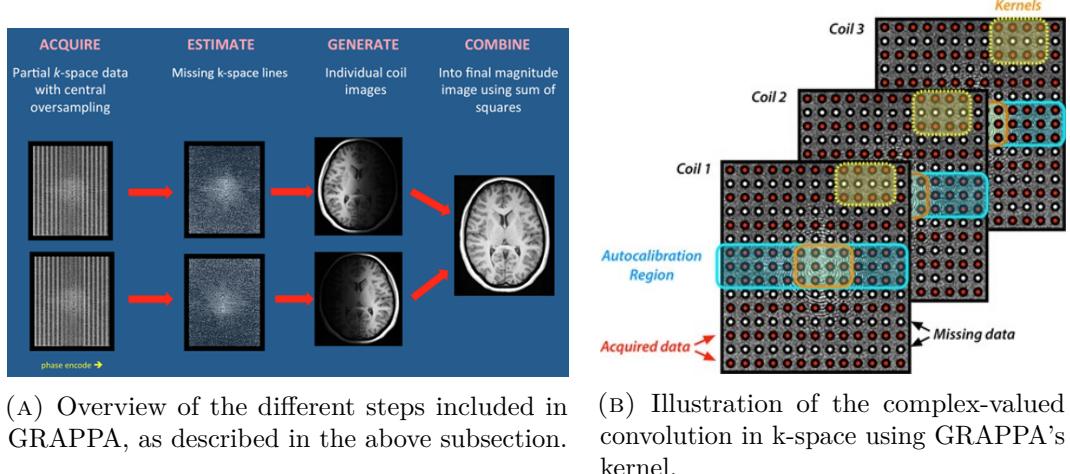


FIGURE 2.2: Illustration of the GRAPPA workflow. Images taken from [32].

2.2.3 SENSE

Sensitivity Encoding (SENSE) is an image domain-based parallel imaging-based reconstruction technique which is mathematically actually very similar to the general, most simple parallel imaging reconstruction method described by equations 2.12 - 2.15, with the key difference being that the mathematics of the SENSE framework is now fully defined in the image space. This contrasts the definition of a relationship between the k-space and the image space.

More specifically, SENSE can in this regard be seen as an 'unfolding' method which 'unfolds' image pixels that are superimposed in the image domain due to aliasing artifacts in each receiver coil's MR image [2][25]. To clarify through an example, consider a parallel imaging acquisition with an acceleration factor $R=3$ and 4 receiver coils (Figure 2.3). Undersampling then diminishes the field of view (FOV) by a factor of three, causing three pixels from the fully sampled MR image to converge onto a single pixel in the aliased MR image of each individual receiver coil. SENSE then relies on prior information regarding the coil sensitivity maps \mathcal{S}_i to separate these amalgamated pixels and restore the complete FOV MR image [25].

In the initial stage of the SENSE reconstruction process, the objective is to construct a sensitivity matrix denoted as \mathcal{S} for a specific pixel location within the aliased single-coil MR images. This matrix possesses dimensions of $N_c \times N_p$, with N_p indicating the number of superimposed full FOV pixels at that pixel location, which is equivalent to the acceleration factor R . To elaborate, the signal received by coil i at position (x, y) in the aliased single-coil image is denoted as $I_i^{aliased}(x, y)$. This signal or pixel value can be described as the cumulative effect of three pixels from the full FOV MR image which all lie on the same line parallel to the aliasing direction. These 'folded' pixels are weighted based on their respective coil sensitivities and are separated by a distance of $\frac{FOV}{R}$ in the full FOV MR image [25].

2.2. Classical reconstruction techniques for undersampled k-space data

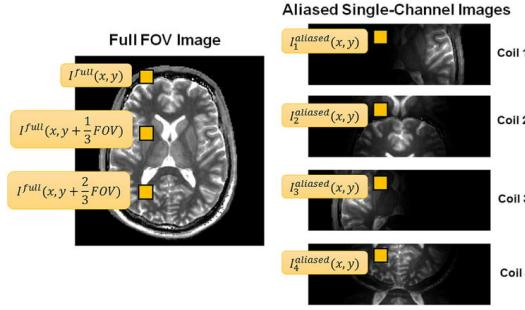


FIGURE 2.3: Illustration of the SENSE reconstruction process for an acceleration factor of $R=3$, utilizing 4 receiver coils. In this scenario, the signals originating from three equidistant pixels within the object (depicted on the left) converge or 'fold' onto a single pixel in each of the downsized (for illustrative purposes) coil images. Image taken from Hamilton *et al.* [25].

This insight into the superposition of pixels in the single-coil MR images leads to a system of equations for each pixel location (x, y) in these single-coil MR images:

$$\mathbf{a} = \mathcal{S}\mathbf{v} \quad (2.18)$$

where \mathbf{v} is an $N_p \times 1$ vector that holds the unfolded MR image pixels and \mathbf{a} is an $N_c \times 1$ vector that holds the signals from all individual receiver coils at a particular location (x, y) within the aliased images. If the sensitivity matrix is square and invertible, equation 2.18 can be solved by multiplying both sides by the inverse of \mathcal{S} . Otherwise, the Moore-Penrose generalized inverse of the sensitivity matrix is employed, very similar to equation 2.15 [25]:

$$\mathbf{v} = (\mathcal{S}^H \mathcal{S})^{-1} \mathcal{S}^H \mathbf{a} \quad (2.19)$$

SENSE is among the two primary parallel imaging techniques commonly employed in clinical settings, with the other being GRAPPA. Most MR scanner manufacturers offer the possibility of a SENSE-based reconstruction and it's common practice to accelerate MR image scans by a factor of two or three, depending on the coil geometry [25]. This means that SENSE's performance is on par with the performance of GRAPPA when it comes to allowed acceleration factors.

2.2.4 Compressed Sensing

A last noteworthy parallel imaging-based MR image reconstruction technique is based on Compressed Sensing (CS) theory. In CS-based MR image reconstruction, CS methods have been integrated with the general framework of parallel imaging explained in subsection 2.2.1 in order to try to attain greater acceleration factors compared to what can be achieved using GRAPPA reconstruction, SENSE reconstruction or general parallel imaging reconstruction alone [24]. This reconstruction method can be explained by looking at equation 2.13 again. As previously shown, this equation can be solved if the system of equations 2.12 is overdetermined. However, the system

2. LITERATURE REVIEW

of equations 2.12 becomes underdetermined if the k-space data is undersampled by an acceleration factor larger than N_c , making the least squares cost function ill-posed, causing it to lack a unique solution. Hence, to address the ill-conditioned least squares cost function, often a regularizer term $R(\mathbf{x})$ is introduced into the optimization problem, in order to incorporate extra assumptions or prior knowledge about the reconstructed MR image into equation 2.13 [2][22]. This regularizer serves the purpose of either isolating the solution or stabilizing it [22]. This results in the following updated cost function:

$$\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x}} \|\mathbf{y} - \mathbf{E}\mathbf{x}\|_2^2 + \lambda R(\mathbf{x}) \quad (2.20)$$

where λ is a hyperparameter [2]. On its own, CS theory illustrates that it's possible to reconstruct a signal obtained through an incoherent sampling pattern with a high level of confidence when the signal displays sparsity in a certain transform domain [33][34]. As MR images exhibit sparsity in some transform domains, CS theory can be used for MR image reconstruction. More specifically, it has been shown that MR images exhibit sparsity in the Discrete Cosine Transform domain and the wavelet domain; the corresponding transform coefficients are more sparse than the original MR images themselves [16]. This means that the regularizer term $R(\mathbf{x})$ can be a regularization function designed to impose a sparsity constraint within a specific transform domain [24]. This is precisely the basis of CS-based MR image reconstruction, modifying equation 2.20 in the following way:

$$\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x}} \|\mathbf{y} - \mathbf{E}\mathbf{x}\|_2^2 + \lambda \|\Psi(\mathbf{x})\|_1 \quad (2.21)$$

where Ψ is an operator which transforms \mathbf{x} into the sparsity domain, followed by the employment of the l_1 norm to encourage sparsity [2].

Although CS-based MR image reconstruction has only recently become commercially available for clinical implementation, its mathematical framework is still very useful as it creates the foundation of some of the state-of-the-art deep learning (DL) methods for MR image reconstruction. Additionally, CS-based MR image reconstruction distinguishes itself from GRAPPA-like and SENSE-like methods, as CS theory supports (even requires) the usage of incoherent undersampling patterns instead of periodic undersampling patterns [33].

2.3 Background on Artificial Neural Networks

Artificial Neural Networks (ANNs), commonly referred to as neural networks, are composed of individual computing units known as artificial neurons (also termed perceptrons). These artificial neurons are organized into layers, which are interconnected, resulting in the formation of a neural network (in its most basic form, also termed a multilayer perceptron) [35].

In a neural network, there are typically two primary layers: an input layer and an output layer. The presence of extra layers positioned between the input and output layers results in the network being labeled as a deep neural network [36]. The network

depth corresponds to the total number of layers it encompasses. Additionally, if the layers lack any cyclic connections, the network is classified as a feedforward neural network. However, if loops or cycles exist within the neural network's layer structure, also called the architecture, it is categorized as a recurrent neural network [37].

Every artificial neuron can be conceptualized as a compact computational entity tasked with calculating the weighted sum of its inputs, incorporating a bias, and subjecting the outcome to a non-linear activation function. The ensuing output is then relayed to the next layer of neurons or to the output of the network [35]. The utility of such non-linear behaviour can be found in the fact that this way, a standard feedforward ANN with an adequate number of neurons in their layers can effectively approximate any continuous function, including arbitrary non-linear ones [38].

The training process of a neural network involves presenting it with input data and comparing the network's output to corresponding reference data. During this process, the weights and biases of the neurons are adjusted to improve the network's alignment with the reference data. To guide these adjustments, a cost function is employed, which assesses the quality of the network's output based on the reference data and certain evaluation metrics. The derivatives of the cost function then provide crucial information on how to optimize the network's weights and biases to enhance its performance [36].

In the following subsections, critical design considerations of ANNs are discussed, with a particular emphasis on their application to MR image reconstruction. Consequently, the discussion will be tailored to address only those design elements that are pertinent to this specific task.

2.3.1 Relevant ANN architectures or frameworks

In the field of DL-based MR image reconstruction, by far the most common ANN architectures or frameworks are the Convolutional Neural Networks (CNNs) and the Generative Adversarial Networks (GANs).

CNNs: CNNs represent an important advancement in the field of ANNs, particularly when dealing with grid-like data, such as natural images and, indeed, such as k-space data and MR images. Because of this, CNNs have emerged as one of the most fitting choices in the context of image processing [39]. These networks operate by employing sets of neurons that function as filter kernels, convolving over the grid-like data while it propagates through the network [36]. CNNs are characterized by a multilayer architecture, typically comprising multiple convolutional layers, activation layers, normalization layers, and pooling layers [35]. Within each of the convolutional layers, a collection of filter kernels, possessing specific dimensions, extracts intricate features from the input image or data. Importantly, the complexity of these features grows proportionally to the CNN's depth, enabling CNNs to learn progressively more intricate representations of the input data [40]. It is worth mentioning that the growing depth of CNNs, partly fueled by the increased availability of computational resources, presents a challenge in the form of more intricate training processes.

2. LITERATURE REVIEW

Therefore, careful consideration must be given to the design of these networks to ensure optimal performance and prevent overfitting [41]. To conclude, one of the distinctive strengths of CNNs is their ability to address a common limitation of traditional neural networks, namely, the immense number of parameters associated with fully connected layers. In the case of multilayer perceptrons, images are flattened into extensive input vectors, leading to an overwhelming number of neural network parameters necessary to accommodate such large input data. CNNs provide a solution to this issue by adopting a locally connected approach, capitalizing on the inherent local correlations present in images. They recognize that pixels in images are locally strongly interrelated, often making it unnecessary to consider relationships between pixels that are far apart from each other [35].

GANs: Training a GAN entails the training of two competing neural networks, with any possible architecture, in a zero-sum game [35]. A standard setup for GANs involves a dynamic interaction between two ANNs: the generator, denoted as $G(\mathbf{z})$, and the discriminator, denoted as $D(\mathbf{x})$, which are parameterized by θ_g and θ_d respectively. The generator operates by generating data from a latent input vector \mathbf{z} and producing the output \mathbf{x}_{gen} . In contrast, the discriminator evaluates its input, which can be either \mathbf{x} from real data or \mathbf{x}_{gen} from the generator, and provides a binary decision such as {real, fake} indicating whether the input comes from the real data distribution or was generated by $G(\mathbf{z})$ [2]. These neural networks' parameters are optimized using an adversarial objective based on game theory, where they strive to outperform each other:

$$\begin{aligned} \{\theta_g, \theta_d\} &= \arg \min_{\theta_g} \max_{\theta_d} \mathcal{L}_{\theta_g, \theta_d} \\ &= \arg \min_{\theta_g} \max_{\theta_d} \mathbb{E}_{\mathbf{x} \sim p_{data}} (\log(D(\mathbf{x}|\theta_d))) + \\ &\quad \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}} (\log(1 - D(G(\mathbf{z}|\theta_g)))) \end{aligned} \quad (2.22)$$

where the first term of $\mathcal{L}_{\theta_g, \theta_d}$ represents a discriminator payoff term and the second term represents a generator payoff term [2][35]. Perhaps at first sight, it may be unclear why GANs are considered for the MR image reconstruction task. The reasoning behind the employment of GANs for this is as follows: as the training advances, the generator network gradually acquires the ability to produce realistic MR data from undersampled MR data, while the discriminator network becomes more proficient at distinguishing between generated (in this case, reconstructions from undersampled MR data) and authentic MR data [2]. This way, the generator can learn the mapping relationship between undersampled MR data and fully sampled MR data, which ultimately is very useful for MR image reconstruction [42].

2.3.2 Evaluation metrics

Commonly employed evaluation metrics in the context of DL-based MR image reconstruction include the Normalized Mean Squared Error (NMSE), the Peak Signal to Noise Ratio (PSNR), the Structural Similarity (SSIM) index, and an l_1 or l_2 loss [2][7][43][44]. Recently, new perceptual losses have also emerged as evaluation metrics, such as the VGG loss [2]. Some of these metrics will now briefly be elucidated.

2.3. Background on Artificial Neural Networks

NMSE: In the context of MR image reconstruction, the NMSE is a metric employed to assess the disparity between reconstructed MR data and fully sampled reference MR data. In general, smaller NMSE values indicate superior reconstruction outcomes. The NMSE can be defined in both image and k-space domains, making it a versatile tool for evaluating MR data fidelity, with the two possible formulae being as follows:

$$NMSE(\mathbf{x}, \hat{\mathbf{x}}) = \frac{\sum_{i=1}^n (\hat{\mathbf{x}}_i - \mathbf{x}_i)^2}{\sum_{i=1}^n (\mathbf{x}_i)^2} \quad (2.23)$$

$$NMSE(\mathbf{y}, \hat{\mathbf{y}}) = \frac{\sum_{i=1}^n (\hat{\mathbf{y}}_i - \mathbf{y}_i)^2}{\sum_{i=1}^n (\mathbf{y}_i)^2} \quad (2.24)$$

where \mathbf{x} and \mathbf{y} correspond to fully sampled MR data and $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ correspond to the noisy reconstructed MR data in the image and k-space domain respectively, and n corresponds to the number of pixels or k-space samples [43]. The NMSE is a commonly adopted metric, and it is suggested to present it as the primary indicator of reconstruction quality [7]. Nonetheless, due to the NMSE's inherent limitations, which can incline it toward favoring smoothness over sharpness causing small anatomical details to be lost as the NMSE penalizes large deviations more than smaller ones, it is highly recommended to supplement the reconstruction evaluation with additional metrics as detailed below [2][7].

PSNR: The PSNR is closely related to the NMSE, as it is calculated based on the mean squared error with the formula:

$$PSNR(\mathbf{x}, \hat{\mathbf{x}}) = 10 \log_{10} \left(\frac{\max(\mathbf{x})^2}{\frac{1}{n} \sum_{i=1}^n (\hat{\mathbf{x}}_i - \mathbf{x}_i)^2} \right) \quad (2.25)$$

where $\max(\mathbf{x})$ represents the highest pixel value in the image \mathbf{x} [43]. Again, an analogous formulation exists for k-space data. A better reconstruction is indicated by higher PSNR values, in contrast to lower NMSE values [7].

SSIM: The SSIM index serves as a comparative metric for evaluating the contrast, brightness and structural similarity between two image samples [43]. Recently, it has gained more and more attention, making it one of the state-of-the-art evaluation metrics to evaluate image quality according to multiple authors [45][46]. In their original work, Wang et al. [45] propose the following form of the SSIM index:

$$SSIM(\mathbf{x}, \hat{\mathbf{x}}) = \frac{(2\mu_x\mu_{\hat{x}} + c_1)(2\sigma_{x\hat{x}} + c_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + c_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2)} \quad (2.26)$$

where μ_x and $\mu_{\hat{x}}$ represent the mean pixel values of images \mathbf{x} and $\hat{\mathbf{x}}$, while $\sigma_{x\hat{x}}$ denotes the covariance between these images. Additionally, σ_x and $\sigma_{\hat{x}}$ correspond to the standard deviation of the pixel values in the images \mathbf{x} and $\hat{\mathbf{x}}$ respectively. The constants c_1 and c_2 are incorporated to mitigate potential instabilities that may arise when $(\mu_x^2 + \mu_{\hat{x}}^2)$ or $(\sigma_x^2 + \sigma_{\hat{x}}^2)$ approaches zero [45]. The SSIM index, ranging from -1 to 1, equals 1 when $\hat{\mathbf{x}}$ is identical to \mathbf{x} [45]. A higher SSIM index thus indicates a better MR image reconstruction. It is important, however, to note that the statistical characteristics of images frequently exhibit significant spatial variation. Hence, it

2. LITERATURE REVIEW

proves beneficial to employ the SSIM index on a local scale rather than a global one [45]. This approach gives rise to the concept of the mean-SSIM (MSSIM) index when comparing two images, $\hat{\mathbf{x}}$ and \mathbf{x} :

$$MSSIM(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{M} \sum_{j=1}^M SSIM(\mathbf{x}_j, \hat{\mathbf{x}}_j) \quad (2.27)$$

where the MSSIM index is computed using a windowing approach, involving the calculation of M structural similarity indices based on smaller, local segments of the images. For this, a Gaussian window is often used [45].

VGG loss: As demonstrated by Johnson et al. [47], the deeper layer feature maps within a CNN, such as a VGG-16 network (which is a network comprising 16 layers), can also be utilized to assess the perceptual similarity between images. This concept gives rise to the VGG loss, serving as an alternative to pixel-wise losses, with the aim of better capturing perceptual similarity [48]. Considering $\phi_{i,j}$ as the j -th feature map (stemming from the j -th filter of the previous layer) of the i -th layer, the VGG loss can be calculated as the mean euclidean distance between the feature representations of a reference or ground-truth (MR) image $\phi_{i,j}(I_{ref})$ and a reconstructed (MR) image $\phi_{i,j}(I_{recon})$ [48]. A lower VGG loss thus indicates a better MR image reconstruction.

2.3.3 Activation functions

A next key design parameter is the activation function that is used in the ANN. It was noted in previous research that CNNs using rectified linear units (ReLUs) as their activation function outperformed those using hyperbolic tangent activation functions in terms of training speed [49]. An alternative activation function, introduced by another group of researchers, is the parametric ReLU (PReLU). This activation function dynamically adjusts the parameters of the rectifiers, leading to improved accuracy and performance with minimal additional computational overhead [50].

2.3.4 Initialization method

Finally, a last crucial design parameter concerns the network parameter initialization method. The concept of normalized initialization, as introduced in [51], involves setting initial biases to zero and selecting initial weights from a uniformly distributed range that is appropriately scaled. However, it's worth noting that this approach assumes linear activations at each layer, which may not hold true for ReLU and PReLU activations, as pointed out by He et al. [50]. They advocate for a different initialization approach, wherein initial biases are initialized to zero and initial weights are sampled from a Gaussian distribution with zero mean and a standard deviation of $\sqrt{\frac{2}{n_j}}$, with n_j denoting the amount of connections to the j -th layer [50]. This particular initialization technique is commonly referred to as the MSRA filler or the He initializer.

2.4 DL-based reconstruction techniques for undersampled k-space data

The field of DL-based MR image reconstruction has known tremendous growth in the past few years [2]. Between 2018 and 2022 alone, around 250 papers on this research subject have been published [44]. Because of this, it is impossible to cover all recent and noteworthy advancements within this field in this thesis. Hence, this work primarily focuses on the state-of-the-art DL methods that introduce new perspectives on the MR image reconstruction problem and come up with new, innovative leaps in implementation methods (not considering simple changes in model architecture). For more detailed, in-depth information about state-of-the-art DL methods for MR image reconstruction, please refer to [2] and [44].

2.4.1 DL-based reconstruction method categories

As previously described, many different approaches to the DL-based MR image reconstruction problem were already created. The author found that this led to the existence of three DL-based reconstruction method categories: methods which operate completely within the k-space (GRAPPA-like methods), methods which completely operate within the image space (SENSE-like methods), and methods which operate in both spaces (dual methods). The operational domain is indicated for each well-discussed framework cited in the subsequent subsections, whilst it is also mentioned whether single-coil or multi-coil data was used for their training and evaluation. Multi-coil models are more relevant due to increased clinical usability.

2.4.2 Trained networks

Conventional trained networks

In the most simple case, an image domain U-Net with an l_1 or l_2 loss is used for the MR image reconstruction task. U-Nets are a specific type of CNN with skip connections between matching downsampling and upsampling layers, and they were already successfully used for image segmentation tasks [7]. Because of this, an image domain U-Net trained with an l_1 loss, taking a single-coil image reconstructed from single-coil undersampled k-space as the input, is conventionally regarded as the baseline DL method for the MR image reconstruction problem. This convention allows for performance comparisons with a consistent DL reference [7]. Reconstructions from this baseline network for the single-coil fastMRI knee dataset are reported to display an NMSE of 0.0406, a PSNR of 30.7, and a MSSIM index of 0.699 on average across different acceleration factors ranging from R=4 to R=8 [7]. This already represents a 40-50% relative improvement compared to the classical method discussed in the baseline U-Net's paper, applied to the same dataset, as measured by NMSE [7]. Note that this aforementioned baseline network falls outside the parallel imaging framework as it only uses single-coil data, aiming to create a DL baseline for even the most rudimentary MR datasets and their reconstruction methods [7].

2. LITERATURE REVIEW

A first truly state-of-the-art DL-based reconstruction method is termed **AMRI**. This method aimed to exploit GANs to formulate an improvement to the existing, more simple l_1 or l_2 loss-based (MR) image reconstruction methods, as utilized in the above described baseline U-Net [52]. This is because minimizing these norms tends to yield a less detailed solution, as they frequently struggle to capture fine high-frequency details. This is due to these losses favoring a smoother, less detailed outcome over one with highly accurate textures. The rationale behind this preference lies in the fact that predicting the mean of a distribution minimizes the average pixel-wise error effectively, albeit at the cost of producing a blended, less distinct image [53]. The authors of AMRI argue that this issue can be mitigated by introducing an adversarial loss to the l_1 or l_2 loss problem formulation, resulting in a composite adversarial loss for the generator, e.g.: $\mathcal{L}_{\theta_g} = \alpha \mathbb{E}_{z \sim p_z} (\log(1 - D(G(z|\theta_g)))) + \beta \| (1 - \mathcal{M}) \odot (\hat{\mathbf{y}} - \mathbf{y}^{full}) \|_2^2$, where α and β are tunable hyperparameters [52]. As the previous equation suggests, AMRI operates completely within the k-space. Its generator and discriminator networks are both CNNs [52]. The generator takes two-channel k-space data as its input, with each channel encoding the real and imaginary parts of the partially sampled k-space data respectively. Additionally, each of the missing k-space samples is initially filled with uniform independent and identically distributed noise, \mathbf{z} , at the input of this network [52]. AMRI's generator has a residual network structure, meaning that the sampled k-space data points are preserved while AMRI only estimates the missing data points in k-space. Mathematically, this means that $\hat{\mathbf{y}} = \mathbf{y} + (1 - \mathcal{M}) \odot G_{out}$ in this framework [52]. The performance of AMRI was evaluated on a single-coil MR dataset with R=2, which resulted in a PSNR of 37.95 (± 0.6) [52].

A subsequent work, however, found that CNNs exhibit a good but still suboptimal reconstruction performance due to the finite, often small size of the convolution kernels, causing long-/medium-range dependencies in the data to be unused [2][42]. This may be solved by including a self-attention layer in the CNN architecture, which is precisely what is suggested in the work where the state-of-the-art DL-based reconstruction method **SARA-GAN** is proposed [42]. Just like AMRI, SARA-GAN is a GAN method which uses CNNs as its generator and discriminator, although SARA-GAN completely operates within the image domain. The generator in particular is a residual CNN which includes a self-attention layer in its upsampling block, allowing the network to extend its receptive field without much additional cost. This typically leads to network outputs displaying more accurate textures [42].

In the self-attention layer, the output feature maps \mathbf{x} of the previous hidden layer with C channels are taken as the input. First, \mathbf{x} is reshaped to be $\mathbf{x} \in \mathbb{R}^{C \times N}$ with N being the product of all other dimensions. Here, $N = \text{height} \times \text{width}$ of the feature maps. Next, three 1×1 convolutions are performed to obtain the new feature spaces $\mathbf{f}(\mathbf{x}) = \mathbf{W}_f \mathbf{x}$, $\mathbf{g}(\mathbf{x}) = \mathbf{W}_g \mathbf{x}$ and $\mathbf{h}(\mathbf{x}) = \mathbf{W}_h \mathbf{x}$ with weight matrices $\mathbf{W}_i \in \mathbb{R}^{C' \times C}$. In SARA-GAN, C' is chosen to be $\frac{C}{8}$ [42]. After this, an attention map can be calculated via:

$$\beta_{j,i} = \frac{\exp(s_{i,j})}{\sum_{i=1}^N \exp(s_{i,j})} \quad (2.28)$$

2.4. DL-based reconstruction techniques for undersampled k-space data

where $s_{i,j} = \mathbf{f}(\mathbf{x}_i)^T \mathbf{g}(\mathbf{x}_j)$ [42][54]. Termed as the attention map, these weights essentially serve to quantify the relevance of pixel j relative to pixel i (represented as columns in \mathbf{x}) in the feature maps of the MR image at the input, forming a proxy to autocorrelation. Because these weights are calculated across the entire height and width of the feature set, the receptive field is no longer confined to the dimensions of a small kernel [54]. Lastly, the output of the self-attention layer is defined as:

$$\mathbf{o} = (\mathbf{o}_1, \dots, \mathbf{o}_j, \dots, \mathbf{o}_N) \in \mathbb{R}^{C \times N}, \text{ with } \mathbf{o}_j = \mathbf{v} \left(\sum_{i=1}^N \beta_{j,i} \mathbf{h}(\mathbf{x}_i) \right) \quad (2.29)$$

where \mathbf{v} is the result of another 1×1 convolution $\mathbf{v}(\mathbf{x}) = \mathbf{W}_v \mathbf{x}$ with weight matrix $\mathbf{W}_v \in \mathbb{R}^{C \times C'}$ [42][54].

For SARA-GAN specifically, the output \mathbf{o} is then modified one last time before being passed to the next layer. To facilitate the generator in learning both the local and long-range global dependencies of the data, the vectors \mathbf{o}_j of the self-attention layer are each multiplied by a weight coefficient γ . This result is then added to the attention layer's corresponding input feature map values \mathbf{x}_j to produce the final output of the self-attention module, denoted here as \mathbf{y}_j :

$$\mathbf{y}_j = \gamma \mathbf{o}_j + \mathbf{x}_j \quad (2.30)$$

The parameter γ is initialized to be 0 and is learnable. Its role is to allow SARA-GAN to adaptively learn how much of the global dependency should be incorporated into the eventual feature map of the MR image [42].

In addition to this innovation regarding the generator architecture, SARA-GAN also introduces a modification to the training process of the discriminator by implementing a relativistic discriminator instead of a vanilla discriminator [2][42]. In a conventional GAN, the discriminator evaluates the likelihood that the input data is real, while the generator's goal is to increase the probability that fake data is perceived as real. However, from an intuition point of view, it is suggested in [55] that the generator should actually aim to increase the probability that fake data is identified by the discriminator as more authentic than real data. This can be achieved through the use of a relativistic discriminator, which during training assesses the probability that for a given real/fake MR image pair the fully sampled MR image is more faithful to reality than the reconstructed MR image, instead of simply assessing the probability that a given MR image is real [42].

This is done by modifying equation 2.22:

$$\{\theta_g, \theta_d\} = \arg \min_{\theta_g} \max_{\theta_d} \mathbb{E}_{\mathbf{x} \sim p_{data}} (\text{sigmoid}(D(\mathbf{x}) - D(G(\mathbf{z})))) + \mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}} (\text{sigmoid}(D(G(\mathbf{z})) - D(\mathbf{x}))) \quad (2.31)$$

where $\text{sigmoid}()$ is the sigmoid activation function [2]. SARA-GAN performed very well on the single-coil MICCAI 2013 grand challenge datasets, generally displaying a PSNR larger than 40 and a MSSIM index larger than 0.95 whilst showing a good robustness against introduced noise [42].

2. LITERATURE REVIEW

Another work proposes **RefineGAN**, a GAN method which operates in the image space and which aims to bridge the gap between frequency losses (defined in the k-space) and image losses (defined in the image domain) by introducing cyclic losses to the MR image reconstruction problem [56]. Its discriminator and generator have a U-Net architecture. Taking inspiration from already defined cyclic losses for natural images, the cyclic loss for the generator G can be defined as: $\mathcal{L}_{cyclic} = \alpha \mathcal{L}_{freq} + \beta \mathcal{L}_{image}$, with α and β as hyperparameters, making $\mathcal{L}_{total} = \mathcal{L}_{adversarial}(G, D) + \mathcal{L}_{cyclic}(G)$ [56]. The loss components \mathcal{L}_{freq} and \mathcal{L}_{image} are calculated in the following way: The generator takes single-coil MR images \mathbf{x} , reconstructed from single-coil undersampled k-space data \mathbf{y} through an IFT, as its input and maps this to de-aliased, high resolution images $\hat{\mathbf{x}}$ as if these were reconstructed from fully sampled k-space data $\hat{\mathbf{y}}$. Then, $\mathcal{L}_{image} = NMSE(\mathbf{x}, \hat{\mathbf{x}})$ and $\mathcal{L}_{freq} = NMSE(\mathbf{y}, \mathcal{M} \odot \mathcal{F}\hat{\mathbf{x}})$ [56]. The performance of RefineGAN is quite good, with the paper’s reporting showing significant improvements compared to other DL methods and with another paper reporting that it exhibits an average PSNR of 34 and an average MSSIM index of 0.9 for single-coil reconstructions where R=4 [2][56].

Similarly, another model, **PIC-GAN**, also seeks improvements by mainly extending the loss function definition to cover multiple representations of the data flowing through the model. However, in contrast to RefineGAN, PIC-GAN goes one step further by extending the loss function definition in a spatial manner, using separate multi-coil MR image components, as well as in a spectral, cyclic manner as discussed before [57]. Again, PIC-GAN is a GAN method operating in the image domain with its discriminator and generator having a deep residual U-Net architecture. It also takes a single, multi-coil-combined MR image \mathbf{x} as its generator’s input and maps this to a de-aliased, high resolution MR image $\hat{\mathbf{x}}$ at its generator’s output. After this, PIC-GAN decomposes $\hat{\mathbf{x}}$ back into its supposed multi-coil components $\hat{\mathbf{x}}_i = \mathcal{S}_i \odot \hat{\mathbf{x}}$, allowing the definition of a multi-coil cyclic loss function [57]. For multi-coil data where R=4, PIC-GAN’s performance is comparable to that of RefineGAN.

Next, an innovative paper which offers an additional new perspective on the DL-based MR image reconstruction problem introduces **SUBGAN**, another GAN method based on residual U-Nets which, this time, operates in both the k-space and the image space. The main key feature of this state-of-the-art reconstruction method lies in its usage of the shared information between adjacent k-space slices. This supplemental input enables the exploitation of potential temporal (in time) or spatial (along the z-axis) k-space data correlations [4]. In order to leverage this spatial (or temporal) interdependence amongst k-space slices, two neighboring subsampled k-space slices, \mathbf{y}^{i-1} and \mathbf{y}^{i+1} , are provided at the k-space generator’s input for the reconstruction of k-space slice \mathbf{y}^i , along with the k-space slice \mathbf{y}^i itself [4]. Similar to the workflow of AMRI, the absent k-space samples are first replaced with noise sampled from a uniform distribution before reconstruction begins. Each k-space slice comprises two channels representing the real and imaginary components of the k-space [4]. SUBGAN outperformed other state-of-the-art GAN methods by achieving approximately a 2.5-point increase in PSNR, all while utilizing only 20% of the original k-space samples on the public single-coil IXI dataset [2].

Trained hybrid networks

An additional significant work puts **GrappaNet** forward, a DL-based MR image reconstruction method specifically designed for multi-coil MR data from parallel imaging acquisitions. Here, the authors argue that an intersection between classical and DL reconstruction methods outperforms either approach alone [24].

Incorporating parallel imaging in the framework of DL-based MR data reconstruction is challenging. The varying relations between the captured views of each parallel imaging scan, which depend on the detector configuration relative to the imaged object, may pose a difficulty in training an ANN capable of performing MR image reconstruction based on multi-coil inputs. The different multi-coil MR scans in the training set may each display different relations between their captured views, which may possibly obscure certain relationships the model is trying to learn and can limit the extent to which neural networks can perform multi-coil reconstruction [24]. Classical parallel imaging methods address this issue by estimating separate sensitivity maps or GRAPPA kernels for each scan. Taking inspiration from those concepts, this motivates the need to incorporate a more classical, scan-specific element within the ANN that can adapt to variations between the sensitivity profiles of different scans [24]. From a deep learning point of view, this comes down to increasing the capacity of the model to deal with variance in the training data without increasing the number of trainable model parameters and thus without increasing the need for increasingly larger training datasets. This is achieved by including a classical GRAPPA layer in the ANN, which can be seen as a constrained layer that enforces intermediate data representations to adhere to the expected GRAPPA reconstruction specific to the multi-coil MR scan that is provided to the model at the input.

Because of this, GrappaNet is actually an end-to-end trained cascaded ANN consisting of two serial CNNs which are connected to each other via a GRAPPA layer [2][24]. GrappaNet thus essentially computes the following reconstruction function: $\hat{\mathbf{x}} = h(f_2(\mathbf{K} \circledast f_1(\mathbf{y})))$, where f_1 and f_2 represent the two CNNs and where h represents the application of an IFT to the multi-coil data followed by a root-sum-of-squares operation [24]. From the previous information, it may already be clear that GrappaNet is a CNN method which operates both in the k-space and in the image space. The CNNs are actually residual U-Nets which both take undersampled multi-coil k-space data as their inputs, employing two real-valued channels to represent the real and imaginary parts of the data. Mappings in both k-space and image space are then performed within each CNN. This is because convolution, pooling, and up-sampling operations have distinct effects in image space and k-space [24]. Ultimately, the individual U-Nets return k-space data as their outputs. They are trained using a composite loss function based on the MSSIM index and the l_1 loss: $\mathcal{L}_{total} = -MSSIM(\mathbf{x}, \hat{\mathbf{x}}) + \lambda \|\mathbf{x} - \hat{\mathbf{x}}\|_1$. In the original work, λ was set to 0.001 [24]. Another innovative trait of GrappaNet is that, unlike many prior methods, the different multi-coil views are not merged until they arrive at the final layer of the network [24]. The overarching architecture and framework of GrappaNet is shown in Figure 2.4.

2. LITERATURE REVIEW

GrappaNet has demonstrated exceptional performance on the public multi-coil fastMRI dataset, achieving results with a PSNR of 40.74 and a MSSIM index of 0.957 for R=4. It is one of the top-performing methods in this context [2][24].

With the above concepts, GrappaNet gave rise to a recent, small, new family of DL models; models which elegantly try to form the intersection between classical and DL-based MR image reconstruction methods. The author proposes to term these models as '**hybrid**' models. Note that these hybrid models always require multi-coil data, as classical reconstruction techniques need data from multiple coils for their interpolation or estimation methods. This is not the case for conventional DL methods, which, as shown by the previously discussed state-of-the-art models, can use either single- or multi-coil data.

A second noteworthy hybrid model is **DeepMRIRec**. Its workflow is very similar to that of GrappaNet: a residual U-Net is combined with GRAPPA methods, although DeepMRIRec completely operates within the image domain instead of functioning as a dual method [58]. Another key difference between DeepMRIRec's architecture and GrappaNet's architecture is that the classical GRAPPA layer has been moved to the input of the model, instead of functioning as a hidden layer. On top of this, ReLU activation layers were changed into PReLU activation layers and the DeepMRIRec study looked at the network depth, learning rate, dropout rate, and the number of filters in the base layer of the U-Net as hyperparameters in a more rigorous manner, as a Bayesian optimizer was employed to identify the most effective values for these hyperparameters with the goal of obtaining an even better U-Net architecture for the task at hand [58]. The final architecture and workflow of DeepMRIRec is shown in Figure 2.5. DeepMRIRec also employs an extended version of GrappaNet's loss function, using an extra VGG loss term and re-optimized loss function hyperparameters, and the training process was extended by using augmented multi-coil training data [58]. In general, as shown for the public multi-coil fastMRI dataset, DeepMRIRec reconstructions display higher MSSIM index values but lower PSNR values compared to reconstructions performed by GrappaNet.

2.4.3 Untrained networks

Up to this point, the literature review has covered different state-of-the-art DL architectures and the methods employed to train them with a specified training dataset. Nevertheless, it's worth mentioning that there also exist a few progressive approaches that can utilize the ACS lines from a single undersampled k-space measurement \mathbf{y} to directly train a DL network without a separate training dataset, diverging from the conventional reliance on a training dataset comprising multiple MR images as a ground truth [2]. The two most notable of these 'untrained' methods are **RAKI** and **SPARK** [2][59][60]. These methods are both CNN methods, but as untrained networks aren't really relevant for this thesis, they won't be extensively discussed here. The take-away from this brief subsection is that, although these untrained networks do not perform as well as highly trained state-of-the-art methods, they do generalize a lot better to unknown scenarios or input data [2].

2.4. DL-based reconstruction techniques for undersampled k-space data

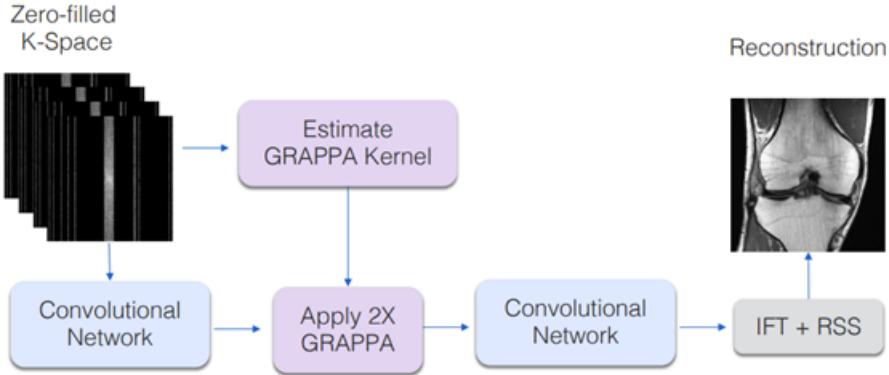


FIGURE 2.4: Illustration of the full GrappaNet model. The first network, referred to as f_1 , processes the multi-coil k-space data that underwent R-fold undersampling and transforms it into k-space data mimicking R'-fold undersampling with the same number of coils ($R' < R$). Subsequently, the GRAPPA kernel \mathbf{K} , separately estimated via the ACS region, is applied to this result. This process then fills in the last missing k-space data points. The value of R' is deliberately kept small enough to ensure accurate image reconstruction using conventional parallel imaging methods like GRAPPA. In the original work, $R'=2$. Image taken from Sriram *et al.* [24].

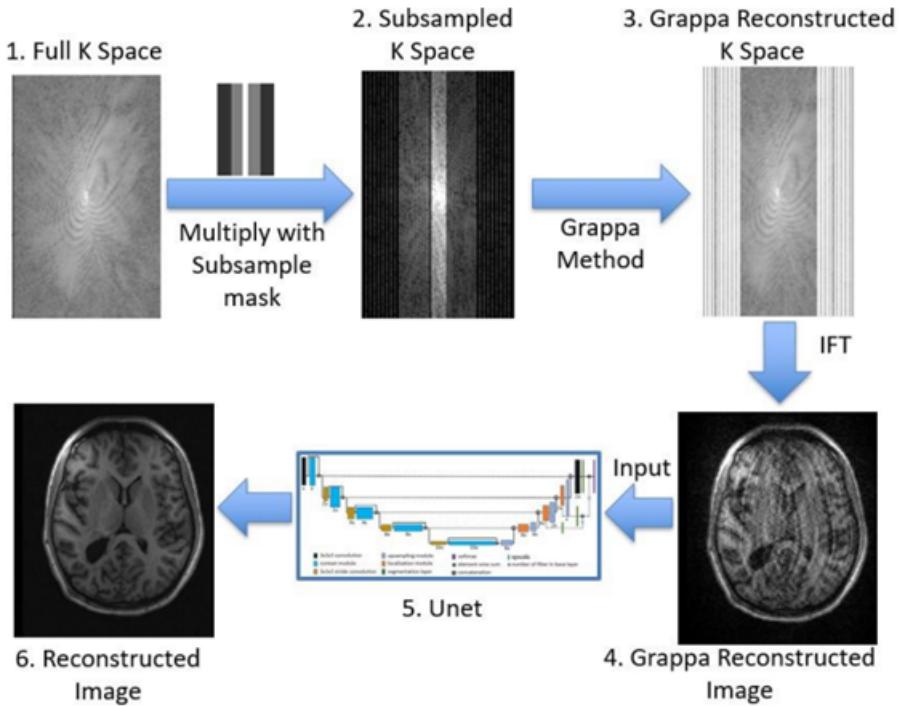


FIGURE 2.5: Illustration of DeepMRIRec. Image taken from Alam *et al.* [58].

2.4.4 Extensions to the MR image reconstruction problem

Before concluding this section, it might be interesting to enumerate some other interesting works which form extensions to the MR image reconstruction problem, just to make the reader aware of other recent advances in this field even though these works may not really be relevant for the research scope of this thesis. These DL methods won't be discussed in detail, but the most innovative aspect(s) about them is mentioned.

A first interesting DL framework is called **DeepcomplexMRI**, which uses a CNN with complex-valued kernels to perform complex convolutions with the complex-valued MR data avoiding the need to work with separate input channels for the real and imaginary parts of the MR data [2][61]. The complex-valued convolution network has the capability to attain a performance that is comparable or even superior to that of real-valued networks, all while utilizing only half the size of the latter [61].

A second notable DL framework for MR image reconstruction is the active acquisition framework, with its most illustrious examples being **LOUPE** and **J-MoDL** [62][63]. All prior DL methods have consistently involved a static k-space sampling mask, set in advance by the user and treated as a separate entity from the reconstruction pipeline. These works, however, have explored the possibility of integrating the sampling process into the reconstruction optimization framework, i.e. the k-space sampling mask itself is treated as learnable. It was determined that this increase in research question domain size resulted in a better reconstruction quality, although this comes at the cost of an increased method complexity and a possibly low practical utility of the optimized k-space sampling mask [2].

A last riveting DL framework is called **CollaGAN**, a multi-input single-output GAN method which exploits information shared between different images from different MR sequences of the same subject by considering T_1 -FLAIR, T_2 -weighted, T_2 -FLAIR and T_2^* -FLAIR images for the reconstruction of undersampled MR data through collaborative mappings [46]. In this approach, rather than relying on cyclic consistency between the k-space and image domain of a single image, consideration is given to a variety of domains, such as these T_1 -weighted and T_2 -weighted data domains [2].

2.4. DL-based reconstruction techniques for undersampled k-space data

2.4.5 Overview

In this last subsection, a final overview of the most extensively discussed DL-based MR image reconstruction methods is given, whilst categorizing them like and comparing them to the previously bespoke classical MR image reconstruction methods from section 2.2 (Table 2.1). The reconstruction methods are grouped based on their model class and for each model any necessary data requirements are displayed as well. Hybrid models are marked in bold.

Optimally, however, one would also try to directly compare their performances, but due to the heterogeneity of the used training datasets this becomes very difficult to do in a reliable and consequent manner. It should be noted that, despite this, some review papers made efforts to do so, such as [2]. There, as already mentioned before, it was concluded that GrappaNet was one of the best performing models [2].

Model class	Model	Requirements
Classical methods	GRAPPA	<i>ACS</i>
	SENSE	\mathcal{S}_i
	CS	Incoherent sampling, \mathcal{S}_i
GRAPPA-like DL models	AMRI	Training dataset
	RAKI	<i>ACS</i>
	SPARK	<i>ACS</i>
SENSE-like DL models	Baseline U-Net	Training dataset
	SARA-GAN	Training dataset
	RefineGAN	Training dataset
	PIC-GAN	Training dataset, \mathcal{S}_i
	DeepMRIRec	Training dataset, <i>ACS</i>
Dual DL models	SUBGAN	Training dataset
	GrappaNet	Training dataset, <i>ACS</i>

TABLE 2.1: Overview of the most important MR image reconstruction methods found in the current literature. *ACS* = autocalibration signal, \mathcal{S}_i = coil sensitivity map estimations.

2.5 Conclusion

As became clear, MR images are actually acquired in the frequency domain, called the k-space. It is then also this relationship between the k-space and the image domain which imposes constraints on any accelerations made during the MR data acquisition, causing the appearance of aliasing and other artifacts when not dealt with correctly [20][21]. Because of this, several classical reconstruction techniques, viewed as the golden standard in clinical practice, have already been developed to try to mitigate these issues as much as possible [2]. These techniques are also referred to as parallel imaging techniques, with the three most notable ones recognized as being GRAPPA [14], SENSE [15], and Compressed Sensing [16][17]. Unfortunately, however, these techniques still only allow accelerations with an acceleration factor two or three depending on the used imaging set-up, and they often suffer from long iteration times [19][22][25]. Consequently, it is interesting to try to perform this combination of MR image reconstruction and artifact suppression through the use of neural networks.

In literature, many different DL models have already been proposed for this, often being CNNs or GANs. Though many architecture and training paradigm improvements have already been suggested, including the usage of different loss functions, attention layers, multi-slice inputs, and a relativistic discriminator, another new, niche idea has recently emerged in the current literature regarding the enhancement of MR image reconstruction performance: the usage of hybrid modelling, where a classical reconstruction technique is essentially fused with a DL-based reconstruction technique. A first model which proposed this strategy, is GrappaNet [24]. Interestingly, initial results obtained via the GrappaNet architecture seem promising, with GrappaNet demonstrating exceptional performance on the public multi-coil fastMRI dataset. It is even one of the top-performing methods in this context [2].

This is because different multi-coil MR scans from a multi-coil training dataset may each display different relations between their captured views, which may possibly obscure certain relationships a DL model is trying to learn and can limit the extent to which neural networks can perform multi-coil reconstruction [24]. Traditional parallel imaging techniques tackle this challenge by computing distinct sensitivity maps or GRAPPA kernels for individual scans. Drawing inspiration from these principles, this underscores the necessity of integrating a classical, scan-specific component into the neural network architecture, capable of adjusting to variations among the sensitivity profiles of diverse scans [24]. From a deep learning point of view, this comes down to increasing the capacity of the model to deal with variance in the training data without increasing the number of trainable model parameters and thus without increasing the need for increasingly larger training datasets.

Unfortunately, however, no well-founded proof supporting the idea that this hybrid modelling systematically improves MR image reconstruction performance can be found in literature yet. As of now, only two models employing hybrid modelling have been published: GrappaNet and DeepMRIRec [24][58].

Chapter 3

Methods: Data Analysis

The comparative MR image reconstruction experiments presented in this work will make use of trained hybrid convolutional neural networks. These will be discussed extensively in the next chapter, Chapter 4, and they essentially aim to perform a reconstruction function $\hat{\mathbf{x}} = f_2(f_1(\mathbf{y}))$, where f_1 represents a classical reconstruction operation layer at the input of the hybrid model, f_2 represents a CNN, $\hat{\mathbf{x}}$ represents the reconstructed MR image, and \mathbf{y} represents the undersampled input k-space data.

As mentioned before, however, it is important to be aware that these models are comprised of millions of trainable parameters, necessitating a substantial amount of data for their training. The role of this data in the performance of these trained networks is thus quite significant. As a result, it's essential to prioritize the quality of this data. While training a model with high-quality data doesn't automatically ensure optimal performance, conversely, a model trained with low-quality data will certainly result in subpar performance. This principle is recognized in computer science as *garbage in, garbage out*.

A closer examination of the data to be utilized for model training and evaluation is undertaken in this chapter. First, the structure and the contents of the used dataset are discussed. After this, a second section is provided focusing on used preprocessing steps, as the raw data is not yet suitable for network training.

This second section consists of two subsections. The first of these two subsections discusses preprocessing steps applied to the raw data before it is passed to the function f_1 . The second subsection discusses preprocessing steps applied to $f_1(\mathbf{y})$ before it is passed to the function f_2 .

A last section then discusses how reconstruction evaluation is performed in Chapter 5 of this work.

3. METHODS: DATA ANALYSIS

3.1 Used dataset

Three specific dataset properties must be met when deciding which (fully sampled) data to use for the intended purposes of this work. First of all, physical complex-valued raw data should be contained within the dataset. If this is not the case, then one could naturally question the practical utility and the validity of any obtained results. Unfortunately, however, most public Magnetic Resonance (MR) image datasets do consist of synthetic k-space data, i.e. k-space data that is derived from already-reconstructed MR images through the use of post-processing methods [7]. Secondly, the used dataset should be of sufficient size and contain sufficient variability for the training of a deep learning (DL) network. This criterion also remains unmet for a lot of public datasets, as they often contain less than 100 scan volumes [7]. Lastly, multi-coil data should be included in the used dataset, as classical reconstruction algorithms require parallel imaging data (see Chapter 2).

Because of a lack of public datasets exhibiting all three of the properties listed above, New York University (NYU) has recently collaborated with Facebook AI Research to create a new, adequate public dataset: the fastMRI dataset [7]. Consequently, as the fastMRI dataset is the first and only large-scale MR dataset which meets all requirements put forward by the previous paragraph, the fastMRI dataset is the dataset that will be used in this work.

This data can be obtained from the NYU fastMRI Initiative database (fastmri.med.nyu.edu) [7][64]. Please note that the NYU fastMRI investigators provided the data but did not participate in the analysis or writing of this work.

3.1.1 The fastMRI dataset

In total, the fastMRI dataset consists of three sub-datasets: a knee dataset (1.594 scan volumes), a brain dataset (6.790 scan volumes), and a prostate dataset (312 scan volumes) [7][65]. Each of these sub-datasets contains three types of fully sampled data from in vivo examinations from a range of MR scanners: raw multi-coil k-space data, emulated single-coil k-space data, and root-sum-of-squares ground-truth images. Additionally, within each of these sub-datasets additional DICOM images (not counted amongst the above-mentioned scan volumes) for which the raw k-space data was discarded can be found. These images remain included, however, as they may still prove useful as additional training or testing data [7]. Raw data is already anonymized through a conversion to the ISMRMRD format and DICOM data is already anonymized using an RSNA clinical trial processor [7].

In order to reduce the computational load of the envisioned experiments, and in order to retain the possibility of performing transfer learning experiments with different types of MR data in the future (e.g. by training a hybrid DL model on brain data and evaluating it on knee data), only the fastMRI brain dataset consisting of axial 2D image volumes will be used in this thesis. Furthermore, as k-space data is required for some of the investigated classical reconstruction methods, the DICOM images of the brain dataset will also be disregarded.

The remaining 6.970 brain scan volumes were acquired at different locations using many different Siemens scanners with field strengths of either 1.5T or 3T [7]. In the brain dataset, T1-weighted, T2-weighted, and FLAIR images are present, with T1-weighted images sometimes using contrast agents (images indicated as T1 POST). The precise distribution of contrast presences and field strengths can be found in Table 3.1. The raw data from these brain scans was recorded with a varying number of receiver coils, ranging anywhere from 2 coils to 28 coils, with the most used number of coils equal to 16. Corresponding MR image slice matrices also have varying sizes, ranging anywhere from 512x214 to 768x396, with the most used matrix size equal to 640x320. The raw brain data also includes some metadata for each scan, including e.g. the scan’s slice field of view (FOV), which can be used for image cropping to compensate for oversampling in the frequency encoding direction [7]. A visual example of fully sampled 16-coil raw MR data part of the brain dataset with a slice matrix size of 640x320 and a FOV of 320x320 is shown in Figure 3.1.

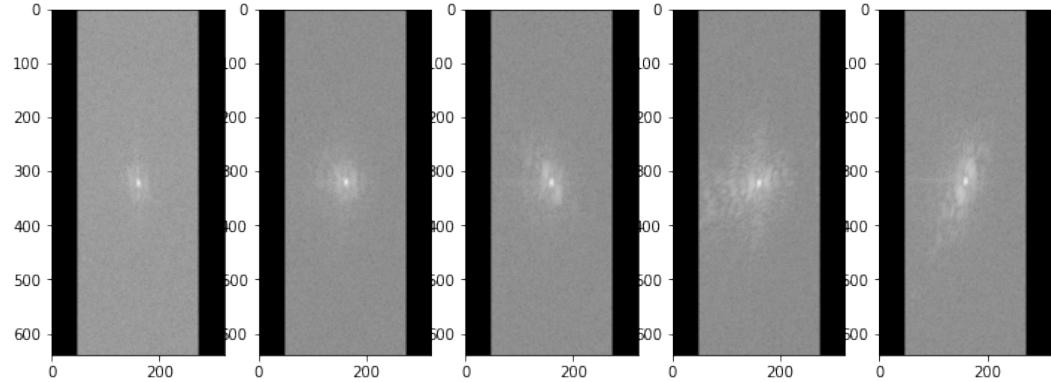
Field Strength	1.5T	3T
T1	375	407
T1 POST	849	641
T2	1.651	2.515
FLAIR	126	406
Total	3.001	3.969

TABLE 3.1: The different sequences part of the fastMRI brain dataset [7].

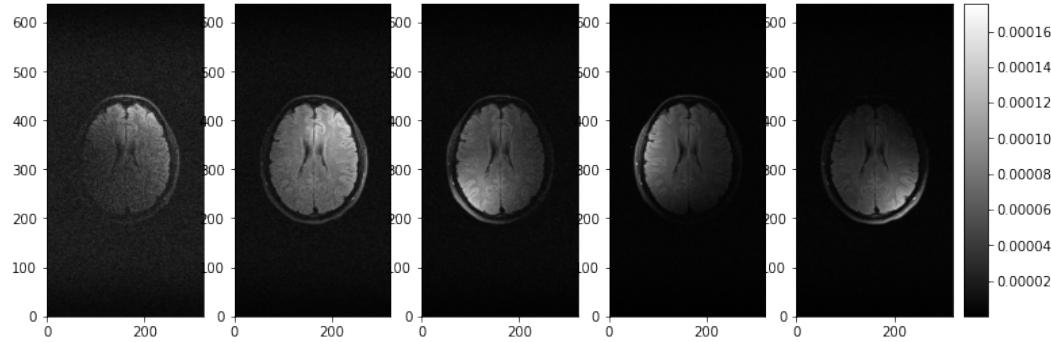
3.1.2 Dataset split

In DL frameworks, it is important to partition the used dataset into (usually) three parts, often of different sizes: a training dataset, a validation dataset, and a test dataset. The training dataset is used to let the DL model find the optimal network parameter adjustments for the loss function and dataset at hand. The validation dataset, on the other hand, is then used to find the optimal values of any relevant hyperparameters and to monitor the degree of overfitting during training. Lastly, the test dataset consists of data the DL model hasn’t seen before during training and is used for the final evaluation of the model’s performance. Typically, such dataset splits happen in a 60-20-20 or an 80-10-10 fashion, but for the fastMRI brain dataset there is actually an official dataset split available. Here, NYU researchers have split the scan volumes in 4 different subsets: a training set, a validation set, a test set, and a challenge set [7]. For the training set, validation set, and test set, fully sampled MR data remains available. Additionally, for the test set, official undersampled versions of the data are included as well. This is achieved by employing retrospective 2D Cartesian undersampling in the phase encoding direction, replicating clinically feasible accelerations in 2D MR data acquisitions [7]. An identical undersampling mask is used for all slices within a given scan volume, and the global acceleration factor R is randomly assigned as either 4 or 8, with equal probability for each. When

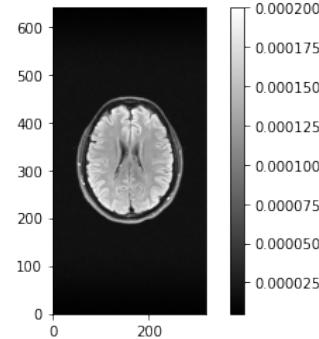
3. METHODS: DATA ANALYSIS



(A) Magnitude of the slice's fully sampled, raw k-space data of coils 1, 3, 6, 8 and 11.



(B) Magnitude of the slice's corresponding coil images.



(C) Root-sum-of-squares reconstruction of this slice.
This is put forward as the ground-truth image.

FIGURE 3.1: Exemplary raw data of an MR scan volume slice, part of the fastMRI brain dataset. The displayed data was acquired through a FLAIR imaging sequence and measured with 16 coils and an MR image slice matrix size of 640x320. Note that no preprocessing was applied here yet, and that the slice has a FOV of 320x320.

$R=4$, the central autocalibration signal (ACS) region encompasses 8% of all k-space lines, while for $R=8$, 4% of all k-space lines are included within the ACS region. Subsequently, after introducing a random offset, the remaining lines are uniformly sampled at equidistant intervals to achieve the targeted acceleration factor [7]. These and other, self-implemented undersampling masks will be more extensively discussed and visualized in subsection 3.1.3.

A special case is formed, however, by the challenge set. For the scan volumes of the challenge set, no fully sampled data is released and only the undersampled version of the MR data is available (undersampled in an identical manner as the test set) as this data is meant to be used for any official fastMRI challenge leaderboard submissions [7]. As a result of the lack of fully sampled data, evaluations on the challenge set become impossible. Therefore, the challenge set will be disregarded in this work, but besides this, the official fastMRI brain dataset split will be used. An overview of the official fastMRI brain dataset split is given in Table 3.2.

Type	Scan Volumes
Training	4.469
Validation	1.378
Test	558
Challenge	565
Total	6.970

TABLE 3.2: The official fastMRI brain dataset split.

3.1.3 Undersampling

Of course, in order to train a DL model, a loss function needs to be defined using input-reference data pairs as explained in Chapter 2. Initially, the fastMRI brain dataset mainly contains fully sampled reference data, so any required undersampled input data still needs to be created for the training of the different hybrid models from Chapter 4. For this processing step, it is important to already appreciate that three different hybrid model pipelines will be evaluated later on in this work: a GRAPPA hybrid model, a SENSE hybrid model, and a CS hybrid model.

Naturally, as it is the most interesting and useful to compare these hybrid implementations of the associated classical reconstruction techniques based on how the latter are actually implemented in clinical practice, this means that different sampling masks are required for the training of each of these hybrid DL models.

Regarding GRAPPA acquisitions, hospitals and literature are pretty clear [66]; the sampled k-space data contains an ACS region of a given size inversely proportional to the acceleration factor R , and outside of the ACS region the remaining k-space lines are sampled in an equidistant manner until the desired acceleration factor R is reached. This is identical to how the fastMRI brain test data is (under)sampled.

3. METHODS: DATA ANALYSIS

SENSE acquisitions also have a clear type of sampling mask used in clinical practice [66]; the whole k-space is symmetrically sampled in a uniform and equidistant way, starting from the $k=0$ line of k-space, in such a manner that the desired acceleration R is reached without the presence of an ACS region.

For CS acquisitions, however, hospitals and literature are less clear about the used k-space sampling patterns. Consequently, the author and his supervisors manually collected some exemplary raw CS data from one of the MR scanners of the UZ Leuven in order to obtain some real examples of CS k-space masks. This way, the chances of using clinically relevant CS sampling patterns increase.

Again, each time, the same 2D undersampling mask is applied uniformly to all slices in a given scan volume. Some visualizations of the three above-discussed sampling mask types are shown in Figure 3.2.

The only remaining undersampling design consideration now pertains to the acceleration factor(s) used to undersample the fully sampled fastMRI brain data with these sampling mask types. As the fastMRI test data is officially undersampled with acceleration factors $R=4$ and $R=8$, it would make sense to train the models on training data containing accelerations of both $R=4$ and $R=8$ as well. However, when looking at the fastMRI GitHub repository, it becomes apparent that any of their original implementations are actually solely trained on data where $R=4$ by default [67]. Due to this, a brief comparative experiment was performed, where a conventional convolutional neural network (CNN), described in Chapter 4 as the baseline DL model, was trained on two separate occasions; once with training data solely containing accelerations where $R=4$, and once with training data containing a mixture of $R=4$ and $R=8$ accelerations, each with an equal probability of occurring. GRAPPA sampling masks were used. The results of this experiment are displayed and discussed more extensively in Chapter 5 (Table 5.1), but based on these results, it was decided to indeed undersample the training and validation data as a mixture of $R=4$ and $R=8$ accelerations, each with an equal probability of occurring.

For the test data, which initially comprises only the official GRAPPA-based undersampling, additional undersampling processes based on SENSE and CS masks were also conducted, employing the same acceleration strategy as mentioned above.

This way, all three types of training, validation, and test data pairs required for the desired training and evaluation of each of the three different hybrid models are obtained.

3.2 Preprocessing

As explained in the introduction of this chapter, this work's investigated hybrid neural network pipelines each try to form a composite function $\hat{\mathbf{x}} = f_2(f_1(\mathbf{y}))$. Here, f_1 represents a classical reconstruction operation layer at the input of the hybrid model pipeline, and f_2 represents a CNN whose input layer takes the output of the f_1 layer as their input.

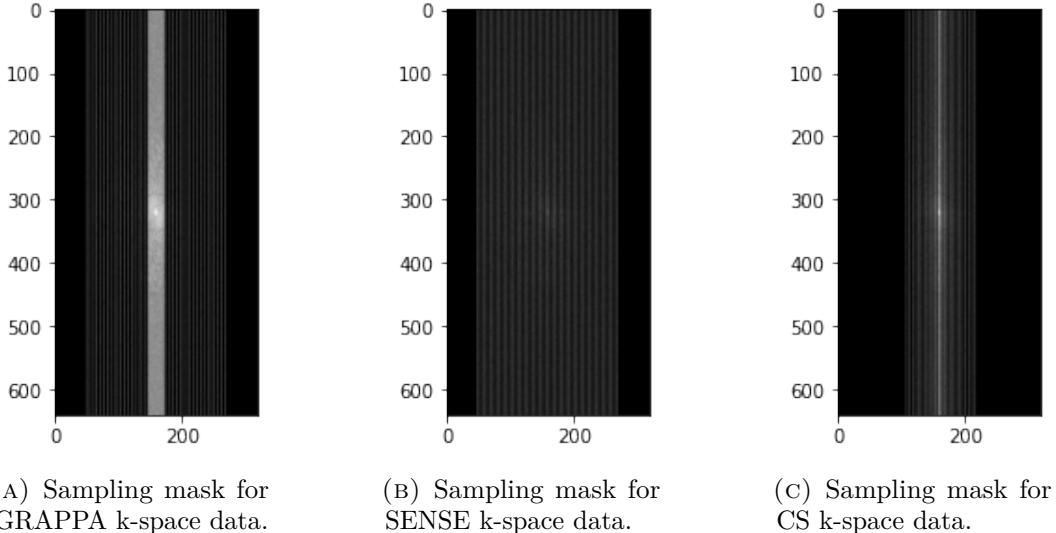


FIGURE 3.2: Illustration of the different k-space sampling mask types used to train the three different hybrid model pipelines from Chapter 4. Here, R=4.

However, this doesn't mean that the currently envisioned

$$\text{raw undersampled data } \mathbf{y} \longrightarrow f_1 \longrightarrow f_2 \longrightarrow \hat{\mathbf{x}} \quad (3.1)$$

pipeline will immediately work, as some transitions between the different pipeline components require some additional preprocessing of the raw and intermediate data.

More concretely, different classical reconstruction operation layers f_1 may require differently preprocessed raw data at their input, with e.g. some function versions of f_1 just requiring ACS data while others require estimated coil sensitivity maps \mathcal{S}_i , whilst their output also needs some additional preprocessing steps such that the function f_2 can learn and converge its learnable parameters optimally and correctly. Because of this, two additional preprocessing layers are added to the pipeline presented in equation 3.1:

$$\text{raw undersampled data } \mathbf{y} \xrightarrow{p_1} f_1 \xrightarrow{p_2} f_2 \rightarrow \hat{\mathbf{x}} \quad (3.2)$$

The preprocessing layers or functions p_1 and p_2 may vary along with varying f_1 layers, with possible f_1 layers being a GRAPPA layer, a SENSE layer, or a CS layer.

As mentioned before, the total hybrid neural network pipelines will be discussed more in detail in Chapter 4, while in the current chapter special attention is given to the layers p_1 and p_2 in the following two subsections.

3. METHODS: DATA ANALYSIS

3.2.1 Preprocessing layer p_1

The nature of the raw undersampled k-space data \mathbf{y} was already discussed in subsection 3.1.3. Although this data already underwent some processing, this data isn't always readily compatible with each of the different classical reconstruction operation layers at the input of the different hybrid models.

Looking back at Table 2.1, it can be seen that GRAPPA requires ACS data whilst SENSE and CS require the availability of estimated coil sensitivity maps \mathcal{S}_i . As the used GRAPPA sampling masks already include ACS data in \mathbf{y} , however, this means that p_1 is an identity transformation layer for the GRAPPA hybrid model. Nonetheless, the situation becomes a bit more complex for the SENSE and CS hybrid models.

As mentioned earlier in Chapter 2, coil sensitivity maps \mathcal{S}_i are usually obtained through the use of either complete or partial autocalibration measurements. Complete autocalibration measurements directly allow the estimation of \mathcal{S}_i in the following manner: If the k-space of the different coils is fully sampled then surface coil images \mathbf{x}_i of high quality can be obtained through the direct employment of an Inverse Fourier Transform. A proxy to the result of an acquisition using a coil with homogeneous sensitivity can then also easily be obtained via the root-sum-of-squares reconstruction \mathbf{x} of the data. Using these obtained images, the coil sensitivity maps can then easily be calculated as [29][68]:

$$\mathcal{S}_i = \mathbf{x}_i / \mathbf{x} \quad (3.3)$$

Alternatively, for partial autocalibration measurements, more intricate estimation methods are required, with the most popular one being ESPIRiT [28]. The advantage of ESPIRiT over estimation methods requiring complete autocalibration measurements is that ESPIRiT only requires an ACS region for its estimations, which can easily be integrated into the actual imaging protocol and thus avoids the need for a separate, full scan before the imaging of the actual patient can start, saving time. In brief, the data of the ACS region is utilized by ESPIRiT to generate a calibration matrix A , constructed by employing a sliding window technique across the ACS data, where each data block within the ACS region is reshaped and treated as a row in A . The resulting columns of A are then actually shifted versions of the ACS area, resulting in a matrix configuration recognized as Block-Hankel [28]. Based on a singular value decomposition of A , the coil sensitivity maps \mathcal{S}_i can then be obtained. For a full derivation, please refer to [28], as the mathematics are not really relevant for this thesis and off-the-shelf implementations of this method exist.

Note that, since fully sampled data is present in the fastMRI brain dataset, both ESPIRiT and equation 3.3 can be used to calculate any necessary coil sensitivities. Nevertheless, it is desired that the created hybrid neural networks from this work remain generally compatible with all possible types of MR data acquisitions, from both clinical practice and research applications.

Because of this, p_1 is chosen to be an ESPIRiT layer for the SENSE and CS hybrid DL models. The reasoning behind this is the following: If complete autocalibration measurements, and thus the resulting coil sensitivity map estimations, are readily available, one can always simply remove the p_1 layer from the network and feed the available sensitivities \mathcal{S}_i directly to the function f_1 to start making inferences. However, if only partial autocalibration measurements are available, then p_1 remains absolutely necessary for network predictions and ESPIRiT is required.

The following workflow is then followed during network training in the case p_1 is an ESPIRiT layer: For any incoming undersampled training data, the corresponding fully sampled reference data is taken and masked with a GRAPPA mask containing an ACS region with a size related to the data's acceleration factor R , creating a realistic ACS region as required by ESPIRiT. On this result, ESPIRiT is applied using an off-the-shelf toolbox; the BART toolbox [69]. This then finally creates the coil sensitivities \mathcal{S}_i which need to be passed to the next layer of the hybrid neural network pipeline. An example of such coil sensitivity estimations, calculated for the data visualized in Figure 3.1, is shown in Figure 3.3. Note that coil sensitivity maps \mathcal{S}_i don't convey any anatomical information, as they display intensity ratios rather than absolute pixel intensities.

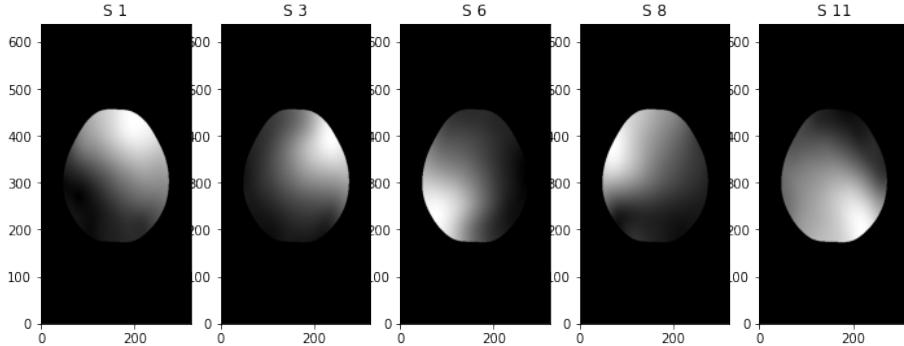


FIGURE 3.3: Illustration of the magnitude of the coil sensitivities \mathcal{S}_i calculated by the BART toolbox for the data shown in Figure 3.1.

3.2.2 Preprocessing layer p_2

As will be discussed in Chapter 4, the output of f_1 is a single complex-valued image representing the final result of that f_1 's corresponding classical reconstruction algorithm. Despite the already major simplification of the data structure compared to the initial data structure of \mathbf{y} , there is still some preprocessing required before the training of the DL part of the pipeline can commence. First, p_2 converts f_1 's output to a real-valued magnitude image, as these are readily compatible with real-valued networks and since standard diagnostic MR image analyses rely on signal intensity alone [70]. Additionally, it is known that neural networks trained on normalized data converge much better during training, even for radiology images [71]. However, the question then becomes: How should the data be normalized?

3. METHODS: DATA ANALYSIS

As it is the most common to normalize MR data via Z-Score normalization, and as the fastMRI researchers originally perform normalization at the slice-level, p_2 also becomes a Z-Score normalization layer at the slice-level for all the hybrid models [67][72]. Additionally, before this result is passed to f_2 , p_2 lastly performs an image cropping operation. This means that the images are cropped to their FOV, compensating for oversampling in the frequency encoding direction of the initial k-space measurements, after which they are either further cropped or padded to have a standard size of 320x320. This is important for two reasons. Firstly, this ensures that all images have the same size before being passed to the neural network f_2 , but more importantly, this ensures that the DL model doesn't try to maximize the reconstruction performance by learning to accurately reconstruct background pixels, given that it should mainly focus on the actual anatomies it is presented with.

3.3 Reconstruction evaluation metrics

After the models from Chapter 4 are trained, model inference is used to make reconstruction predictions on the test dataset. Based on these inference results, as mentioned in subsection 3.1.2, the performance of the trained models can then be studied in relation to the ground truth reconstruction from fully sampled data. The evaluation metrics by which this comparison is done are a crucial choice, as a visual comparison in all test data is subjective and overly time-consuming.

As explained in subsection 2.3.2, the Normalized Mean Squared Error (NMSE) is often seen as the primary indicator of reconstruction quality. Nonetheless, as already mentioned before, the NMSE metric favors smoothness over sharpness as the NMSE penalizes large deviations more than smaller ones. Because of this, the evaluations made in Chapter 5 will also include the Peak Signal to Noise Ratio (PSNR), the mean-Structural Similarity (MSSIM) index, and the VGG loss, in an attempt to also consider noise levels, contrast, brightness, structure, and perceptual similarity more explicitly. Note, however, that since the PSNR is indirectly related to the NMSE, the PSNR will also likely share some of its bias. In case of the MSSIM index, an 11x11 Gaussian window with a filter width of $\sigma=1.5$ is used, as these are the default values for the MSSIM index calculation. For the calculation of the VGG loss, the feature maps of the first four convolutional blocks of a VGG19 network are chosen, exactly as used in [58]. All evaluations employ a brain mask, as calculated by the BART toolbox, and are performed at the image volume-level, calculating the metrics in 3D if possible (NMSE, PSNR) or in 2D averaged across the image volume's slices if not (MSSIM index, VGG loss) given that 2D evaluations are also appropriate since the fastMRI data was collected using 2D sequences.

However, in spite of these efforts to make the evaluation process as multifaceted as possible, results that will be presented in Chapter 5 illustrate that the NMSE, PSNR, and MSSIM may not be the best means to reflect reconstruction quality in MR image applications, where perceptual similarity is more important than pure error reduction.

Therefore, a new evaluation metric is proposed and used in this work, tailored specifically to the evaluation of structure in residual error maps: the Singular Value Decomposition Metric (SVDM).

The intuition behind the SVDM is as follows: If a reconstructed MR image $\hat{\mathbf{x}}$ is of high quality and thus highly perceptually similar to the corresponding ground truth image \mathbf{x} , then its residual error map $(\hat{\mathbf{x}} - \mathbf{x})^2$ will be very noisy. In that case, the singular values of that residual error map's singular value decomposition $(\hat{\mathbf{x}} - \mathbf{x})^2 = \mathbf{U}\Sigma\mathbf{V}^T$ will start to approach a Marchenko-Pastur distribution around said noise level, which is the statistical distribution that describes the singular value distribution of a matrix that contains pure Gaussian noise [73]. In the case of a bad reconstruction, the opposite will be true: the residual error map will contain a lot of structure, and its singular value decomposition will display multiple singular values that are much larger than those expected within the Marchenko-Pastur distribution. This means that reconstruction quality can be assessed based on the singular value distribution of the residual error map of a given reconstructed MR image, avoiding biases introduced by only looking at error magnitudes through the consideration of error structure as well.

To translate this mathematical intuition to a metric for the 2D image case, the following formula is proposed:

$$\text{SVDM} = \min_{k \in \mathbb{N}} \frac{k}{n} \quad \text{s.t.} \quad \sum_{i=1}^k \Sigma_{ii} \geq 0.9 \sum_{i=1}^n \Sigma_{ii} \quad (3.4)$$

where Σ_{ii} are the singular values (sorted in decreasing order) of the residual error map $(\hat{\mathbf{x}} - \mathbf{x})^2$, and n is the rank of the 2D MR image matrix. As k can never be lower than 0 or higher than n , the SVDM will always have a value between 0 and 1.

Given that the singular value distribution of a residual error map containing structure will be dominated by high singular values, as explained above, the SVDM will be low, and vice versa. This means that MR image reconstructions with a higher reconstruction quality will display a higher SVDM value. In the 3D case, where an MR image volume is considered instead of an MR image slice, the SVDM is calculated for each image volume slice and averaged across the different slices.

Although singular value decompositions have already been used as evaluation tools for the sharpness and noisiness of images [74][75], this evaluation technique has, to the best of the author's knowledge, not yet been applied to the residual error maps of regression tasks such as the MR image reconstruction task as described in this work. Moreover, the aforementioned existing evaluations are based on the trace of just the first few singular values of a given image. This means that a much smaller portion of the singular value distribution is taken into account compared to the here-proposed SVDM.

A verification of the SVDM's desired behaviour will be reported and discussed in Chapter 5 (section 5.2).

3.4 Conclusion

A closer examination of the data and evaluation metrics to be utilized for model training and evaluation was undertaken in this chapter. In this work, the fastMRI brain dataset will be used, consisting of 6.970 brain scan volumes acquired via 2D acquisitions performed at different locations. This dataset is quite heterogeneous, as it doesn't only contain scan volumes from different types of Siemens scanners but also from different sequences (Table 3.1). Additionally, the official fastMRI brain dataset split will be used (Table 3.2).

This chapter also revealed that three hybrid models will be trained and evaluated in this thesis: a GRAPPA hybrid model, a SENSE hybrid model, and a CS hybrid model. This means that three different versions of the training, validation, and test data need to be generated, as these reconstruction paradigms all require different k-space sampling masks. In addition, based on an experiment of which the results are presented in section 5.1, it was decided to train these models on training data containing acceleration factors of both R=4 and R=8. Because of this, all datasets were undersampled as a mixture of R=4 and R=8 accelerations, each with an equal probability of occurring, using the three different mask types displayed in Figure 3.2.

Inside the hybrid model pipelines, before the data arrives at the deep learning function, the raw and intermediate data still require some additional preprocessing. This preprocessing becomes possible via the inclusion of two preprocessing layers p_1 and p_2 inside the hybrid models; one before (p_1) and one after (p_2) the classical reconstruction layer of the hybrid model. For the GRAPPA hybrid model, p_1 is an identity transform operation layer, while for the SENSE and CS hybrid models p_1 is a coil sensitivity estimation layer employing the ESPIRiT algorithm. For all models, p_2 is a magnitude calculation, Z-Score normalization, and image cropping layer.

Lastly, this chapter stated which evaluation metrics will be used for model evaluation. Moreover, a new evaluation metric is proposed and used in this work, tailored specifically to the evaluation of structure in residual error maps; the Singular Value Decomposition Metric, in hopes of avoiding biases introduced by only looking at error magnitudes through the consideration of error structure as well.

Chapter 4

Methods: The Hybrid Models

When it comes to the training of a neural network, part of the performance outcome of the model is determined by the collection of details of the employed architecture and training paradigm. As a result, this chapter discusses these aspects of the hybrid convolutional neural network pipelines investigated in this thesis. A first section examines the pipeline architecture, while a second section discusses the used training paradigm.

4.1 Hybrid Neural Network pipeline architecture

As it might already be appreciated from Chapter 3, each of the final hybrid neural network (HNN) pipelines proposed in this work forms a specific version of the following general, mathematical composite function: $f_2 \circ p_2 \circ f_1 \circ p_1$. Here, the exact function of f_1 and p_1 depends on the HNN type, whilst the functions f_2 and p_2 remain the same for all reviewed (hybrid) deep learning (DL) pipelines. Component f_2 makes up most of the pipeline architectures and contains the pipelines' actual learnable parameters, as this is the conventional convolutional neural network (CNN) component of the HNNs, while the functions f_1 , p_1 , and p_2 are the HNN components consisting of non-learnable parameters. Here, f_1 represents the result of the integration of a classical reconstruction algorithm into the DL pipelines, i.e. a classical reconstruction operation layer, transforming the original CNN architectures into HNN architectures, while p_1 and p_2 are preprocessing functions that were already discussed in Chapter 3.

To answer the research questions posed in the first chapter of this thesis, three HNN pipelines or architectures are investigated: a GRAPPA hybrid model, a SENSE hybrid model, and a CS hybrid model. This thus creates three different HNNs, each with a different f_1 layer applying the correct mathematical operations on its input data exactly as described by subsections 2.2.2, 2.2.3, or 2.2.4. In the case of CS specifically, f_1 performs wavelet-based CS with $\lambda=0.005$ and a maximal number of iterations equal to 50, as these are the standard values suggested by the BART toolbox. For each HNN model, p_1 ensures that the correct inputs are delivered to f_1 which then outputs a single, complex-valued reconstruction $\hat{x}_{\text{intermediate}}$ that is passed to p_2 and eventually to the CNN f_2 .

4. METHODS: THE HYBRID MODELS

f_1 was implemented using the BART toolbox [69] for the SENSE hybrid model and CS hybrid model, while for the GRAPPA hybrid model the GRAPPA python package pygrappa [76] was used.

Considering that many of the DL frameworks discussed in Chapter 2 use U-Nets, and considering that the DL component of the HNN is not the focus of this work, a rather unambiguous real-valued feedforward U-Net is taken as the HNNs’ CNN f_2 . It is a new, modified version of the Baseline U-Net discussed in subsection 2.4.2, which in turn was based on the work of Ronneberger et al. [77].

The selected U-Net architecture, denoted as f_2 , is structured around two distinct (but connected) convolutional sub-networks. One network handles down-sampling and feature extraction through the use of several max-pooling layers with a 2x2 kernel and stride of 2, resulting in halved spatial dimensions. Feature extraction is further facilitated by employing blocks consisting of a pair of consecutive 3x3 convolutions interleaved with instance normalization layers and Parametric Rectified Linear Unit (PReLU) activation functions. Mirroring the structure of this down-sampling network, the other network (the up-sampling network) then increases the spatial resolution again using bilinear up-sampling layers with a scale factor of 2, doubling the resolution between blocks. Additionally, unlike the down-sampling network, the up-sampling network incorporates the concatenation of two inputs to the first convolution of each block: the up-sampled features from the preceding block and the features propagated through the skip connection from the corresponding block in the down-sampling network at the same resolution. Finally, a series of 1x1 convolutions is employed at the end of the up-sampling path to reduce the number of channels to one while retaining spatial resolution. This effectively ensures that a single, real-valued magnitude image with the same spatial resolution as the input image $p_2(\hat{\mathbf{x}}_{\text{intermediate}})$ is returned by the network f_2 .

Of course, the number of top-level channels and the number of max-pooling (and up-sampling) operations remain to be selected as hyperparameters for the above-described CNN architecture. In order to make the HNNs expressive enough for the Magnetic Resonance image reconstruction task, yet also ensure that the HNNs can be trained on the available hardware, 32 top-level channels and 4 max-pooling operations were used in the final CNN architecture of f_2 .

Given that all of the different HNNs have the same global, composite architecture $\hat{\mathbf{x}} = f_2(p_2(f_1(p_1(\mathbf{y}))))$ which is now fully known and explained, this hybrid convolutional neural network pipeline architecture can be visualized in a general way, as shown in Figure 4.1.

One factor remains missing, however, before every single one of the proposed research questions may be answered; a conventional DL baseline method (not containing any classical reconstruction method layers) with which the HNNs can be compared. For this, this work’s Zero-Filling (ZF) DL model was created, consisting of a version of the general HNN pipeline where p_1 is an identity transformation layer and f_1 is a simple Inverse Fourier Transform + root-sum-of-squares operation

4.2. Training paradigm

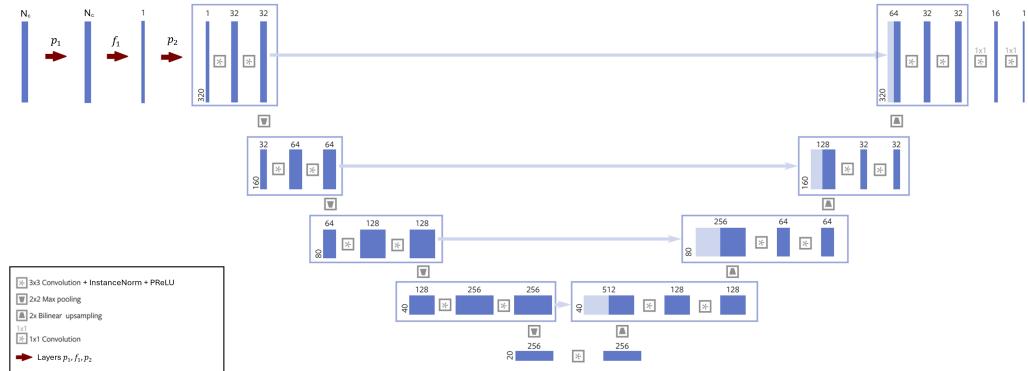


FIGURE 4.1: The general architecture of the investigated HNN pipelines.

combination. This preprocessing effectively leaves all initially missing k-space samples as zeroes, justifying the name of the proposed baseline DL model, and converts the HNN pipeline back to a conventional CNN pipeline. The ZF DL model pipeline was trained and evaluated twice; once for data where an autocalibration signal (ACS) region is present, and once for data where an ACS region is absent, as the presence of an ACS region already inherently affects reconstruction quality. In the first case, a GRAPPA sampling mask as described in Chapter 3 was used as the ZF sampling mask, as this mask always contains ACS data. In the second case, a SENSE sampling mask was used as the ZF sampling mask, as this mask never contains ACS data. Besides the above mentioned changes, the rest of the pipeline (i.e. f_2 and p_2) remains the same as for the HNNs, to make the experiments as comparative as possible.

4.2 Training paradigm

The training paradigm of a neural network consists of a loss function, an optimizer, and some training settings. Again, as the DL part of the HNNs is not the focus of this thesis, a simple l_1 loss was chosen as the HNNs' loss function:

$$\hat{\theta} = \operatorname{argmin}_{\theta} \sum_{i=0}^{n_{data}} \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_1 \quad (4.1)$$

It is important, however, to mention that, for the training of the investigated networks, L2 regularization with a penalty hyperparameter value of $\lambda=0.0005$ was used. The purpose of this is two-fold. Firstly, this prevents overfitting and thus might help boost the performance of the model on new, unseen data. Secondly, this prevents numerical instabilities and exploding gradients from occurring. This makes the actual loss function equal to:

$$\hat{\theta} = \operatorname{argmin}_{\theta} \sum_{i=0}^{n_{data}} \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_1 + \sum_{j=0}^{n_{params}} \theta_j^2 \quad (4.2)$$

4. METHODS: THE HYBRID MODELS

The value of λ was chosen based on the results of a study related to regularization for image-based neural networks [78].

The loss function is then optimized by the numerical RMSprop algorithm each time an HNN is trained. The RMSprop and Adam optimizers are usually regarded as the most (equally) suitable optimizers for DL training purposes given that they both allow the usage of an adaptive learning rate, but based on a small comparative experiment presented in [79], RMSprop was chosen over Adam. Models were each trained for 50 epochs, with an initial learning rate of 0.001 which decayed to 0.0001 after 40 epochs. A batch size of 1 was used, meaning that each batch contains a single image slice. The experiments were seeded in order to make the network parameter initializations deterministic for all the HNNs and normalized initializations were used as this is the default initialization method of PyTorch. After each training experiment, the model weights and biases for which the performance on the validation data was the best were saved. Model training was performed on an Asus GTX 1080 Ti Founders Edition 11GB GPU system. Training times of the HNNs were rather long due to the long iteration times of the classical reconstruction operation layers, with the training of a complete HNN pipeline usually taking 3 weeks to complete on the aforementioned hardware.

4.3 Conclusion

In this chapter, the Hybrid Neural Networks which will be utilized in the following chapter were presented. Their network architecture is summarized in Figure 4.1. Exact implementation methods were also mentioned, with GRAPPA being implemented using pygrappa [76] and SENSE and Compressed Sensing being implemented using the BART toolbox [69]. Additionally, two versions of the used conventional deep learning baseline network were also introduced, with one Zero-Filling deep learning model using data where an autocalibration signal region is present in k-space, and the other using data where said region is absent.

All deep learning-based models are trained using an l_1 loss function and L2 regularization, in hopes of preventing overfitting and stabilizing the training processes. Training persists for 50 epochs, with an initial learning rate of 0.001 before decaying to 0.0001 after 40 epochs. RMSprop was chosen as the optimizer used to train the models.

Chapter 5

Results & Discussion: Model Training And Evaluation

Hybrid modelling sounds promising, but is this complex strategy really worth using? If so, which classical reconstruction method has a higher synergy with deep learning models? And what is the largest acceleration factor that can reasonably be achieved with the proposed models? Those are the questions this chapter will try to answer.

Here, experiment results are presented and discussed in three parts. In the first section, the investigation focuses on whether training deep learning models solely on data with a single acceleration factor R is preferable, or if using a mixture of acceleration factors leads to better performance (see Chapter 3). The second section focuses on the properties and the verification of the behaviour of the SVDM (see Chapter 3). The third section presents the core of this chapter and thesis. It compares the performance evaluations of different hybrid models to those of their conventional DL model, classical model, and model-absent counterparts.

The investigated paradigms are the GRAPPA paradigm, the SENSE paradigm, and the CS paradigm. The corresponding hybrid models are referred to as GRAPPA U-Net, SENSE U-Net, and CS U-Net respectively, while the two versions of the baseline conventional DL model are referred to as the ZF (ACS) U-Net and the ZF (no ACS) U-Net.

5.1 Model training

5.1.1 Results

All deep learning-based models from this work are trained on a mixture of data containing a combination of R=4 and R=8 accelerations. This decision was made based on the results of a comparative experiment (see subsection 3.1.3), for which ZF (ACS) U-Net was used. The average values of the evaluation metrics calculated for the test scan reconstructions made by ZF (ACS) U-Net trained on R=4 data and made by ZF (ACS) U-Net trained on R=4+8 data are displayed in Table 5.1.

R _{training}	R _{evaluation}	NMSE	PSNR	MSSIM	VGG	SVDM
4	4+8	0.06352	30.91	0.8876	0.09488	0.2906
	4	0.01206	35.73	0.9449	0.05324	0.3329
	8	0.11743	26.16	0.8326	0.13659	0.2499
4+8	4+8	0.02087	33.81	0.9285	0.06899	0.317
	4	0.01211	35.72	0.9449	0.05418	0.3326
	8	0.02946	31.96	0.9135	0.08256	0.3024

TABLE 5.1: Results of the comparative experiment described in subsection 3.1.3 investigating the influence of R_{training} on model performance. The model performances were evaluated on test scans with R=4, R=8, or on all test scans together. Average metric values are displayed. Performance differences for R_{evaluation}=4+8 are statistically significant for every metric (Wilcoxon signed-rank test: p_{corrected}<.0001).

For evaluations on the entire test set (R=4+8), it was tested whether the performance differences are statistically significant. For this, a two-sided paired samples Wilcoxon signed-rank test was used, as the test concerns matched samples and the differences between the pairs are not approximately normally distributed for any of the metrics (Shapiro-Wilk test: p<.0001). For every metric one test was performed, after which Bonferroni-corrected p-values were calculated. In this case, this returned a p_{corrected}<.0001 for all tests, indicating that the performance improvement (lower NMSE, higher PSNR, higher MSSIM index, lower VGG loss, and higher SVDM) as indicated by the values in Table 5.1 is significant for every metric when considering the entire test dataset.

The learning curves of the models are shown in Figure 5.1. It shows model performance on the validation dataset, tracked throughout training by calculating the average 'validation loss' for a given epoch. From these figures, it can be seen that no overfitting is occurring yet. Additionally, ZF (ACS) U-Net trained on a mixture of accelerations seems to be converging at a much faster pace throughout its learning process compared to that of ZF (ACS) U-Net trained on a single acceleration factor. From training step 500.000 onwards, ZF (ACS) U-Net trained on a single acceleration factor manages to decrease its average validation loss by a value of 0.012 while ZF (ACS) U-Net trained on a mixture of accelerations manages to decrease its average validation loss by a value of 0.024, doubling the aforementioned convergence speed.

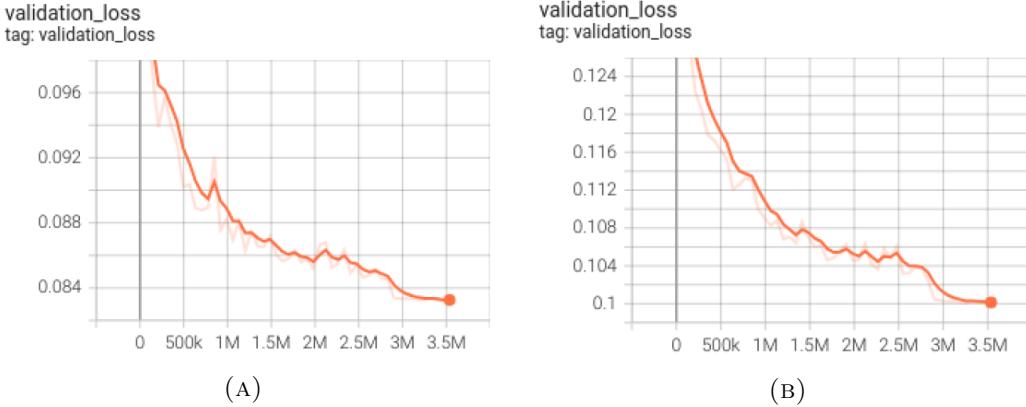


FIGURE 5.1: Learning curves displaying the evolution of model performance on the validation dataset during training, once for ZF (ACS) U-Net trained on R=4 training data (A) and once for ZF (ACS) U-Net trained on a mixture of R=4 and R=8 training data (B). The x-axis displays the training epochs converted into training steps (here, equal to the number of image slices in the training dataset times the number of epochs) and the learning curve itself is slightly smoothed for visual purposes, though the original curve is still shown with a lower opacity. The y-axis displays the loss values, calculated and averaged across the validation dataset.

5.1.2 Discussion

Analysis of the Normalized Mean Squared Error (NMSE), Peak Signal to Noise Ratio (PSNR), mean-Structural Similarity (MSSIM) index, VGG loss, and Singular Value Decomposition metric (SVDM) distributions for the reconstructions of the two networks reveals the advantage of training deep learning (DL) models on a mixture of acceleration factors.

When specifically looking at the subsets of the test set containing only R=4 or R=8 scans, it appears that the inclusion of R=8 training scans not only aids in the reconstruction of R=8 test scans but that it also prevents large performance degradations for R=4 test scan reconstructions for every metric.

One could hypothesize that these observations may be explained by a combination of (partly) shared information between different accelerations and the presence of a Pareto optimal in the statistically learned reconstruction function for a given acceleration factor (e.g. R=4) and used dataset.

Firstly, it's plausible that the artifacts produced by R=8 scans actually form a larger set of which the artifacts produced by R=4 scans are a subset. This would mean that the broader artifact sets produced by higher acceleration factors like R=8 share informational traits with lower accelerated (but still aliased) Magnetic Resonance (MR) scans (where e.g. R=4). Secondly, given that the training dataset consists of 4.469 scan volumes, each consisting of multiple slices and each representing different MR acquisition parameters, anatomies, and coil geometries, it could very

well be that the model reaches a training point where it's learning certain pixel relationships which improve the reconstruction performance for one type of $R=4$ scan but decrease the reconstruction performance for another type of $R=4$ scan. Of course, given that the fastMRI dataset is heterogeneous and unbalanced, this might still result in net improvements of the loss function calculated and averaged across the entire training and validation dataset in an epoch allowing the model to still keep converging throughout the 50 epochs of training, but with a slower pace and less drastic performance gains on the test dataset.

However, if the above is true, then it is also not very efficient to employ all 4,469 training scans to let the model slowly hone in on this one distant Pareto optimal for this one acceleration factor, especially if artifacts from other, higher acceleration factors are partly informative about the artifacts from this one acceleration factor as well. Thus, using e.g. merely half of the training data would then only minorly reduce reconstruction performance for this one acceleration factor, exactly as observed in Table 5.1, whilst simultaneously freeing up available data to let the model learn artifact patterns as produced by other higher acceleration factors, resulting in a more drastic increase in performance for those accelerated scans, as can also be observed from the same table.

This hypothesis gains additional support from the observation of a clear slower convergence in the training regime featuring solely $R=4$ scans, in contrast to the training regime encompassing both $R=4$ and $R=8$ scans, as depicted in Figure 5.1. This indicates that, indeed, a combination of an intra-acceleration factor Pareto optimal and an inter-acceleration information sharing mechanism might possibly be at play here.

5.2 Model evaluation process

Several conventionally used evaluation metrics already exist, but each has its limitations. It is believed that a metric based on singular value decompositions can offer added value, avoiding a reliance on error magnitudes alone. Therefore, the new SVDM evaluation metric was introduced to reflect potential structure in the residual error maps of reconstructed MR images (see Chapter 3). This section now aims to verify its behaviour in different situations, to ensure that the metric behaves as desired, at least in the specific evaluation situations that are relevant for this thesis.

More specifically, the SVDM should at least be able to correctly and clearly detect noise, blurring, changes in the acceleration factor, and changes in the autocalibration signal (ACS) region size.

Four experiments were performed where a randomly chosen Zero-Filled (ZF) MR image from the fastMRI brain test dataset was taken and altered with each of the aforementioned image modifications, plotting the SVDM for varying degrees of severity. In the case of additive Gaussian noise, the standard deviation of the noise was scaled to the mean of the image's pixel values through the use of a noise factor in order to yield a signal-to-noise ratio (SNR) metric that is easy to interpret.

5.2.1 Results

The results of these experiments are shown in Figure 5.2.

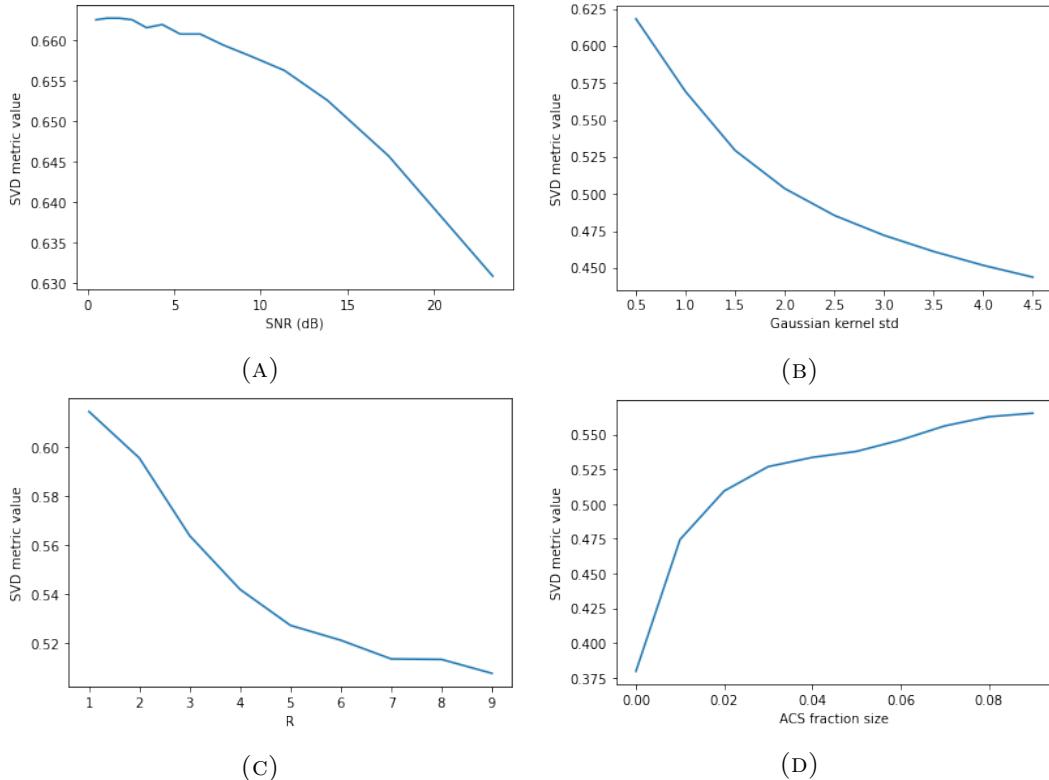


FIGURE 5.2: Results of the SVDM verification experiments described in section 5.2. These graphs display the influence of Gaussian noise (A), Gaussian blurring (B), the acceleration factor R (C), and the ACS fraction size (D) on the SVDM. Std = standard deviation.

In the case of Gaussian blurring, higher kernel standard deviations lead to error maps with more structure as the image becomes more and more dissimilar to the ground truth image, which in turn consistently leads to lower SVDM values. Similarly, higher acceleration factors and lower ACS fraction sizes generate more outspoken image artifacts and thus also generate more structured error maps with lower SVDM values. However, for increasing levels of noise (decreasing SNR), the SVDM seems to increase.

5.2.2 Discussion

Based on these results, the SVDM adequately serves the intended purpose for which this metric was introduced in this study, namely, to mitigate biases arising from solely considering reconstruction error magnitudes by also looking at reconstruction error structure via the SVDM. The SVDM may thus be relied upon in this work's analyses.

Considering the cases of blurring and undersampling artifacts, the SVDM values are proportional to how the human visual system would evaluate these images. However, one important, but expected, pitfall of the SVDM comes forward in the experiment where Gaussian noise is added to the ZF MR image. Visually, as the SNR decreases, the image also becomes more and more dissimilar to the ground truth image. This would thus lead to a lower quality score given by the human visual system. In spite of this, the SVDM seems to increase instead. Although a bit counter-intuitive, this behaviour is expected by design. As the noise starts to dominate the MR image, its error map contains increasing relative amounts of noise compared to the already present amount of error structure. This means that noise will start to dominate over structure in a relative manner in the error map of the image, leading to higher SVDM values. Though this shows that the SVDM works as intended and can be used in this work, this also shows that the SVDM should still be complemented by metrics which look at error magnitude. This way, the complete nature of the errors can be investigated. That is also the reason why, in total, 5 evaluation metrics are used in this thesis: the NMSE, PSNR, MSSIM index, VGG loss, and SVDM.

Interestingly, the plots also reveal that the SVDM behaves in a non-linear manner. This is likely caused by the asymptotic situations where error maps either contain a maximum amount of relative structure or a maximum amount of relative noise, leading to plateaus in the SVDM evaluation curves presented here.

Finally, it does remain important to mention that the ideal scenario where the error maps of reconstructed MR images only contain spatially homoscedastic Gaussian noise, thus maximizing the SVDM of those reconstructions, likely doesn't exist. This is because in multi-coil applications, each receiver coil will display different amounts of thermal and other background noise, leading to a root-sum-of-squares MR image which exhibits locally homoscedastic Gaussian noise, but globally heteroscedastic Gaussian noise [73]. This means that the residual error maps have implicit structure due to the coil geometry. Consequently, this might then create additional high singular value components that affect the SVDM. Fortunately, even though this makes evaluations through the SVDM in an absolute manner more difficult in some situations (e.g. when data collected from different scanners is directly compared), relative comparisons in this work remain valid as, here, there is only a primary interest in a comparison across reconstruction methods, which are all equally affected by the coil design, at least on the level of individual images.

5.3 Model evaluation outcomes

Following the above two sections, the core of this work can now be presented next. In this section, the results of the performance evaluations related to the three different reconstruction paradigms of interest (being GRAPPA, SENSE, and CS) are summarized and discussed.

Evaluations of the investigated reconstruction methods were done through the calculation of each method's NMSE distribution, PSNR distribution, MSSIM index distribution, VGG loss distribution, and SVDM distribution calculated across its reconstructions made on the test dataset (see Chapter 3).

To analyze the influence of hybrid modelling on a given reconstruction method as reliably as possible, there is a primary focus on intra-paradigm comparisons, each comparing four related reconstruction methods: (1) the ZF method of that reconstruction paradigm (e.g. ZF (GRAPPA mask)), (2) the classical reconstruction method of that reconstruction paradigm (e.g. GRAPPA), (3) the conventional DL-based reconstruction method related to that reconstruction paradigm (e.g. ZF (ACS) U-Net), and (4) the Hybrid Neural Network (HNN) of that reconstruction paradigm (e.g. GRAPPA U-Net). For the DL-based methods, architectures and training configurations were already discussed in Chapter 4.

5.3.1 Results

All obtained results are presented in Table 5.2. However, to make the desired comparisons more easy, the results from this table are also schematically visualized in Figure 5.3.

For evaluations on the entire test set ($R=4+8$), it was checked if the HNNs produced statistically significant performance gains or drops. Thus, statistical tests were performed for the three different HNNs, and the evaluation metric distributions (NMSE, PSNR, MSSIM index, VGG loss, and SVDM) of these HNN reconstructions were compared once to those of their conventional DL model counterparts and once to their classical model counterparts. Again, for this, a two-sided paired samples Wilcoxon signed-rank test was used. Once more, for every metric-model combination one test was performed, after which Bonferroni-corrected p-values were calculated. Perhaps amazingly, this returned a $p_{\text{corrected}} < .0001$ for all tests, indicating that the average performance differences displayed in Table 5.2 and Figure 5.3 are statistically significant for every metric and HNN model when considering the entire test dataset.

The fact that none of the metric (difference) distributions approximate normal distributions is not surprising. After all, only two discrete acceleration factors are present in the test dataset: $R=4$ and $R=8$. All metric distributions are thus bimodal distributions. Because of this, it might also be interesting to visualize some actual MR image reconstructions produced by the different investigated reconstruction methods for each of these two distribution modes.

5. RESULTS & DISCUSSION: MODEL TRAINING AND EVALUATION

$R_{\text{evaluation}}$	Model	NMSE	PSNR	MSSIM	VGG	SVDM
4+8	ZF (SENSE mask)	0.404	20.3	0.695	0.144	0.237
	ZF (CS mask)	0.225	23.3	0.746	0.132	0.260
	ZF (GRAPPA mask)	0.0728	28.3	0.835	0.129	0.252
	SENSE	0.0982	28.6	0.802	0.103	0.372
	CS	0.181	27.7	0.788	0.107	0.314
	GRAPPA	0.0537	29.7	0.796	0.102	0.384
	ZF (no ACS) U-Net	0.131	25.3	0.843	0.120	0.241
	ZF (ACS) U-Net	0.0209	33.8	0.929	0.0690	0.317
	SENSE U-Net	0.0285	33.7	0.927	0.0691	0.323
	CS U-Net	0.0413	32.0	0.91	0.0876	0.297
4	GRAPPA U-Net	0.0201	33.9	0.920	0.0781	0.340
	ZF (SENSE mask)	0.381	20.6	0.702	0.142	0.243
	ZF (CS mask)	0.130	25.3	0.778	0.123	0.271
	ZF (GRAPPA mask)	0.0437	30.1	0.868	0.120	0.263
	SENSE	0.0473	32.1	0.842	0.0682	0.410
	CS	0.0174	34.3	0.866	0.0730	0.385
	GRAPPA	0.0398	31.1	0.803	0.0821	0.418
	ZF (no ACS) U-Net	0.131	25.3	0.841	0.120	0.242
	ZF (ACS) U-Net	0.0121	35.7	0.945	0.0542	0.333
	SENSE U-Net	0.0141	36.6	0.947	0.0505	0.354
8	CS U-Net	0.0127	35.7	0.941	0.0703	0.330
	GRAPPA U-Net	0.0120	35.9	0.937	0.0628	0.369
	ZF (SENSE mask)	0.430	20.1	0.690	0.145	0.233
	ZF (CS mask)	0.321	21.3	0.715	0.140	0.252
	ZF (GRAPPA mask)	0.101	26.4	0.804	0.138	0.243
	SENSE	0.141	25.2	0.767	0.134	0.334
	CS	0.341	21.1	0.711	0.139	0.245
	GRAPPA	0.0646	28.6	0.796	0.118	0.350
	ZF (no ACS) U-Net	0.131	25.4	0.845	0.120	0.242
	ZF (ACS) U-Net	0.0295	32.0	0.914	0.0826	0.302
	SENSE U-Net	0.0401	30.8	0.908	0.0863	0.290
	CS U-Net	0.0678	28.3	0.879	0.104	0.263
	GRAPPA U-Net	0.0291	32.0	0.906	0.0907	0.313

TABLE 5.2: The results of all performed reconstruction method evaluations. The model performances were evaluated on test scans with $R=4$, $R=8$, or on all test scans together. Average metric values are displayed. Performance differences for $R_{\text{evaluation}}=4+8$ are statistically significant for every metric and HNN model (Wilcoxon signed-rank test: $p_{\text{corrected}}<.0001$). Values are rounded to have three significant figures.

5.3. Model evaluation outcomes

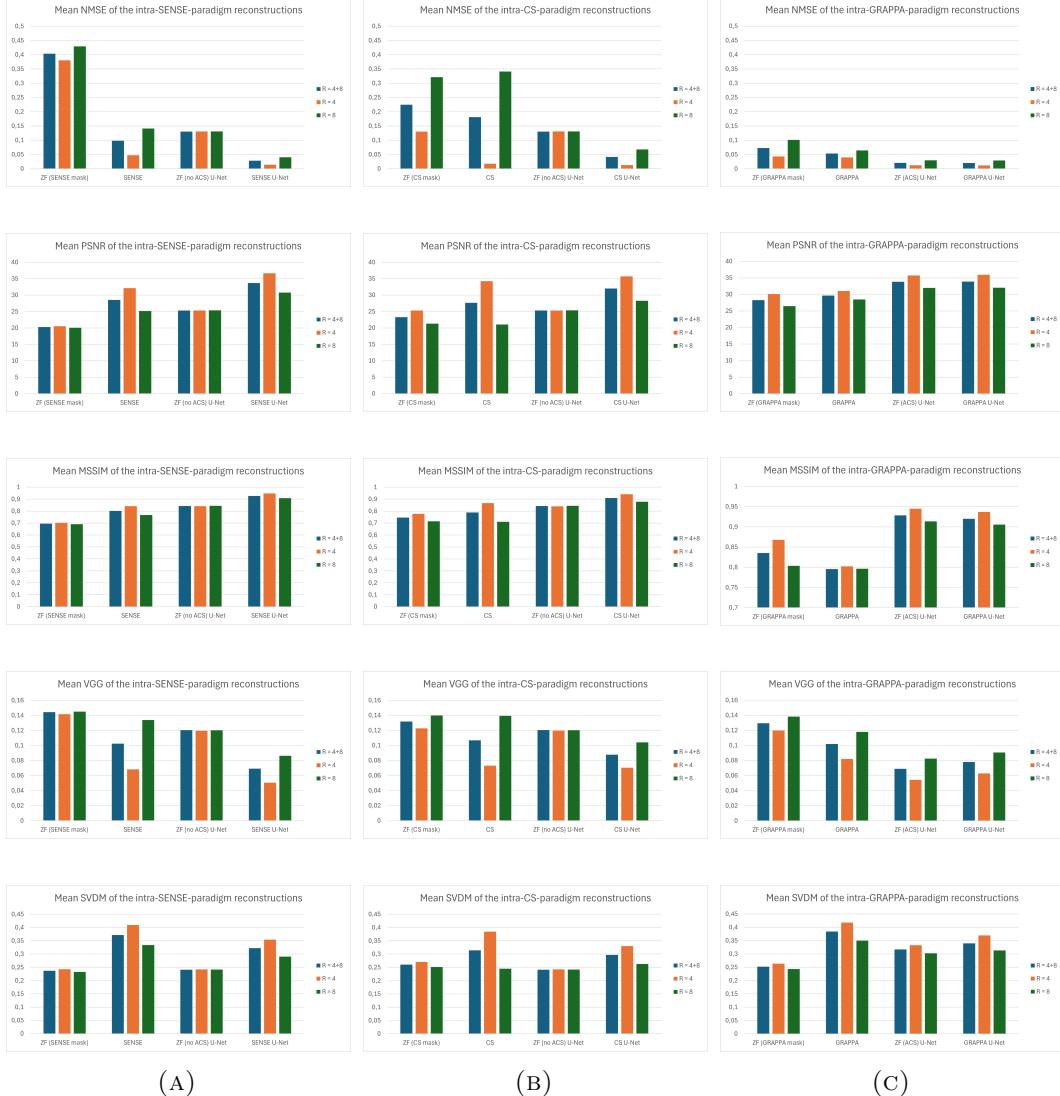


FIGURE 5.3: Figures facilitating the desired intra-paradigm comparisons based on the data displayed in Table 5.2. Each column of figures visualizes an intra-paradigm comparison, with each column-figure analyzing one evaluation metric. In each figure, data is displayed for test scans with $R=4$ (orange), $R=8$ (green), or all test scans together (blue). The left column (A) focuses on the SENSE reconstruction paradigm, the middle column (B) focuses on the CS reconstruction paradigm, and the right column (C) focuses on the GRAPPA reconstruction paradigm.

5. RESULTS & DISCUSSION: MODEL TRAINING AND EVALUATION

Representative examples of different types of MR slice reconstructions are shown in Figures A.1-A.8 (inserted in Appendix A), each displaying reconstructions alongside their residual error maps and their evaluation metric values. These were chosen by selecting test scan volumes which consistently appear to lie near the 50th percentile of their mode as measured by the SVDM across the different reconstruction methods.

For the CS paradigm, the ZF (no ACS) U-Net is chosen as the conventional DL model baseline, since the CS mask doesn't really have an ACS region. For R=4, the CS mask does sample three central lines in k-space (see Figure 3.2 (C)), but this is not nearly enough to act as an ACS region. Moreover, the R=8 CS mask only samples one central line in k-space, just like the SENSE mask does.

5.3.2 Discussion

When looking at the results of the **intra-SENSE and intra-CS comparisons**, it becomes clear that hybrid modelling causes huge performance improvements with respect to the conventional DL baseline in case there is no ACS region present in the raw input data. Irrespective of the acceleration factor, the HNN models score better than the ZF (no ACS) U-Net for every metric. No small feat, especially when considering that all performance gains are relatively big as well. The worse performance of ZF (no ACS) U-Net is also confirmed visually in Figures A.2, A.3, A.6, and A.7, where it appears that it has mainly learned to reconstruct the shape of the head whilst remaining unsuccessful in the de-aliasing task inside the brain region itself, both for R=4 and R=8. An explanation for this might be that, in the case of the absence of an ACS region, aliasing artifacts span across the entire width of the image. If such artifacts need to be de-aliased, one would need to consider relationships between pixels which lie very far away from each other, which is something the current U-Net architectures can't do due to the finite and small size of their convolutional kernels. In the future, however, this problem might be mitigated via the use of an attention mechanism, as mentioned in the discussion of SARA-GAN in Chapter 2.

Similarly, it becomes evident that the SENSE and CS HNN models also outperform their corresponding classical reconstruction methods (and by extension, their corresponding ZF reconstruction methods). The NMSE, PSNR, MSSIM index, and VGG loss values all improve. There is only one exception: the SVDM worsens. Interestingly, this means that the U-Net of the SENSE and CS HNN models decreases the error magnitudes at the cost of slightly increasing the error structure in a given reconstruction. This also becomes visibly apparent from the error maps displayed in Figures A.2 and A.3. Multiple explanations are possible for this.

Firstly, it is possible that this is a sign of slight overfitting, as overfitting typically induces structural errors in the output of a given neural network, with the error severity depending on the severity of the overfitting. Secondly, it might also be possible that this phenomenon is caused by the usage of a simple l_1 loss as the loss function during training. As mentioned in Chapter 2, the l_1 loss favors a smoother, less detailed outcome over one with highly accurate textures.

This is because predicting the mean of a distribution minimizes the average pixel-wise error effectively, albeit at the cost of producing a more blended image [53]. If this effect is at play here, then this would equally explain the current decrease in error magnitude, with an increasing error structure. A third and last explanation is that this phenomenon can be a side effect of the nature of the task the U-Net of the SENSE and CS HNNs is performing. Based on Figures A.2 and A.3, it looks like the U-Net’s residual de-aliasing function is partly related to de-noising. This is especially clear from the CS paradigm figures. As this de-noising is done through some learned smoothing operation, this smoothing can also cause error structures to appear in the network’s outputs, as backed up by the results presented and discussed in the previous section (Figure 5.2 (B)).

Which of these three explanations are actually at play here is hard to say with certainty, but it can be hypothesized that it will likely be explanation two and three which cause the current observations, as regularization was used during the training of the HNNs making overfitting less likely.

If the results of the **intra-GRAPPA comparisons** are reviewed, it becomes clear that the added value of hybrid modelling is less outspoken in the case there is an ACS region of sufficient size present in the raw input data. Compared to the conventional DL baseline ZF (ACS) U-Net, some evaluation metric values improve, some worsen. To be more precise, the NMSE, PSNR, and SVDM improve, while the MSSIM index and VGG loss don’t. Nonetheless, the metrics which improve form a majority, and it is also important to recognize that worse performance based on the MSSIM index and the VGG loss is less critical than worse performance based on the NMSE, PSNR, and SVDM, as these metrics are much more direct representations of error. For example, the VGG19 network used to calculate the VGG loss was trained on the ImageNet dataset, only consisting of natural images. This means that its feature extraction isn’t necessarily tailored to the features of (medical) MR images specifically, and to which of those are the most important for classification tasks or perceptual quality evaluations. Similarly, given that the MSSIM index represents an interplay between luminance, contrast, and structure, all averaged across different image patches, there are a lot of possible factors which may incorrectly cause underestimations of the MSSIM index value. Because of this, some authors even refer to the SSIM or MSSIM index as ‘a bendable ruler’ and ‘unreliable’ [80]. Within this context, it still seems permitted to say that GRAPPA U-Net marginally (and significantly) performs better than ZF (ACS) U-Net.

Performance differences themselves are also smaller than for the intra-SENSE and intra-CS comparisons. This can be explained by the same reasoning used to explain why the ZF (no ACS) U-Net is not successful in the de-aliasing task. If an ACS region is present, aliasing artifacts are only related to high-frequency information. This essentially creates a highly blurred version of the MR image, as the low-frequency information is un- or less corrupted (depending on the ACS region size), as evident from the ZF reconstructions presented in Figures A.4 and A.8. This also means that the de-aliasing goal can be achieved here without a high need

5. RESULTS & DISCUSSION: MODEL TRAINING AND EVALUATION

to consider relationships between pixels from distant locations in the image, and this ACS-dependent de-aliasing is thus something a conventional U-Net is perfectly capable of. The remaining performance improvements made by GRAPPA U-Net are then solely the result of a reduction in dataset variance, created by the GRAPPA layer, but this effect is apparently a bit less influential than one initially might have hypothesized, at least in the current experiment set-up. The not much worse performance of ZF (ACS) U-Net is again confirmed visually in Figures A.4 and A.8.

Nevertheless, it is evident that the GRAPPA HNN model does outperform its corresponding classical reconstruction method (and by extension, its corresponding ZF reconstruction method). Again, the NMSE, PSNR, MSSIM index, and VGG loss values all improve, although, again, the SVDM worsens, likely due to the same possible mechanisms as discussed earlier for SENSE U-Net and CS U-Net.

Now that it is established that hybrid modelling results in a reconstruction performance at least equal to but often better than the reconstruction performance of the conventional DL pipeline considered in this thesis (and by extension, better than the reconstruction performance of classical reconstruction methods), one interesting question remains: **What is the maximally allowed acceleration factor R using these HNN models?** Although an interesting question, it is also a difficult one. Based on Figures A.1-A.8, it can be seen that the GRAPPA U-Net and SENSE U-Net reconstructions may still show image artifacts, even for $R=4$, the lowest considered acceleration factor. In clinical practice, these images would be considered unusable. Based on this, the maximally allowed acceleration factor R is likely somewhere between $R=3$ and $R=4$ for these HNNs, especially if they would be retrained with the inclusion of $R=3$ training data. Though perhaps lower than expected, this would still result in faster acquisitions compared to the $R=2$ and $R=3$ acquisitions most commonly used in the current clinical practice, potentially leading to a few extra patients that may be imaged in one day, decreasing patient waiting times. Contrarily, however, the CS U-Net reconstructions appear to still look artifact-free for $R=4$ (though not all test set reconstructions were manually inspected). This indicates that for CS U-Net, the maximally allowed acceleration factor R is likely at least $R=4$, and thus higher than the maximally allowed acceleration factor of GRAPPA U-Net and SENSE U-Net. However, before any claim of clinical usability can be solidified, one would first have to set up another comparative experiment where e.g. CS U-Net reconstructions are presented to radiologists together with matching MR images stemming from concurrent clinical reconstruction protocols. Only if there is an agreement amongst radiologists that the newly proposed reconstruction method actually delivers equally good- or even better-looking MR images, then it can be said that this new reconstruction method can achieve faster imaging whilst avoiding a compromise in diagnostic quality.

Importantly, however, it needs to be mentioned that the currently trained HNNs were not optimized or trained to their full capacity. This also means that their full potential is not displayed here, and that the actual maximally allowed acceleration factors may be potentially higher. For example, the learning rate and maximal amount

of epochs were arbitrarily chosen. Additionally, the L2 regularization hyperparameter was not fine-tuned due to time constraints. For the classical CS reconstruction layer in CS U-Net, default parameter values were used as suggested by the BART toolbox, but perhaps some other parameter values might also increase performance. In short, while the primary aim of this study was to compare the different hybrid models, the aim of another, future study might be to further optimize them.

To conclude this section, some **inter-paradigm comparisons** are also made. Firstly, it can be noticed that, although CS reconstruction is the best performing classical reconstruction technique for R=4, CS reconstructions seem to degrade much quicker in quality with respect to the acceleration factor R compared to the GRAPPA and SENSE reconstructions. Based on Figure A.7, it seems that a consistent pixel intensity offset is responsible for this. Somehow, for R=8, the BART toolbox creates CS reconstructions with a much lower average pixel intensity than the average pixel intensity of SENSE, GRAPPA, and ground truth scans. It is unknown whether this result is expected or not, as no studies corroborating these results could be found. If this is not the case, then it might be that different CS parameter values would need to be chosen, or that the currently used sampling mask (obtained from one of the scanners in the UZ Leuven) is not the type of CS mask the BART toolbox expects, but unfortunately the BART toolbox' documentation is not clear about this.

Secondly, also from inter-paradigm comparisons, it becomes immediately clear that there are still some challenges related to the usage of numerical evaluation metrics to evaluate the perceptual quality of an MR image, besides the potential pitfalls of the MSSIM index and VGG loss that were already discussed. For example, if one would only consider the data from Table 5.2 and Figure 5.3, one could think based on the NMSE that ZF reconstructions with a GRAPPA mask look better than SENSE reconstructions. Obviously, however, when looking at Figures A.2 and A.4, this is not what the human visual system would indicate. A possible explanation for this, is that the ACS region in the ZF MR image essentially spreads out the contrast/low-frequency information of the scan's k-space across the image in the image domain, effectively lowering the average pixel error of the reconstruction, even though this is a blurred, perceptually bad image. Essentially, this would be a case where the bias of the NMSE metric (inclined toward favoring smoothness over sharpness) as mentioned in Chapter 2 and 3 clearly comes forward. Additionally, as the PSNR is related to the mean squared error of a given reconstruction, which in turn is related to the NMSE, the PSNR likely also shares the same bias (although this bias may be less noticeable for the PSNR because it explicitly accounts for noise levels). In the context of MR imaging and MR image reconstruction, this thus means that the NMSE and PSNR are biased for scans whose raw data exhibits an ACS region. This phenomenon accentuates the importance of additional evaluation metrics such as the SVDM even more.

5.4 Conclusion

This chapter presented and discussed the results of three different experiments, the last of which forming the core of this thesis.

In the first experiment, an investigation was conducted to determine whether training deep learning models exclusively on training data containing accelerations related to a single acceleration factor R is preferable, or if training them with a mixture of acceleration factors yields better performance. From its results, it becomes clear that the inclusion of multiple acceleration factors in the training dataset improves model performance in general. This is likely because of an interplay between an intra-acceleration factor Pareto optimal in the statistically learned reconstruction function and an inter-acceleration information sharing mechanism.

In the second experiment, the nature of the SVDM was verified. There, it became clear that the SVDM behaves as intended and may thus be relied upon in this work's analyses. It does, however, remain important to still complement the SVDM by metrics which look at error magnitude, as the SVDM does display a bias towards situations where noise dominates the reconstructed images.

In the third and most important experiment, various HNN models were evaluated and compared with conventional DL models and classical reconstruction methods. The results demonstrated that HNN models generally provide significant performance improvements over both conventional DL models and classical reconstruction methods across several evaluation metrics (NMSE, PSNR, MSSIM index, VGG loss, and SVDM). HNN models specifically excelled in scenarios lacking an ACS region, outperforming their conventional counterparts more consistently and significantly (all metrics improved), while the added value of hybrid modelling is less pronounced when an ACS region is present (only the NMSE, PSNR, and SVDM improved).

It was also noted that the SVDM tended to worsen for HNN and conventional DL models compared to classical reconstruction methods, indicating a trade-off where error magnitudes decreased but error structures increased. This phenomenon might be attributed to the inherent smoothing effect of the l_1 loss function or the nature of the residual de-aliasing performed by the U-Net architecture.

Visually, inspected GRAPPA U-Net and SENSE U-Net reconstructions still displayed image artifacts at an acceleration of $R=4$, suggesting that the maximally allowed acceleration factor for these models is likely between $R=3$ and $R=4$. In contrast, inspected CS U-Net reconstructions remained artifact-free at $R=4$, indicating a higher maximally allowed acceleration factor for the CS U-Net model.

These results highlight the potential of hybrid modelling to enhance MR image reconstruction, potentially enabling faster imaging without compromising diagnostic quality. However, the chapter also recognized the need for further optimization of the HNN models and suggested that future research should focus on fine-tuning hyperparameters and exploring more advanced architectures like those incorporating attention mechanisms.

Chapter 6

Conclusion

The research questions outlined in Chapter 1 have been addressed in this study, and with the results and conclusions provided in Chapter 5 they can finally be answered. This chapter thus finalizes this thesis by delivering these answers, and by formulating which next steps can be taken in the future.

Conclusion Here, this work's research questions are answered.

Does hybrid modelling have the capacity to increase the performance of the exemplary conventional deep learning-based MRI reconstruction pipeline investigated in this thesis? Yes, hybrid modelling has the capacity to significantly increase the performance of the conventional deep learning-based MRI reconstruction pipeline investigated in this thesis. The HNN models outperform the conventional DL baseline across various metrics, proving to be a promising approach.

Which classical reconstruction method has a higher synergy with deep learning models? The HNN models outperform the conventional DL baseline model especially in scenarios where an ACS region is not available. While improvements are thus substantial (the NMSE, PSNR, MSSIM index, VGG loss, and SVDM all improved) for SENSE and CS acquisitions, the performance gains are less pronounced but still present (the NMSE, PSNR, and SVDM improved) for GRAPPA acquisitions, where an ACS region is available. This phenomenon is likely related to the small and finite size of the convolutional kernels of U-Nets.

What is the largest acceleration factor that can reasonably be achieved with the proposed models? The maximum achievable acceleration factor R with the proposed HNN models varies depending on the specific model. The CS U-Net model demonstrates the highest resilience, handling an acceleration factor of at least R=4 without generating noticeable artifacts in the inspected test scans. However, it is important to note that this conclusion is contingent upon validation through a blind comparison experiment performed by radiologists. Additionally, since these models were not fully optimized, there is potential for achieving even higher acceleration factors with further fine-tuning and development.

6. CONCLUSION

Future work From the previous chapter, it became clear that there is enough future work to do. For example, considering the training of the HNNs, it might be interesting to further investigate the optimal acceleration factor distribution used to undersample the training data. Additionally, the used number of epochs, learning rate, L2 regularization hyperparameter value, loss function, initialization method, and BART toolbox parameter values might be further optimized, allowing the HNN architectures to come to their full potential. On top of this, more complex HNNs could be designed and tested, where the U-Net of the current HNNs is changed into a more complex neural network containing design elements of the more state-of-the-art architectures discussed in Chapter 2, such as e.g. attention layers and multi-slice inputs. The design and validation of evaluation metrics also remains an important area of research. Next to this, transfer learning experiments could be performed, where the HNNs are e.g. tested on the fastMRI knee dataset after being trained on the fastMRI brain dataset, as was the case here. Lastly, having radiologists perform a comparison between the HNN reconstructions and the reconstructions from concurrent clinical practice would also mark a significant step forward.

Code availability The code developed for the analyses in this study is available at <https://github.com/MathijsVanhaverbeke/fastMRI-hybrid-modelling/>. Interested researchers can access the repository to review, replicate, and build upon the work presented in this thesis. For any specific inquiries or issues related to the code, please contact the author at vanhaverbeke.mathijs@gmail.com.

Appendix A

Reconstruction examples

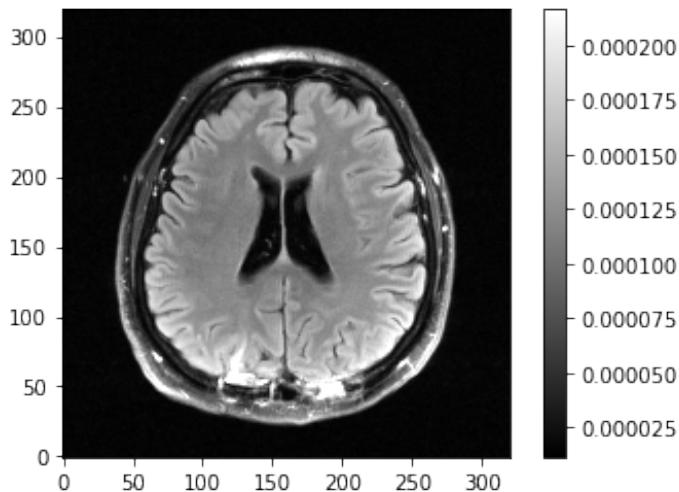


FIGURE A.1: Magnitude of the ground truth of the test scan volume slice of the reconstructions presented in Figures A.2, A.3, and A.4.

A. RECONSTRUCTION EXAMPLES

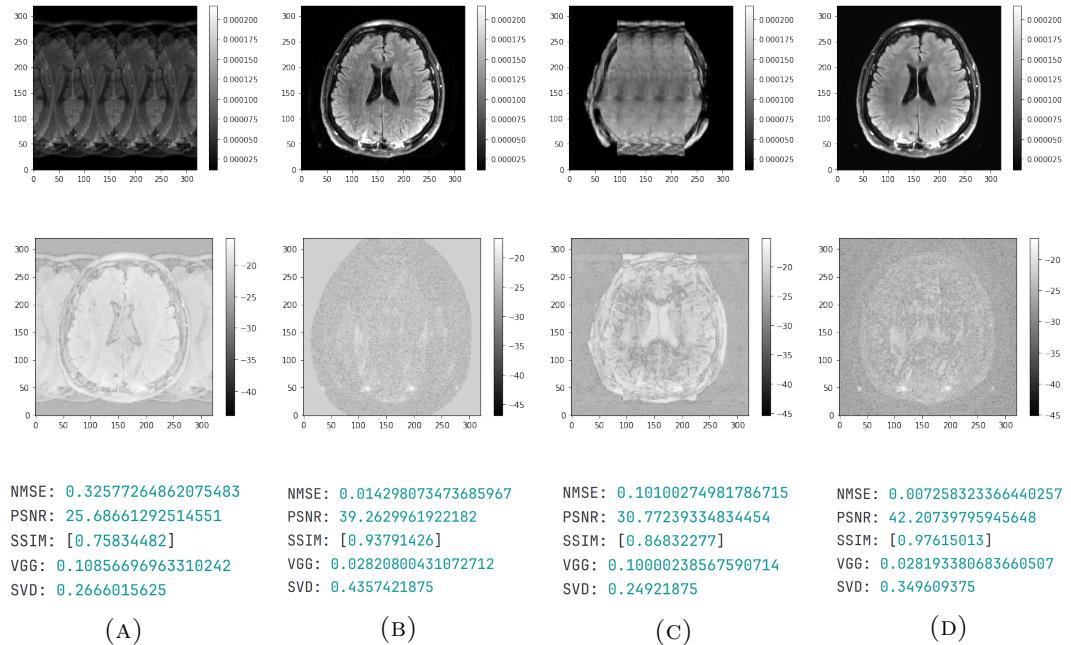


FIGURE A.2: Reconstruction outputs of the different intra-SENSE paradigm reconstruction methods: ZF (SENSE mask) (A), SENSE (B), ZF (no ACS) U-Net (C), and SENSE U-Net (D). This concerns the same test scan volume slice as the one shown in Figure A.1. Here, $R=4$. For each reconstruction, the magnitude of the slice of interest is shown, together with its residual error map and evaluation metric values.

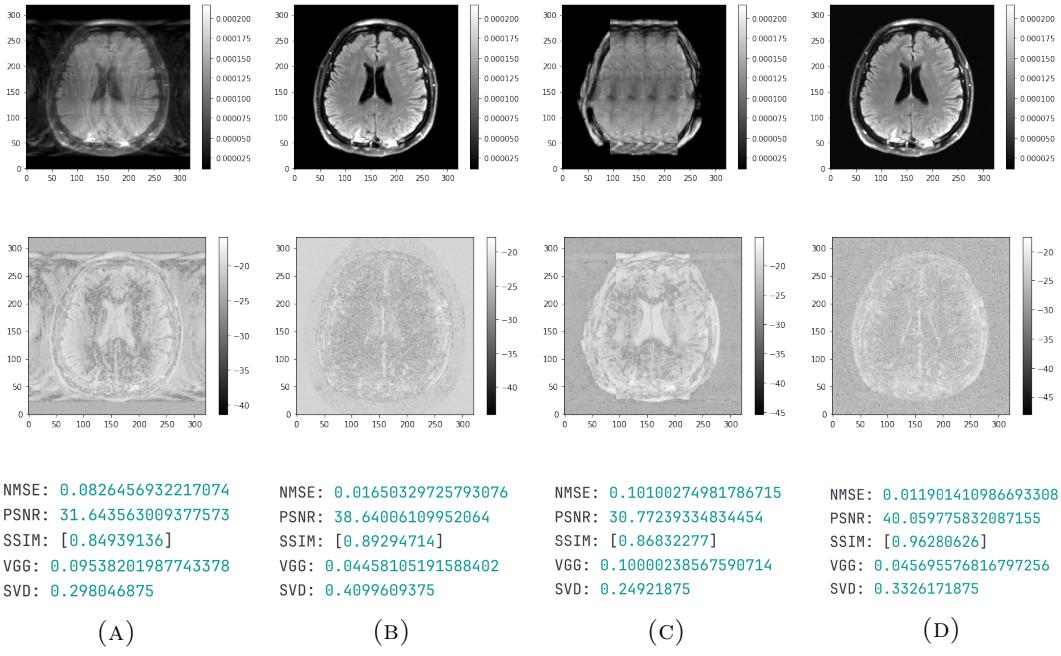


FIGURE A.3: Reconstruction outputs of the different intra-CS paradigm reconstruction methods: ZF (CS mask) (A), CS (B), ZF (no ACS) U-Net (C), and CS U-Net (D). This concerns the same test scan volume slice as the one shown in Figure A.1. Here, $R=4$. For each reconstruction, the magnitude of the slice of interest is shown, together with its residual error map and evaluation metric values.

A. RECONSTRUCTION EXAMPLES

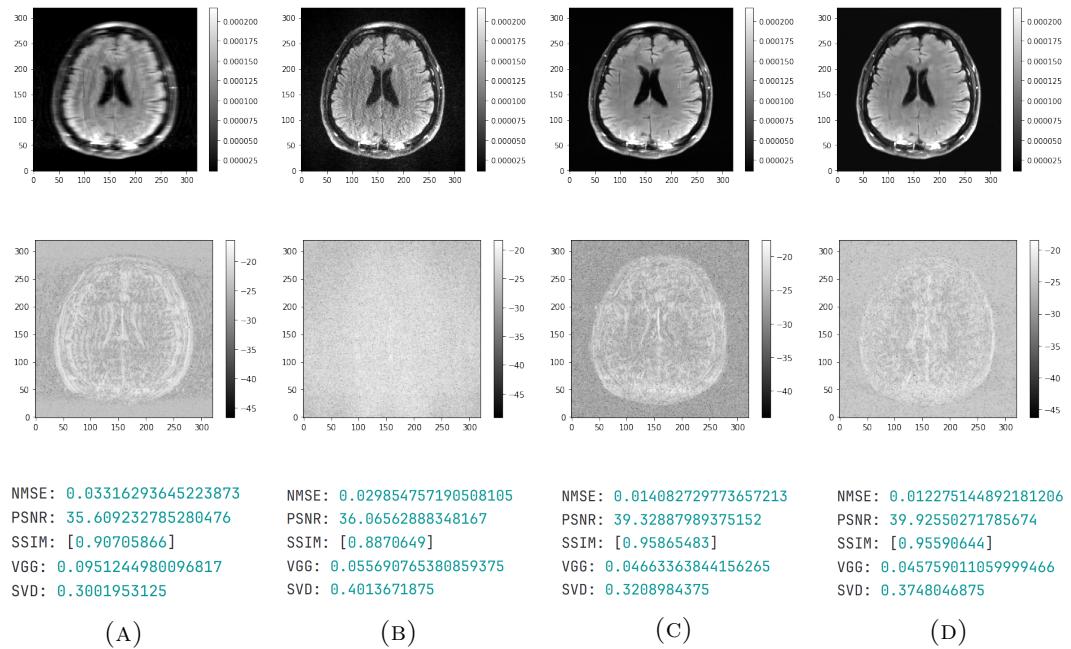


FIGURE A.4: Reconstruction outputs of the different intra-GRAPPA paradigm reconstruction methods: ZF (GRAPPA mask) (A), GRAPPA (B), ZF (ACS) U-Net (C), and GRAPPA U-Net (D). This concerns the same test scan volume slice as the one shown in Figure A.1. Here, R=4. For each reconstruction, the magnitude of the slice of interest is shown, together with its residual error map and evaluation metric values.

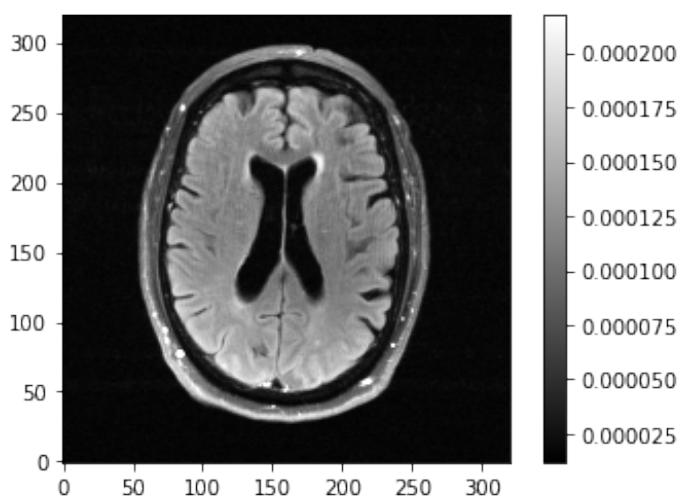


FIGURE A.5: Magnitude of the ground truth of the test scan volume slice of the reconstructions presented in Figures A.6, A.7, and A.8.

A. RECONSTRUCTION EXAMPLES

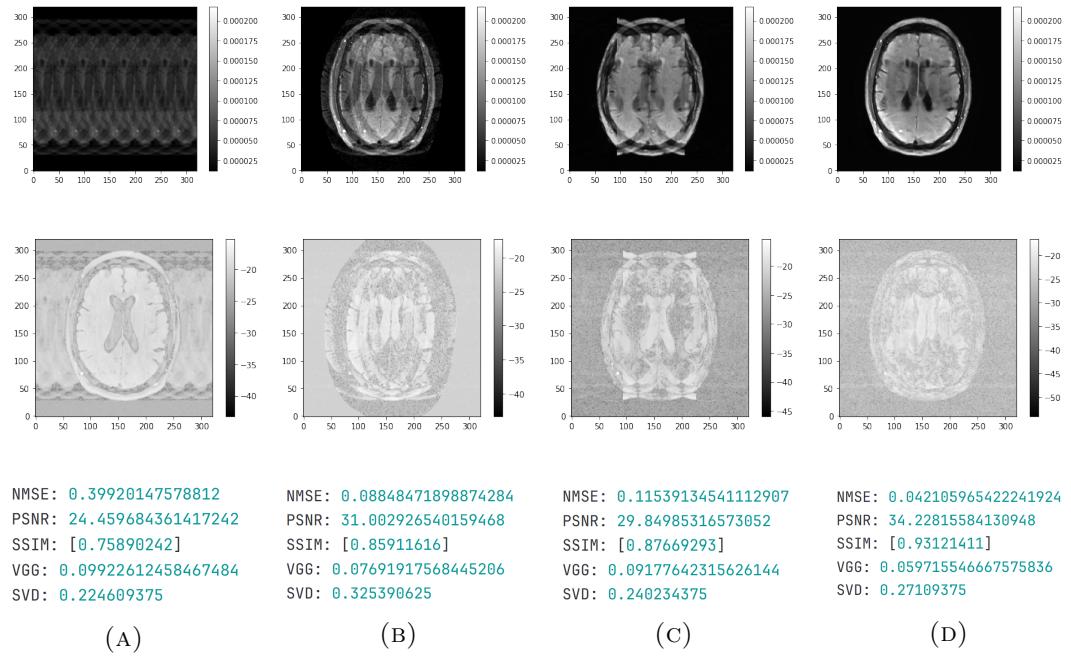


FIGURE A.6: Reconstruction outputs of the different intra-SENSE paradigm reconstruction methods: ZF (SENSE mask) (A), SENSE (B), ZF (no ACS) U-Net (C), and SENSE U-Net (D). This concerns the same test scan volume slice as the one shown in Figure A.5. Here, R=8. For each reconstruction, the magnitude of the slice of interest is shown, together with its residual error map and evaluation metric values.

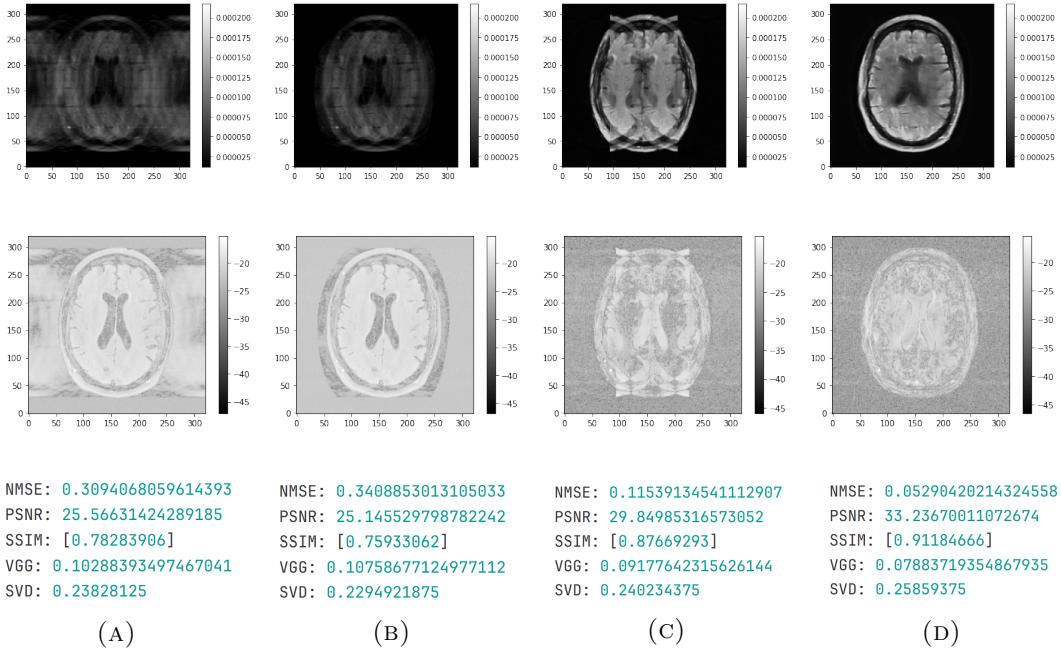


FIGURE A.7: Reconstruction outputs of the different intra-CS paradigm reconstruction methods: ZF (CS mask) (A), CS (B), ZF (no ACS) U-Net (C), and CS U-Net (D). This concerns the same test scan volume slice as the one shown in Figure A.5. Here, $R=8$. For each reconstruction, the magnitude of the slice of interest is shown, together with its residual error map and evaluation metric values.

A. RECONSTRUCTION EXAMPLES

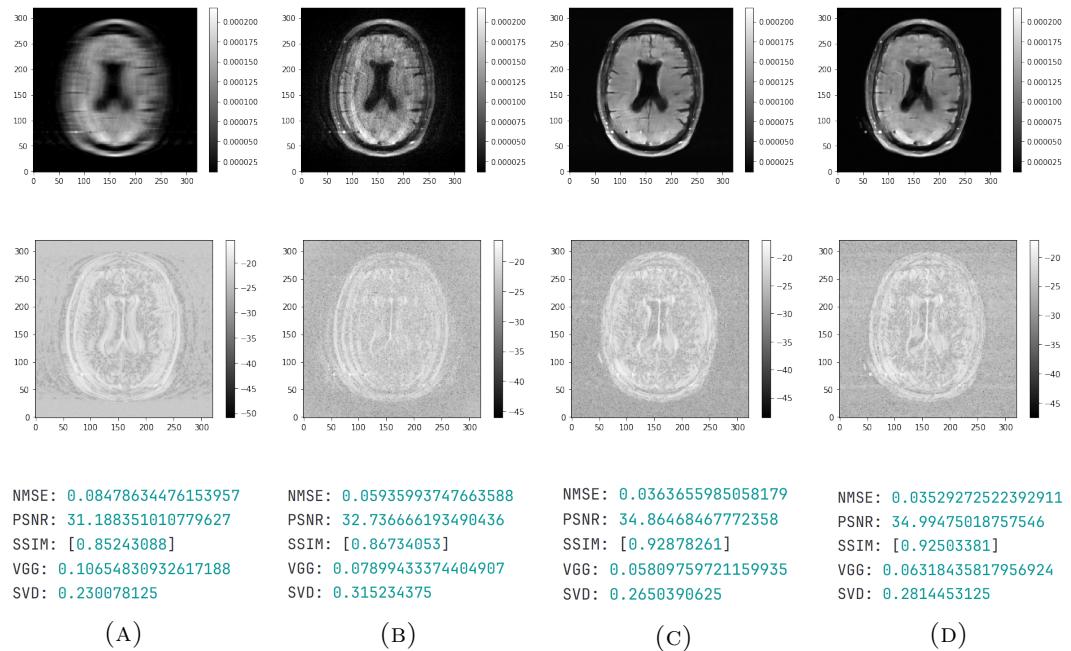


FIGURE A.8: Reconstruction outputs of the different intra-GRAPPA paradigm reconstruction methods: ZF (GRAPPA mask) (A), GRAPPA (B), ZF (ACS) U-Net (C), and GRAPPA U-Net (D). This concerns the same test scan volume slice as the one shown in Figure A.5. Here, R=8. For each reconstruction, the magnitude of the slice of interest is shown, together with its residual error map and evaluation metric values.

Bibliography

- [1] U.S. Food & Drug Administration (FDA) (2017). *MRI (Magnetic Resonance Imaging) - Benefits and Risks*. Available on: <https://www.fda.gov/radiation-emitting-products/mri-magnetic-resonance-imaging/benefits-and-risks> [accessed on 21/10/23].
- [2] Pal A., Rathi Y. (2022). A review and experimental evaluation of deep learning methods for MRI reconstruction. *Journal of Machine Learning for Biomedical Imaging*, 1, 1-58.
- [3] National Kidney Foundation Inc. (2015). *Contrast Dye and the Kidneys*. Available on: <https://www.kidney.org/atoz/content/Contrast-Dye-and-Kidneys> [accessed on 23/05/2024].
- [4] Shaul R., David I., Shitrit O., Riklin Raviv T. (2020). Subsampled brain MRI reconstruction by generative adversarial neural networks. *Medical Image Analysis*, 65, 1017-47.
- [5] Edelstein W.A., Mahesh M., Carrino J.A. (2010). MRI: time is dose - and money and versatility. *Journal of the American College of Radiology*, 7(8), 650-652.
- [6] U.S. Food & Drug Administration (FDA) (2018). *MRI (Magnetic Resonance Imaging)*. Available on: <https://www.fda.gov/radiation-emitting-products/medical-imaging/mri-magnetic-resonance-imaging> [accessed on 21/10/23].
- [7] Zbontar J., Knoll F., Sriram A. *et al.* (2018). fastMRI: An Open Dataset and Benchmarks for Accelerated MRI. Available on: <https://arxiv.org/pdf/1811.08839v1.pdf>.
- [8] Zaitsev M., Maclarens J., Herbst M. (2015). Motion artifacts in MRI: A complex problem with many partial solutions. *Journal of Magnetic Resonance Imaging*, 42(4), 887-901.
- [9] Dillinger H., Kozerke S., Guenthner C. (2022). Direct comparison of gradient Fidelity and acoustic noise of the same MRI system at 3 T and 0.75 T. *Magnetic Resonance in Medicine*, 88(4), 1937-1947.
- [10] Winkler S., Saniour I., Chaudhari A., Robb F., Vaughan J. (2021). MRSaFE: Tissue Heating Prediction for MRI: a Feasibility Study. Available on: <https://arxiv.org/ftp/arxiv/papers/2102/2102.01023.pdf>.

BIBLIOGRAPHY

- [11] Andre J.B., Bresnahan B.W., Mossa-Basha M. *et al.* (2015). Toward Quantifying the Prevalence, Severity, and Cost Associated With Patient Motion During Clinical MR Examinations. *Journal of the American College of Radiology*, 12(7), 689-695.
- [12] Enders J., Zimmermann E., Rief M. *et al.* (2011). Reduction of claustrophobia during magnetic resonance imaging: methods and design of the "CLAUSTRO" randomized controlled trial. *BMC Medical Imaging*, 11(4).
- [13] Tokur S., Lederle K., Terris D.D., Jarczok M.N., Bender S., Schoenberg S.O., Weisser G. (2012). Process analysis to reduce MRI access time at a German University Hospital. *International Journal for Quality in Health Care*, 24(1), 95-99.
- [14] Griswold M. A. *et al* (2002). Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, 47, 1202-1210.
- [15] Pruessmann K. P., Weiger M., Scheidegger M. B., Boesiger P. (1999). SENSE: sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42, 952-962.
- [16] Lustig M., Donoho D., Pauly J.M. (2007). Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6), 1182-1195.
- [17] Lustig M., Donoho D.L., Santos J.M., Pauly J.M. (2008). Compressed sensing MRI. *IEEE Signal Processing Magazine*, 25(2), 72-82.
- [18] Deshmane A., Gulani V., Griswold M.A., Seiberlich N. (2012). Parallel MR imaging. *Journal of Magnetic Resonance Imaging*, 36(1), 55-72.
- [19] Zeng G., Guo Y., Zhan J. *et al.* (2021). A review on deep learning MRI reconstruction without fully sampled k-space. *BMC Medical Imaging*, 21(195).
- [20] Maes F. (2023). *Medical Imaging & Analysis: Magnetic Resonance Imaging* [Class Handout]. Place of Publication: Faculty of Engineering Science, University of KU Leuven.
- [21] Suetens P. (2011). *Fundamentals of medical imaging* 3rd ed., Cambridge University Press.
- [22] Cruz G., Yaman B., Akçakaya M., Doneva M., Prieto C. (2022). Chapter 2 - MRI Reconstruction as an Inverse Problem. *Advances in Magnetic Resonance Technology and Applications*, 7, 37-57.
- [23] Zheng H., Fang F., Zhang G. (2019). Cascaded dilated dense network with two-step data consistency for MRI reconstruction. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Article 156, 1744-1754.

- [24] Sriram A., Zbontar J., Murrell T., Zitnick C.L., Defazio A., Sodickson D.K. (2020). GrappaNet: Combining Parallel Imaging with Deep Learning for Multi-Coil MRI Reconstruction. Available on: <https://arxiv.org/pdf/1910.12325.pdf>.
- [25] Hamilton J., Franson D., Seiberlich N. (2017). Recent advances in parallel imaging for MRI. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 101, 71-95.
- [26] Sriram A. et al. (2020). End-to-End Variational Networks for Accelerated MRI Reconstruction. In: Martel A.L., et al. Medical Image Computing and Computer Assisted Intervention - MICCAI 2020. MICCAI 2020. Lecture Notes in Computer Science, vol 12262.
- [27] Hou L., Gao H., Zhang X. (2016). A Two-Stage Low Rank Approach for Calibrationless Dynamic Parallel Magnetic Resonance Image Reconstruction. *Journal of Scientific Computing*, 69(3), 1014-1032.
- [28] Uecker M., Lai P., Murphy M.J., Virtue P., Elad M., Pauly J.M., Vasanawala S.S., Lustig M. (2014). ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. *Magnetic Resonance in Medicine*, 71(3), 990-1001.
- [29] Griswold M.A., Breuer F., Blaimer M., Kannengiesser S., Heidemann R.M., Mueller M., Nittka M., Jellus V., Kiefer B., Jakob P.M. (2006). Autocalibrated coil sensitivity estimation for parallel imaging. *NMR in Biomedicine*, 19(3), 316-324.
- [30] Bertrand A. (2023). *Selected Topics in Biomedical Signal Processing - Data-driven linear filter design for biomedical sensor arrays* [Class Handout]. Place of Publication: Faculty of Engineering Science, University of KU Leuven.
- [31] Pruessmann K.P. (2006). Encoding and reconstruction in parallel MRI. *NMR in Biomedicine*, 19, 288-299.
- [32] Elster A.D. (2023). *GRAPPA/ARC - How does GRAPPA/ARC work*. Available on: <https://mriquestions.com/grappaarc.html> [accessed on 1/11/2023].
- [33] Ye J.C. (2019). Compressed sensing MRI: a review from signal processing perspective. *BMC Biomedical Engineering*, 1, 8.
- [34] Zhang J., Ghanem B. (2018). ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing. *Conference on Computer Vision and Pattern Recognition*, pp. 1828-1837.
- [35] Suykens J. (2023). *Artificial Neural Networks and Deep Learning* [Class Handout]. Place of Publication: Faculty of Engineering Science, University of KU Leuven.
- [36] Goodfellow I., Bengio Y., Courville A. (2016). *Deep Learning*, MIT Press.

BIBLIOGRAPHY

- [37] Jain A.K., Mao J., Mohiuddin K.M. (1996). Artificial neural networks: a tutorial. *Computer*, 29(3), 31-44.
- [38] Hornik K., Stinchcombe M., White H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2, 359-366.
- [39] Sapijuszko G., Mikhael W.B. (2018). An overview of recent convolutional neural network algorithms for image recognition. *IEEE 61st International Midwest Symposium on Circuits and Systems (MWSCAS)*, pp. 743-746.
- [40] O'Shea K., Nash R. (2015). An Introduction to Convolutional Neural Networks. Available on: <https://arxiv.org/pdf/1511.08458.pdf>.
- [41] Simonyan K., Zisserman A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. Available on: <https://arxiv.org/pdf/1409.1556.pdf>.
- [42] Yuan Z., Jiang M., Wang Y., Wei B., Li Y., Wang P., Menpes-Smith W., Niu Z., Yang G. (2020). SARA-GAN: Self-Attention and Relative Average Discriminator Based Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction. *Frontiers in Neuroinformatics*, 14, 611-666.
- [43] Zhao X., Yang T., Li B. (2022). A review on generative based methods for MRI reconstruction. *Journal of Physics: Conference Series*, 2330(1).
- [44] Singh D., Monga A., de Moura H.L., Zhang X., Zibetti M.V.W., Regatte R.R. (2023). Emerging Trends in Fast MRI Using Deep-Learning Reconstruction on Undersampled k-Space Data: A Systematic Review. *Bioengineering*, 10, 1012.
- [45] Wang Z., Bovik A.C., Sheikh H.R., Simoncelli E.P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600-612.
- [46] Lee D., Kim J., Moon W.J., Ye J.C. (2019). CollaGAN: Collaborative GAN for Missing Image Data Imputation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2482-2491.
- [47] Johnson J., Alahi A., Fei-Fei L. (2016). Perceptual losses for real-time style transfer and super-resolution. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9906, 694-711.
- [48] Ledig C. et al. (2017). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. Available on: <https://arxiv.org/pdf/1609.04802v5.pdf>.
- [49] Krizhevsky A., Sutskever I., Hinton G.E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90.

BIBLIOGRAPHY

- [50] He K., Zhang X., Ren S., Sun J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1026-1034.
- [51] Glorot X., Bengio Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research*, 9, 249-256.
- [52] Shitrit O., Riklin Raviv T. (2017). Accelerated Magnetic Resonance Imaging by Adversarial Neural Network. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Lecture Notes in Computer Science()*, vol 10553.
- [53] Pathak D., Krähenbühl P., Donahue J., Darrell T., Efros A.A. (2016). Context Encoders: Feature Learning by Inpainting. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2536-2544.
- [54] Medium (2021). *Self Attention in Convolutional Neural Networks*. Available on: <https://shorturl.at/dwDF3> [accessed on 7/11/23].
- [55] Jolicoeur-Martineau, A. (2018). The relativistic discriminator: a key element missing from standard GAN. Available on: <https://arxiv.org/pdf/1807.00734.pdf>.
- [56] Quan T.M., Nguyen-Duc T., Jeong W.K. (2018). Compressed Sensing MRI Reconstruction Using a Generative Adversarial Network With a Cyclic Loss. *IEEE Transactions on Medical Imaging*, 37(6), 1488-1497.
- [57] Lv J., Wang C., Yang G. (2021). PIC-GAN: A Parallel Imaging Coupled Generative Adversarial Network for Accelerated Multi-Channel MRI Reconstruction. *Diagnostics (Basel)*, 11(1), 61.
- [58] Alam S., Uh J., Dresner A., Hua C., Khairy K. (2023). Deep-learning-based acceleration of MRI for radiotherapy planning of pediatric patients with brain tumors. Available on: <https://arxiv.org/ftp/arxiv/papers/2311/2311.13485.pdf> [accessed on 31/12/23].
- [59] Akçakaya M., Moeller S., Weingärtner S., Ugurbil K. (2019). Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging. *Magnetic Resonance in Medicine*, 81(1), 439-453.
- [60] Arefeen Y., Beker O., Cho J., Yu H., Adalsteinsson E., Bilgic B. (2022). Scan-specific artifact reduction in k-space (SPARK) neural networks synergize with physics-based reconstruction to accelerate MRI. *Magnetic Resonance in Medicine*, 87(2), 764-780.
- [61] Wang S., Cheng H., Ying L., Xiao T., Ke Z., Zheng H., Liang D. (2020). DeepcomplexMRI: Exploiting deep residual network for fast parallel MR imaging with complex convolution. *Magnetic Resonance Imaging*, 68, 136-147.

BIBLIOGRAPHY

- [62] Bahadir C.D., Dalca A.V., Sabuncu M.R. (2019). Learning-Based Optimization of the Under-Sampling Pattern in MRI. *Information Processing in Medical Imaging. Lecture Notes in Computer Science*, vol 11492.
- [63] Aggarwal H.K., Jacob M. (2020). J-MoDL: Joint Model-Based Deep Learning for Optimized Sampling and Reconstruction. *IEEE Journal of Selected Topics in Signal Processing*, 14(6), 1151-1162.
- [64] Knoll F. *et al.* (2020). fastMRI: A Publicly Available Raw k-Space and DICOM Dataset of Knee Images for Accelerated MR Image Reconstruction Using Machine Learning. *Radiology: Artificial Intelligence*, 2(1).
- [65] NYU (2020). *Welcome to the fastMRI Dataset*. Available on: <https://fastmri.med.nyu.edu/> [accessed on 3/04/24].
- [66] Blaimer M., Breuer F., Mueller M., Heidemann R.M., Griswold M.A., Jakob P.M. (2004). SMASH, SENSE, PILS, GRAPPA: How to choose the optimal method. *Topics in Magnetic Resonance Imaging*, 15, 223-236.
- [67] GitHub (2024). *fastMRI*. Available on: <https://github.com/facebookresearch/fastMRI> [accessed on 4/04/2024].
- [68] Herterich R., Sumarokova A. (2019). Coil Sensitivity Estimation and Intensity Normalisation for Magnetic Resonance Imaging. Available on: <https://www.diva-portal.org/smash/get/diva2:1366830/FULLTEXT01.pdf> [accessed on 5/04/2024].
- [69] BART (n.d.). BART: Computational Magnetic Resonance Imaging. Available on: <https://mrirecon.github.io/bart/> [accessed on 6/04/2024].
- [70] Wymer D.T., Patel K.P., Burke W.F., Bhatia V.K. (2020). Phase-Contrast MRI: Physics, Techniques, and Clinical Applications. *RadioGraphics*, 40(1), 122-140.
- [71] Masoudi S., Harmon S.A., Mehralivand S., Walker S.M., Raviprakash H., Bagci U., Choyke P.L., Turkbey B. (2021). Quick guide on radiology image pre-processing for deep learning applications in prostate cancer research. *Journal of Medical Imaging (Bellingham)*, 8(1).
- [72] Carré A., Klausner G., Edjlali M. *et al.* (2020). Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics. *Scientific Reports*, 10, 12340.
- [73] Dr. ir. Daan Christiaens, personal communication, April 25 2024.
- [74] Wee C.Y., Paramesran R. (2008). Image sharpness measure using eigenvalues. *International Conference on Signal Processing*, 9, 840-843.
- [75] Ameer S. (2020). EQ: An Eigen Image Quality Assessment based on the Complement Feature. *International Journal of Image, Graphics and Signal Processing*, 12(6), 13-18.

BIBLIOGRAPHY

- [76] Python (2024). pygrappa 0.26.2. Available on: <https://pypi.org/project/pygrappa/> [accessed on 6/04/2024].
- [77] Ronneberger O., Fischer P., Brox T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. Lecture Notes in Computer Science()*, 9351.
- [78] Krizhevsky A., Sutskever I., Hinton G.E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90.
- [79] Rajit Sanghvi (2021). A Complete Guide to Adam and RMSprop Optimizer. Available on: <https://medium.com/analytics-vidhya/a-complete-guide-to-adam-and-rmsprop-optimizer-75f4502d83be> [accessed on 6/04/2024].
- [80] Venkataramanan A.K., Wu C., Bovik A. C., Katsavounidis I., Shahid Z. (2021). A Hitchhiker's Guide to Structural Similarity. *IEEE Access*, 9, 28872-28896.