KU LEUVEN

FACULTY OF
ENGINEERING SCIENCE

Deep Learning-Based MRI Reconstruction in Data Heterogeneity: A Hybrid Modelling Approach

Alex Slock

Thesis submitted for the degree of
Master of Science in
Biomedical Engineering, option
Bio-Informatics and AI

Supervisor
Prof. dr. ir. D. Christiaens

Assessors
Prof. dr. ir. F. Maes
F. De Keyzer

Assistant-supervisor
Dr. T. Dresselaers

Academic year 2024 – 2025

Published by Alex Slock,
Faculty of Engineering Science, Kasteelpark Arenberg 1 bus 2200, B-3001 Leuven

Preface

I would like to begin by thanking my promoters, Daan and Tom, for their invaluable guidance throughout this thesis. I am especially grateful for the way they consistently challenged me with sharp, critical questions that pushed me to refine my ideas, clarify my reasoning, and maintain a high scientific standard. Their thoughtful feedback, helpfulness, and availability made this a valuable learning experience for which I am truly thankful.

My thanks also go to the Medical Imaging Research Center for providing the resources, infrastructure, and supportive environment necessary for conducting this research. The access to high-quality data and computing facilities was essential to the development and completion of this thesis.

I would also like to express my appreciation to the psychology students completing their internships at UZ Leuven during the same period. Their company turned every lunch break into a welcome moment of relaxation, humor, and good conversation, something I came to value more than expected during this intense process.

Finally, I am deeply grateful to my family and friends for their constant encouragement, patience, and support throughout this journey.

Alex Slock

Contents

Abstract

Deep Learning-Based MRI Reconstruction in Data Heterogeneity: A Hybrid Modelling Approach

Magnetic Resonance Imaging plays a vital role in modern diagnostics, yet its prolonged acquisition times remain a bottleneck in clinical workflows. To address this, researchers have long relied on acceleration techniques such as parallel imaging and compressed sensing (CS). More recently, deep learning has emerged as a game-changing alternative, offering rapid and high-quality image reconstruction from undersampled data. However, there is still a big problem: many deep learning models fail to generalize when imaging conditions like anatomy, contrast, or scanner hardware are different from their training data. This severely limits their clinical applicability.

This master thesis explores a promising solution through hybrid modelling, combining the robustness of physics-based CS reconstruction with the adaptability of deep neural networks. Specifically, we develop and evaluate a CS-hybrid reconstruction pipeline trained on raw, multi-coil k-space data from the publicly available NYU fastMRI dataset. The study focuses on how well such a hybrid model generalizes across anatomical domains and imaging parameters. A scenario closer to real-world clinical diversity.

This study finds that hybrid models substantially outperform classical CS reconstructions in all tested scenarios, producing images with higher fidelity and fewer artifacts. A key result is that a hybrid model trained on mixed anatomical data (brain and knee) not only generalizes better across domains, but even outperforms a domain-specific model on its own target domain. For example, the mixed-domain model achieves better results on brain scans than a hybrid model trained exclusively on brain data. This highlights that training on heterogeneous data enhances model robustness, not only across, but also within specific imaging tasks.

Even under challenging conditions like high undersampling or contrast variation, the hybrid model trained on mixed-domain data shows strong and consistent performance. These results underscore the potential of hybrid deep learning models as a clinically viable approach for fast, high-quality MRI in diverse and unpredictable imaging environments.

List of Figures and Tables

List of Figures

List of Tables

Chapter 1

Introduction

Magnetic resonance imaging (MRI) is an indispensable clinical tool due to its excellent soft-tissue contrast and non-invasive nature. However, its long acquisition times remain a major challenge for widespread use. To reduce scan time, classical acceleration techniques have been developed. Parallel imaging methods exploit multiple coil receiver data to shorten scans, while compressed sensing (CS) techniques leverage image sparsity to reconstruct from undersampled data (Pal & Rathi, 2022). Although these approaches have achieved impressive speed-ups, they rely on handcrafted models and iterative algorithms. Consequently, their acceleration is often limited by convergence speed or image artifacts, and high acceleration factors remain difficult without sacrificing quality (Pal & Rathi, 2022). In practice, classical methods still involve lengthy reconstructions and moderate acceleration levels.

In recent years, deep learning (DL) has emerged as a powerful new paradigm for MRI reconstruction. Neural networks (e.g. convolutional U-Nets, variational networks) can be trained on large MRI datasets to directly map undersampled k-space or images to high-quality reconstructions. These data-driven models have demonstrated higher reconstruction fidelity and dramatically faster inference compared to traditional iterative solvers (Heckel et al., 2024). For example, deep models can achieve real-time image reconstruction, which is critical for time-sensitive applications such as MR-guided interventions (Zeng et al., 2021). On the other hand, deep networks typically require large amounts of paired training data and may not generalize well when imaging conditions change (e.g. different scanners, contrasts, or anatomies). In practice, most published deep-learning reconstructions are trained and tested on narrow datasets (one anatomy, one contrast), and may degrade under domain shifts.

To combine the strengths of both worlds, hybrid methods have been proposed that embed physics-based MRI operators into deep networks. These models incorporate known reconstruction steps (e.g. GRAPPA or CS reconstruction) as fixed layers or pre-processing, followed by a learned refinement network. In principle, this hybrid architecture brings together the robustness of classical priors with the expressive power of learning. Indeed, prior work has shown that integrating traditional models

and deep learning often produces superior results. For example, (Sriram, Zbontar, Murrell, Zitnick, et al., 2020) demonstrated "GrappaNet," which weaves GRAPPA interpolation into a CNN, yielding better performance at high acceleration than either method alone. However, such hybrid approaches remain relatively unexplored in terms of generalization: it is unknown how well a hybrid model trained on one dataset will perform when imaging conditions vary.

This thesis builds on an earlier study of hybrid architectures (Vanhaverbeke, 2024), in which a CS-based hybrid model showed the most promise under the training conditions tested. The goal of the current work is to rigorously investigate the generalization performance of that CS-hybrid model under realistic, heterogeneous clinical conditions. In other words: *How does the CS-hybrid reconstruction network perform when tested on new anatomies, contrasts, and acquisition parameters not seen during training?* To address this, experiments are designed that vary anatomy (brain vs knee), image contrast, and acquisition parameters, reflecting real-world variability. In doing so, this work seeks to contribute to the understanding of how well hybrid models generalize in realistic, clinically relevant settings, and to identify potential failure modes and areas for further improvement.

The remainder of the thesis is structured as follows. Chapter 2 reviews related work on MRI reconstruction: it covers classical methods (parallel imaging, CS), modern deep-learning approaches, and existing hybrid models, highlighting their strengths and limitations. Chapter 3 describes the methodology, including dataset construction (using the NYU fastMRI raw k-space data to ensure clinical relevance (Knoll et al., 2020; Zbontar et al., 2019)), details of the CS-based hybrid architecture, and the training and evaluation protocols used. Chapter 4 presents experimental results: we compare the hybrid model (trained on brain-only vs multi-anatomy data) to a pure CS baseline across multiple generalization tests, analysing quantitative metrics and visual reconstructions. Finally, Chapter 5 summarizes the findings, critically evaluates the approach, and suggests avenues for future research.

Chapter 2

Literature review

This chapter provides a comprehensive overview of the current state of MRI reconstruction, with a particular focus on deep learning methods that integrate classical parallel imaging techniques. The review is structured to provide both technical foundations and a critical assessment of current research gaps.

The first section begins by outlining the principles of MRI acquisition, emphasizing the role of k-space and the need for acceleration. Next, we discuss classical reconstruction methods such as GRAPPA, SENSE and CS, which serve as clinical standards for MRI reconstruction using parallel imaging. However, while effective, these methods suffer from limitations when accelerating too much.

Alongside classical techniques, deep learning has introduced a fundamentally different reconstruction paradigm that leverages neural networks for enhanced performance. To provide a foundational understanding, basic concepts of neural networks are introduced, followed by an exploration of how deep learning is applied to MRI reconstruction. Special attention is given to a small but promising class of hybrid models that combine classical parallel imaging with deep learning.

The last section identifies a key gap in the current literature: while hybrid methods show promising results in narrow settings, there has been little to no systematic evaluation of their generalization across anatomical regions, contrasts, or sampling protocols. This motivates the primary objective of this thesis: to assess the generalizability of hybrid reconstruction models in more realistic, heterogeneous clinical conditions.

2.1 Principles of Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is a medical imaging technique based on the interaction between atomic nuclei and electromagnetic fields. This section presents the fundamental physical principles that govern MRI acquisition and image formation. The explanation is primarily based on the lecture materials (Maes, 2024), complemented by the theoretical introduction in the first chapter (Tourais et al., 2022) of the book(Akçakaya et al., 2023).

2.1.1 Nuclear spin and magnetic moment

MRI relies on the intrinsic quantum property of spin found in elementary particles. Spin, while not representing physical rotation, manifests as quantized angular momentum $\boldsymbol{J}$ and gives rise to a magnetic dipole moment $\boldsymbol{\mu}$ in charged particles (e.g. proton). This moment allows particles such as hydrogen nuclei (^{1}H, equal to a proton), abundant in tissue and the most commonly imaged nucleus in MRI, to interact with external magnetic fields.

The magnetic dipole moment $\boldsymbol{\mu}$ of a spin is proportional to the angular momentum $\boldsymbol{J}$, and is given by:

$$\boldsymbol{\mu} = \gamma \boldsymbol{J}. \tag{2.1}$$

where γ is the gyromagnetic ratio. For protons, $\gamma/2\pi \approx 42.6\,\mathrm{MHz\,T^{-1}}$, meaning a proton in a 1 T magnetic field will precess at approximately 42.6 MHz.
When placed in an external static magnetic field $\boldsymbol{B} = B_0\hat{z}$, the magnetic moment experiences a torque, resulting in precessional motion around the field direction:

$$\boldsymbol{\tau} = \boldsymbol{\mu} \times \boldsymbol{B}_0, \tag{2.2}$$

which causes the magnetic moment to precess around $\boldsymbol{B}_0$ at the Larmor frequency:

$$\omega_0 = \gamma B_0. \tag{2.3}$$

Quantum mechanically, a proton has two spin states: spin-up $s = +\frac{1}{2}$ and spin-down $s = -\frac{1}{2}$, corresponding to quantized energy levels separated by:

$$\Delta E = \hbar\omega_0. \tag{2.4}$$

Transition between these energy levels occur when the system absorbs or emits photons at the resonance frequency ω_0, equal to the Larmor frequency. This resonance condition underlies the basic mechanism of MRI.

When there is no external magnetic field $\boldsymbol{B}$, both states are equally likely. However, at equilibrium in the presence of an external field $\boldsymbol{B} = B_0\hat{z}$, the lower-energy spin-up state is slightly more populated than the spin-down state. This imbalance results in a net magnetization vector $\boldsymbol{M}_0$ (all spins combined) aligned with the field direction. This forms the basis for MRI signal generation, where controlled electromagnetic fields are used to excite and measure nuclear spins, producing detailed images of biological tissue.

2.1.2 Magnetization dynamics and RF excitation

In a macroscopic sample, the ensemble average of proton magnetic moments yields a bulk magnetization vector $\mathbf{M}(t)$, that itself precesses around the external field:

$$\frac{d\boldsymbol{M}}{dt} = \gamma \boldsymbol{M} \times \boldsymbol{B}. \tag{2.5}$$

Since the static magnetic field $\boldsymbol{B}_0$ is orders of magnitude stronger than $\boldsymbol{M}$, changes in the longitudinal direction ($\hat{z}$-axis) are not practically detectable. Instead, MRI detects the time-varying transverse magnetization $M_{xy}(t)$, which generates a measurable signal in receiver coils.

To acively manipulate this net magnetization and generate imaging contrast, an additional RF magnetic field $\boldsymbol{B}_1(t)$, oscillating at the Larmor frequency ω_0, is applied perpendicular to $\boldsymbol{B}_0$. In a reference frame rotating at ω_0, $\boldsymbol{B}_1(t)$ appears static, causing $\boldsymbol{M}$ to tip away from the $\hat{z}$-axis, a process known as excitation. By controlling the duration Δt and amplitude of the RF pulse, the flip angle can be precisely adjusted. For example, a 90° RF pulse fully rotates the equilibrium magnetization $\boldsymbol{M} = M_0\hat{z}$ into the transverse plane, maximizing M_{xy} while eliminating M_z.

Once the RF pulse is turned off, the magnetization returns to equilibrium via two processes: (1) longitudinal (T_1) relaxation, which describes the recovery of the magnetization along $\hat{z}$ driven by energy exchange of the spins with surrounding molecules (the lattice), and (2) transversal (T_2) relaxation, which characterizes the decay in the xy-plane due to dephasing among spins by spin-spin interactions.
These are modelled by the following exponential decay models:

$$M_z(t) = M_0\left(1 - e^{-t/T_1}\right), \quad M_{xy}(t) = M_{xy}(0)e^{-t/T_2}. \tag{2.6}$$

Typically, $T_2 < T_1$, as spin-spin dephasing occurs faster than energy exchange with the lattice.

2.1.3 Temporal Parameters and Image Contrast

MRI contrast arises from variations in local spin density, relaxation times (T_1 and T_2), and RF pulse timing, all of which influence signal intensity and tissue differentiation. Since spin density and relaxation times are tissue-specific, image contrast in MRI is governed by the interaction of these intrinsic properties with sequence timing parameters. Two of the most important timing parameters are the repetition time (TR), which defines the time between consecutive excitations, and the echo time (TE), which is the time between the RF excitation and the center of the signal readout. Adjusting TE and TR allows control over image contrast, highlighting different tissue properties. Short TR and short TE enhance T_1 contrast by suppressing T_2 effects, producing T_1-weighted images. Conversely, long TR and TE minimize T_1 influence and emphasize differences in T_2 relaxation, yielding T_2-weighted images. Proton density-weighted images are obtained with a long TR (to reduce T_1 contrast) and a short TE (to reduce T_2 effects), so spin density becomes the biggest influence.

2.1.4 Signal detection and image reconstruction

The precessing transverse magnetization $M_{xy}(t)$ induces a voltage in the receiver coils, forming the measurable MR signal. However, this measured signal S arises from the sum of contributions from many voxels s_r:

$$S^{TR,TE} = \sum s_r^{TE,TR} \tag{2.7}$$

MRI achieves spatial encoding by making the precession frequency ω_0 position-dependent, introducing a phase shift across voxels to differentiate their contributions in the final image:

$$S^{TR,TE} = \sum s_r^{TE,TR} e^{-j\phi_r} \tag{2.8}$$

This spatial encoding is achieved by superimposing additional magnetic field gradient $\boldsymbol{G}(t)$ onto the static field $\boldsymbol{B_0}$. By applying linear gradients whose amplitudes and durations are controlled over time, the phase accumulation $\phi(r)$ across space becomes a linear function of spatial position. This ensures that each spatial location contributes a distinct phase to the signal, effectively encoding spatial frequency information. Under these conditions, the measured signal corresponds to a weighted sum of spatial frequencies and can be expressed as:

$$S(\boldsymbol{k}) = \int s(\boldsymbol{r}) e^{-j\phi(\boldsymbol{r})} d\boldsymbol{r} = \int s(\boldsymbol{r}) e^{-i2\pi \boldsymbol{k}\cdot\boldsymbol{r}} d\boldsymbol{r} \tag{2.9}$$

This expression is a 2D Fourier transform, mapping $s(\mathbf{r})$ in image space to the measured signal $S(\boldsymbol{k})$ in k-space (spatial frequency domain), where $\mathbf{k}$ depends on the applied gradient waveforms:

$$\mathbf{k}(t) = \frac{\gamma}{2\pi} \int_0^t \mathbf{G}(\tau)\, d\tau \tag{2.10}$$

The final MR image is obtained by numerically computing the inverse Fourier transform of the measured signal $S(\mathbf{k})$.

To obtain an image with high spatial resolution, k-space must be sufficiently sampled. In Cartesian acquisitions, spatial encoding is typically applied along two orthogonal directions within the imaging plane:
(1) The frequency encoding (readout direction). A gradient G_x is applied along the x-axis during signal acquisition, separating spatial contributions via frequency-dependent precession. This frequency variation allows the simultaneous capture of a full line in k-space along the k_x direction.
(2) Phase encoding. To encode the orthogonal y-direction, a gradient G_y is applied briefly before signal acquisition. This induces a position-dependent phase shift $\phi(y) = \gamma G_y \Delta t \cdot y$, effectively encoding each k-space line with a unique phase. The phase encoding gradient is varied between repetitions of the excitation and readout sequence, so that successive lines in k-space are filled over time. In simple Cartesian acquisitions, typically one phase-encoded line is acquired per excitation, which makes this step the primary bottleneck for scan time. However, some advanced methods can acquire multiple k-space lines per excitation using gradient refocusing or echo trains.

2.1.5 Undersampling and Aliasing

The spatial resolution and overall image fidelity in MRI depend on the density and extent of k-space sampling. The maximum sampled spatial frequency $k_{\max}$ (i.e., the edge of k-space) determines image resolution:

$$\Delta x = \frac{1}{2k_{\max}}. \tag{2.11}$$

Meanwhile, the field of view (FOV) is inversely proportional to the sampling interval Δk in k-space:

$$\text{FOV} = \frac{1}{\Delta k}. \tag{2.12}$$

Aliasing artifacts arise when structures outside the intended FOV appear superimposed in the reconstructed image, obscuring diagnostic details. To prevent this, k-space must be sampled with sufficient density, following the Nyquist criterion:

$$\Delta k \leq \frac{1}{\text{FOV}}. \tag{2.13}$$

Undersampling k-space by skipping phase encoding lines reduces scan time but violates the Nyquist condition, resulting in aliasing artifacts. While sampling fewer lines in k-space shortens scan duration, which is desirable in clinical practice, it is at the cost of image fidelity.

To mitigate these effects, modern accelerated MRI techniques utilize advanced reconstruction algorithms, ranging from classical compressed sensing to deep learning methods that leverage prior knowledge for artifact suppression and recover high-quality images from undersampled k-space data.

2.2 Undersampling and Classical Reconstruction Techniques for MRI

This chapter discusses classical methods to reconstruct MR images from undersampled k-space data. Unless otherwise noted, the content is primarily based on Chapters 6 (Cummings et al., 2022) and 8 (Feng, 2022) of the book by Akçakaya, Doneva, and Prieto (Akçakaya et al., 2023), which cover parallel imaging and compressed sensing, respectively.

As described in the previous section, an MR image $\mathbf{x}$ can be reconstructed from a fully sampled k-space measurement $\mathbf{y}^{\text{full}}$ via the Fourier transform:

$$\mathbf{y}^{\text{full}} = \mathcal{F}\mathbf{x} + \eta \tag{2.14}$$

where $\mathcal{F}$ denotes the Fourier transform and $\eta \sim \mathcal{N}(0, \sum)$ represents the measurement noise, typically assumed to have a Gaussian distribution (Virtue & Lustig, 2017).

To accelerate the MR acquisition process, data can be undersampled by skipping phase-encoding lines in k-space. This can be implemented by applying a binary sampling mask $\mathcal{M}$ that selects a subset of k-space lines in the phase encoding direction:

$$\mathbf{y} = \mathcal{M} \odot \mathcal{F}\mathbf{x} + \eta \tag{2.15}$$

where $\mathbf{y}$ is the undersampled k-space, and $\odot$ is the element-wise multiplication operator (Sriram, Zbontar, Murrell, Defazio, et al., 2020; Zheng et al., 2019).

Direct image reconstruction using an inverse Fourier transform from undersampled data leads to artifacts and resolution loss. The severity of these effects depends on the sampling mask, which may limit the maximum captured spatial frequency or introduce incoherent k-space gaps, resulting in aliasing. To restore high-quality images despite undersampling, several classical reconstruction techniques have been developed. These methods, long established in clinical MRI practice, include parallel imaging approaches such as GeneRalized Autocalibrating Partially Parallel Acquisitions (GRAPPA) (Griswold et al., 2002) , SENSitivity Encoding (SENSE) (Pruessmann et al., 1999), as well as compressed sensing (CS) (Candes et al., 2006; Donoho, 2006; Lustig et al., 2007, 2008).

2.2.1 Parallel MR Imaging

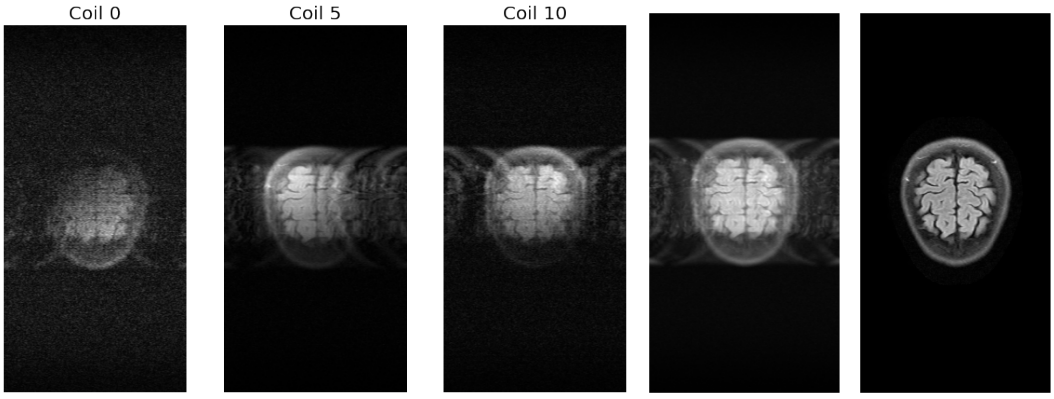
In MRI, data are collected in k-space using receiver coils. Modern MR systems employ arrays of multiple receiver coils, each with spatially localized sensitivity. These coil sensitivity variations can be leveraged to reconstruct undersampled data through parallel imaging techniques. The signal from each coil can be modelled as:

$$\mathbf{y}_i = \mathcal{M} \odot \mathcal{F}(S_i \odot \mathbf{x}) + \eta, \tag{2.16}$$

where S_i is the sensitivity map of the i-th coil, $\mathbf{x}$ is the true image, and $\mathbf{y}_i$ is the undersampled k-space measurement from that coil. Here, $i = 1, 2, \ldots, n_c$, with n_c denoting the number of coils.

As illustrated in Figure 2.1, when k-space is undersampled, individual coil images appear aliased and cannot be simply combined to recover the original image. Instead, parallel imaging techniques exploit the spatial variation in coil sensitivity profiles to resolve the aliasing. If the sensitivity maps S_i are known and the sampling satisfies the Nyquist criterion, then the set of equations across all coils forms an overdetermined linear system, allowing recovery of the true image. In accelerated acquisitions, each coil receives undersampled data, but their differing sensitivities provide complementary information that enables partial signal recovery.

In theory, parallel imaging permits an acceleration factor up to the number of coils in an array (n_c). However, in practice, limits of 2-3 are typical due to noise amplification and ill-conditioning. Direct inversion of Equation (2.16) to compute $\mathbf{x}$ is computationally inefficient and sensitive to noise. Therefore, algorithmic approaches such as GRAPPA, SENSE, and CS are preferred, which are discussed next.



(A) Aliased coil images (3 examples) (B) RSS combination (C) CS reconstruction

FIGURE 2.1: Illustration of parallel imaging reconstruction under undersampled k-space acquisition. (a) Aliased images from three selected coils (undersampled by a factor $R = 4$). (b) The root-sum-of-squares (RSS) combination of all $n_c = 20$ coil images, which does not resolve aliasing because it ignores coil sensitivity encoding. (c) Compressed Sensing reconstruction using coil sensitivity maps, which effectively resolves aliasing artifacts and reconstructs the underlying image.

2.2.2 GRAPPA

GeneRalized Autocalibrating Partially Parallel Acquisitions (GRAPPA) is a parallel imaging technique that operates directly in k-space, where it reconstructs missing data points by modelling their relationships with nearby acquired samples across multiple receiver coils (Griswold et al., 2002). The GRAPPA weight set captures the relationship between neighbouring k-space points. It relies on a fully sampled central region of k-space known as the Autocalibration Signal (ACS) region to learn the interpolation weights. This same weight set is then applied to the remaining undersampled data to estimate missing k-space points.

A GRAPPA kernel defines a structured neighbourhood in k-space, consisting of known source points and the corresponding target point to be estimated. A standard GRAPPA kernel includes a few source points positioned along fully sampled phase-encoding lines near the target point and extends across the coil dimension. This pattern is repeated across the dataset and captures the local spatial dependencies induced by the coil sensitivities. To determine the weights, source points and target points are collected for each occurrence of the kernel in the ACS data. Let $\mathbf{S}_{\text{ACS}}$ and $\mathbf{T}_{\text{ACS}}$ denote the matrices of source and target samples from the ACS region. Then the GRAPPA weight matrix $\mathbf{W}$ is learned via:

$$\mathbf{T}_{\text{ACS}} = \mathbf{S}_{\text{ACS}}\mathbf{W}, \tag{2.17}$$

and the weights are computed using the Moore–Penrose pseudo-inverse:

$$\mathbf{W} = \mathbf{S}_{\text{ACS}}^{\dagger}\mathbf{T}_{\text{ACS}}. \tag{2.18}$$

Once trained, these weights are applied to the undersampled data to estimate the missing target points:

$$\mathbf{T} = \mathbf{SW}. \tag{2.19}$$

The ACS region is usually positioned in the central k-space, where the signal is strongest and the signal-to-noise ratio is highest. This placement enhances the stability of the GRAPPA weight estimation by reducing the impact of noise. ACS data can be collected either during the regular scan, or as a separate scan either before or after undersampled data acquisition.

One of GRAPPA's strengths in clinical practice is its robustness to imperfections in the data. By averaging over a large number of kernel instances within the ACS region, the algorithm minimizes the influence of localized artifacts. Moreover, GRAPPA operates without requiring explicit coil sensitivity maps, which simplifies calibration and avoids potential sources of error when maps are inaccurate or corrupted by motion.

2.2.3 SENSE

SENSitivity Encoding (SENSE) (Pruessmann et al., 1999) is an image-domain parallel imaging technique that explicitly uses coil sensitivity profiles to disentangle aliasing caused by undersampling. When the k-space is undersampled by a factor R, the reconstructed image contains aliasing: each observed pixel in the aliased image corresponds to a mixture of R distinct spatial locations in the original full-FOV image. Each of these spatial contributions is modulated by the sensitivity of the respective coil at that position.

Let $\mathbf{y}$ be the vector of aliased measurements (length n_c) from all coils at a given image location, and $\mathbf{C}$ the $n_c \times R$ sensitivity matrix corresponding to the R aliased locations. The aliased measurements are then given by:

$$\mathbf{y} = \mathbf{Cx}, \tag{2.20}$$

where $\mathbf{x}$ contains the R true pixel values to be recovered. Solving this system via the pseudo-inverse:

$$\mathbf{x} = \mathbf{C}^{\dagger}\mathbf{y}, \tag{2.21}$$

yields the unaliased pixel values. This reconstruction is applied across all aliased locations to recover the full image.

In theory, exact recovery in SENSE requires the number of effectively independent coil elements along the phase encoding direction to match or exceed the acceleration factor ($n_{c,PE} \geq R$). For instance, in a 3×4 coil array, if acceleration is applied along the vertical axis, only the 3 vertically distinct coils contribute to resolving aliasing. Due to imperfect coil sensitivity profiles and noise amplification during inversion, the practical acceleration limit is typically lower than this theoretical bound.

Accurate estimation of the coil sensitivity maps S_i is essential. One approach is the ratio method from the original SENSE work (Pruessmann et al., 1999), where each coil image is divided by a reference body coil image to approximate S_i. A more robust and widely adopted method is ESPIRiT (Uecker et al., 2014), which estimates smooth, consistent sensitivity maps directly from the ACS region using an eigenvalue-based approach. ESPIRiT integrates well with iterative reconstruction techniques and reduces sensitivity to noise and boundary artifacts.

Although SENSE can surpass GRAPPA in achievable acceleration under ideal conditions, its performance strongly depends on accurate coil sensitivity estimation. Errors in sensitivity maps can lead to noise amplification and residual aliasing artifacts, particularly in low-SNR regions or areas where the encoding matrix is poorly conditioned.

2.2.4 CS

Compressed sensing (CS) exploits the sparsity of MR images in a transform domain (e.g., wavelets or finite differences) to enable recovery from incoherently undersampled data (Candes et al., 2006; Donoho, 2006; Lustig et al., 2007, 2008). The key idea behind compressed sensing is that many natural images, including MR images, can be efficiently represented using only a small number of significant coefficients in an appropriate transform domain. This observation leads to a natural question: if most coefficients contribute little to the final image, can we avoid acquiring them in the first place?

Consider a simplified measurement model, adapted from equation (2.15):

$$\mathbf{y} = E\mathbf{x} \tag{2.22}$$

where the vector $\mathbf{x}$ of size N represents the MR image, the encoding operator $E = \mathcal{M} \odot \mathcal{F}$ of size $M \times N$ represents the operation between image and k-space, and $\mathbf{y}$ represents the k-space measurements concatenated into a vector of size M. Suppose the image $\mathbf{x}$ is known to be K-sparse, meaning it contains only $K \ll N$ significant values. If the locations of these non-zero elements were known in advance, the measurement system could be restricted to only these components using a reduced encoding matrix E_K of size $M \times K$.

$$\mathbf{y} = E_K\mathbf{x}_K + \eta. \tag{2.23}$$

Under ideal conditions, K measurements suffice to recover $\mathbf{x}_K$. In reality, since the locations of the sparse coefficients are unknown, more measurements (typically $2K$) are required. To ensure a unique solution to Equation (2.23), any set of $2K$ columns of E must be linearly independent—a requirement known as incoherent sampling.

For compressed sensing to succeed, two key conditions must be met: (1) the signal must be sparse or compressible in some transform domain and (2) the sampling

must be incoherent with respect to this sparse basis.

First, while MR images are not inherently sparse in the image domain, they are typically compressible in transform domains such as wavelets or finite differences. These domains compact most of the image energy into a small number of significant coefficients.

Second, to maximize incoherence, random undersampling patterns are typically used. The incoherence of a sampling pattern can be analysed through its point spread function (PSF), computed by applying a 2D Fourier transform to a zero-filled sampling mask. Low PSF sidelobes correspond to high incoherence. High incoherence is obtained using random undersampling patterns, while regular undersampling patterns give high coherence. A common strategy is variable-density random sampling, which allocates more samples to the center of k-space (where most signal energy resides) and fewer to the periphery. This balances energy preservation with artifact incoherence.

Next, the algorithm needs to reconstruct the image from this undersampled data. In compressed sensing, image reconstruction from undersampled data is framed as an inverse problem where one seeks the sparsest signal that is still consistent with the acquired measurements. Since the locations of the K non-zero coefficients of the signal are unknown in practice, the reconstruction must involve a search over all possible sparse representations. This leads to the formulation of the reconstruction problem as an optimization:

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{y} = E\mathbf{x}, \tag{2.24}$$

where the zero-norm $\|\mathbf{x}\|_0$ counts the number of non-zero entries in the signal. The goal is to select the sparsest solution that matches the measurements.

In practice, this formulation is refined to improve computational efficiency and robustness. First, the L_0-norm minimization is computationally expensive, and the convex L_1-norm is used instead, which also promotes sparsity but enables efficient optimization. Second, few signals are truly sparse. Instead, CS uses compressible signals, meaning they have a representation in a different domain (e.g. finite differences or wavelets) where they are sparse. The compressible signal $\mathbf{x}$ is expressed as $\mathbf{d} = \Phi\mathbf{x}$, using a sparsifying transform Φ. Third, signal measurements are contaminated by noise, therefore the L_2-norm is used to quantify the difference between the measurements and the estimated solution. Therefore, the optimization problem can be modified to:

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - E\mathbf{x}\|_2 < \epsilon \tag{2.25}$$

Using Lagrange multipliers this becomes:

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - E\mathbf{x}\|_2^2 + \lambda\|\Phi\mathbf{x}\|_1, \tag{2.26}$$

where λ is a regularization parameter balancing data consistency and sparsity. In this formulation, the L_2-norm ensures consistency with measured data, while the L_1-norm enforces a prior on sparsity in the transform domain.

The reconstruction is typically performed using iterative algorithms such as the iterative soft-thresholding algorithm (ISTA), its accelerated variant FISTA, nonlinear conjugate gradient methods, or the alternating direction method of multipliers (ADMM) (Akçakaya et al., 2023). These algorithms progressively enforce data consistency and shrink insignificant coefficients in the sparse domain.

CS can be used as a standalone algorithm to improve MRI reconstruction, or in combination with parallel imaging. The encoding operator E from equation (2.16) then collapses the operation $\mathcal{M} \odot \mathcal{F}(C \odot \mathbf{x})$ into $E\mathbf{x}$, now including $\mathbf{C}$, the matrix of the coil sensitivity maps S_i. This extended model forms the basis for joint compressed sensing and parallel imaging reconstructions, enabling further acceleration by exploiting both spatial sparsity and coil sensitivity diversity.

Collectively, SENSE, GRAPPA, and CS form the foundation of modern accelerated MRI techniques. Each leverages a different type of structure: spatial encoding, coil diversity, or signal sparsity to reconstruct images from incomplete data. These methods have enabled substantial reductions in scan time but also face limitations, particularly under extreme undersampling or when their respective assumptions are violated. Additionally, classical reconstruction approaches typically rely on iterative solvers that require substantial computation time, resulting in long reconstruction delays. This can limit their practicality in time-sensitive clinical workflows.

2.3 Background on Deep Learning

Unless stated otherwise, this section draws primarily from the course handout on Artificial Neural Networks and Deep Learning (Suykens, 2024).

In recent years, artificial intelligence (AI) has become increasingly influential in medical imaging, particularly in applications such as image classification, segmentation, and, more recently, the reconstruction of under-sampled MRI data.

The motivation to apply deep learning to MRI reconstruction stems from the limitations of classical methods. Traditional approaches, such as GRAPPA, SENSE, and Compressed Sensing (CS), have played an essential role in enabling accelerated imaging by exploiting coil sensitivity profiles and sparsity priors. However, they often require lengthy reconstruction times, careful manual tuning of hyperparameters, and typically rely on fixed mathematical priors that may not capture the complex variability found in clinical data. Deep learning presents a powerful alternative, learning directly from examples to model intricate signal relationships and context-dependent features. This data-driven approach enhances reconstruc-

tion quality while significantly reducing acquisition time (Huang et al., 2025; Pal & Rathi, 2022; Singh et al., 2023).

2.3.1 Conceptual Foundations: From AI to Deep Learning

While the terms artificial intelligence, machine learning, and deep learning are often used interchangeably in both academic and clinical literature, they each carry specific and important distinctions. AI refers broadly to computational systems capable of performing tasks that typically require human intelligence. Within this domain, machine learning (ML) describes a subset of methods that learn patterns or rules from data without being explicitly programmed. Deep learning (DL) is a specialized subset of machine learning that employs multi-layered artificial neural networks (ANNs) to extract increasingly abstract representations from input data.

These distinctions are not merely semantic. They reflect underlying differences in data requirements, model capacity, and the degree of human intervention needed. Classical ML methods often rely on hand-crafted features and shallow architectures, while DL models can learn features automatically and at multiple levels of abstraction. This has made deep learning particularly suitable for problems involving high-dimensional data and complex mappings, such as the transformation of incomplete MRI measurements into high-quality anatomical images.

2.3.2 Introduction To Artificial Neural Networks

At the core of deep learning models are artificial neural networks, composed of layers of interconnected units (neurons). Each neuron performs a weighted sum of its inputs followed by a non-linear activation function:

$$y_i = \sigma\left(\sum_{j=1}^{n} w_{ij}x_j + b_i\right) \tag{2.27}$$

Here, x_j are the input features, w_{ij} are the weights, b_i is the bias term, and σ is the non-linear activation function (e.g., ReLU, sigmoid). The neural network consists of an input layer, output layer, and layer(s) of interconnected neurons in between that are not visible from the outside, referred to as hidden layer(s). The depth of a neural network is equal to the number of hidden layers.

A key property of neural networks is their ability to approximate any continuous non-linear function over a compact interval. This was proven by Sonoda and Murata, who showed that feedforward networks with unbounded activation functions, such as ReLU, serve as universal approximators (Sonoda & Murata, 2017).

2.3.3 Training and testing of Neural Networks

A neural network learns, or trains, by adapting the weights w_{ij} based on examples. Initially, input data flows through the network, producing an output through

a process called forward propagation. Next, the loss is computed by comparing the output to the ground truth using a loss function (e.g., mean squared error (MSE)). This is followed by backpropagation, where the loss is propagated backward through the network using the chain rule to calculate the gradients of the loss with respect to all weights. Finally, the weights are then optimized according to these gradients.

Effective training requires careful data splitting to ensure that a model generalizes well to unseen data. In supervised learning settings, datasets are commonly divided into training, validation, and test sets. The validation set is used to monitor the model's performance during training and to detect overfitting—when the model's parameters become too specialized to the training data and perform poorly on new data. Without a validation set, it is difficult to determine when overfitting occurs, as the training loss will typically decrease monotonically due to continuous optimization of the model's weights on the training data.

To detect overfitting, the loss on the validation set is computed after each training epoch. An increase in validation loss while training loss continues to decrease indicates that the model is starting to overfit, as it no longer generalizes well to unseen examples. The model weights corresponding to the epoch with the lowest validation loss are therefore selected as the optimal set. This behaviour is illustrated in figure 2.2, where epochs to the right of the minimum validation loss indicate the onset of overfitting. While the validation set is used during training to tune model

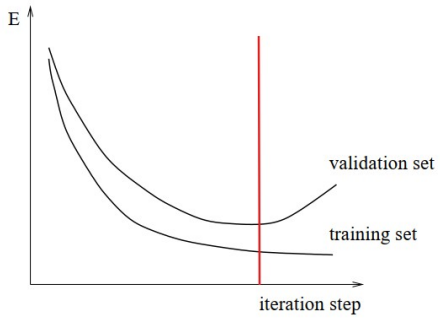


FIGURE 2.2: Loss on training and validation set, vertical line represents moment of minimum validation loss, adapted from (Suykens, 2024)

parameters and detect overfitting, the test set serves a different purpose: it is used only after training is complete to provide an unbiased evaluation of the model's generalization to unseen data.

Importantly, the test set must remain untouched throughout the training and validation process to ensure that performance metrics are not overestimated. In medical imaging, special attention must be given to how data is split, as multiple slices can originate from a single patient. Random splitting is generally discouraged in this context due to the risk of information leakage between training, validation, and test

sets.

2.4 Deep Learning in MRI Reconstruction of undersampled k-space data

This section provides a detailed review of the principal categories of deep learning models employed in the reconstruction of undersampled MRI data. It emphasizes the diversity of architecture types, explains their strategies for solving the reconstruction problem, discusses common training paradigms, and highlights notable architectural innovations.

2.4.1 Reconstruction Strategy: From Physics-Based to Hybrid Architectures

The effectiveness of a deep learning model for MRI reconstruction relies not only on architectural design, but equally on the strategy employed to solve the inverse problem. Reconstruction strategies can be broadly categorized by the degree to which they incorporate physical knowledge of the MRI acquisition process. This results in three major classes: model-based, data-driven, and hybrid approaches that combine the strengths of classical and learned components.

Model-Based Networks

Model-based methods embed knowledge of the MRI acquisition physics directly into the architecture. These approaches aim to bridge the gap between data-driven learning and traditional iterative reconstruction techniques. Specifically, they mimic optimization algorithms such as gradient descent or the alternating direction method of multipliers (ADMM), unrolling their iterative steps into trainable network layers. For this reason, they are also often referred to as "unrolled optimization networks". A key feature of these models is the incorporation of a data consistency step, ensuring that reconstructions remain faithful to the measured k-space data, followed by a learned regularization module that replaces classical priors with deep neural networks (Kim et al., 2024; Wang et al., 2021). These architectures exploit the known physics of the MRI acquisition process, allowing for improved interpretability and stability compared to purely data-driven networks.

Data-Driven Networks

In contrast to model-based methods, data-driven strategies rely entirely on learning the inverse mapping from undersampled inputs to fully reconstructed images. These models make no explicit assumptions about the MRI physics, but instead learn the reconstruction function through supervised learning on paired examples. Data-driven networks can operate in the image domain, frequency domain, or both. They often use U-Nets, residual networks, or generative adversarial networks (GANs) to learn an end-to-end transformation from undersampled inputs to fully reconstructed

images. While being able to reach higher accuracies than model-based networks, purely data-driven models may struggle with generalization to unseen sampling patterns or anatomical variations unless trained on large and diverse datasets (Wang et al., 2021).

Hybrid Methods

Hybrid networks seek to combine the physical interpretability of model-based methods with the representational power of data-driven models. This is accomplished by embedding known operators, classical reconstruction techniques, or acquisition constraints directly into trainable neural networks, allowing the network to leverage both domain knowledge and learned priors. Studies on unrolled networks suggest that incorporating domain knowledge, rather than solely relying on extensive training data, enhances generalization (Monga et al., 2020). Thus, by embedding classical techniques in hybrid networks, possibly allows them to generalize better to unseen data while maintaining the adaptability of deep learning. These approaches can be particularly well-suited for clinical translation, where interpretability, generalization, and performance are of highest priority.

2.4.2 Deep Learning Architectures and innovations

The landscape of deep learning models for MRI reconstruction can be broadly categorized into five major classes: convolutional neural networks (CNNs), unrolled optimization networks (model-based), generative adversarial networks (GANs), transformer-based models, and diffusion models. Each of these architectures introduces distinct assumptions, design choices, and trade-offs tailored to the reconstruction task. In this section, we describe each model category and highlight relevant architectural innovations that enhance their reconstruction performance, including skip and residual connections, attention mechanisms, and dual-domain strategies.

Convolutional Neural Networks (CNNs)

CNNs represent one of the earliest and most prevalent architectures in DL-based MRI reconstruction. These models typically learn a direct mapping from undersampled image-domain or k-space inputs to fully reconstructed images. The U-Net architecture, for example, applies a downsampling encoder to extract hierarchical features, followed by an upsampling decoder that reconstructs the image. Skip connections preserve spatial resolution and detail during reconstruction. The U-Net is a canonical example that has been widely used as a baseline due to its simplicity and consistent performance (Ronneberger et al., 2015).

Image-domain CNNs have demonstrated competitive reconstruction quality with minimal architectural complexity. A standard U-Net trained using an L1 or L2 loss on zero-filled reconstructions already offers substantial improvements over classical compressed sensing approaches (Zbontar et al., 2019). Enhancements such as residual connections, multiscale processing, and channel-wise attention have been used to

improve their expressive power. Channel attention, in particular, selectively emphasizes informative features while suppressing noise, improving fidelity in multi-coil settings (Schlemper et al., 2018).

A key architectural innovation in CNNs is the use of skip connections, which were originally introduced in the U-Nets and ResNets. These connections preserve spatial detail by allowing high-resolution feature maps from the encoder to bypass downsampling and directly contribute to decoder outputs (Ronneberger et al., 2015). Residual connections extend this idea by enabling layers to learn residual functions with respect to their inputs, rather than direct mappings. Such techniques are particularly effective in cascaded networks like RefineGAN, where each module incrementally refines the reconstruction (Quan et al., 2018). These mechanisms improve both PSNR and SSIM, especially in high-undersampling regimes (Schlemper et al., 2018; Zbontar et al., 2019), and help suppress vanishing gradient issues during backpropagation, making them indispensable in training very deep or cascaded models.

Despite their strengths, conventional CNNs are inherently limited by their local receptive field, which restricts their ability to model spatial relationships between distant regions in the image, known as long-range dependencies. In MRI reconstruction, this includes capturing coherent anatomical structures (e.g., edges or textures spanning large areas) or resolving aliasing artifacts that manifest across non-local regions of k-space and image space. To overcome this, recent architectures incorporate global context through more advanced mechanisms such as attention layers and transformer blocks, as discussed below (Aggarwal et al., 2019; Huang, Fang, et al., 2022).

Unrolled Optimization Networks

Unrolled networks imitate the iterative steps of traditional optimization algorithms by casting them into fixed-depth trainable architectures (see 2.4.1). A representative model is the Variational Network (VN), which unrolls the gradient descent steps used in variational inference. Each stage of the network includes a CNN functioning as a trainable prior, alongside a fixed data consistency module that ensures alignment with acquired measurements. The entire network is trained end-to-end, learning optimal update rules for both data fidelity and image regularization from data (Hammernik et al., 2018).

Another key example is MoDL (Model-based Deep Learning). This architecture follows a similar pattern, alternating between a data consistency operation and a learned CNN denoiser. What sets MoDL apart is its modular training strategy and its ability to handle coil sensitivity maps internally, allowing adaptation to single-coil and multi-coil data. The regularization module is trained independently and then embedded in the iterative loop, offering better generalization across datasets (Aggarwal et al., 2019).

More recent developments in this category include deep equilibrium models that simulate infinite unrolling via fixed-point iterations and implicit layers. These models aim to reduce memory overhead while maintaining high accuracy and convergence reliability (Gilton et al., 2021).

Generative Adversarial Networks (GANs)

GANs introduce a fundamentally different approach by learning to generate high-fidelity images that are indistinguishable from real data. In the context of MRI reconstruction, a generator network is trained to produce images from undersampled data, while a discriminator network attempts to distinguish the output from ground truth images. The adversarial training framework encourages the generator to produce visually realistic and high-detail reconstructions (Mardani et al., 2019).

Early models such as DAGAN (Deep De-Aliasing GAN) utilize a U-Net generator and combine adversarial loss with pixel-wise loss functions (e.g., L1 or perceptual loss) to enhance image sharpness and texture detail (Yang et al., 2018). Later innovations, such as SARA-GAN, introduced attention mechanisms like self-attention into the upsampling path, allowing the model to capture long-range spatial dependencies (Yuan et al., 2020).

The use of attention mechanisms in GANs plays a crucial role. These modules dynamically reweight feature representations based on spatial or channel-wise relevance, allowing the network to focus on informative structures such as tissue boundaries, lesions, or coil sensitivity patterns. Channel attention often relies on global statistics like average or max pooling to reweight feature channels, while spatial attention emphasizes localized anatomical structures (Liu et al., 2023).

Despite their high perceptual fidelity, GANs can be challenging to train and are prone to hallucination artifacts. Careful design is needed to ensure diagnostic reliability. To address this, hybrid GAN models, such as RefineGAN, incorporate data consistency layers or cyclic loss terms that enforce agreement between the reconstructed and measured k-space data (Quan et al., 2018).

Transformer-Based Architectures

Transformers, originally developed for sequence modelling in natural language processing, have recently been adapted for medical imaging tasks. In MRI reconstruction, transformers offer the advantage of capturing long-range spatial dependencies via self-attention mechanisms (Huang et al., 2025), a capability that is often lacking in CNNs. Therefore, these models are frequently used in combination with CNNs to balance local feature extraction with global context modelling (Huang, Xing, et al., 2022).

For instance, SwinMRI introduces hierarchical Swin transformer blocks into a cascaded network, outperforming CNN architectures on several benchmark tasks (Huang, Fang, et al., 2022). Similarly, DSFormer uses dual-domain Swin transformer blocks trained in a self-supervised fashion, improving robustness across different sampling patterns and anatomies (Zhou et al., 2022).

The core innovation, self-attention, computes pairwise relationships across spatial locations, effectively enlarging the receptive field without increasing kernel size. Transformers are especially valuable for high-resolution MRI applications, though their computational demands remain a challenge.

Diffusion Models

Diffusion models represent an emerging class of generative models for MRI reconstruction. They define a forward process that progressively adds noise to an image and a reverse process that denoises the data step-by-step, guided by a learned score function (Huang et al., 2025). In MRI, diffusion models have been employed to reconstruct images by conditioning the reverse diffusion process on the measured k-space data. This enables sampling from a learned posterior distribution that respects both the image prior and the measurement constraints (Song et al., 2022).

Recent works demonstrate that diffusion models produce high-detail, artifact-free images in low-SNR and highly undersampled settings (Chung & Ye, 2022). Their probabilistic framework makes them particularly useful for uncertainty quantification, a crucial factor in clinical applications. However, long inference times and large data requirements remain significant challenges. Ongoing work aims to address these limitations through architectural and algorithmic refinements (Huang et al., 2025).

Hybrid networks

GrappaNet (Sriram, Zbontar, Murrell, Zitnick, et al., 2020) exemplifies a hybrid approach that tightly integrates classical GRAPPA-based parallel imaging with deep learning. The model incorporates the GRAPPA operator as a differentiable but fixed component embedded within a sequence of neural network transformations. The design of GrappaNet demonstrates that integrating classical parallel imaging with deep neural networks enables higher acceleration factors than either method can achieve independently.

The GrappaNet pipeline can be described as:

$$\mathbf{x} = h \circ f_2(\mathbf{G} * f_1(\mathbf{k})) \tag{2.28}$$

where $\mathbf{k}$ is the undersampled multi-coil k-space input, f_1 and f_2 are CNNs composed of two U-Nets operating in both k-space and image domains, $\mathbf{G}$ is the GRAPPA interpolation operator (calculated from ACS lines), and h denotes the coil combination and inverse Fourier transform operations done on the multi-coil k-space output

of f_2. Notably, both f_1 and f_2 include a sequence of operations: a U-Net in k-space, a hard data consistency step (re-inserting measured samples), a domain transform to image space (via inverse FT), another U-Net in image space, and a return to k-space via FT, followed by a second data consistency operation. Each U-Net maps n_c complex-valued input channels (corresponding to the number of coils) to $2n_c$ output channels, with convolution applied separately to real and imaginary components (Sriram, Zbontar, Murrell, Zitnick, et al., 2020).

A distinguishing feature of GrappaNet is that the network maintains all coil-specific data separately throughout the reconstruction process, unlike many previous approaches that combine coil views early into a single image. Instead, GrappaNet postpones coil combination until the final stage of the pipeline. This design choice allows the network to fully exploit the spatial diversity and complementary parallel imaging information across coils during learning. Additionally, the CNNs alternate between k-space and image-domain refinement, leveraging complementary domain-specific properties (Sriram, Zbontar, Murrell, Zitnick, et al., 2020).

The first subnetwork, f_1, serves a critical preprocessing role. It transforms the highly undersampled k-space data (acceleration R) into a pseudo-densified k-space with lower undersampling (e.g., R' = 2), making it more suitable for GRAPPA interpolation, which struggles with high acceleration factors. GRAPPA is then applied as a non-trainable, scan-specific operator that interpolates missing lines using coil-by-coil convolution based on calibration data. This operator is implemented as a differentiable layer to enable backpropagation during end-to-end training. The second subnetwork, f_2, follows a similar dual-domain structure and refines the interpolated k-space output to remove residual artifacts and enhance image quality. Finally, h converts the refined multi-coil k-space into a single-channel image using inverse FT and a coil combination method such as root-sum-of-squares (RSS). The model is trained end-to-end using an L1 image-domain loss, while the GRAPPA weights remain fixed and scan-specific. This hybrid structure combines the interpretability and reliability of GRAPPA with the expressive capacity of deep networks trained to improve both pre- and post-interpolation stages (Sriram, Zbontar, Murrell, Zitnick, et al., 2020).

DeepMRIRec (Alam et al., 2023), a second hybrid reconstruction network, follows a different philosophy. It takes as input a magnitude image reconstructed from undersampled multi-coil k-space data using conventional GRAPPA with ACS-based interpolation weights. Although GRAPPA ensures a physics-consistent reconstruction, its output often suffers from residual artifacts and reduced perceptual quality under high acceleration.

To address these limitations, DeepMRIRec applies a CNN with a U-Net-style encoder-decoder backbone and residual connections inspired by ResNet. The network includes batch normalization and is trained using a compound loss function combining L1 loss, SSIM, and perceptual loss based on VGG19 features. Extensive data

augmentation and coil compression are used to enhance generalization and reduce memory load.

Unlike GrappaNet, which learns in both k-space and image space with explicit data consistency steps, DeepMRIRec operates solely in the image domain and refines GRAPPA-based reconstructions post hoc. While it does not enforce explicit physical constraints, grounding the network on GRAPPA outputs provides a structured input that incorporates physical priors, which may support improved perceptual and anatomical fidelity (Alam et al., 2023).

In addition to GrappaNet and DeepMRIRec, a last hybrid model that integrates classical PI with deep learning is GRAPPA-GAN (Tavaf et al., 2021). In this architecture, GRAPPA is used to generate an initial reconstruction, which is then refined by a conditional GAN trained to enhance visual fidelity and suppress artifacts. This two-stage approach mirrors the structure of DeepMRIRec, though it relies on adversarial training rather than conventional loss terms. While the model demonstrates improved perceptual quality over GRAPPA, it inherits the stability challenges of GANs.

Dual-Domain

Finally, several architectures leverage the dual representation of MRI data in both k-space and image domains. These "Dual-domain" networks alternate between k-space and image-domain processing, enabling models to benefit from the sparsity and local smoothness in image space while exploiting the structured redundancy and calibration properties of k-space (Eo et al., 2018).

KIKI-Net exemplifies this approach by integrating an image-domain U-Net and a k-space CNN within each stage, connected by forward and inverse Fourier transforms. This hybrid design enables more effective artifact correction compared to single-domain models (Eo et al., 2018).

2.4.3 Training Paradigms: supervised vs self-supervised learning

An equally important consideration is the choice of training paradigm. Most deep learning models for MRI reconstruction are trained in a supervised fashion, using fully sampled images as ground truth. The loss function typically penalizes pixel-wise errors (e.g., L1 or L2), structural dissimilarity (SSIM), or perceptual differences derived from pretrained feature extractors (Mardani et al., 2019; Yang et al., 2018). Supervised training remains the dominant approach due to its effectiveness and straightforward formulation.

However, supervised methods face practical challenges. High-quality, fully sampled MRI datasets are expensive and time-consuming to acquire, and may not be available for all anatomies, contrasts, or scanner configurations. Moreover, models

trained on specific sampling masks or anatomical regions may generalize poorly to other settings.

To mitigate these limitations, a number of recent works have explored self-supervised or unsupervised training schemes. One popular strategy is k-space splitting, wherein the available undersampled data is partitioned into two disjoint subsets: one used to reconstruct an image, the other serving as a pseudo-ground truth for loss computation (Zhou et al., 2022). For example, DSFormer uses such a masking strategy to enable self-supervised training while retaining high reconstruction quality (Zhou et al., 2022).

Another notable direction is scan-specific training, where a neural network is trained or fine-tuned directly on the undersampled data from a single scan. This approach is exemplified by Deep Image Prior (DIP), which fits a randomly initialized CNN to one image only, exploiting the network's inductive bias to recover structure without external data (Ulyanov et al., 2020). While not competitive with supervised models in terms of peak performance, these methods offer robustness and adaptability, especially in settings where training data is scarce or heterogeneous.

While untrained and self-supervised approaches hold promise for data-efficient and generalizable models, they currently lag behind state-of-the-art supervised methods in performance.

2.4.4 Overview and comparative strengths

The rapid evolution of deep learning techniques for MRI reconstruction has resulted in a diverse ecosystem of models, each tailored to address the fundamental challenges of aliasing suppression, data fidelity, and high-fidelity image restoration. While the various model classes and architectural strategies discussed above offer different advantages, they also embody important trade-offs in terms of interpretability, computational efficiency, robustness, and clinical applicability.

CNN-based models, particularly U-Nets and their residual variants, remain the workhorse of MRI reconstruction due to their simplicity, fast inference times, and competitive performance across a variety of benchmarks. They are particularly well suited for scenarios where large, well-curated training datasets are available, and acquisition protocols are relatively consistent (Ronneberger et al., 2015; Zbontar et al., 2019). However, their reliance on local receptive fields can limit their ability to recover globally coherent structures, especially at high undersampling rates (Huang, Fang, et al., 2022; Zhou et al., 2022).

Generative models such as GANs and diffusion models bring significant benefits in perceptual quality. They are uniquely capable of producing anatomically plausible and visually detailed reconstructions, even in severely undersampled conditions (Mardani et al., 2019). GANs, in particular, are skilled at capturing high-frequency

textures, while diffusion models offer a probabilistic framework that can quantify uncertainty. However, these models are more prone to hallucinations. The absence of explicit data fidelity enforcement in pure GAN frameworks necessitates careful integration of consistency constraints to ensure diagnostic reliability (Huang et al., 2025). Training stability is another obstacle, GANs and diffusion models require careful tuning of loss functions, optimization schedules, and regularization techniques to ensure convergence. The computational cost of training and utilization also limits their accessibility in resource-constrained environments.

Transformer-based models are gaining traction for their capacity to model long-range dependencies, with emerging evidence suggesting their superiority in handling complex anatomical variations and multi-contrast settings (Huang, Fang, et al., 2022; Zhou et al., 2022). Nonetheless, their computational demands are substantial, and their performance often depends heavily on large-scale training and architectural tuning.

Most of the above-mentioned architectures are considered purely data-driven models. One significant drawback is their tendency to function as "black boxes," making it challenging to diagnose reconstruction errors or understand failure mechanisms. This lack of interpretability remains a key concern, especially in contexts requiring regulatory approval and clinical adoption, where explainable AI is generally more trusted.

Unrolled optimization networks, such as MoDL and Variational Networks (Aggarwal et al., 2019; Hammernik et al., 2018), offer a compelling alternative by embedding domain knowledge directly into the architecture. This way, they provide more transparency, than "black-box" models. Their iterative framework enforces stricter data consistency and often excels in limited-data or domain-shifted scenarios (Wang et al., 2021).

Hybrid methods represent a promising compromise. They combine the generalizability and expressivity of learning-based models with the structure and reliability of physics-based approaches. Architectures like GrappaNet demonstrate that embedding known operators within learnable systems can significantly enhance performance and robustness without sacrificing interpretability (Sriram, Zbontar, Murrell, Zitnick, et al., 2020).

2.5 Generalization in MRI Reconstruction

In the context of deep learning for MRI reconstruction, generalization refers to a model's ability to perform reliably on data that differ from those seen during training. Unlike controlled benchmark settings, clinical MRI scans are inherently heterogeneous, encompassing variations in anatomy, scanner hardware, field strength, noise levels, and acquisition protocols. A particularly critical challenge is the pres-

ence of undersampled k-space data, which is central to accelerating MRI but introduces significant aliasing artifacts. Since the hybrid models developed in this work are specifically designed to reconstruct from heavily undersampled data, robust generalization to diverse patient anatomies and acquisition settings under such undersampling is essential for clinical applicability (Heckel et al., 2024; Huang et al., 2025).

Despite considerable progress in DL-based MRI reconstruction, studies have demonstrated that even high-performing models struggle to maintain accuracy across distribution shifts, such as moving from knee to brain scans or changing the contrast or undersampling factor (Darestani et al., 2021; Hammernik et al., 2021; Huang, Wang, et al., 2022; Knoll et al., 2019). Even the top-performing models in the fastMRI challenge (Zbontar et al., 2019) struggled with generalization, as analyzed in (P. M. Johnson et al., 2021). This variability undermines the clinical viability of models that may excel on one dataset but fail in real-world, cross-domain scenarios. Therefore, understanding and improving generalization is essential for advancing DL models from research prototypes to reliable clinical tools.

2.5.1 Challenges in Achieving Generalization

The principal challenge in generalization arises from the high heterogeneity of MRI data. Differences in scanner manufacturers (e.g., Siemens, GE, Philips), field strengths (e.g., 1.5T, 3T, or 7T), imaging sequences (e.g., T1-weighted, T2-weighted, or fluid-attenuated inversion recovery [FLAIR]), noise levels, and patient-specific anatomy can drastically alter image characteristics (Huang et al., 2025). Motion artifacts and physiological fluctuations vary across anatomical regions. For instance, brain imaging is particularly sensitive to patient motion and physiological processes such as respiration and cardiac pulsation, whereas knee scans often suffer from mechanical joint movement. In addition, variations in FOV and matrix size influence spatial resolution and sampling in k-space, thereby affecting image dimensions and aliasing behaviour. All these factors introduce domain shifts, where the test data distribution diverges from the training distribution, leading to performance degradation (Heckel et al., 2024).

Undersampling schemes and noise levels also vary widely in clinical settings. A model trained exclusively on Cartesian undersampling patterns may struggle when applied to different k-space trajectories, such as radial or spiral, which have distinct sampling characteristics and aliasing patterns. Similarly, networks trained on low-noise data may fail to generalize in the presence of higher noise levels commonly encountered in fast or low-SNR acquisitions (Fujita et al., 2024).

Finally, models also face limitations due to training data scarcity. MRI reconstruction networks typically require large volumes of high-quality, fully sampled data. However, access to such datasets remains limited due to privacy concerns, acquisition time, and storage constraints. This data bottleneck constrains the diversity of

training sets, further hampering generalization.

2.5.2 Strategies to Improve Generalization

A number of strategies have emerged to enhance the generalizability of DL-based MRI reconstruction models. These include diverse training data, realistic data augmentation, transfer learning, self-supervised learning and architectural innovations.

Diverse Training Data

Training on heterogeneous datasets improves robustness by exposing models to a broad range of imaging conditions. Studies show that networks trained on mixed-domain data—e.g., multiple anatomies, contrasts, and acceleration factors—perform significantly better on cross-domain tests (Fujita et al., 2024; Knoll et al., 2019). However, acquiring such comprehensive datasets remains a logistical and ethical challenge, especially across institutions.

Data Augmentation

Realistic data augmentation can simulate variability not present in limited datasets. Conventional augmentations such as rotation or intensity scaling are often insufficient due to the physics-based nature of MRI. Therefore, physics-aware augmentations—like simulating motion artifacts, variable undersampling masks, or noise injection—are more effective in promoting generalization (Hammernik et al., 2022). These augmentations also help prevent overfitting to specific acquisition patterns, a common failure mode in DL reconstruction models.

Transfer Learning

Transfer learning (TL) has been increasingly explored as a means to reduce data requirements and boost generalization. By pretraining on large datasets and fine-tuning on domain-specific data, TL can adapt models to new anatomical structures, scanner types, or acquisition settings (Hossain et al., 2024). For example, models trained on 1.5T brain data have been successfully adapted to 3T scanners or cardiac anatomy using TL with minimal retraining time and data (Singh et al., 2023).

One of the earliest examples of TL success is AUTOMAP (Zhu et al., 2018), which was trained on natural images but demonstrated strong generalization to MR image reconstruction. Despite being trained on nonmedical data, AUTOMAP's domain-agnostic feature learning enabled it to reconstruct (low-resolution) MR images effectively, highlighting the potential of transferable representations for medical imaging tasks.

Self-Supervised Learning

In traditional supervised MRI reconstruction, neural networks are trained using paired data: undersampled k-space inputs and fully sampled, ground truth images

as targets. However, acquiring fully sampled data is often impractical or impossible in many clinical settings.

When fully sampled ground truth data are unavailable or scarce in the target domain, self-supervised learning techniques offer an effective alternative by leveraging the intrinsic structure of the undersampled k-space data itself. A notable example is SSDU (Self-Supervised Deep Learning via Data Undersampling) (Yaman et al., 2020), which eliminates the need for fully sampled reference images by exploiting the redundancy within the acquired undersampled data itself. The key idea is to partition the undersampled k-space data for each training example into two disjoint subsets: an input subset and a loss subset. The input subset, representing a more aggressively undersampled version of the original data, is fed into the network. The model reconstructs an image based solely on this partial information. Meanwhile, the withheld loss subset remains unseen by the network during reconstruction and is used exclusively to compute the training loss by comparing the predicted k-space values against these held-out measurements. This approach enables the network to learn to reconstruct missing k-space data by enforcing data consistency on the unseen loss subset, effectively providing a self-supervised training signal. Consequently, SSDU eliminates the need for fully sampled ground truth images during training, relying instead on the internal consistency within the undersampled data itself.

These self-supervised approaches offer several advantages: they reduce overfitting to specific training datasets, enable training in scenarios where fully sampled data acquisition is infeasible, and improve robustness to domain shifts by learning directly from the acquired data's statistical properties. Benchmark studies have shown that SSDU and related methods can achieve performance comparable to or exceeding that of supervised models in cross-domain evaluations (Singh et al., 2023; Yaman et al., 2020).

Architectural Design

Architectural choices play a critical role in robustness. Physics-informed models, such as VarNet and MoDL, incorporate domain knowledge by embedding data consistency layers and optimization-inspired unrolling, improving stability under distribution shifts (Hammernik et al., 2018).

The unrolled optimization-based model, VarNet, has demonstrated superior robustness across contrast, SNR, and sampling pattern variations. Its integration of data consistency steps helps it maintain performance even under domain shifts (Knoll et al., 2019). MoDL, which blends iterative model-based reconstruction with CNN priors, effectively balances flexibility and constraint, enabling it to generalize across sampling patterns and noise levels (Aggarwal et al., 2019).

Recurrent Inference Machines (RIMs) extend this principle by emulating iterative

inference. They update reconstruction predictions over several recurrent steps, refining the output based on learned dynamics. Lonning et al. (Lønning et al., 2019) showed that RIMs generalize well across different anatomies, field strengths, and resolutions, including real-world under-sampled data, due to their recursive refinement mechanism.

Physically-primed DNNs (Avidan & Freiman, 2023) embed the undersampling mask directly into the network architecture, helping the model adapt to acquisition variability. This design significantly boosts robustness across domains by explicitly conditioning the network on sampling characteristics.

Score-based generative models (Jalal et al., 2021) offer an entirely different perspective by modelling the image distribution itself rather than direct mappings from k-space to image space. These models sample from a learned posterior using Langevin dynamics, achieving high-fidelity reconstructions with built-in uncertainty quantification. Crucially, they decouple the reconstruction from the measurement process, enabling domain-agnostic performance across anatomies and sampling schemes.

This trend of embedding physical knowledge into DL models underscores a key development in MRI reconstruction. These type of architectures demonstrate superior generalization performance. The model presented in this thesis builds upon this insight, integrating physics-based priors (in the form of classical reconstruction techniques) to enhance robustness and reliability across heterogeneous MRI conditions.

2.6 Conclusion

This chapter provided an overview of the progression from classical MRI reconstruction techniques to modern deep learning-based approaches. Traditional methods such as GRAPPA, SENSE, and CS have established powerful frameworks for accelerating MRI acquisition by exploiting coil sensitivity profiles and sparsity priors. However, their reliance on handcrafted priors and iterative solvers limits both their reconstruction quality, especially in challenging undersampling regimes, and their clinical applicability due to often lengthy reconstruction times, which remain a significant bottleneck in routine practice.

Deep learning methods have emerged as a promising alternative, learning complex data-driven priors from examples and enabling end-to-end mappings from undersampled k-space to high-quality image reconstructions. While these methods often requiring substantial offline training times, their inference speed is typically much faster than classical iterative solvers. This rapid reconstruction capability holds great potential to alleviate clinical bottlenecks associated with delayed image availability.

Architectures such as convolutional neural networks, unrolled optimization methods, GANs, transformers, and diffusion models have shown impressive results on standardized datasets. Nevertheless, each class of models introduces unique assumptions and trade-offs, balancing performance, interpretability, and clinical applicability.

Among these, hybrid deep learning models, which integrate classical reconstruction operators with neural networks, have emerged as a novel but relatively unexplored approach. Only a few hybrid models exist (GrappaNet, DeepMRIRec, and GRAPPA-GAN) that embed classical parallel imaging operators directly into deep networks. These methods combine the robustness and inductive bias of parallel imaging with the expressivity of deep networks, achieving strong reconstruction results even at high acceleration factors.

Despite their theoretical appeal and promising early results, the generalization capabilities of hybrid models remain largely unexplored. Models trained on a specific dataset or scanner protocol may fail to generalize to new clinical settings. Existing studies of hybrid models evaluate on narrow, single-domain datasets with limited anatomical regions (e.g., knee or brain), fixed contrasts (e.g., T1-weighted), and uniform acquisition protocols. Consequently, it remains unclear how hybrid models perform under domain shifts, including variations in anatomy, contrast, scanner hardware, field strength, and other acquisition parameters. This open question is particularly relevant for clinical translation, where variability in imaging conditions necessitates models that can robustly adapt without retraining.

Chapter 3

Methods

This chapter presents the methodological framework for developing and evaluating a hybrid MRI reconstruction pipeline that combines classical compressed sensing with deep learning. The goal is to leverage the strengths of both approaches to enhance image quality in accelerated MRI.

It begins by defining data requirements and explaining the choice of the fastMRI dataset, which provides clinically relevant multi-coil k-space data across diverse anatomical regions, contrast types, and acquisition protocols. This ensures the model is trained and tested under heterogeneous conditions, improving generalization.

Next, it details the knee and brain data subsets, ground truth image generation, and undersampling strategies. The dataset is split to maintain anatomical balance and variability across training, validation, and test sets.

The second part outlines the hybrid model architecture, covering preprocessing, compressed sensing reconstruction via the BART toolbox, and the design and training of the U-Net used for refinement. Together, these components form a modular and extensible pipeline for accelerated MRI reconstruction.

Finally, the chapter describes evaluation metrics used to assess reconstruction quality, incorporating both classical pixel-based measures and perceptually motivated criteria.

3.1 Dataset

In order to develop and evaluate a hybrid MRI reconstruction pipeline that combines classical compressed sensing with deep learning, the choice of dataset is critical. Several key requirements must be met to ensure both the clinical relevance and technical feasibility of the approach.

First, the dataset must contain a sufficient volume of training data. Deep learning models, particularly convolutional neural networks, require large and diverse datasets to generalize effectively and avoid overfitting. This is especially important in medical imaging, where anatomical variability, scanner differences, and acquisition protocols can significantly affect model performance.

Second, the dataset must include raw k-space data. Unlike image-based datasets, raw k-space data allows for the realistic simulation of undersampling and the application of classical reconstruction techniques such as compressed sensing. This ensures that the reconstruction pipeline can be evaluated in a manner that closely mirrors real-world clinical workflows, where raw data is the starting point for image formation.

Third, the dataset must support parallel imaging, which requires multi-coil acquisitions. As discussed in Chapter 2, classical reconstruction methods rely on spatial encoding provided by multiple receiver coils. Without multi-coil data, these classical methods, and by extension hybrid models that incorporate them, cannot be applied.

The NYU fastMRI dataset, developed by Facebook AI Research and NYU Langone Health, is currently the largest publicly available dataset that satisfies all of these criteria. It provides raw, multi-coil k-space data for a variety of anatomies and contrasts, along with corresponding ground-truth images reconstructed from fully sampled acquisitions. The dataset is specifically designed to support research in accelerated MRI and has become a widely adopted benchmark in the field. This data can be obtained from the NYU fastMRI Initiative database (fastmri.med.nyu.edu) (Knoll et al., 2020)(Zbontar et al., 2019). Note that NYU fastMRI investigators provided the data but did not participate in the analysis or writing of this work.

An important motivation for selecting the fastMRI dataset is its high degree of heterogeneity, which is central to the goals of this study. This heterogeneity is expressed across multiple dimensions, including anatomy, contrast type, scanner hardware, field strength, and acquisition protocol. Specifically, the knee and brain subsets include data from different anatomical regions, acquired with different sequence parameters, collected on both 1.5T and 3T scanners from various Siemens models. The number of receiver coils ranges from 2 to 28, and matrix sizes differ significantly across scans, especially in the brain dataset. This diversity reflects real-world clinical variability and makes the dataset particularly well-suited for evaluating the generalization performance of MRI reconstruction models trained on heterogeneous

data. By explicitly preserving this heterogeneity across the training, validation, and test splits, this work aims to assess the robustness and adaptability of the proposed hybrid reconstruction pipeline.

3.1.1 The NYU fastMRI dataset

The fastMRI dataset is structured into four anatomical sub-datasets: knee, brain, prostate, and breast MRI. Specifically, it includes over 1,500 fully sampled knee scans, 6,970 brain scans, 312 prostate exams, and 300 breast exams with raw k-space data. Each sub-dataset contains four types of data: raw multi-coil k-space measurements, emulated single-coil k-space data derived from the multi-coil acquisitions, ground-truth images reconstructed using the root-sum-of-squares method, and DICOM images. All data have been thoroughly de-identified. Raw data were converted to the ISMRMRD format, and DICOM images were processed using the RSNA clinical trial processor. Manual inspection was also performed to ensure the removal of any protected health information (PHI) (Zbontar et al., 2019).

The single-coil data and DICOM images, were excluded from this study. The single-coil data lacks the spatial encoding benefits of parallel imaging, while the DICOM images do not contain raw k-space data, rendering them unsuitable for reconstruction tasks. This study focuses on the knee and brain subsets, which offer the largest volumes of raw, multi-coil k-space data and are widely used in benchmark studies. Their specific characteristics are described in detail next.

The knee MRI subset of the fastMRI dataset consists of 1,594 fully sampled scans, each comprising approximately 36 slices on average. These scans were acquired using a conventional Cartesian 2D proton-density-weighted (PDW) turbo spin echo (TSE) protocol in the coronal plane. Two variations of the sequence are included: with and without fat suppression (PDFS and PD, respectively), with the dataset being nearly evenly split between the two (Zbontar et al., 2019).

The scans were acquired on four different Siemens clinical MRI systems. Three of these are 3T systems—Magnetom Skyra (663 scans), Prisma (83 scans), and Biograph mMR (153 scans)—accounting for a total of 899 scans. The remaining 695 scans were acquired on a 1.5T Magnetom Aera system. Although the timing parameters varied slightly between systems, the sequence parameters were matched as closely as possible. All scans used an echo train length of 4, a matrix size of 320 $\times$ 320, an in-plane resolution of 0.5 mm $\times$ 0.5 mm, and a slice thickness of 3 mm with no inter-slice gap. Repetition times (TR) ranged from 2200 to 3000 milliseconds, and echo times (TE) ranged from 27 to 34 milliseconds. (Zbontar et al., 2019).

For the purposes of training and validation, the fastMRI researchers gave an official dataset split. It is divided into 973 training volumes (484 PD, 489 PDFS) and 199 validation volumes (100 PD, 99 PDFS) (Zbontar et al., 2019). The test set, consisting of 118 volumes (59 PD, 59 PDFS), and an additional 196 volumes reserved for

a future image reconstruction challenge, were excluded from this study. These sets do not include fully sampled k-space data or ground-truth RSS reconstructions, and thus cannot be used for supervised learning or quantitative evaluation. An overview of the knee dataset is given in table 3.1. The brain MRI subset comprises 6,970 fully

Subset	**PD Volumes**	**PDFS Volumes**	**Total Volumes**
Train	484	489	973
Validation	100	99	199
Test	59	59	118 (excluded)
Challenge	–	–	196 (excluded)
Total	–	–	1,594

TABLE 3.1: Knee dataset split by sequence type

sampled axial 2D scans, each containing approximately 16 slices on average. This dataset was curated to ensure de-identification by including only axial acquisitions (Zbontar et al., 2019). The scans were collected across five clinical sites using 11 different Siemens MRI systems, with field strengths of either 1.5T or 3T. The number of receiver coils varied between 2 and 28, with 16-channel coils being the most common.

The dataset includes a diverse set of contrasts, namely T1-weighted, T2-weighted, FLAIR (fluid-attenuated inversion recovery), and T1-weighted post-contrast (T1 POST) (Zbontar et al., 2019). This diversity reflects a broad range of clinical imaging protocols and enhances the generalizability of models trained on this data. Unlike the knee dataset, the brain subset exhibits a wide variety of reconstruction matrix sizes, further increasing its heterogeneity.

The data are split into 4,469 training files, 1,378 validation files, and 558 test files. An additional 565 files are reserved for a future challenge. Unlike the knee dataset, the test set for the brain subset includes fully sampled k-space data and ground-truth reconstructions, and is therefore useful in this study (Zbontar et al., 2019). The challenge set, which lacks ground-truth data, is excluded. An overview of the brain dataset is visualized in table 3.2. For the multi-coil datasets used in this study,

Subset	**T1**	**T1 POST**	**T2**	**FLAIR**	**Total Files**
Train	498	949	2,678	344	4,469
Validation	169	287	815	107	1,378
Test	65	122	322	49	558
Challenge	–	–	–	–	565 (excluded)
Total	–	–	–	–	6,970

TABLE 3.2: Brain dataset split by contrast type

ground truth images are generated using the *root-sum-of-squares (RSS)* reconstruc-

tion method applied to the fully sampled k-space data. This approach is one of the most widely used coil combination techniques in clinical MRI (Zbontar et al., 2019). The RSS method involves two main steps. First, an inverse Fourier transform is applied to the k-space data from each individual coil:

$$\tilde{m}_i = \mathcal{F}^{-1}(y_i) \tag{3.1}$$

where y_i denotes the k-space data from the i-th coil, and $\tilde{m}_i$ is the corresponding image in the spatial domain. Second, the coil images are combined voxel-wise using the root-sum-of-squares formula:

$$\tilde{m}_{\text{rss}}(\mathbf{r}) = \left(\sum_{i=1}^{N_c} |\tilde{m}_i(\mathbf{r})|^2 \right)^{1/2} \tag{3.2}$$

where $\tilde{m}_{\text{rss}}(\mathbf{r})$ is the final magnitude image at spatial location $\mathbf{r}$, and N_c is the number of receiver coils.

To ensure consistency across the dataset and to compensate for oversampling in the frequency-encoding direction, all ground truth images are cropped to a central region of 320×320 pixels (Zbontar et al., 2019).

3.1.2 Dataset Split

Although the fastMRI dataset provides predefined training, validation, and test splits for both the brain and knee subsets, not all of this data was used in this study. The goal was to train the deep learning model on a balanced dataset, ensuring that the U-Net would not become biased toward either anatomical region. To achieve this, an equal number of image slices from the brain and knee datasets were selected, rather than an equal number of volumes. This decision was motivated by the fact that brain volumes contain, on average, 16 slices per scan, whereas knee volumes contain approximately 36 slices. Since the U-Net is trained on a per-slice basis with a batch size of one, balancing the number of slices was the most appropriate strategy.

All available knee volumes from the fastMRI training and validation sets were included, totalling 1,172 files and 41,877 slices. An equal number of brain slices were then selected to match this total. However, a challenge arose due to the absence of a predefined test set with ground truth for the knee data. To address this, a custom test set was constructed by partitioning the knee validation set. The proportions of the final train, validation, and test sets were chosen to mirror the original fastMRI splits: 70% for training, 20% for validation, and 10% for testing.

The selection process was performed slice-wise, while ensuring that only complete volumes were included to prevent data leakage. For the training set, the first 70% of the total knee slices were selected from the knee training set, and an equivalent number of brain slices were taken from the beginning of the brain training set. To

construct the validation set, the remaining knee slices from the training set were combined with the first portion of the knee validation set, until the desired number of slices (20% of the total) was reached. A matching number of brain slices was selected from the beginning of the brain validation set. The test set, representing the final 10% of slices, was created by selecting the last portion of the knee validation set and the first portion of the brain test set.

In total, the mixed training configuration consisted of 59,108 slices for training, 16,856 slices for validation, and 8,420 slices for testing, evenly split between brain and knee data. A detailed overview of the slice and volume counts in this setup is provided in Table 3.3. Note that the files were selected in such a way that the distribution of contrasts within each subset preserved the original proportions defined by the fastMRI dataset (as given in table 3.2 and table 3.1), thereby maintaining the heterogeneity of the data across all splits.

Subset	**Modality**	**Files**	**Slices**
Train	Knee	821	29,556
	Brain	1,846	29,552
Validation	Knee	234	8,424
	Brain	527	8,432
Test	Knee	117	4,212
	Brain	263	4,208
Total	**Knee**	**1,172**	**41,877**
	Brain	**2,636**	**41,856**

TABLE 3.3: **Overview of the dataset split used for the mixed brain–knee setup.** The table lists the number of volumes and slices for both brain and knee data in the training, validation, and test sets. This table does not include the purely brain-based configuration, which used the standard train and validation subset of the brain fastMRI dataset, see Table 3.2.

In addition to this mixed brain–knee dataset, a second training configuration was evaluated using a purely brain-based dataset to serve as a comparison in the generalization experiments presented in Chapter 4. For this setup, all available brain data of the fastMRI dataset is used: 4,469 brain volumes were used for training and 1,378 for validation. These volumes were selected directly from the fastMRI training and validation sets, preserving the original contrast distribution and also maintain an equal split between acceleration factors $R = 4$ and $R = 8$.

Importantly, both training configurations, mixed (brain+knee) and brain-only, were evaluated using the same test set described above. This design allows for a direct and fair comparison of model generalization across different anatomical domains.

3.1.3 Undersampling

To simulate accelerated MRI acquisitions, retrospective Cartesian undersampling was applied to the fully sampled k-space data from the fastMRI dataset. This process is critical for creating realistic training, validation, and test inputs for the deep learning reconstruction models. Undersampling was performed exclusively in the phase-encoding direction, preserving the frequency-encoding axis. This mirrors typical clinical practice and ensures compatibility with existing MRI system constraints.

Two types of undersampling masks were employed in this study: GRAPPA-style masks and CS-style masks, each serving a specific purpose within the hybrid reconstruction pipeline. Note that the same two-dimensional undersampling mask was applied uniformly across all slices in a given scan volume.

The GRAPPA-style masks are used primarily to extract an auto-calibration signal (ACS) region, which is required to estimate coil sensitivity maps using the ESPIRiT algorithm during the preprocessing stage of the reconstruction pipeline. A more detailed explanation of this step and the role of the ESPIRiT algorithm is provided in Section 3.2.1. To construct a GRAPPA-style mask, a fully sampled central region is included by retaining a fixed percentage of the lowest spatial frequency k-space lines: specifically, 8% of lines are preserved for $R = 4$, and 4% for $R = 8$. The remaining k-space lines are then selected equidistantly, starting from a random offset, in a manner that meets the desired acceleration factor R. The choice of equidistant spacing is motivated by its ease of implementation on standard MRI hardware and its alignment with current clinical practices (Zbontar et al., 2019).

The second type, the CS-style undersampling masks, was used to generate the actual input data fed to the hybrid reconstruction pipeline. As discussed in Chapter 2, compressed sensing requires incoherent sampling, which is achieved using (variable-density) random undersampling patterns. However, unlike GRAPPA or SENSE, where the sampling strategies are well-established in clinical practice and literature, compressed sensing does not follow a universally standardized k-space sampling scheme. Therefore, representative CS patterns were manually selected from real-world scanner data acquired at UZ Leuven. This approach ensures that the applied sampling patterns simulate realistic, clinically relevant acquisition settings.

Each scan was randomly assigned an acceleration factor of either $R = 4$ or $R = 8$, with equal probability. These values are consistent with those used in the official fastMRI test set and are common benchmarks in the MRI reconstruction literature. Although the fastMRI baseline models are trained using only $R = 4$ data by default, a preliminary experiment in prior work showed that training on a mixture of $R = 4$ and $R = 8$ data leads to better performance in general (Vanhaverbeke, 2024). Based on those findings, the same strategy was adopted in this study. This design also aims to improve robustness and performance across varied acquisition scenarios. Figure 3.1 visualizes the k-space trajectories of CS and GRAPPA masks for both $R = 4$

and $R = 8$.

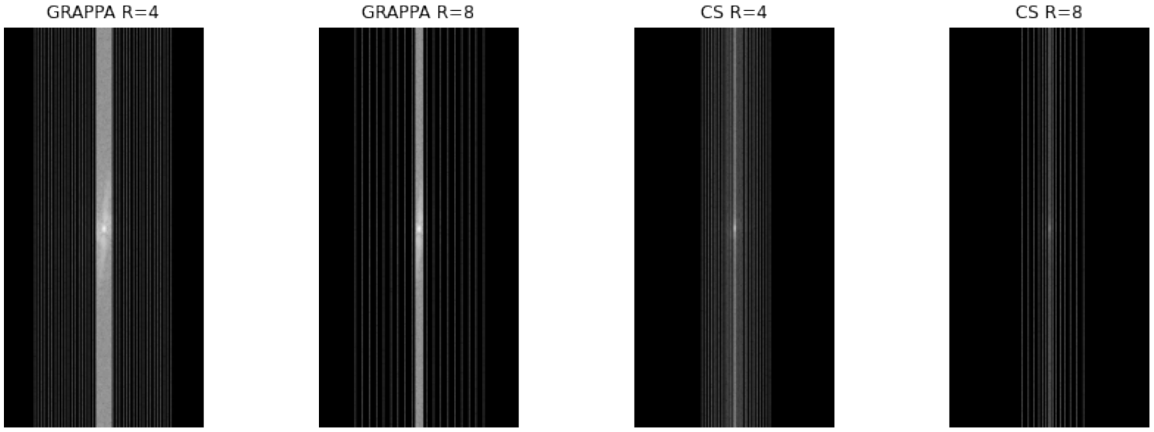


FIGURE 3.1: Undersampling masks used in this study for GRAPPA and CS at two acceleration factors ($R = 4$ and $R = 8$). All masks are applied uniformly across slices within a scan volume.

3.2 Hybrid model architecture

The reconstruction pipeline developed in this work follows a hybrid approach that combines classical compressed sensing (CS) with deep learning (DL), leveraging the strengths of both methodologies. This choice of CS as the classical component is motivated by its strong empirical synergy with DL-based refinement, as demonstrated in previous year (Vanhaverbeke, 2024). However, it is important to note that other reconstruction strategies, such as GRAPPA or SENSE, are also widely used in clinical settings and represent a relevant form of acquisition heterogeneity. To accommodate this, the pipeline is intentionally designed with modularity in mind: the classical reconstruction module can be readily replaced with an alternative technique, provided it outputs a coil-combined image in a compatible format. This flexibility makes the architecture adaptable to various clinical and research settings, and in principle, allows it to generalize across different acquisition strategies.

Formally, the model is structured as a composition of functions, where the final reconstructed image $\hat{x}$ is obtained by applying a classical reconstruction method f_1 to the undersampled k-space data y, followed by a convolutional neural network f_2 that refines the intermediate reconstruction:

$$\hat{x} = f_2 \circ p_2 \circ f_1 \circ p_1(y) \tag{3.3}$$

Here, p_1 and p_2 are preprocessing layers that ensure compatibility between the pipeline components. The intermediate output $\hat{x}_{intermediate} = f_1 \circ p_1(y)$ is a complex-

valued image, which is subsequently transformed by p_2 into a format suitable for the CNN f_2.

3.2.1 Preprocessing layer p_1

The first preprocessing stage p_1 prepares the undersampled k-space data for the classical CS reconstruction done in f_1. A key requirement of the CS reconstruction is accurate coil sensitivity maps S_i , which characterize the spatial sensitivity profiles of the multichannel receiver coils. This preprocessing layer thus calculates S_i and passes it on to f_1.

To estimate these sensitivity maps, the ESPIRiT algorithm (Uecker et al., 2014), implemented within the BART toolbox ("BART Toolbox", n.d.), is employed. ESPIRiT is chosen because it is a robust and widely used method that estimates coil sensitivities directly from the acquired k-space data, using only the ACS region. This eliminates the need for a separate calibration scan and makes ESPIRiT particularly practical for clinical applications.

Technically, ESPIRiT constructs a calibration matrix by applying a sliding window over the ACS region to extract overlapping k-space blocks, which are reshaped into rows. The resulting matrix has a block-Hankel structure, and a singular value decomposition (SVD) is performed to extract the dominant signal subspace. From this, ESPIRiT derives eigenvalue maps and corresponding eigenvectors, which represent the coil sensitivities. Only components associated with eigenvalues close to one are retained, ensuring that the maps are consistent with the measured data (Uecker et al., 2014).

In this work, the ACS region is extracted using a GRAPPA-style mask (see Subsection 3.1.3), since the provided mask from UZ Leuven does not include one. Because this approach involves retrospectively applying both a CS mask and a GRAPPA-style mask to fully sampled data, it is important to note that such a setup is only feasible in simulation. In real-world acquisitions, a single sampling pattern (typically a variable-density mask) is used that already includes an ACS suitable for ESPIRiT.

A useful feature of this design is that it is flexible: if reliable coil sensitivity maps are already available (for example, pre-calculated or provided by the vendor), the p_1 preprocessing step can be skipped completely. This makes it easy for the model to adapt to different research or clinical settings.

3.2.2 Classical Reconstruction f_1

The classical reconstruction component f_1 performs compressed sensing using the BART toolbox ("BART Toolbox", n.d.). The reconstruction minimizes an objective function of the form:

$$\hat{x}_{intermediate} = \arg\min_x ||Ex - y||_2^2 + \lambda||\Psi x||_1 \tag{3.4}$$

where E is the encoding operator incorporating coil sensitivities (computed in p_1) and undersampling, Ψ is a (sparsity promoting) wavelet transform, and $\lambda = 0.005$ is the regularisation parameter, controlling the trade-off between data fidelity and sparsity. The optimization is performed using an iterative algorithm with a maximum of 50 iterations, as recommended by the BART developers. The output of this stage is a coil-combined, complex-valued image $\hat{x}_{intermediate}$, which serves as the input to the second preprocessing layer p_2

Preprocessing Layer p_2

The second preprocessing stage p_2 transforms the intermediate reconstruction into a format compatible with the deep learning model. First, the image is cropped to match the dimensions of the ground-truth RSS images provided in the fastMRI dataset. This cropping step compensates for oversampling in the frequency-encoding direction, which often introduces background pixels with no anatomical content. Removing these pixels ensures that the deep learning model focuses on reconstructing relevant anatomical structures.

Next, the complex-valued image is converted to a real-valued format. This is necessary because the convolutional neural network f_2 operates on real-valued inputs. While this transformation discards phase information, it is justified by the fact that most diagnostic MRI analyses rely solely on signal magnitude, and the U-Net architecture used here is not designed to process complex-valued data.

Finally, the image is normalized using Z-score normalization at the slice level. This normalization strategy, which standardizes each slice to have zero mean and unit variance, is commonly used in MRI reconstruction tasks and has been shown to improve convergence during training. It is also consistent with the preprocessing pipeline used in the original fastMRI baseline models.

3.2.3 Deep Learning Component f_2

The final stage of the pipeline is a convolutional neural network f_2, which refines the intermediate reconstruction. Given the widespread adoption of U-Net architectures in MRI reconstruction literature, and designing a novel deep learning model is not the primary focus here, a previously developed U-Net model from last year's work was reused (Vanhaverbeke, 2024). This model is based on a modified version of the fastMRI baseline U-Net (Zbontar et al., 2019), which itself derives from the architecture originally proposed by (Ronneberger et al., 2015).

This choice allows focusing resources on integrating and evaluating the hybrid reconstruction pipeline rather than on neural network architecture design. Nonetheless, the selected U-Net is well-established and appropriate for the task. Its structure is described here for transparency and reproducibility (Vanhaverbeke, 2024).

The architecture is organized into two primary paths: a contracting (downsampling) path and an expansive (upsampling) path. The contracting path captures multiscale features by repeatedly applying convolutional operations with 3×3 kernels, each followed by instance normalization and Parametric Rectified Linear Unit (PReLU) activations, which enhance learning stability and non-linearity. Spatial resolution is reduced by max-pooling layers with 2×2 kernels and strides of 2, allowing the network to encode increasingly abstract representations.

Complementing this, the expansive path gradually restores the original spatial dimensions through bilinear upsampling by a factor of two at each stage. Critically, feature maps from corresponding levels in the contracting path are concatenated via skip connections, preserving spatial details lost during downsampling and enabling precise reconstruction of anatomical structures. The final layer applies 1×1 convolutions to condense the multichannel feature maps into a single-channel output, producing a real-valued magnitude image, aligned with the input dimensions of $p_2(\hat{x}_{intermediate})$.

In configuring this network, 32 initial feature channels were selected, with four downsampling and upsampling steps, balancing the need for model expressiveness against computational resources. This U-Net design offers a robust and efficient framework for enhancing the intermediate reconstructions generated by the classical CS step, facilitating improved image quality in the overall hybrid pipeline.

3.2.4 Training Procedure

The training setup for the deep learning component f_2 follows the approach established in previous work (Vanhaverbeke, 2024). An L1 loss function is used to minimize the mean absolute error between the predicted images and the ground truth. Additionally, L2 regularization (weight decay) with a penalty parameter $\lambda = 0.0005$ is applied to mitigate overfitting and improve numerical stability by preventing exploding gradients. The overall loss function optimized during training is given by:

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{N} \|x_i - \hat{x}_i\|_1 + \lambda \sum_{j=1}^{M} \theta_j^2, \tag{3.5}$$

where N is the number of training samples, M is the number of model parameters, x_i are the ground truth images ($\tilde{m}_{\text{rss}}(\mathbf{r})$, described in subsection 3.1.1), and $\hat{x}_i$ are the model predictions.

Optimization is performed using the RMSprop algorithm, selected based on its adaptive learning rate and favourable comparative performance. Training was conducted for 50 epochs with an initial learning rate of 0.001, which was decreased to 0.0001 after 40 epochs. A batch size of one (single slice per batch) was used. Deterministic seed initialization and PyTorch's default normalized weight initialization were employed to ensure reproducibility.

The model weights yielding the best validation performance were saved for final evaluation. Training was performed on an Asus GTX 1080 Ti GPU. Due to the integration of the computationally intensive classical reconstruction step, involving iterative optimization, training the hybrid pipeline required extended durations, typically at least one week.

3.3 Reconstruction Evaluation Metrics

Evaluating the fidelity of reconstructed MR images is a crucial aspect of assessing the effectiveness of deep learning models. Given the inherent trade-offs in MRI reconstruction, such as the balance between noise suppression, structural preservation, and perceptual quality, relying on a single metric may yield a biased or incomplete picture of performance. Therefore, a combination of complementary evaluation metrics is used: Normalized Mean Squared Error (NMSE), Peak Signal-to-Noise Ratio (PSNR), Multi-Scale Structural Similarity Index (MSSIM), and a perceptual loss based on feature differences from a pretrained VGG network (VGG loss).

Normalized Mean Squared Error (NMSE). The NMSE is defined as the squared ℓ_2 norm of the error between the reconstructed image $\hat{x}$ and the fully sampled reference image x, normalized by the energy of the reference image:

$$\mathrm{NMSE}(\hat{\mathbf{x}}, \mathbf{x}) = \frac{\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} \tag{3.6}$$

A lower NMSE indicates better reconstruction. NMSE serves as a primary indicator of reconstruction fidelity and is applicable both in the image and k-space domains. However, NMSE tends to favor overly smooth reconstructions, as large pixel-wise deviations are penalized more heavily than small ones. Consequently, NMSE may obscure the loss of fine anatomical detail, and should not be used in isolation (Alam et al., 2023; Pal & Rathi, 2022).

Peak Signal-to-Noise Ratio (PSNR). PSNR is a logarithmic measure of the peak signal intensity relative to reconstruction noise:

$$\mathrm{PSNR}(\hat{\mathbf{x}}, \mathbf{x}) = 20 \cdot \log_{10}\left(\frac{\mathrm{MAX}_{\mathbf{x}}}{\sqrt{\mathrm{MSE}(\hat{\mathbf{x}}, \mathbf{x})}}\right) \tag{3.7}$$

Here, $\mathrm{MAX}_{\mathbf{x}}$ denotes the maximum pixel intensity in $\mathbf{x}$, and $\mathrm{MSE}(\hat{\mathbf{x}}, \mathbf{x}) = \frac{1}{N}\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2$ is the mean squared error, therefore making it closely related to NMSE. Higher PSNR values indicate greater similarity. Although PSNR provides a more interpretable scale than NMSE, it inherits many of the same drawbacks, including sensitivity to global brightness and contrast variations (Alam et al., 2023; Pal & Rathi, 2022).

Structural Similarity Index Measure (SSIM and Mean SSIM). To better account for perceptual quality, the Structural Similarity Index (SSIM) compares local patterns of pixel intensities, incorporating luminance, contrast, and structural information. For images x and $\hat{x}$, SSIM is defined as (Zhou Wang et al., 2004):

$$\mathrm{SSIM}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{(2\mu_{\mathbf{x}}\mu_{\hat{\mathbf{x}}} + C_1)(2\sigma_{\mathbf{x}\hat{\mathbf{x}}} + C_2)}{(\mu_{\mathbf{x}}^2 + \mu_{\hat{\mathbf{x}}}^2 + C_1)(\sigma_{\mathbf{x}}^2 + \sigma_{\hat{\mathbf{x}}}^2 + C_2)} \tag{3.8}$$

where μ, σ, and $\sigma_{x\hat{x}}$ are local means, standard deviations, and covariance terms, and c_1, c_2 are small constants to stabilize division when denominators are near zero.

To evaluate the full image or volume, SSIM is computed over a sliding window across the image, and the Mean SSIM (MSSIM) is then reported as the average SSIM value across all windows. In this work, the mean SSIM (MSSIM) is computed by averaging local SSIM values over the image using a 11×11 Gaussian window with $\sigma = 1.5$, the default values for MSSIM index calculation.

SSIM is especially suited for MRI because it evaluates structural integrity rather than pure intensity differences. It captures localized artifacts, such as blurring or ghosting, that may be invisible to metrics like NMSE or PSNR. For instance, in undersampled MRI, reconstructions may achieve high PSNR while lacking fine anatomical detail. MSSIM mitigates this by directly assessing structure-preserving quality. While MSSIM is more aligned with human visual perception than NMSE or PSNR, it is not without limitations. It can underestimate edge distortions, show reduced sensitivity in high-intensity regions, and behave unstably in areas with low local variance (Alam et al., 2023). Despite these limitations, MSSIM remains a commonly used benchmark for structural fidelity.

VGG Loss. In addition to traditional pixel-wise metrics such as NMSE, PSNR, and MSSIM, perceptual loss functions have become increasingly popular for evaluating and training MRI reconstruction models. These losses measure differences between images not only at the pixel level but also at higher-level feature representations, which better capture perceptual and structural similarity relevant to human visual assessment.

The perceptual loss is typically computed using feature maps extracted from convolutional neural networks pretrained on large-scale image datasets. In particular, this work, based on (Alam et al., 2023), uses the first four convolutional blocks of the VGG-19 network to compute the Euclidean distance between the intermediate feature maps of the reconstructed image $\hat{x}$ and the reference image x:

$$\mathcal{L}_{\mathrm{VGG}}(\hat{\mathbf{x}}, \mathbf{x}) = \sum_{l} \|\boldsymbol{\phi}_l(\hat{\mathbf{x}}) - \boldsymbol{\phi}_l(\mathbf{x})\|_2^2 , \tag{3.9}$$

where $\boldsymbol{\phi}_l(\cdot)$ denotes the feature map obtained from the l-th layer of the pretrained network. where $\phi_i(\cdot)$ denotes the activation in the i-th block of the VGG

network. This perceptual loss emphasizes high-level texture and semantic information, enabling better characterization of subtle differences that may be visually significant but undetectable to NMSE or PSNR (J. Johnson et al., 2016; Simonyan & Zisserman, 2015).

Although all four metrics provide valuable information, no single one fully captures the quality of an MR reconstruction. NMSE and PSNR are sensitive to photometric changes and outliers, and may penalize reconstructions that preserve sharp edges or subtle structures (Alam et al., 2023; Pal & Rathi, 2022). MSSIM improves upon this by modelling perceptual features more directly, but still exhibits spatial insensitivity and potential artifacts at image borders. VGG loss complements the suite by reflecting higher-level perceptual similarity, but requires pre-trained CNNs and introduces dependency on the chosen network architecture and layer selection.

All evaluations employed an anatomical mask generated by the BART toolbox, which suppresses background regions (containing zero values) and focuses the metric computation on the imaged anatomy. NMSE and PSNR were computed at the volume level using 3D metrics, while MSSIM and VGG loss were calculated slice-wise in 2D and averaged over the volume. Given the 2D nature of fastMRI acquisitions, this approach is appropriate and consistent with prior literature (Zbontar et al., 2019).

3.4 Conclusion

This chapter has presented the complete methodological setup for implementing and evaluating a hybrid MRI reconstruction pipeline. The use of the fastMRI dataset ensures that the model is trained and tested on clinically relevant data with high inherent heterogeneity, including variations in anatomical regions, contrast types, scanner hardware, and acquisition protocols. By explicitly preserving this diversity across all dataset splits, the methodology is well-suited to investigate the model's ability to generalize under realistic clinical variability. The carefully balanced dataset split further prevents anatomical bias and supports fair performance comparisons.

The hybrid model architecture combines wavelet-based compressed sensing, implemented by the BART toolbox ("BART Toolbox", n.d.), with a U-Net-based deep learning module, connected through well-defined preprocessing stages. This design allows for flexible experimentation and future extensions, such as replacing individual components or adapting the pipeline to other anatomical regions.

A carefully designed training procedure, coupled with appropriate dataset balancing and varying acceleration factors, further reinforces the robustness of the approach. The inclusion of a diverse set of reconstruction evaluation metrics ensures that both objective fidelity and perceptual quality are accounted for, addressing known limitations of individual metrics in isolation.

The hybrid model outlined in this chapter forms the foundation for the experiments and analyses presented in Chapter 4.

Chapter 4

Results & Discussion: Model training and evaluation

This chapter presents the experimental results obtained from evaluating the hybrid MRI reconstruction pipeline described in Chapter 3. The primary objective of these experiments is to assess the model's ability to generalize across heterogeneous conditions, including variation in anatomical region, image contrast, and acceleration factor. Through systematic evaluation under increasingly diverse test scenarios, this chapter provides a comprehensive analysis of reconstruction fidelity, robustness, and clinical relevance.

To this end, a series of comparative experiments were conducted using three reconstruction approaches: (1) a classical compressed sensing baseline (CS), (2) a hybrid deep learning model trained exclusively on brain MRI data (further called: Hybrid-Brain), and (3) a hybrid deep learning model trained on a dataset containing both brain and knee MRI scans (further called: Hybrid-Multi). All models were evaluated using the same test sets and under consistent preprocessing and sampling conditions, enabling direct comparison across settings.

The remainder of this chapter is structured as follows. Section 4.1 provides a brief summary of the training results for the hybrid models. Section 4.2 presents the evaluation of the models through three distinct experiments, each designed to isolate a specific aspect of generalization. The experimental setup is first described, followed by a presentation of quantitative results and an integrated discussion of findings. Finally, Section 4.3 concludes the chapter with a synthesis of the key results.

4.1 Training of the Hybrid Models

The hybrid models were trained using the datasets and configuration described in Chapter 3. To briefly recall: the Hybrid-Brain model was trained on 4,469 brain volumes and validated on 1,378 brain volumes. The Hybrid-Multi model was trained on a mixed dataset comprising 1,846 brain volumes and 821 knee volumes, and validated on a corresponding set of 527 brain and 234 knee volumes. For both models, the training data included an equal distribution of acceleration factors $R = 4$ and $R = 8$ as well as all available contrast types. The same network architecture and training hyperparameters were applied to both models, including an L_1 loss function with L_2 regularization ($\lambda = 0.0005$), optimization via RMSprop, and a learning rate schedule that decreased from 10^{-3} to 10^{-4} after epoch 40.

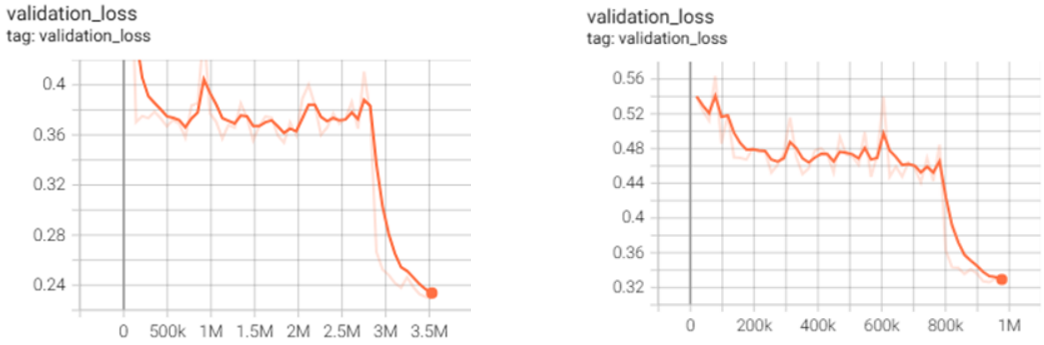
The classical reconstruction component (compressed sensing) was implemented using the BART toolbox with fixed parameters and was not updated during training. Only the deep learning component (the U-Net) was trained.

4.1.1 Results

Figure 4.1 shows the evolution of the validation loss for the Hybrid-Brain and Hybrid-Multi models over the course of training. These learning curves were obtained by computing the validation loss (as described above) on the entire validation set at the end of each training epoch. Both models exhibit a two-phase learning behaviour: an initial plateau during the first 40 epochs, followed by a sharp drop in validation loss that coincides with the first scheduled learning rate reduction. Despite differences in dataset size and anatomical diversity, the learning curves display comparable transitions in performance. The Hybrid-Brain model consistently achieves a slightly lower final validation loss compared to the Hybrid-Multi model.

However, both models converge to similar reconstruction performance as measured by NMSE, MSSIM, and PSNR, which were also computed on the validation set after each epoch. Plots for these metrics throughout training are shown in Appendix A. These reconstruction metrics improve analogously to the validation loss during training, and their final values are nearly identical for both models, indicating that the small difference in validation loss does not translate into a significant difference in perceptual or pixel-wise reconstruction accuracy on the validation set.

No signs of overfitting are observed: the validation loss continues to decrease or remain stable throughout training, indicating that the networks maintain good generalization to unseen validation data. The close alignment between the evolution of the reconstruction metrics and the validation loss further confirms the suitability of the chosen loss function and training procedure for optimizing perceptual and pixel-wise image quality.



(A) (B)

FIGURE 4.1: Validation loss curves for the Hybrid-Brain model (A) and Hybrid-Multi model (B) over the course of training. The loss was computed after each epoch on the full validation set using the combined ℓ_1 loss with ℓ_2 regularization ($\lambda = 0.0005$).

4.1.2 Discussion

Although both models initially show a slow improvement in performance, the sharp drop in validation loss following the first learning rate reduction suggests that a lower learning rate facilitates a second, more effective optimization phase. This two-phase learning behaviour likely reflects an initial stage where the networks capture broad anatomical features, followed by a fine-tuning phase in which more subtle structural details are learned.

Interestingly, while the Hybrid-Multi model consistently achieves a lower validation loss than the Hybrid-Brain model, this advantage does not translate into better reconstruction performance when evaluated using standard metrics. NMSE, MSSIM, and PSNR values on the validation set remain nearly identical across the two models. This discrepancy suggests that the ℓ_1 loss function with L_2 regularization, although useful for guiding optimization, may be sensitive to low-level signal differences that are not adequately captured by perceptual or structural similarity metrics.

Overall, the consistency in performance across models supports the robustness of the hybrid reconstruction approach. The alignment in learning dynamics across both models further underscores the effectiveness of the learning rate schedule in driving convergence. The slightly lower loss of the Hybrid-Multi model, despite comparable validation metrics, may reflect a tighter fit to the objective function rather than a meaningful gain in generalization.

4.2 Hybrid Model Evaluation

4.2.1 Experimental Setup

To fairly evaluate the reconstruction performance and generalization capacity of the proposed models, we designed a unified experimental setup across three distinct test-time evaluation scenarios. Each experiment compares three reconstruction approaches under identical input conditions, using standard evaluation metrics and statistical testing procedures.

Used models

All experiments evaluate the same three reconstruction methods. The first is a **classical CS** baseline, which performs a wavelet-regularized reconstruction using the BART toolbox. This method solves a sparsity-promoting optimization problem with a fixed regularization parameter ($\lambda = 0.005$) and a maximum of 50 iterations. The second method, **Hybrid-Brain**, is a hybrid deep learning model trained exclusively on brain MRI data. It consists of a classical CS reconstruction followed by a learned U-Net enhancement. The third model, **Hybrid-Multi**, uses the same hybrid architecture but is trained on a multi-domain dataset combining both brain and knee scans. Full architectural and training details of these models are provided in Section

Evaluation process

Quantitative comparison is based on standard image quality metrics, introduced in Section 3.3, including Peak Signal-to-Noise Ratio (PSNR), Mean Structural Similarity Index Measure (MSSIM), Normalized Mean Squared Error (NMSE), and the VGG perceptual loss. All metrics are computed at the volume level for analysis. To ensure that comparisons focus on meaningful image content, all metrics were computed over an anatomical mask, suppressing zero-pixels in the background region.

All models were evaluated on the same test sets to enable direct comparison. Each test set consists of fully sampled multi-coil k-space volumes that were retrospectively undersampled using the same Cartesian acceleration masks and sampling strategies described in Section 3.1.3. The full test set consisted of 380 volumes in total: 263 brain volumes (4,208 slices) and 117 knee volumes (4,212 slices). Each anatomical subset included a representative mix of image contrasts and an equal distribution of acceleration factors R = 4 and R = 8. This ensured that evaluations captured a broad spectrum of clinical variability while maintaining balance across test conditions.

To determine the statistical significance of observed performance differences, paired tests were conducted on a per-volume basis across all evaluation metrics. First, the distribution of differences between paired samples was assessed for normality using the Shapiro-Wilk test. Due to consistent violations of the normality assumption, the

non-parametric Wilcoxon signed-rank test was appliedfor all pairwise model comparisons.The tests were conducted in a one-sided manner, reflecting the expected direction of improvement (e.g., increase in PSNR, decrease in NMSE). To correct for multiple comparisons across metrics and experiments, Bonferroni correction was applied to the resulting p-values. Statistical significance was defined at a corrected p-value threshold of 0.05.

Performed experiments

To systematically assess both in-domain performance and cross-domain generalization, three distinct experiments were conducted, each defined by the anatomical composition of its test set.

- Experiment 1 evaluates all models on brain data only, allowing us to assess how well the Hybrid-Brain model performs on its training domain, and whether Hybrid-Multi can match or exceed its performance despite training on a broader dataset.
- Experiment 2 evaluates all models on knee data only, which serves as an out-of-distribution test for Hybrid-Brain but an in-domain test for Hybrid-Multi. This comparison is used to investigate whether training on multiple anatomies improves generalization to new anatomies.
- Experiment 3 uses a balanced mix of brain and knee test volumes. It simulates a clinical setting in which anatomical variation is present at inference time, and evaluates how robustly each model generalizes to this diversity.

To support the quantitative findings reported in each experiment, representative reconstruction examples for all evaluated methods are included in Appendix C. These visualizations illustrate one slice per evaluation subset, selected around the median SSIM value (49th–51st percentile) across all three models. The appendix provides side-by-side comparisons of the ground truth, model reconstructions, and residual error maps, along with the corresponding quantitative metrics. These examples highlight differences in perceptual and structural reconstruction quality between models across varying anatomical regions and acceleration factors.

4.2.2 Experiment 1: Brain data evaluation

Experiment 1 evaluates model performance on brain MRI data only, providing an in-domain test for the Hybrid-Brain model. The goal is twofold: first, to determine whether Hybrid-Brain model improves upon the CS baseline; and second, to assess whether the Hybrid-Multi model can match or exceed the performance of a domain-specific model, or whether anatomical diversity during training introduces a performance trade-off in this setting.

Quantitative evaluation was conducted on 263 fully sampled brain MRI volumes, totalling 4,208 slices. The dataset included a balanced distribution of image contrasts and acceleration factors ($R = 4$ and $R = 8$), enabling a thorough analysis

of reconstruction fidelity under varying conditions. Table 4.1 reports the average image quality metrics obtained for each model across the entire brain test set.

The CS baseline yields the lowest reconstruction quality across all metrics. Hybrid-Brain improves substantially over CS, and Hybrid-Multi outperforms both. All pairwise differences between models were statistically significant across all metrics, with Bonferroni-corrected p-values below 0.001.

Model	SSIM	PSNR	NMSE	VGG
CS	0.755	24	0.259	0.11
Hybrid-Brain	0.873	26.4	0.113	0.0926
Hybrid-Multi	**0.904**	**31.1**	**0.0561**	**0.0885**

TABLE 4.1: Mean reconstruction metrics across brain test volumes (Experiment 1)

When analysing performance by acceleration factor, all models degraded at R=8, the relative degradation was steepest for the CS baseline and least severe for the Hybrid-Brain model. The performance gap between Hybrid-Multi and Hybrid-Brain narrowed. Nevertheless, Hybrid-Multi consistently retained superior absolute performance at both acceleration levels over the other 2 models ($p_{corrected} < 0.001$). Although VGG loss followed the same trend in global averages, differences between Hybrid-Multi and Hybrid-Brain at R = 8 were not statistically significant for this metric. A complete breakdown of metric performance per contrast and acceleration is included in Appendix B.

4.2.3 Experiment 2: Knee Data Evaluation

Experiment 2 evaluates the models on knee MRI data, a domain unseen by the Hybrid-Brain model during training and thus serving as an out-of-distribution test for this network. For the Hybrid-Multi model, knee data was part of the training domain, and the experiment thus constitutes an in-domain test. The objective is to investigate whether learned priors from brain data alone are sufficient for generalization to the structurally distinct knee anatomy, and to evaluate whether the inclusion of anatomically diverse training data enables more robust performance,

The knee test set comprised 117 volumes, totaling 4,212 slices, consisting of a balanced distribution of contrasts (Proton-Density weighted (PD) and Proton-Density weighted with Fat Suppression (PDFS)) and acceleration factors ($R = 4$ and $R = 8$). Table 4.2 presents the mean reconstruction metrics for the three models. The classical CS baseline achieves the lowest scores across all metrics. The Hybrid-Brain model, despite being trained exclusively on brain images, consistently outperformed the CS baseline. The Hybrid-Multi model, achieved the best performance overall. Its reconstructions are superior based on all metrics compared to both alternatives. Statistical testing confirmed that all pairwise differences between models were significant ($p < 0.001$) for SSIM, PSNR, and NMSE. For the VGG per-

ceptual loss, significance held for all comparisons except that between Hybrid-Brain and CS, where the difference did not reach the corrected significance threshold.

Model	SSIM	PSNR	NMSE	VGG
CS	0.687	22.7	0.256	0.0321
Hybrid-Brain	0.761	24.4	0.151	0.0314
Hybrid-Multi	**0.838**	**31**	**0.0421**	**0.0281**

TABLE 4.2: Mean reconstruction metrics across knee test volumes (experiment 2)

Contrast-specific analysis and breakdowns by acceleration factor further confirmed the superior performance of the Hybrid-Multi model across acquisition types and undersampling conditions. The Hybrid-Brain model, while consistently outperforming CS, showed limitations under stronger undersampling (R = 8) and greater anatomical deviation from its training data. Full breakdowns and tables are provided in Appendix B.

4.2.4 Experiment 3: Combined Evaluation on Brain + Knee

In Experiment 3, both brain and knee MRI data were combined to form a mixed-domain test set. This evaluation simulates a clinically realistic deployment scenario in which reconstruction models must generalize across multiple anatomical structures. The goal of this experiment is to assess model robustness in the face of anatomical heterogeneity at inference time, and to determine whether performance advantages observed in domain-specific tests translate to a mixed-anatomy setting.

The combined test set included 380 volumes, aggregating the brain and knee datasets from Experiments 1 and 2. Table 4.3 reports the mean reconstruction metrics computed across all volumes. The Hybrid-Multi model achieves the best overall reconstruction quality, outperforming both the CS baseline and the Hybrid-Brain model on all reported metrics. The Hybrid-Brain model improves substantially over CS but is outperformed by Hybrid-Multi, which benefits from training on diverse anatomies. Statistical testing confirms that the differences in all four metrics between all model pairs are highly significant ($p_{corrected} < 0.001$)

Model	SSIM	PSNR	NMSE	VGG
CS	0.734	23.6	0.258	0.0857
Hybrid-Brain	0.839	25.8	0.125	0.0737
Hybrid-Multi	**0.884**	**31.1**	**0.0518**	**0.0699**

TABLE 4.3: Mean reconstruction metrics across all brain & knee test volumes (Experiment 3)

A stratified analysis by acceleration factor is shown in Table 4.4. At R = 4, Hybrid-Multi exhibited particularly strong performance. In contrast, CS and

Hybrid-Brain lagged considerably behind. At R = 8, performance declined for all models, with the CS baseline suffering the most substantial degradation. The Hybrid-Brain model showed moderate resilience, but the Hybrid-Multi model again retained the best absolute performance. While nearly all differences were statistically significant, the difference in VGG-loss between CS and Hybrid-Brain at R = 4 was not, indicating greater variance in performance under milder undersampling.

R	Model	SSIM	PSNR	NMSE	VGG
4	CS	0.812	27.3	0.0837	0.0645
	Hybrid-Brain	0.875	27.6	0.0779	0.0634
	Hybrid-Multi	**0.926**	**35.2**	**0.0137**	**0.0577**
8	CS	0.65	19.6	0.444	0.108
	Hybrid-Brain	0.8	23.9	0.175	0.0847
	Hybrid-Multi	0.839	26.7	0.0923	0.0829

TABLE 4.4: Mean reconstruction metrics per acceleration factor (R=4 and R=8) for knee and brain data, averaged across acquisitions (Experiment 3)

Across all six evaluated contrasts (AXFLAIR, AXT1, AXT1POST, AXT2, PD, and PDFS), the Hybrid-Multi model consistently outperformed both the CS baseline and the Hybrid-Brain model across all quantitative metrics. While the VGG perceptual loss showed less variation across models, Hybrid-Multi achieved the lowest or near-lowest VGG scores in most contrasts, indicating good perceptual similarity. The Hybrid-Brain model performed better than CS across all contrasts, but its improvements were comparatively modest, particularly in non-brain contrasts such as PD and PDFS. A breakdown of reconstruction performance across different MRI contrast types at R=4 is included in B.

4.2.5 Discussion

This subsection provides an integrated interpretation of the results from all experiments, with a focus on understanding the behaviour of hybrid deep learning models across varying anatomical domains and acceleration factors. The findings are analysed in the context of domain-specific versus multi-domain training, performance consistency, and clinical applicability.

Generalization across anatomies

A central finding of this study is the strong generalization capability of the Hybrid-Multi model, which consistently outperformed both the domain-specific Hybrid-Brain model and the traditional CS baseline across all test settings: across anatomical domains, acceleration factors, and acquisition types. Notably, even in the brain-only test (Experiment 1), the Hybrid-Multi model achieved higher reconstruction quality than the Hybrid-Brain model, despite the latter being trained solely on brain data. This counterintuitive result suggests that training on heterogeneous data may

induce beneficial regularization, allowing the model to capture more general anatomical features and avoid overfitting to a narrow domain, improving performance even within a single domain.

In contrast, the Hybrid-Brain model exhibited clear limitations when evaluated on knee data, confirming that deep learning models trained exclusively on a single anatomy may struggle to generalize across anatomically distinct domains. While the Hybrid-Brain model did outperform the CS baseline in out-of-distribution settings, the performance gap relative to Hybrid-Multi was substantial and statistically robust.

The results across all three experiments also collectively underscore the consistent superiority of hybrid deep learning models over the classical compressed sensing (CS) baseline in accelerated MRI reconstruction.

Performance across Contrasts

MRI reconstruction models must contend with the heterogeneity introduced by different acquisition protocols and image contrasts, which affect signal intensity, texture, and anatomical boundaries. Robust performance across multiple contrasts is therefore essential for clinical generalization.

In Experiment 3 at $R = 4$, the Hybrid-Multi model consistently outperformed both Hybrid-Brain and the CS baseline across all six evaluated contrasts (AXFLAIR, AXT1, AXT1POST, AXT2, PD, and PDFS). These results suggest that training on a diverse set of contrasts enables the model to develop contrast-invariant representations or adaptively handle varying contrast-specific features, contributing to its superior generalization capability.

Performance under undersampling

As expected, all models showed reduced performance under higher undersampling (R = 8), consistent with the greater difficulty of recovering high-frequency information from more sparsely sampled k-space data. The CS baseline, where performance degraded sharply, was most affected. Interestingly, the Hybrid-Brain model showed the smallest relative performance drop between R = 4 and R = 8, particularly in the brain domain. This suggests that domain-specific training may provide stronger priors that are more resistant to undersampling-induced artifacts.

Nevertheless, the Hybrid-Multi model maintained the best absolute performance under both acceleration regimes. Its ability to reconstruct high-fidelity images under strong undersampling highlights the value of hybrid modelling with diverse training data, especially in real-world scenarios where acquisition conditions vary.

An important practical question arising from these results is the maximal accelera-

tion factor R that hybrid models can feasibly support while still producing artifact-free, diagnostically usable images. Classical CS-based reconstruction methods typically achieve acceptable quality only at modest acceleration (no greater than R=2 or R=3). In contrast, the present results indicate that hybrid models can reliably reconstruct images at higher acceleration factors. Reconstructions at R=4 using the Hybrid-Multi model appeared, in most cases, visually artifact-free and structurally faithful. While not all reconstructed volumes were subjected to detailed manual inspection, the observed absence of overt artifacts in many representative cases, combined with strong performance across multiple quantitative metrics, supports this conclusion.

At R=8, however, artifacts were commonly observed across test images. These artifacts included residual aliasing, subtle geometric distortion, and local intensity irregularities. Although hybrid models were clearly superior to CS under these conditions, the presence of visible imperfections at this acceleration rate suggests that R=8 exceeds the reliably usable limit for most clinical tasks.

Before claiming that R=4 is certainly viable for clinical application, careful evaluation beyond standard quantitative metrics is required. Although multiple evaluation metrics were used to provide a comprehensive assessment of reconstruction quality, these metrics may not fully capture diagnostic utility or subtle perceptual features relevant to radiological assessment. To evaluate the clinically meaningful fidelity of the reconstructed images, expert radiologist assessments and uncertainty quantification methods must be incorporated. Such approaches would bridge the gap between technical image quality and clinical utility, and are essential for regulatory approval and translational adoption.

Perceptual quality and metric sensitivity

While multiple image quality metrics were employed in this study to provide a comprehensive evaluation of reconstruction performance, the VGG perceptual loss demonstrated particular limitations in its ability to consistently reflect reconstruction fidelity across all settings. In general, trends in VGG loss followed those observed in conventional metrics such as SSIM, PSNR, and NMSE. However, its discriminative power diminished in more granular analyses. For example, in the comparison between Hybrid-Multi and Hybrid-Brain at the higher acceleration factor (R = 8), differences in VGG loss did not reach statistical significance despite clear differences in PSNR and NMSE. This suggests that perceptual loss does not always reflect structural or pixel-level discrepancies, particularly under challenging undersampling conditions.

A critical factor underlying this behaviour is the origin and nature of the VGG perceptual loss itself. The VGG loss is computed by measuring the Euclidean distance between feature representations extracted from intermediate layers of a pre-trained VGG19 convolutional neural network. While this approach captures high-level se-

mantic differences that may correlate with perceptual similarity in natural image contexts, its applicability to medical imaging—particularly MRI—is inherently limited. The VGG19 model was originally trained on the ImageNet dataset, which consists exclusively of natural images such as animals, vehicles, and landscapes. As a result, the features it encodes are tailored to object categories and textures typical of natural scenes, rather than to anatomical structures, tissue boundaries, or pathological variations present in MR images. Future research may benefit from developing perceptual loss functions based on feature extractors trained explicitly on medical datasets.

This observation is important for future work relying on perceptual similarity measures, as it underscores the need to combine multiple evaluation metrics when assessing reconstruction fidelity. It also highlights the need for further investigation into alternative, more trustworthy reconstruction metrics.

Implications for clinical deployment

Experiment 3, which combined brain and knee test data, provides the most realistic proxy for clinical deployment. The strong and consistent performance of the Hybrid-Multi model in this setting demonstrates its practical robustness and supports the hypothesis that hybrid architectures trained on heterogeneous anatomical data are better suited for real-world use. The inferior performance of the Hybrid-Brain model in this setting highlights the potential risks of domain-specific training when applied to diverse clinical populations.

The results thus advocate for the inclusion of anatomically diverse data in training pipelines for deep learning-based MRI reconstruction. By combining physics-based regularization (here with CS) with learned priors from multiple domains, hybrid models can achieve both high reconstruction quality and strong generalization. Clinically, deploying models trained on mixed anatomy data may reduce the need for retraining per protocol or scanner setting, streamlining integration into diverse imaging pipelines.

Limitations and Future work

While the presented experiments and analyses provide strong evidence in support of hybrid deep learning models trained on heterogeneous data, several limitations of the current study must be acknowledged. These limitations pertain to the scope of anatomical coverage, dataset characteristics, model architecture choices, and evaluation methodology.

First, although the Hybrid-Multi model was trained on multiple anatomical domains, the training dataset was limited to brain and knee MRI scans. As a result, the generalization claims made in this study are currently restricted to anatomies with at least partial representation in the training data. Consequently, further work

is necessary to determine whether the observed generalization behaviour persists in even more diverse or atypical anatomical settings, including cardiac, abdominal, or musculoskeletal regions beyond the knee. An immediate extension of the current work could be expanding the anatomical diversity of the training dataset. This would allow for a more comprehensive assessment of cross-domain generalization and may enhance the robustness of hybrid models in real-world clinical workflows. Furthermore, evaluating the models on pathological cases, rather than only healthy subjects or anatomical variability, would allow for understanding whether hybrid models can reconstruct diagnostically relevant features with high fidelity.

Second, although the dataset was balanced in terms of acceleration factors and included various contrast types, it was still derived from a retrospective reconstruction setting. All undersampling was performed synthetically on fully sampled data. While this is a common and necessary approach for controlled evaluation, it does not fully account for variability introduced by prospective undersampling in a clinical scanner environment. Artifacts introduced by real-world acquisition processes, such as gradient nonlinearities, coil miscalibration, motion, but also metal implants of patients, were not explicitly considered and may affect reconstruction performance.

Third, the deep learning components of both Hybrid-Brain and Hybrid-Multi were instantiated using a U-Net architecture. While this choice reflects a widely adopted and empirically effective baseline, it may not be optimal for all reconstruction tasks or domains. The hybrid framework itself is agnostic to the exact architecture of the learned module, and different network designs, such as attention mechanisms, transformer-based models, or architectures that incorporate complex-valued inputs by explicitly handling both real and imaginary parts of the MRI signal, may further improve generalization or robustness.

Finally, beyond architectural considerations, several key training and implementation choices were adopted for simplicity and comparability, rather than for optimality. The training protocol for the U-Net, including the number of epochs, learning rate, and batch size, followed standard values reported in prior work but were not tailored or tuned to the specific data characteristics or reconstruction objectives of this study. Additionally, the L2 regularization term used in the loss function was fixed at a commonly used value and was not subjected to systematic hyperparameter tuning. As such, it is possible that alternative regularization strategies or parameter settings could have yielded superior reconstructions.

On the classical reconstruction side, the compressed sensing component implemented via the BART toolbox was also treated as a fixed preprocessing step. The regularization parameter (λ) and number of iterations were set to default values recommended in the BART documentation and were not calibrated for the particular datasets used in this work. While this choice enabled consistency and reproducibility, it may not represent the best achievable performance of the classical CS baseline, nor the optimal integration with the subsequent deep learning stages.

More fundamentally, the current pipeline treats classical CS reconstruction as a sequential and decoupled module. This makes it easy to swap out with other classical techniques, making this architecture more versatile in different clinical settings. However, One compelling future direction would be to reframe the entire reconstruction process as a single, trainable end-to-end system. In such a design, not only the weights of the U-Net but also the parameters of the classical CS stage, such as the sparsity threshold, regularization strength, and iterative update strategy, could be learned jointly during training. Even preprocessing steps, such as normalization, coil sensitivity estimation, or k-space regridding, could potentially be integrated and made differentiable. This holistic approach could allow the entire pipeline to co-adapt its components, thereby achieving more globally optimal reconstruction quality and robustness. The results presented here should therefore be seen as a lower bound on the performance that could be achieved by future, fully optimized hybrid frameworks.

4.3 Conclusion

The experimental results presented in this chapter provide comprehensive evidence for the superior reconstruction performance and generalization capacity of the Hybrid-Multi model compared to both a classical compressed sensing baseline and a domain-specific hybrid model trained only on brain data. Across all evaluated experiments, Experiment 1 (brain-only), Experiment 2 (knee-only), and Experiment 3 (mixed brain and knee data), the Hybrid-Multi model consistently achieved the highest reconstruction fidelity, as measured by SSIM, PSNR, NMSE, and VGG perceptual loss. These improvements were statistically significant in nearly all comparisons and were robust across anatomical regions, acquisition contrasts, and acceleration factors.

Notably, the Hybrid-Multi model outperformed the Hybrid-Brain model even within the brain domain, despite the latter being trained exclusively on brain data. This indicates that anatomical diversity in training not only supports cross-domain generalization, but may also enhance in-domain performance through more comprehensive feature learning. Furthermore, the model maintained strong reconstruction quality up to an acceleration factor of R=4, with most images appearing artifact-free. In contrast, reconstructions at R=8 still exhibited residual artifacts, suggesting that while hybrid deep learning enables higher acceleration than traditional methods, a practical upper limit for routine clinical use lies around R=4.

Together, these findings establish that hybrid models trained on heterogeneous data can reconstruct high-quality MR images across a broad spectrum of clinical conditions. The results emphasize the value of integrating both physics-based and data-driven priors, and highlight training diversity, not just architectural complexity, as a key determinant of reconstruction performance and generalization in accelerated

MRI.

Chapter 5

Conclusion

This work evaluated a CS-based hybrid MRI reconstruction model in the face of realistic clinical heterogeneity. The experiments consistently found that hybrid deep-learning models outperformed the classical CS baseline in all scenarios. Notably, a multi-domain hybrid model (trained jointly on brain and knee data) achieved the best generalization. It reconstructed both brain and knee images with higher fidelity than a domain-specific hybrid model, even when tested on brain-only data. In fact, training on mixed anatomies did not hurt performance on any single domain; rather it improved robustness. For example, at 4× acceleration the Hybrid-Multi model produced visually clean, artifact-free images in both brain and knee, whereas the brain-only hybrid model left more residual errors when applied to knee data. Across different contrasts and acceleration factors, the Hybrid-Multi model showed superior quantitative metrics (e.g. PSNR, SSIM) and consistently higher image quality.

The study also revealed limitations. All models' performance dropped under extreme undersampling (8×), where residual aliasing and distortions became visible in the reconstructions. The CS baseline in particular degraded sharply at high acceleration. These findings suggest a practical limit around R=4 for fully reliable reconstructions under the used settings; higher accelerations may require further innovations. From a methodological standpoint, the controlled experiments demonstrated that explicitly preserving data heterogeneity during training (multiple anatomies, contrasts, field strengths) is beneficial. The carefully balanced dataset split and diverse sampling regime were effective in testing model robustness. Nevertheless, this work used retrospectively undersampled data; prospective subsampling and real scanner imperfections (motion, field variations, hardware differences) were not fully modelled and could affect performance in deployment.

Critically assessing the methodology, the main components were appropriate: the choice of the fastMRI raw dataset provided realistic multi-coil k-space data across anatomies, and the wavelet+U-Net hybrid architecture offered a flexible pipeline. The use of a known toolkit (BART) for the CS step ensured reproducibility. However, using a single U-Net for refinement, while a standard baseline, may not exploit

the latest architectures (e.g. transformer-based or diffusion models) that could further improve results. The evaluation relied on classical image quality metrics; no reader study or clinical scoring was performed, which limits conclusions about diagnostic utility. Also, while we tested domain shifts in anatomy and contrast, other shifts (different scanners, field strengths, patient populations) remain to be explored.

Critically assessing the methodology, the main components were appropriate: the choice of the fastMRI raw dataset provided realistic multi-coil k-space data across anatomies, and the wavelet+U-Net hybrid architecture offered a flexible pipeline. The use of a known toolkit (BART) for the CS step ensured reproducibility. However, using a single U-Net for refinement, while a standard baseline, may not exploit the latest architectures (e.g. transformer-based or diffusion models) that could further improve results. The evaluation relied on classical image quality metrics; no reader study or clinical scoring was performed, which limits conclusions about diagnostic utility.

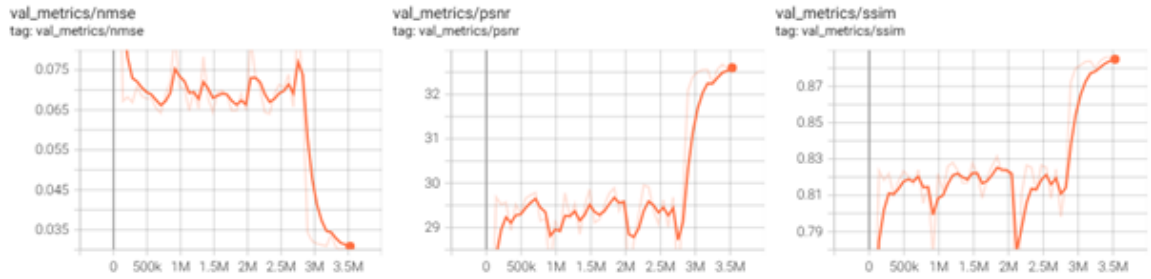
Several directions are recommended for future work. First, expand the anatomical and clinical diversity of training data beyond brain and knee. As this work observed, training on varied data tends to yield more robust models, so including other body parts or pathological cases could strengthen generalization. Second, investigate realistic acquisition variations: one could include prospectively undersampled clinical scans with realistic noise and motion, or simulate nonidealities (field inhomogeneities, coil miscalibration, metal implant artifacts) during training. Third, explore alternative hybrid architectures: the CS+U-Net pipeline could be extended with recent advances (e.g. learned proximal operators, or score-based models) to further improve quality and uncertainty quantification. Fourth, incorporating model uncertainty and expert review is crucial. Future studies should involve radiologist assessment of reconstruction fidelity and measure the impact on diagnostic tasks.

In summary, the thesis demonstrates that hybrid deep learning models can generalize under anatomical and contrast variability, especially when trained on heterogeneous data. The CS-based hybrid approach shows clear advantages over classical CS alone. By systematically evaluating under domain shifts, this work provides evidence that combining physics-based reconstruction with learning is a promising path toward clinically robust accelerated MRI. Continued research along the suggested avenues will help bridge the gap between technical performance and real-world deployment.

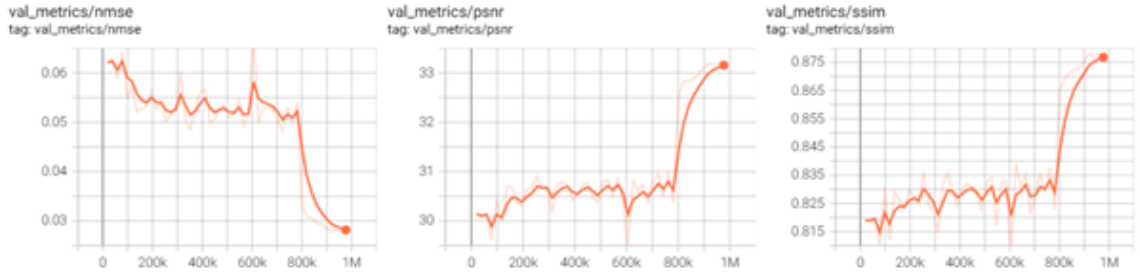
Appendices

Appendix A

Training of U-Net



(A) From left to right: NMSE, PSNR and MSSIM on validation set during training of the Hybrid-Brain model



(B) From left to right: NMSE, PSNR and MSSIM on validation set during training of the Hybrid-Multi model

FIGURE A.1: Validation reconstruction metrics for the Hybrid-Brain and Hybrid-Multi models tracked throughout training. Both plots show metric evolution per epoch on the validation set.

Appendix B

Detailed experiment results

B.1 Experiment 1: Brain data evaluation

Acquisition	Model	MSSIM	PSNR	NMSE	VGG
AXFLAIR	CS	0.787	24.7	0.128	0.047
	Hybrid-Brain	0.866	25.1	0.118	0.0512
	Hybrid-Multi	0.929	35.3	0.0123	**0.038**
AXT1	CS	0.827	26.2	0.102	0.0736
	Hybrid-Brain	0.904	26.3	0.0995	0.0767
	Hybrid-Multi	**0.952**	**36.4**	0.0123	0.0632
AXT1POST	CS	0.84	26.6	0.102	0.0816
	Hybrid-Brain	0.898	26.7	0.0998	0.086
	Hybrid-Multi	0.94	36.2	**0.0117**	0.0788
AXT2	CS	0.823	28.7	0.0665	0.0843
	Hybrid-Brain	0.901	28.9	0.0636	0.0776
	Hybrid-Multi	0.943	34.9	0.0143	0.0733

Table B.1: Mean reconstruction metrics per brain acquisition type at fixed acceleration factor (R=4). This table presents the model-wise average reconstruction metrics for each individual brain acquisition type (AXT1, AXT2, AXT1POST, AXFLAIR), keeping the acceleration factor fixed at R=4. This isolates the effect of acquisition contrast on reconstruction performance.

B.2 Experiment 2: Knee Data Evaluation

B.3

R	Model	MSSIM	PSNR	NMSE	VGG
4	CS	0.824	27.7	0.0815	0.0798
	Hybrid-Brain	0.898	27.9	0.0782	0.0771
	Hybrid-Multi	**0.942**	**35.4**	**0.0134**	**0.0706**
8	CS	0.678	19.8	0.455	0.142
	Hybrid-Brain	0.845	24.7	0.152	0.11
	Hybrid-Multi	0.862	26.5	0.103	0.108

TABLE B.2: Mean reconstruction metrics per acceleration factor (R=4 and R=8) for brain data, averaged across acquisitions. This table shows the average reconstruction performance for each model at acceleration factors R=4 and R=8, computed across all brain acquisitions. This analysis isolates the effect of acceleration factor on reconstruction quality.

Acquisition	Model	MSSIM	PSNR	NMSE	VGG
PDFS	CS	0.722	26.2	0.121	0.00939
	Hybrid-Brain	0.791	27.6	0.0909	0.0106
	Hybrid-Multi	0.85	**34.9**	0.0172	**0.00937**
PD	CS	0.818	26.2	0.0707	0.0384
	Hybrid-Brain	0.838	26.4	0.069	0.0424
	Hybrid-Multi	**0.907**	**34.9**	**0.0129**	0.0373

TABLE B.3: Mean reconstruction metrics per knee acquisition type at fixed acceleration factor (R=4). This table presents model-wise average reconstruction performance on the PDFS and PD knee acquisitions separately, using only volumes acquired at acceleration factor R=4. This allows analysis of contrast influence in knee imaging.

R	Model	MSSIM	PSNR	NMSE	VGG
4	CS	0.783	26.2	0.0888	0.0279
	Hybrid-Brain	0.821	26.8	0.077	0.0309
	Hybrid-Multi	**0.886**	**34.9**	**0.0145**	**0.0272**
8	CS	0.592	19.2	0.421	0.0362
	Hybrid-Brain	0.703	22	0.224	0.0318
	Hybrid-Multi	0.791	27.2	0.0694	0.029

TABLE B.4: Mean reconstruction metrics per acceleration factor (R=4 and R=8) for knee data, averaged across acquisitions. This table reports the average reconstruction performance of each model at R=4 and R=8, computed over all knee acquisitions. It allows assessment of the effect of undersampling rate on knee reconstruction quality.

Acquisition	Model	MSSIM	PSNR	NMSE	VGG
AXFLAIR	CS	0.787	24.7	0.128	0.047
	Hybrid-Brain	0.866	25.1	0.118	0.0512
	Hybrid-Multi	0.929	35.3	0.0123	0.038
AXT1	CS	0.827	26.2	0.102	0.0736
	Hybrid-Brain	0.904	26.3	0.0995	0.0767
	Hybrid-Multi	**0.952**	**36.4**	0.0123	0.0632
AXT1POST	CS	0.84	26.6	0.102	0.0816
	Hybrid-Brain	0.898	26.7	0.0998	0.086
	Hybrid-Multi	0.94	36.2	**0.0117**	0.0788
AXT2	CS	0.823	28.7	0.0665	0.0843
	Hybrid-Brain	0.901	28.9	0.0636	0.0776
	Hybrid-Multi	0.943	34.9	0.0143	0.0733
PDFS	CS	0.722	26.2	0.121	0.00939
	Hybrid-Brain	0.791	27.6	0.0909	0.0106
	Hybrid-Multi	0.85	34.9	0.0172	**0.00937**
PD	CS	0.818	26.2	0.0707	0.0384
	Hybrid-Brain	0.838	26.4	0.069	0.0424
	Hybrid-Multi	0.907	34.9	0.0129	0.0373

TABLE B.5: Mean reconstruction metrics per knee or brain acquisition type at fixed acceleration factor (R=4). This table presents model-wise average reconstruction performance on the AXFLAIR, AXT1, AXT1POST, AXT2 (brain); and PDFS and PD (knee)-contrasts separately, using only volumes acquired at acceleration factor R=4. This allows analysis of contrast influence in general imaging.

Appendix C

Reconstruction examples

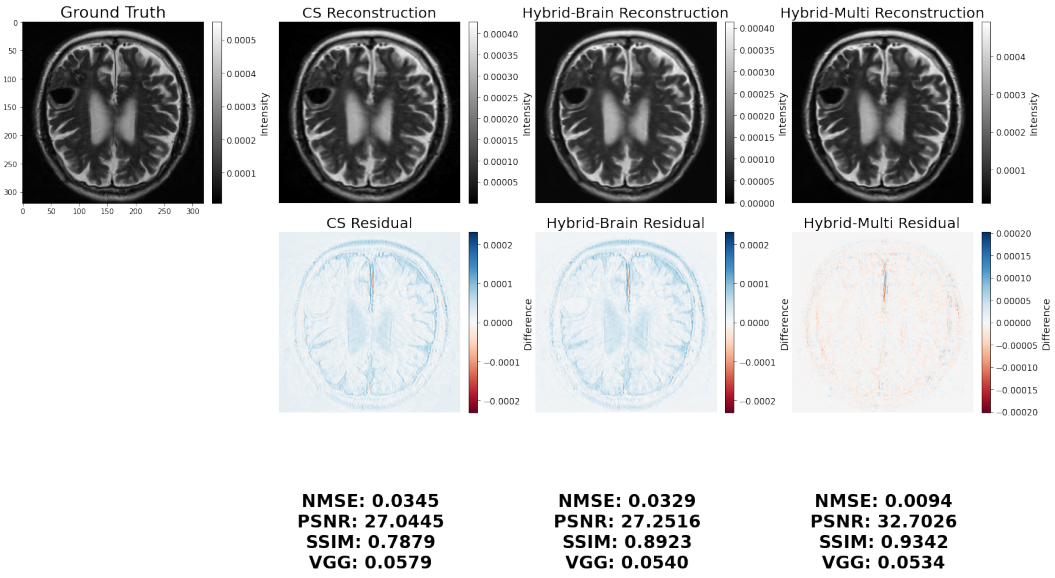


FIGURE C.1: Qualitative comparison of reconstruction performance for brain MRI (experiment 1) at acceleration factor R=4. The figure shows the ground truth (left) and one representative slice for all three models: Classical CS, Hybrid-Brain and Hybrid-Multi (left to right). For each method, the reconstructed image, its corresponding residual error map (difference from ground truth), and evaluation metrics are displayed.

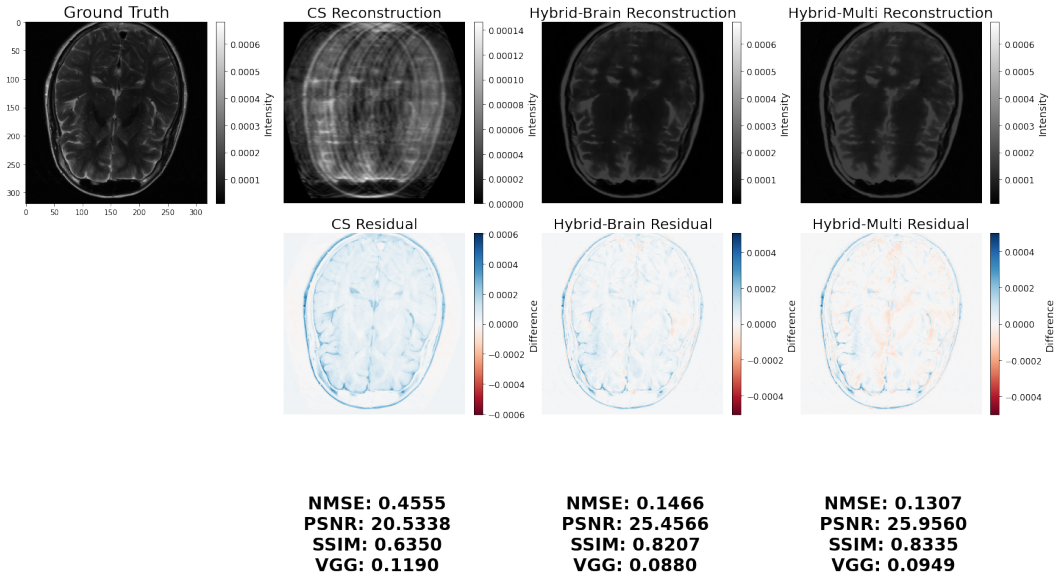


FIGURE C.2: Qualitative comparison of reconstruction performance for brain MRI (experiment 1) at acceleration factor R=8. The figure shows the ground truth (left) and one representative slice for all three models: Classical CS, Hybrid-Brain and Hybrid-Multi (left to right). For each method, the reconstructed image, its corresponding residual error map (difference from ground truth), and evaluation metrics are displayed.

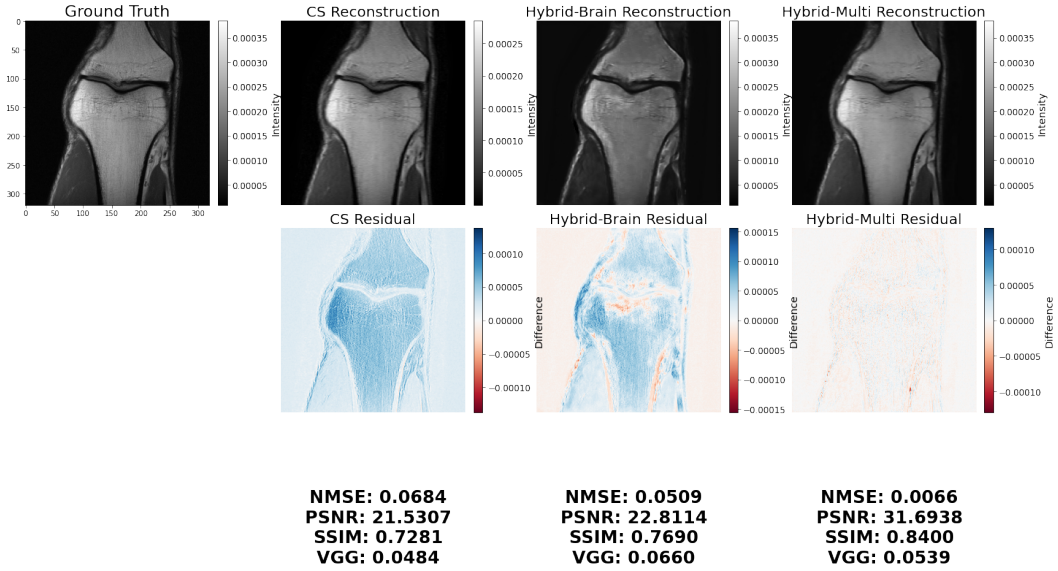


FIGURE C.3: Qualitative comparison of reconstruction performance for knee MRI (experiment 2) at acceleration factor R=4. The figure shows the ground truth (left) and one representative slice for all three models: Classical CS, Hybrid-Brain and Hybrid-Multi (left to right). For each method, the reconstructed image, its corresponding residual error map (difference from ground truth), and evaluation metrics are displayed.

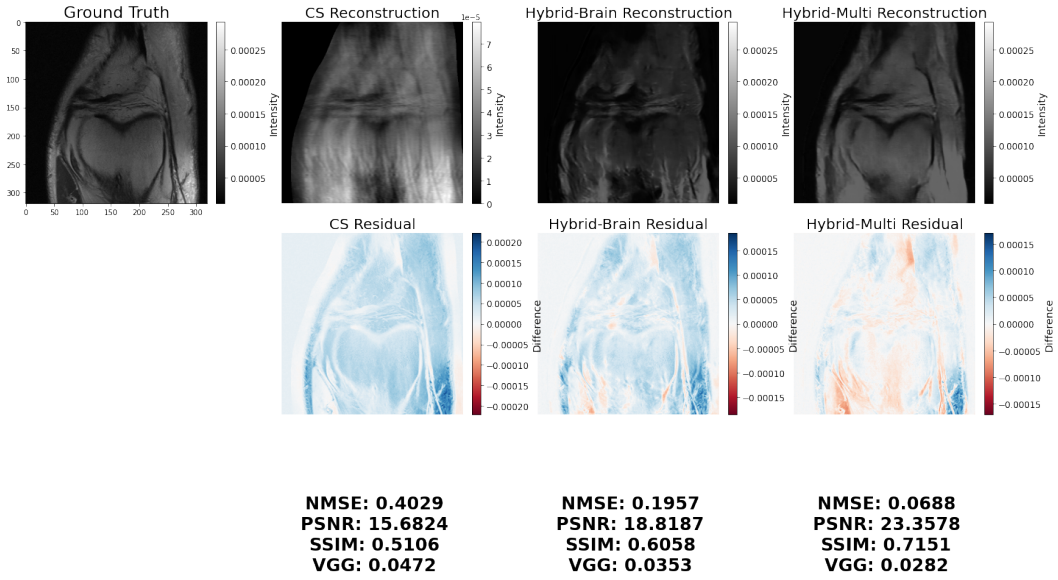


FIGURE C.4: Qualitative comparison of reconstruction performance for knee MRI (experiment 2) at acceleration factor R=8. The figure shows the ground truth (left) and one representative slice for all three models: Classical CS, Hybrid-Brain and Hybrid-Multi (left to right). For each method, the reconstructed image, its corresponding residual error map (difference from ground truth), and evaluation metrics are displayed.

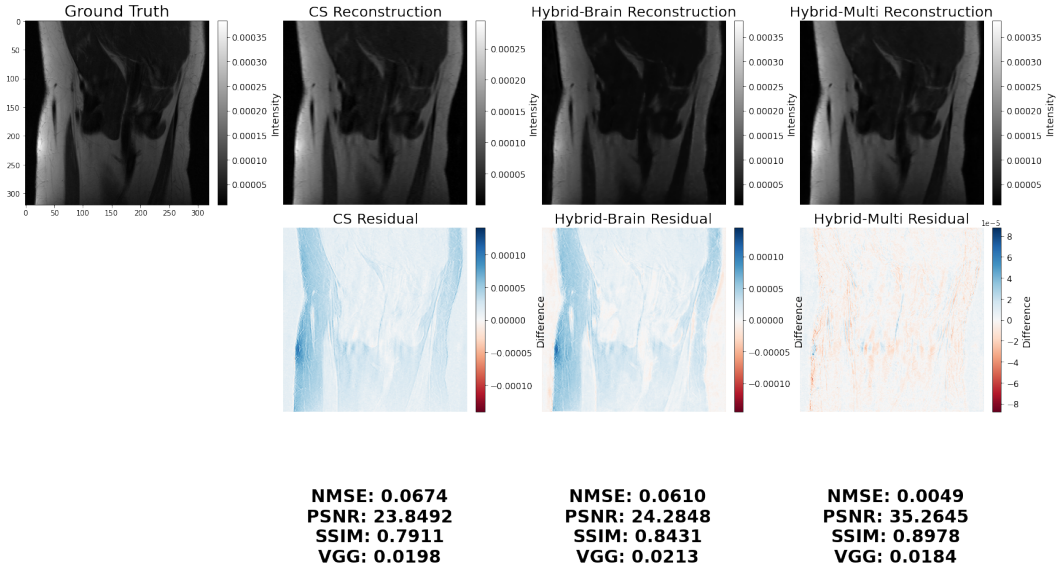


FIGURE C.5: Qualitative comparison of reconstruction performance for knee and brain MRI (experiment 3) at acceleration factor R=4. The figure shows the ground truth (left) and one representative slice for all three models: Classical CS, Hybrid-Brain and Hybrid-Multi (left to right). For each method, the reconstructed image, its corresponding residual error map (difference from ground truth), and evaluation metrics are displayed.

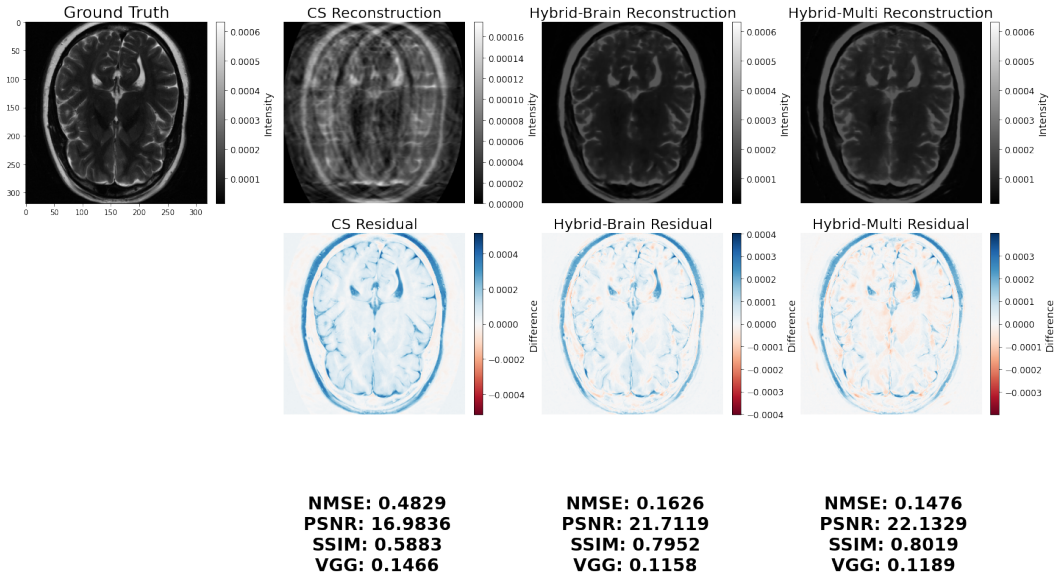


FIGURE C.6: Qualitative comparison of reconstruction performance for knee and brain MRI (experiment 3) at acceleration factor R=8. The figure shows the ground truth (left) and one representative slice for all three models: Classical CS, Hybrid-Brain and Hybrid-Multi (left to right). For each method, the reconstructed image, its corresponding residual error map (difference from ground truth), and evaluation metrics are displayed.

Bibliography

Aggarwal, H. K., Mani, M. P., & Jacob, M. (2019). MoDL: Model-Based Deep Learning Architecture for Inverse Problems. *IEEE Transactions on Medical Imaging*, *38*(2), 394–405. https://doi.org/10.1109/TMI.2018.2865356

Akçakaya, M., Doneva, M., & Prieto, C. (Eds.). (2023). *Magnetic resonance image reconstruction: Theory, methods, and applications* (Vol. 7). Academic Press.

Alam, S., Uh, J., Dresner, A., Hua, C.-h., & Khairy, K. (2023, November 22). *Deep-learning-based acceleration of MRI for radiotherapy planning of pediatric patients with brain tumors*. arXiv: 2311.13485 [eess]. https://doi.org/10.48550/arXiv.2311.13485

Avidan, N., & Freiman, M. (2023, November 26). *MA-RECON: Mask-aware deep-neural-network for robust fast MRI k-space interpolation*. arXiv: 2209.00462 [eess]. https://doi.org/10.48550/arXiv.2209.00462

BART Toolbox. (n.d.). Retrieved November 28, 2024, from https://mrirecon.github.io/bart/

Candes, E., Romberg, J., & Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, *52*(2), 489–509. https://doi.org/10.1109/TIT.2005.862083

Chung, H., & Ye, J. C. (2022, July 16). *Score-based diffusion models for accelerated MRI*. arXiv: 2110.05243 [eess]. https://doi.org/10.48550/arXiv.2110.05243

Cummings, E., Macdonald, J. A., & Seiberlich, N. (2022, January 1). Chapter 6 - Parallel Imaging. In M. Akçakaya, M. Doneva, & C. Prieto (Eds.), *Advances in Magnetic Resonance Technology and Applications* (pp. 129–157, Vol. 7). Academic Press. https://doi.org/10.1016/B978-0-12-822726-8.00016-6

Darestani, M. Z., Chaudhari, A. S., & Heckel, R. (2021, June 11). *Measuring Robustness in Deep Learning Based Compressive Sensing*. arXiv: 2102.06103 [eess]. https://doi.org/10.48550/arXiv.2102.06103

Donoho, D. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, *52*(4), 1289–1306. https://doi.org/10.1109/TIT.2006.871582

Eo, T., Jun, Y., Kim, T., Jang, J., Lee, H.-J., & Hwang, D. (2018). KIKI-net: Cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. *Magnetic Resonance in Medicine*, *80*(5), 2188–2201. https://doi.org/10.1002/mrm.27201

Feng, L. (2022, January 1). Chapter 8 - Sparse Reconstruction. In M. Akçakaya, M. Doneva, & C. Prieto (Eds.), *Advances in Magnetic Resonance Technology*

and Applications (pp. 189–221, Vol. 7). Academic Press. https://doi.org/10.1016/B978-0-12-822726-8.00018-X

Fujita, N., Yokosawa, S., Shirai, T., & Terada, Y. (2024). Numerical and Clinical Evaluation of the Robustness of Open-source Networks for Parallel MR Imaging Reconstruction. *Magnetic resonance in medical sciences: MRMS: an official journal of Japan Society of Magnetic Resonance in Medicine*, *23*(4), 460–478. https://doi.org/10.2463/mrms.mp.2023-0031

Gilton, D., Ongie, G., & Willett, R. (2021, June 3). *Deep Equilibrium Architectures for Inverse Problems in Imaging*. arXiv: 2102.07944 [eess]. https://doi.org/10.48550/arXiv.2102.07944

Griswold, M. A., Jakob, P. M., Heidemann, R. M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., & Haase, A. (2002). Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, *47*(6), 1202–1210. https://doi.org/10.1002/mrm.10171

Hammernik, K., Klatzer, T., Kobler, E., Recht, M. P., Sodickson, D. K., Pock, T., & Knoll, F. (2018). Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, *79*(6), 3055–3071. https://doi.org/10.1002/mrm.26977

Hammernik, K., Küstner, T., & Rueckert, D. (2022, January 1). Chapter 11 - Machine Learning for MRI Reconstruction. In M. Akçakaya, M. Doneva, & C. Prieto (Eds.), *Advances in Magnetic Resonance Technology and Applications* (pp. 281–323, Vol. 7). Academic Press. https://doi.org/10.1016/B978-0-12-822726-8.00021-X

Hammernik, K., Schlemper, J., Qin, C., Duan, J., Summers, R. M., & Rueckert, D. (2021). Systematic evaluation of iterative deep neural networks for fast parallel MRI reconstruction with sensitivity-weighted coil combination. *Magnetic Resonance in Medicine*, *86*(4), 1859–1872. https://doi.org/10.1002/mrm.28827

Heckel, R., Jacob, M., Chaudhari, A., Perlman, O., & Shimron, E. (2024). Deep learning for accelerated and robust MRI reconstruction. *Magnetic Resonance Materials in Physics, Biology and Medicine*, *37*(3), 335–368. https://doi.org/10.1007/s10334-024-01173-8

Hossain, M. B., Shinde, R. K., Oh, S., Kwon, K.-C., & Kim, N. (2024). A Systematic Review and Identification of the Challenges of Deep Learning Techniques for Undersampled Magnetic Resonance Image Reconstruction. *Sensors (Basel, Switzerland)*, *24*(3), 753. https://doi.org/10.3390/s24030753

Huang, J., Fang, Y., Wu, Y., Wu, H., Gao, Z., Li, Y., Ser, J. D., Xia, J., & Yang, G. (2022, April 10). *Swin Transformer for Fast MRI*. arXiv: 2201.03230 [eess]. https://doi.org/10.48550/arXiv.2201.03230

Huang, J., Wu, Y., Wang, F., Fang, Y., Nan, Y., Alkan, C., Abraham, D., Liao, C., Xu, L., Gao, Z., Wu, W., Zhu, L., Chen, Z., Lally, P., Bangerter, N., Setsompop, K., Guo, Y., Rueckert, D., Wang, G., & Yang, G. (2025). Data- and Physics-Driven Deep Learning Based Reconstruction for Fast MRI: Fundamentals and Methodologies. *IEEE Reviews in Biomedical Engineering*, *18*, 152–171. https://doi.org/10.1109/RBME.2024.3485022

Huang, J., Xing, X., Gao, Z., & Yang, G. (2022). Swin Deformable Attention U-Net Transformer (SDAUT) for Explainable Fast MRI. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI*, 538–548. https://doi.org/10.1007/978-3-031-16446-0_51

Huang, J., Wang, S., Zhou, G., Hu, W., & Yu, G. (2022). Evaluation on the generalization of a learned convolutional neural network for MRI reconstruction. *Magnetic Resonance Imaging*, *87*, 38–46. https://doi.org/10.1016/j.mri.2021.12.003

Jalal, A., Arvinte, M., Daras, G., Price, E., Dimakis, A. G., & Tamir, J. I. (2021, December 6). *Robust Compressed Sensing MRI with Deep Generative Priors*. arXiv: 2108.01368 [cs]. https://doi.org/10.48550/arXiv.2108.01368

Johnson, J., Alahi, A., & Fei-Fei, L. (2016, March 27). *Perceptual Losses for Real-Time Style Transfer and Super-Resolution*. arXiv: 1603.08155 [cs]. https://doi.org/10.48550/arXiv.1603.08155

Johnson, P. M., Jeong, G., Hammernik, K., Schlemper, J., Qin, C., Duan, J., Rueckert, D., Lee, J., Pezzotti, N., De Weerdt, E., Yousefi, S., Elmahdy, M. S., Van Gemert, J. H. F., Schülke, C., Doneva, M., Nielsen, T., Kastryulin, S., Lelieveldt, B. P. F., Van Osch, M. J. P., … Knoll, F. (2021). Evaluation of the Robustness of Learned MR Image Reconstruction to Systematic Deviations Between Training and Test Data for the Models from the fastMRI Challenge. In N. Haq, P. Johnson, A. Maier, T. Würfl, & J. Yoo (Eds.), *Machine Learning for Medical Image Reconstruction* (pp. 25–34). Springer International Publishing. https://doi.org/10.1007/978-3-030-88552-6_3

Kim, S., Park, H., & Park, S.-H. (2024). A review of deep learning-based reconstruction methods for accelerated MRI using spatiotemporal and multi-contrast redundancies. *Biomedical Engineering Letters*, *14*(6), 1221–1242. https://doi.org/10.1007/s13534-024-00425-9

Knoll, F., Hammernik, K., Kobler, E., Pock, T., Recht, M. P., & Sodickson, D. K. (2019). Assessment of the generalization of learned image reconstruction and the potential for transfer learning. *Magnetic Resonance in Medicine*, *81*(1), 116–128. https://doi.org/10.1002/mrm.27355

Knoll, F., Zbontar, J., Sriram, A., Muckley, M. J., Bruno, M., Defazio, A., Parente, M., Geras, K. J., Katsnelson, J., Chandarana, H., Zhang, Z., Drozdzalv, M., Romero, A., Rabbat, M., Vincent, P., Pinkerton, J., Wang, D., Yakubova, N., Owens, E., … Lui, Y. W. (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), e190007. https://doi.org/10.1148/ryai.2020190007

Liu, J., Qin, C., & Vaighan, M. Y. (2023). High-Fidelity MRI Reconstruction Using Adaptive Spatial Attention Selection and Deep Data Consistency Prior. *IEEE Transactions on Computational Imaging*, *9*, 298–313. https://doi.org/10.1109/TCI.2023.3258839

Lønning, K., Putzky, P., Sonke, J.-J., Reneman, L., Caan, M. W. A., & Welling, M. (2019). Recurrent inference machines for reconstructing heterogeneous MRI

data. *Medical Image Analysis*, *53*, 64–78. https://doi.org/10.1016/j.media.2019.01.005

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. https://doi.org/10.1002/mrm.21391

Lustig, M., Donoho, D. L., Santos, J. M., & Pauly, J. M. (2008). Compressed Sensing MRI. *IEEE Signal Processing Magazine*, *25*(2), 72–82. https://doi.org/10.1109/MSP.2007.914728

Maes, F. (2024). Medical imaging & analysis: Magnetic resonance imaging. *Faculty of Engineering Science, KU Leuven*.

Mardani, M., Gong, E., Cheng, J. Y., Vasanawala, S. S., Zaharchuk, G., Xing, L., & Pauly, J. M. (2019). Deep Generative Adversarial Neural Networks for Compressive Sensing MRI. *IEEE transactions on medical imaging*, *38*(1), 167–179. https://doi.org/10.1109/TMI.2018.2858752

Monga, V., Li, Y., & Eldar, Y. C. (2020, August 7). *Algorithm Unrolling: Interpretable, Efficient Deep Learning for Signal and Image Processing*. arXiv: 1912.10557 [eess]. https://doi.org/10.48550/arXiv.1912.10557

Pal, A., & Rathi, Y. (2022). A review and experimental evaluation of deep learning methods for MRI reconstruction. *The journal of machine learning for biomedical imaging*, *1*, 001.

Pruessmann, K. P., Weiger, M., Scheidegger, M. B., & Boesiger, P. (1999). SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, *42*(5), 952–962.

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. https://doi.org/10.1109/TMI.2018.2820120

Ronneberger, O., Fischer, P., & Brox, T. (2015, May 18). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. arXiv: 1505.04597 [cs]. https://doi.org/10.48550/arXiv.1505.04597

Schlemper, J., Caballero, J., Hajnal, J. V., Price, A. N., & Rueckert, D. (2018). A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction. *IEEE Transactions on Medical Imaging*, *37*(2), 491–503. https://doi.org/10.1109/TMI.2017.2760978

Simonyan, K., & Zisserman, A. (2015, April 10). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. arXiv: 1409.1556 [cs]. https://doi.org/10.48550/arXiv.1409.1556

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 (Basel, Switzerland)*, *10*(9), 1012. https://doi.org/10.3390/bioengineering10091012

Song, J., Meng, C., & Ermon, S. (2022, October 5). *Denoising Diffusion Implicit Models*. arXiv: 2010.02502 [cs]. https://doi.org/10.48550/arXiv.2010.02502

Sonoda, S., & Murata, N. (2017). Neural network with unbounded activation functions is universal approximator. *Applied and Computational Harmonic Analysis*, *43*(2), 233–268. https://doi.org/10.1016/j.acha.2015.12.005

Sriram, A., Zbontar, J., Murrell, T., Defazio, A., Zitnick, C. L., Yakubova, N., Knoll, F., & Johnson, P. (2020, April 15). *End-to-End Variational Networks for Accelerated MRI Reconstruction*. arXiv: 2004.06688 [eess]. https://doi.org/10.48550/arXiv.2004.06688

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. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 14303–14310. https://doi.org/10.1109/CVPR42600.2020.01432

Suykens, J. (2024). Artificial Neural Networks and Deep Learning. *Faculty of Engineering Science, KU Leuven*.

Tavaf, N., Torfi, A., Ugurbil, K., & Moortele, P.-F. V. de. (2021, February 15). *GRAPPA-GANs for Parallel MRI Reconstruction*. arXiv: 2101.03135 [eess]. https://doi.org/10.48550/arXiv.2101.03135

Tourais, J., Coletti, C., & Weingärtner, S. (2022, January 1). Chapter 1 - Brief Introduction to MRI Physics. In M. Akçakaya, M. Doneva, & C. Prieto (Eds.), *Advances in Magnetic Resonance Technology and Applications* (pp. 3–36, Vol. 7). Academic Press. https://doi.org/10.1016/B978-0-12-822726-8.00010-5

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. https://doi.org/10.1002/mrm.24751

Ulyanov, D., Vedaldi, A., & Lempitsky, V. (2020). Deep Image Prior. *International Journal of Computer Vision*, *128*(7), 1867–1888. https://doi.org/10.1007/s11263-020-01303-4

Vanhaverbeke, M. (2024). *The influence of hybrid modelling on deep learning-based MRI reconstruction performance*.

Virtue, P., & Lustig, M. (2017). The Empirical Effect of Gaussian Noise in Undersampled MRI Reconstruction. *Tomography (Ann Arbor, Mich.)*, *3*(4), 211–221. https://doi.org/10.18383/j.tom.2017.00019

Wang, S., Xiao, T., Liu, Q., & Zheng, H. (2021). Deep learning for fast MR imaging: A review for learning reconstruction from incomplete k-space data. *Biomedical Signal Processing and Control*, *68*, 102579. https://doi.org/10.1016/j.bspc.2021.102579

Yaman, B., Hosseini, S. A. H., Moeller, S., Ellermann, J., Uğurbil, K., & Akçakaya, M. (2020). Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magnetic Resonance in Medicine*, *84*(6), 3172–3191. https://doi.org/10.1002/mrm.28378

Yang, G., Yu, S., Dong, H., Slabaugh, G., Dragotti, P. L., Ye, X., Liu, F., Arridge, S., Keegan, J., Guo, Y., & Firmin, D. (2018). DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Recon-

struction. *IEEE Transactions on Medical Imaging*, *37*(6), 1310–1321. https://doi.org/10.1109/TMI.2017.2785879

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*. https://doi.org/10.3389/fninf.2020.611666

Zbontar, J., Knoll, F., Sriram, A., Murrell, T., Huang, Z., Muckley, M. J., Defazio, A., Stern, R., Johnson, P., Bruno, M., Parente, M., Geras, K. J., Katsnelson, J., Chandarana, H., Zhang, Z., Drozdzal, M., Romero, A., Rabbat, M., Vincent, P., … Lui, Y. W. (2019, December 11). *fastMRI: An Open Dataset and Benchmarks for Accelerated MRI*. arXiv: 1811.08839 [cs]. https://doi.org/10.48550/arXiv.1811.08839

Zeng, G., Guo, Y., Zhan, J., Wang, Z., Lai, Z., Du, X., Qu, X., & Guo, D. (2021). A review on deep learning MRI reconstruction without fully sampled k-space. *BMC Medical Imaging*, *21*(1), 195. https://doi.org/10.1186/s12880-021-00727-9

Zheng, H., Fang, F., & Zhang, G. (2019). Cascaded Dilated Dense Network with Two-step Data Consistency for MRI Reconstruction. *Advances in Neural Information Processing Systems*, *32*.

Zhou, B., Dey, N., Schlemper, J., Salehi, S. S. M., Liu, C., Duncan, J. S., & Sofka, M. (2022, August 17). *DSFormer: A Dual-domain Self-supervised Transformer for Accelerated Multi-contrast MRI Reconstruction*. arXiv: 2201.10776 [eess]. https://doi.org/10.48550/arXiv.2201.10776

Zhou Wang, Bovik, A., Sheikh, H., & Simoncelli, E. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612. https://doi.org/10.1109/TIP.2003.819861

Zhu, B., Liu, J. Z., Cauley, S. F., Rosen, B. R., & Rosen, M. S. (2018). Image reconstruction by domain-transform manifold learning. *Nature*, *555*(7697), 487–492. https://doi.org/10.1038/nature25988