

ABSTRACT

Title of thesis: TOWARDS FASTER MRI ACQUISITION
 THROUGH DEEP NEURAL NETWORKS
 Cheng Peng
 Master of Science, 2019

Thesis directed by: Professor Rama Chellappa
 Department of Electrical and Computer Engineering

Extensive research on accelerating Magnetic Resonance Imaging (MRI) has been done along two fronts: (i) hardware acceleration, and (ii) image post processing. We present the results of our work on image post processing, where the inputs are sparsely sampled volumetric images, and our deep-learning-based models seek to output the original, densely sampled images. Specifically, we propose two different methods for accelerating MRI, corresponding to two different aspects of the MR acquisition process.

First, We propose a marginal super-resolution (MSR) approach based on 2D convolutional neural networks (CNNs) for interpolating an anisotropic brain magnetic resonance scan along the highly under-sampled direction. Previous methods for slice interpolation only consider data from pairs of adjacent 2D slices. The possibility of fusing information from the direction orthogonal to the 2D slices remains unexplored. Our approach performs MSR in both sagittal and coronal directions, which provides an initial estimate for slice interpolation. The interpolated slices are then fused and refined in the axial direction for improved consistency. Since

MSR consists of only 2D operations, it is more feasible in terms of GPU memory consumption and requires fewer training samples compared to 3D CNNs. Our experiments demonstrate that the proposed method outperforms traditional linear interpolation and baseline 2D/3D CNN-based approaches. We conclude by showcasing the method’s practical utility in estimating brain volumes from under-sampled brain MR scans through semantic segmentation.

Secondly, although undersampled MR image recovery has been widely studied for accelerated MR acquisition, it has been mostly studied under a single sequence scenario, despite the fact that multi-sequence MR scan is common in practice. We aim to optimize multi-sequence MR image recovery from undersampled k-space data under an overall time constraint while considering the difference in acquisition time for various sequences. We first formulate it as a constrained optimization problem and then show that finding the optimal sampling strategy for all sequences and the best recovery model at the same time is combinatorial and hence computationally prohibitive. To solve this problem, we propose a blind recovery model that simultaneously recovers multiple sequences, and an efficient approach to find the near-optimal combination of sampling strategy and recovery model. Our experiments demonstrate that the proposed method not only outperforms sequence-wise recovery, but also sheds light on how to optimally undersample the k-space for each sequence within an overall time budget.

TOWARDS FASTER MRI ACQUISITION
THROUGH DEEP NEURAL NETWORKS

by

Cheng Peng

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Science
2019

Advisory Committee:
Professor Rama Chellappa, Chair/Advisor
Professor Behtash Babadi
Professor ShaoHua Kevin Zhou

© Copyright by
Cheng Peng
2019

Acknowledgments

I owe my gratitude to all the people who have made this thesis possible and because of whom my graduate experience has been one that I will cherish forever.

First and foremost I would like to thank my advisor, Professor Rama Chellappa for supporting me and giving me an invaluable opportunity to work on challenging problems over the past two years. I first got to know him as his undergraduate teaching assistant, and have always enjoyed learning from such a brilliant mind. Furthermore, he always has an upbeat, infectious attitude, which relieved me of a lot of anxiety in tackling difficulties during my graduate studies. It is absolutely astonishing for me to know someone who is as accomplished as Prof. Chellappa, while as friendly and understanding.

I would like to thank Prof. Shaohua Kevin Zhou, who gave me the opportunity to work with him when I first contacted him about my interest in medical imaging and computer vision. He has been extremely helpful in guiding me through the process of top notch research, from the formulation of ideas all the way to the completion of the writings.

I also thank Prof. Betash Babadi, who along with Prof. Rama Chellappa taught me my first machine learning course, which has been instrumental in shaping my interests in research; I am grateful that he agreed to be a member of my thesis committee. I would like to thank Prof. Romel Gomez and Prof. Julius Goldhar for giving me an opportunity to work as a teaching assistant both as an undergraduate and graduate. It gave me a lot of perspectives in understanding the

joy and importance of learning while teaching.

I owe my deepest thanks to my family - my mother and father, who are also professors, in giving me encouragement and warning me about possible adversities. Without their advice I would not be in the position I am right now.

Lastly, I would like to thank Haofu Liao, Wei-an Andy Lin, Jingxiao Zheng for their friendship and helpful discussions on solving research problems. I would like to thank all the fellow teaching assistants for ENEE101, with whom I shared a wonderful time when teaching the class. I would like to thank Wenyan Li, for her companionship and constant support.

Table of Contents

Acknowledgements	ii
List of Tables	vi
List of Figures	vii
1 Introduction	1
2 Background	7
2.1 Traditional slice interpolation methods	7
2.2 Learning based super-resolution methods	8
2.3 2D MRI Recovery	9
3 Problem Formulation	11
3.1 Slice Interpolation	11
3.2 Multi-sequence MR Recovery	12
4 Proposed Algorithms	15
4.1 Deep Slice Interpolation	15
4.1.1 Marginal Super-Resolution	16
4.1.2 Two-View Fusion and Refinement	17
4.2 Multi-sequence Recovery	18
4.2.1 Blind recovery model	18
4.2.2 Sampling strategy searching	19
4.2.3 An optimization point-of-view	19
4.2.4 Single-sequence Training vs Multi-sequence Training	20
4.2.5 Network Architecture	21
5 Experimentation	22
5.1 Settings	22
5.1.1 Implementation Details	22
5.2 Deep Slice Interpolation	22
5.2.1 Data sets	22

5.2.2	Evaluation metrics	23
5.2.3	Quantitative Evaluations	24
5.2.4	Visual Comparisons	24
5.2.5	Ablation study	26
5.3	Multi-Sequence Recovery	29
5.3.1	Data sets	29
5.3.2	Acquisition time and undersampling settings	30
5.3.3	Evaluation metrics	31
5.3.4	Results	32
6	Conclusion	38
6.1	Deep Slice Interpolation	38
6.2	Multi-sequence Recovery	38
6.3	Future Work	39
	Bibliography	41

List of Tables

5.1	Quantitative evaluations for different slice interpolation approaches. For DICE and HD performance metrics, we present results on gray matter (GM)/white matter (WM) segmentation. The best results are in bold and the second best are <u>underlined</u>	25
5.2	Quantitative ablation study. Baseline numbers are also included for comparison. The best results are in bold and the second best are <u>underlined</u>	28
5.3	Quantitative evaluations for the top performing $\lambda_{1:S}$ under different acquisition time assumption. The performance numbers presented here are PSNR (dB) and SSIM.	33

List of Figures

1.1	The axial, coronal, and sagittal views of an anisotropic MR volume are fitted to isotropic resolution through (Left) linear interpolation and (Right) our proposed slice-interpolation method.	4
4.1	Marginal Super-Resolution Pipeline.	15
4.2	Two-view Fusion Pipeline.	17
4.3	Refinement Pipeline.	18
4.4	Multi-sequence recovery pipeline with the masks M_s randomly selected.	21
5.1	Visual comparisons of slice interpolation approaches. For 4x sparsity, the second of three interpolated MR slices is presented. For 8x sparsity, the third of seven interpolated slices is presented.	26
5.2	Visual comparison of gray matter (Green)/white matter (Blue) segmentation over different methods, with respective DICE scores listed under the images.	27
5.3	Visual comparison for the proposed components.	29
5.4	Quantitative recovery performance comparison. The Pearson correlation coefficient between Dedicated and MIMO vs between Dedicated and ZF is 0.85 vs -0.33 in the selected range	31
5.5	Visual comparison of different recovery methods on real data, with PSNR and SSIM values listed under the images. After successful recovery, the images become sharper with more visible details.	34
5.6	Additional visual comparison of different recovery methods on real data.	35
5.7	Additional visual comparison of different recovery methods on real data, with a different time setting.	36
5.8	Visual comparison of different recovery methods on simulated data. Note that BraTS sequences are interpolated for registration; therefore the image quality is not as good as the real data.	36
5.9	Visual comparison of different recovery methods on simulated data, with a different time setting.	37

Chapter 1: Introduction

Magnetic Resonance Imaging (MRI) is a medical imaging technique. Its fundamental principle is based on Nuclear Magnetic Resonance (NMR), a property of atom first observed by Bloch [1] and Purcell [2]. In general, the MRI process can be concisely explained in the following steps (for a more comprehensible review of MRI principles, please refer to the survey by Wright [3]):

- First, a strong magnetic field, denoted as \mathbf{B}_0 , is generated within the MRI machine, allowing all the atoms to be aligned to a uniform direction. The time required for all atoms to align is typically referred to as the *longitudinal relaxation time*.
- Secondly, Radio Frequency (RF) pulses, denoted as \mathbf{B}_1 , are sent by the machine, giving the atoms energy and knocking them off the axis induced by \mathbf{B}_0 . The specific amount that the atoms are knocked off from their original axis is depended on the resonance between the atoms and the designed \mathbf{B}_1 .
- Lastly, after the RF pulses are stopped, the atoms realign with \mathbf{B}_0 , in the meantime emitting energy provided by \mathbf{B}_1 as RF signals. The time for atoms to re-align to \mathbf{B}_0 is typically referred to as the *transverse relaxation time*.

- The emitted RF signals are captured by sensors within the machine, and contain information about the nature of the atoms. These signals are mapped to the k-space, which can be thought of as a measurement of the Fourier Transform of the MR image. The k-space is then used to create the image by performing inverse Fourier Transform. The result in image space highlights the contrasts between human tissues.

As a lot of diseases manifest themselves as an increase of water (e.g. inflammation) or fat (e.g. tumors) in certain regions, MRI can choose specific atom to target and match the resonance of its RF pulse, thus creating disease-specific images to aid doctors for diagnostics. As the result, the resolution and contrast of the images generated by MRI is often considerably better than those generated by other imaging techniques (e.g. Ultrasound). Furthermore, sending magnetic fields, which is similar in nature to RF signals sent by mobile phones, through patients posts no significant health risks. In comparison, X-Ray and Computational Tomography, which are other forms of popular medical imaging techniques, expose patients to ionizing radiation, which may be hazardous to human health.

Due to the stated advantages, MRI is widely used in the medical domain. MRI also has a few drawbacks. The MR machines, as well as the maintenance of them, are very expensive. This is the result of the high demand of power, quality of material, and quality of engineering needed to create and maintain strong and consistent magnetic fields, requiring coils to perform as close to superconductors as possible. Additionally, the speed of MRI acquisition is physically constrained by

the longitudinal and transverse relaxation time, as explained previously. Depending on the type of targeted atom, a session of MRI acquisition can take much longer than X-Ray or CT. This not only limits patients' access to MRI machines, but also introduces technical difficulties such as motion blur and limitations in dynamic, real-time imaging.

Consequently, MRI acquisition acceleration has been an active research area. On the hardware side, efforts have been made to introduce redundancy by adding multiple receiver coil [4] to detect the RF signals. On the image processing front, there is a long history of recovery undersampled MR images, from applying the theory of compressed sensing [5], to the more data-driven dictionary learning [6]. With the advent of deep learning, recently many new methods have been proposed by using deep convolutional neural networks (DCNN) [7-9].

Although these previous works have made great strides in recovering undersampled MR images, there still are major areas that remain relatively unexplored. Firstly, most of the literature focuses on recovering undersampled 2D images. In reality, due to the lack of assurance on the quality of the recovery, doctors and technicians still fully sample the entire 2D k-space for diagnostics, and attempt to cut time by reducing the number of slices taken. For example, Figure 1.1 shows a brain MR scan which is sparsely sampled along the axial direction. As a result, image quality suffers when viewing from Coronal and Sagittal directions.

It is desirable to have a consistent resolution across all dimensions, both for visualization and for medical analysis tasks such as brain volume estimation. Compared to the traditional perspective on accelerating MRI, the ability to upsample 3D

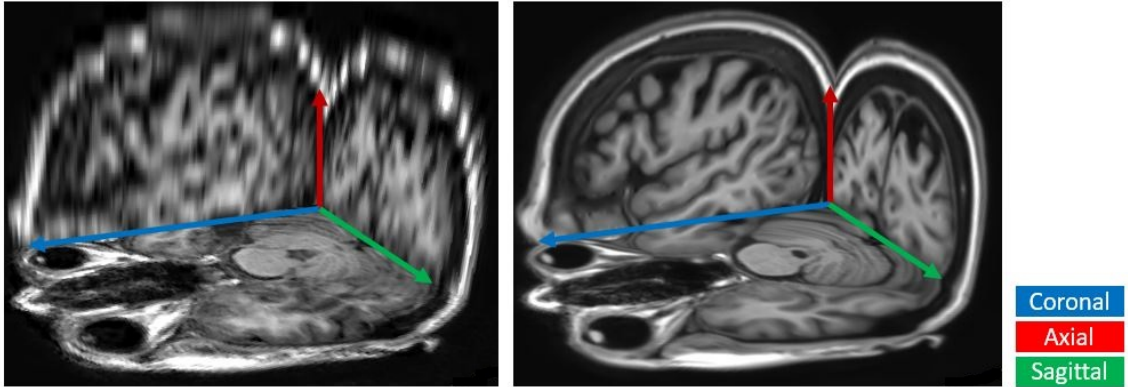


Figure 1.1: The axial, coronal, and sagittal views of an anisotropic MR volume are fitted to isotropic resolution through (Left) linear interpolation and (Right) our proposed slice-interpolation method.

MR images provides another, perhaps more realistic, way to speed up acquisition.

Another interesting area that is relatively less explored is the consideration of multiple MR sequences instead of one. Most patients take multiple sequences at a time, each of which targets and suppresses the signals of specific atoms. For instance, brain tumor patients can take more than three different sequences at a time, each of which gives different clues about the locations of the tumors and the area that is affected by them. While diagnostically desirable, taking multiple sequences drastically increase the overall acquisition time. There is much shared information across them, and in general to the best strategy to sample the most important information is not known, given the limited time.

This thesis attempts to solve these two problems. In short summary, the contributions that we make can be summarized as follows:

For increasing the resolution of 3D MR images,

- We propose a novel marginal super-resolution approach to break down the 3D slice interpolation problem into several 2D problems, which is more feasible in terms of GPU memory consumption and the amount of data available for training.
- We propose a two-view fusion approach to incorporate the 3D anatomical structure. The interpolated slices after fusion achieve high structural consistency. The final refinement further recovers fine details.
- We perform extensive evaluations on a large-scale MR data set, and show that the proposed method outperforms all the competing CNN models, including 3D CNNs, in terms of quantitative measurement, visual quality, and brain matter segmentation.

For analyzing multi-sequence super-resolution and the strategy for sampling them,

- We formulate a constrained optimization problem, where given a limited acquisition time, we seek to find the best strategy to undersample the k-spaces of multiple sequences to achieve the best overall recovery;
- We propose a novel CNN-based blind recovery model that extrapolates the shared information across different sequences and simultaneously recover them, as well as an efficient approach to find the near-optimal combination of sampling strategy and recovery model;
- We perform extensive evaluation on a large amount of simulated k-space data,

which shows that the proposed model outperforms the method of recovering each sequence on its own and sheds light on how to near-optimally undersample the k-spaces of multiple sequences.

Chapter 2: Background

2.1 Traditional slice interpolation methods.

Early work on interpolating volumetric medical data dates back to 1992, when Goshtasby et al. [10] proposed to leverage the small and gradual anatomic differences between consecutive slices, and find correspondence between pixels by searching through small neighborhoods. A slew of methods were proposed in the subsequent years, focusing on finding more accurate deformation fields, including shape-based methods [11], morphology-based methods [12], registration-based methods [13], etc. Linear interpolation can be regarded as a special example, which essentially assumes no deformation between slices.

An important assumption made in the above-mentioned methods is that adjacent slices contain similar anatomical structures, i.e., the changes in the structures have to be sufficiently small such that a dense correspondence can be found between two slices. This assumption largely limits the applicability of slice interpolation methods especially when slices are sparsely sampled. Furthermore, these methods did not utilize the information outside the two adjacent slices.

2.2 Learning based super-resolution methods.

Slice interpolation can be viewed as a special case of 3D super-resolution. Here we review the literature of 2D Single Image Super-Resolution (SISR), especially those approaches based on CNNs. Dong et al. [14] first proposed SRCNN, for learning a mapping that optimally transforms low-resolution (LR) images to high-resolution (HR) images. Many subsequent studies explored strategies to improve SISR by using deeper architectures and weight-sharing [15–17]. However, these methods require bilinear upsampling as a pre-processing step, which drastically increases computational complexity [18]. To address this issue, Dong et al. [18] proposed to apply deconvolution layers for the LR image to be directly upsampled to finer resolution. Furthermore, many studies have shown that residual learning provided better performance in SISR [19–21]. Specifically, Zhang et al. [21] incorporated both residual learning and dense blocks [22], and introduced Residual Dense Blocks (RDB) to allow for all layers of features to be seen directly by other layers, achieving state-of-the-art performance.

Generative Adversarial Networks (GAN) [23] have also been incorporated in SISR to improve the visual quality of the generated images. Ledig et al. pointed out that training SISR networks solely by L_1 or L_2 loss intrinsically leads to blurry estimations, and proposed SRGAN [20], which generated much sharper and realistic images compared to other approaches, despite having lower peak signal to noise ratios.

Though available computation capacity has been increasing, 3D CNNs are still

limited by memory capacity due to a considerable increase in the size of network parameters and input data. A common compromise is to extract small patches from 3D volume to reduce the input size [24]; however, this also limits the effective receptive field of the network. In practice, 3D CNNs are also limited by the amount of available training data to ensure generalization.

2.3 2D MRI Recovery

There has been a long history of research focused on methods to undersample MR k-space data while maintaining image quality, in the hope of accelerating acquisition. Since undersampling in the Fourier domain leads to aliasing in the image domain, the problem is typically described as de-aliasing. Lustig et al. [5] first proposed to use Compressed Sensing in MRI (CSMRI), assuming that the undersampled MR images have a sparse representation in some transform domain, where noise can be discarded through minimizing the \mathcal{L}_0 norm of the representation. This method was shown to yield much better results than zero-filling the missing k-space samples (ZF); however, presuming a fixed sparse transform for every MR image often leads to secondary artifacts and limits recovery performance in practice. Extending on CSMRI, Ravishankar et al. [6] applied more adaptive sparse modelling through Dictionary Learning (DLMRI), where the transformation is optimized through specific sets of data, resulting in improved sparsity encoding. To further explore redundancy within the MR data, Huang et al. [25] found the anatomical similarity between T_1 and T_2 -weighted MR images by considering the group sparsity, while Hirabayashi et

al. [26] attempted to extrapolate the redundancy from adjacent slices.

Recently, Convolutional Neural Network (CNN) have been shown to obtain superior performance in many computer vision tasks due to its ability to generate efficient hierarchical features. This has inspired further research in applying CNN in the domain of MR reconstruction. Schlemper et al. [7] proposed a cascade of CNNs that incorporates data consistency layers to de-noise MRI in image domain while maintaining consistency in the k-space, and showed that the results are significantly better than produced by DLMRI. Yang et al. [8] proposed DAGAN, which recovers undersampled MR images through a U-Net structure with perceptual and adversarial loss in addition to L_1 loss in image space and frequency space. Quan et al. [9] proposed RefineGAN, which performs reconstruction and refinement through two different networks, and enforces an image space/frequency space cyclic loss.

Although all the above mentioned CNN-based methods have obtained impressive results, they focus on single sequence reconstruction. Few studies have been on exploring the multi-sequence scenario, which is common in practice. Xiang et al. [27] showed that a highly undersampled T_2 sequence, given a fully sampled T_1 sequence, can still be well-recovered through a Dense U-Net that takes two images as inputs and outputs the recovered T_2 image. Despite this work, there has not been a quantitative study done with regard to the best strategy at undersampling k-spaces over a range of sequences for image recovery.

Chapter 3: Problem Formulation

3.1 Slice Interpolation

Let $I(x, y, z) \in \mathbb{R}^{N \times N \times N}$ denote an isotropic MR volume. By convention, we refer the x axis as the “sagittal” axis, the y axis as the “coronal” axis, and the z axis as the “axial” axis. Accordingly, there are three types of slices:

- the sagittal slice for a given x : $I^x(y, z) = I(x, y, z), \forall x$;
- the coronal slice for a given y : $I^y(x, z) = I(x, y, z), \forall y$;
- the axial slice for a given z : $I^z(x, y) = I(x, y, z), \forall z$.

We also define a slab of s slices, say along the x axis, as

$$\mathbb{I}^{x,s} = \left\{ I^{x+l}(y, z) \mid l = -\frac{s-1}{2}, \dots, 0, \dots, \frac{s-1}{2} \right\}. \quad (3.1)$$

$\mathbb{I}^{y,s}$ and $\mathbb{I}^{z,s}$ are defined similarly. Without loss of generality, in this work we consider slice interpolation along the axial axis. From $I(x, y, z)$, the corresponding anisotropic MR volume is defined as

$$I_{\downarrow k}(x, y, z) = I(x, y, k \cdot z), \quad (3.2)$$

where k is the sparsity factor. The *goal of slice interpolation* is to find a transformation $\mathcal{T}: \mathbb{R}^{N \times N \times \frac{N}{k}} \rightarrow \mathbb{R}^{N \times N \times N}$ that can optimally transform $I_{\downarrow k}(x, y, z)$ back to $I(x, y, z)$.

There are two possible baseline realizations of \mathcal{T} using CNNs:

- **2D CNN.** More in line with traditional methods, a 2D CNN takes two adjacent slices $I_{\downarrow k}^z(x, y)$ and $I_{\downarrow k}^{z+1}(x, y)$ as inputs, and directly estimates the in-between missing slices. One major drawback of this approach is that a simple 2D CNN has limited capabilities of modeling the variations in highly anisotropic volumes.
- **3D CNN.** A 3D CNN is learned as a mapping from the sparsely sampled volume $I_{\downarrow k}(x, y, z)$ to a fully sampled volume $I(x, y, z)$. This straightforward approach, however, suffers from training memory issue and insufficient training data.

In this thesis, we present our proposed algorithm that retains the advantages of the baseline CNN models discussed above while mitigating their disadvantages.

3.2 Multi-sequence MR Recovery

We first note that the most popular MR k-space sampling method is through Cartesian trajectory, where a series of acquisition is performed along equally-spaced parallel lines, which is conventionally called *phase encoding lines*. This leads to a natural implementation for MR undersampling, where the technicians can drop certain phase encoding lines from the sampling grid [5]. In this chapter, we focus

on undersampling with 1D masks along the phase encoding direction, which is the method that Xiang et al [27] used.

Consider multiple MR sequences with full k-space spectra $\{F_s\}_{s=1}^S$, with each spectrum sampled by N phase encoding lines. For each F_s , the unit time for sampling a phase encoding line is denoted by t_s . We define 1D sampling masks $\mathcal{M}_s \in \{0, 1\}^N$ which selects a subset of encoding lines $\mathcal{M}_s \odot F_s$ for faster acquisition. By applying the inverse Fourier transform \mathcal{F}^{-1} , an undersampled MR image for sequence s is reconstructed as

$$I_{M_s} = \mathcal{F}^{-1}(\mathcal{M}_s \odot F_s). \quad (3.3)$$

When fully sampled, the MR image is reconstructed by $I_s = \mathcal{F}^{-1}(F_s)$. If we denote the number of selected encoding lines by $|\mathcal{M}_s|$, the total time needed to acquire all the sequences is

$$T = \sum_{s=1}^S t_s \times |\mathcal{M}_s|. \quad (3.4)$$

Although undersampled MR is shorter to acquire, it exhibits degraded quality compared to fully sampled MR. In this work, we aim to search for an optimal sampling strategy $\{\mathcal{M}_s\}_{s=1}^S$ and a deep neural network f_θ that optimally recovers fully sampled $\{I_s\}_{s=1}^S$ from $\{I_{M_s}\}$ with a time constraint $T \leq T_{max}$. This constrained optimization problem can be formulated as follows:

$$\min_{\theta, \{\mathcal{M}_s\}} \sum_{s=1}^S E_{I_s \sim p(I_s)} [\|f_\theta(I_{M_s}) - I_s\|_1] \quad \text{s.t.} \quad \sum_{s=1}^S t_s |\mathcal{M}_s| \leq T_{max}. \quad (3.5)$$

In (3.5), we use the L_1 loss; however, other loss functions can be used too.

The problem defined in (3.5) is *combinatorial* in nature. First, the set $\{\mathcal{M}_s\}_{s=1}^S$ has a total of 2^{NS} possible combinations. Secondly, the best recovery model depends

on the choice of sampling strategy. As a result, the optimal solution to (3.5) is in general difficult to find. As a preliminary attempt, we assume a fixed candidate set $\mathcal{C} \in \{m_1, \dots, m_F\}$ for each \mathcal{M}_s . The number of possible sampling strategies becomes F^S instead. However, even with the simplification, a straightforward approach to (3.5), which is

$$\min_{\mathcal{M}_{1:S} \in \mathcal{C}^S} \left(\min_{\theta} \sum_{s=1}^S E_{I_s \sim p(I_s)} [\|f_{\theta}(I_{\mathcal{M}_s}) - I_s\|_1] \right) \quad \text{s.t.} \quad \sum_{s=1}^S t_s |\mathcal{M}_s| \leq T_{max}, \quad (3.6)$$

still requires training F^S models and then choosing the one with minimum loss.

In this thesis, we propose an efficient approach that finds a reasonable $(\theta, \{\mathcal{M}_s\}_{s=1}^S)$ while circumventing the computational cost in training excessive number of models. Conceptually, we propose to first train a Blind Recovery Model (BRM), which takes randomly undersampled MR sequences as inputs, and recovers them to fully sampled MR sequences. The trained BRM can then be used as an MR sequence quality estimator to search for the optimal Sampling Strategy (SS) $\{\mathcal{M}_s^*\}_{s=1}^S$. Finally, with $\{\mathcal{M}_s^*\}_{s=1}^S$, we can proceed to solve (3.6) by fine-tuning on the existing BRM. In total, the proposed method only requires training one CNN, which significantly reduces the computational cost.

Chapter 4: Proposed Algorithms

4.1 Deep Slice Interpolation

We propose to break down the 3D slice interpolation problem into a series of 2D tasks, and interpolate the contextual information from all three anatomical views to achieve structurally consistent reconstruction and improved memory efficiency. The two stages are as follows:

- Marginal super-resolution (MSR), where we provide high-quality estimates of the interpolated slices by extrapolating context from sagittal and coronal axes.
- Two-view Fusion and Refinement (TFR), where we fuse the estimations and further refine with information from the axial axis.

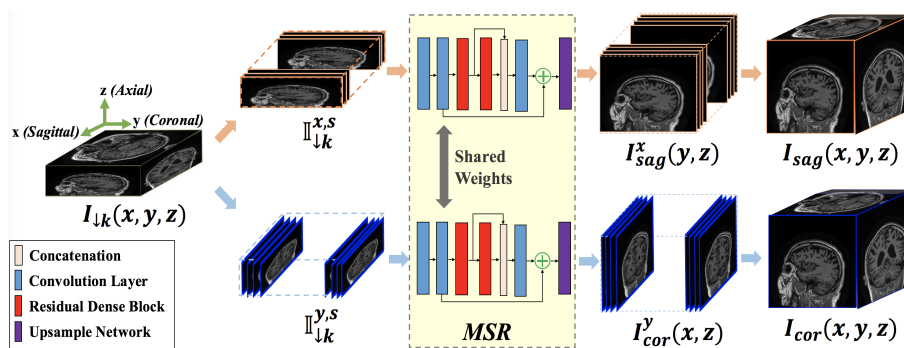


Figure 4.1: Marginal Super-Resolution Pipeline.

4.1.1 Marginal Super-Resolution

Fig. 4.4 demonstrates the pipeline of MSR. Given $I_{\downarrow k}(x, y, z)$, we view it as a sequence of 2D sagittal slices $I_{\downarrow k}^x(y, z)$ marginally from the sagittal axis. The same volume can also be treated as $I_{\downarrow k}^y(x, z)$ from the coronal axes. We observe that super-resolving $I_{\downarrow k}^x(y, z)$ to $I^x(y, z)$ and $I_{\downarrow k}^y(x, z)$ to $I^y(x, z)$ are equivalent to applying a sequence of 2D super-resolution along the x axis and y axis, respectively. Therefore, we apply a residual dense network (RDN) [21] \mathcal{M}_θ to upsample $I_{\downarrow k}^x(y, z)$ and $I_{\downarrow k}^y(x, z)$ as follows:

$$I_{sag}^x(y, z) = \mathcal{M}_\theta(\mathbb{I}_{\downarrow k}^{x,s}(y, z)), \quad I_{cor}^y(x, z) = \mathcal{M}_\theta(\mathbb{I}_{\downarrow k}^{y,s}(x, z)). \quad (4.1)$$

Notice that instead of super-resolving 2D slices independently, we propose to take a slab of s slices as input and estimate a single SR output. Using a larger s allows more context to be modelled. The MSR process is repeated for all x and y . Finally, the super-resolved slices can be reformatted as sagittally and coronally super-resolved volumes, $I_{sag}(x, y, z)$ and $I_{cor}(x, y, z)$, respectively. We apply the following L_1 loss to train the RDN:

$$\mathcal{L}_{MSR} = \|\mathcal{M}_\theta(\mathbb{I}_{\downarrow k}^{x,s}) - I_{gt}^x\|_1 + \|\mathcal{M}_\theta(\mathbb{I}_{\downarrow k}^{y,s}) - I_{gt}^y\|_1, \quad (4.2)$$

where $I_{gt}^x = I^x(y, z)$ and $I_{gt}^y = I^y(x, z)$ in the isotropic MR volume.

From the axial perspective, $I_{sag}(x, y, z)$ and $I_{cor}(x, y, z)$ provide line-by-line estimates for the missing axial slices. However, since no constraint is enforced on the estimated axial slices, inconsistent interpolations lead to noticeable artifacts (See Section 5.2.5). We resolve this problem in the second TFR stage of the proposed

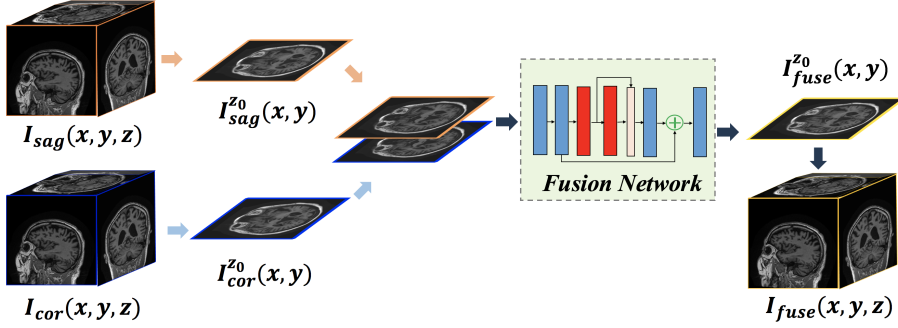


Figure 4.2: Two-view Fusion Pipeline.

pipeline.

4.1.2 Two-View Fusion and Refinement

The TFR stage is the counterpart of MSR which further improves the quality of slice interpolation by learning the structural variations along the axial direction.

As shown in Fig. 4.2, we first resample the sagittally and coronally super-resolved volumes $I_{sag}(x, y, z)$ and $I_{cor}(x, y, z)$ from the axial direction to obtain $I_{sag}^z(x, y)$ and $I_{cor}^z(x, y)$, respectively. A fusion network \mathcal{F}_ϕ takes the two slices as inputs and combines information from the two views. The objective function for training the fusion network is:

$$\mathcal{L}_{fuse} = \|I_{fuse}^z(x, y) - I_{gt}^z\|_1, \quad (4.3)$$

where $I_{fuse}^z(x, y) = \mathcal{F}_\phi(I_{sag}^z, I_{cor}^z)$ is the output of the fusion network, and $I_{gt}^z = I^z(x, y)$ in the isotropic MR volume. After training, the fusion network is applied to all the *interpolated* slices $\{I_{sag}^z \mid (z \bmod k) \neq 0\}$ and $\{I_{cor}^z \mid (z \bmod k) \neq 0\}$, yielding an MR volume $I_{fuse}(x, y, z)$.

After fusion, the interpolated slices already have visually pleasing qualities.

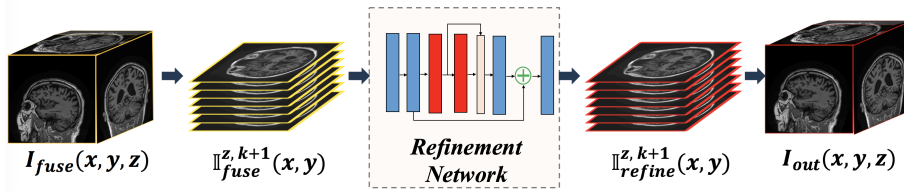


Figure 4.3: Refinement Pipeline.

Finally, to improve between-slice consistency along the axial axis, a refinement network \mathcal{R}_ψ takes a slab of $k + 1$ slices $\mathbb{I}_{fuse}^{z, k+1}$ as input and generates a consistent output slab $\mathbb{I}_{refine}^{z, k+1}$. The size is selected as $k + 1$ to make sure the refinement network has information from one or two *observed* slices. The pipeline is illustrated in Fig. 4.3. The loss function used in the network is given by:

$$\mathcal{L}_{refine} = \|\mathbb{I}_{refine}^{z, k+1} - \mathbb{I}_{gt}^{z, k+1}\|_1. \quad (4.4)$$

4.2 Multi-sequence Recovery

4.2.1 Blind recovery model

A blind recovery model (BRM) is a CNN f_θ which recovers I_s by fusing information from different undersampled MR sequences $\{I_{\mathcal{M}_s}\}_{s=1}^S$, $\mathcal{M}_s \in \mathcal{C}$. We adopt a data augmentation approach, which randomly selects sampling masks from \mathcal{C} , under the following *unconstrained optimization problem*:

$$\theta^* =_\theta \sum_{s=1}^S E_{I_s \sim p(I_s), \mathcal{M}_s \sim p(\mathcal{C})} [\|f_\theta(I_{\mathcal{M}_s}) - I_s\|_1]. \quad (4.5)$$

Our intuition is that for MR sequences, the more structural information discarded through harsh undersampling, the more difficult it is for a CNN to recover I from I_m , which leads to larger reconstruction loss. Therefore, the reconstruction loss of

an MR sequence can be used as a quality measure.

4.2.2 Sampling strategy searching

Given a trained BRM f_{θ^*} , we propose to search for the optimal sampling strategy by finding the one with minimum loss:

$$\mathcal{M}_{1:S}^* = \mathcal{M}_{1:S} \sum_{s=1}^S E_{I_s \sim p(I_s)} [\|f_{\theta^*}(I_{\mathcal{M}_s}) - I_s\|_1] \quad \text{s.t.} \quad \sum_{s=1}^S t_s |\mathcal{M}_s| \leq T_{max}. \quad (4.6)$$

The exhaustive search requires F^S forward passes, which is significantly less computationally demanding than training F^S CNNs. The solution θ^* can be further improved by learning a refined model specific to \mathcal{M}_s^* :

$$\hat{\theta} =_{\theta} \sum_{s=1}^S E_{I_s \sim p(I_s)} [\|f_{\theta}(I_{\mathcal{M}_s^*}) - I_s\|_1]. \quad (4.7)$$

4.2.3 An optimization point-of-view

In this section, we show that the proposed approach can be understood as a projected optimization method. For a general constrained optimization problem:

$$\min_x f(x) \quad \text{s.t.} \quad x \in C, \quad (4.8)$$

the projected gradient descent first updates the candidate solution x_t by

$$y_{t+1} = x_t - \eta \cdot \nabla f(x_t), \quad (4.9)$$

y_{t+1} is then projected back to the feasible set C by

$$x_{t+1} =_{x \in C} \|x - y_{t+1}\|. \quad (4.10)$$

When training BRM using (4.5), we are effectively performing a coordinate descent for (3.5) with respect to θ . In the subsequent sampling strategy searching, the solution is projected back onto the feasible set using (4.6) and (4.7).

4.2.4 Single-sequence Training vs Multi-sequence Training

Since the BRM takes multiple images from different sequences as inputs, one has the option of training (a) multiple SISO (single input single output) CNNs, with one per sequence, or (b) one monolithic MIMO (multiple input multiple output) CNN for all sequences. We believe that the latter option holds several advantages over the former. First, option (a) does not consider the complementary information across different sequences. As both Xiang et al. [27] and Huang et al. [25] have shown, there exists strong correlation between sequences on the same patient, as they share the underlying anatomic structures. If a particular sequence is severely undersampled, leading to the loss of some anatomic detail, such information may be present in other less severely undersampled sequences. Secondly, option (b) only requires training one model, while option (a) requires S models. As all the models attempt to eliminate distortions due to undersampling, they should learn similar features. Consequentially, the models in option (a) either share mostly similar features, leading to inefficiency, or learn features that are tuned to particular sequences, leading to less generality in features.

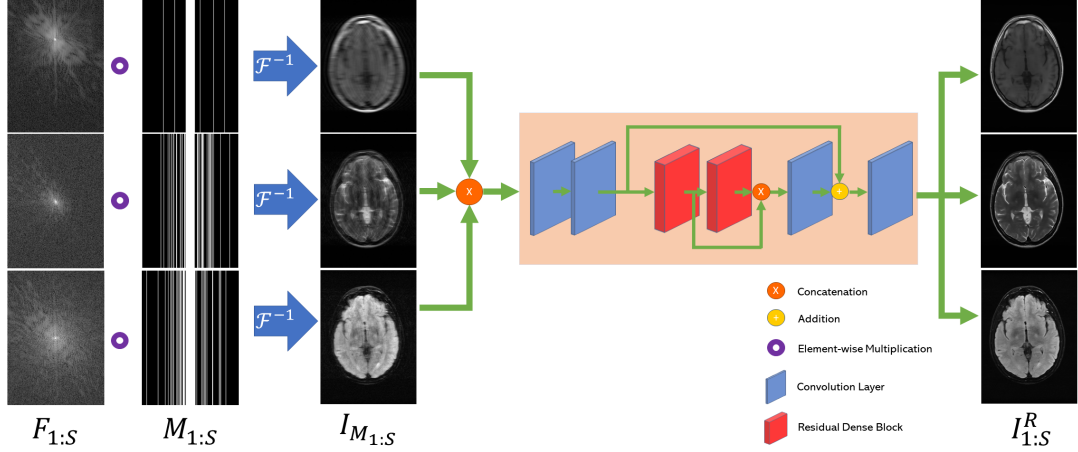


Figure 4.4: Multi-sequence recovery pipeline with the masks M_s randomly selected.

4.2.5 Network Architecture

Our multi-sequence simultaneous recovery approach is shown in Fig 4.4. The approach is based on Residual Dense Block (RDB) [21], which incorporates the idea of residual learning and Dense Block [22], allowing all layers of features to be seen directly by other layers. It has been shown that RDB achieves state-of-the-art performance in the domain of Super-Resolution. Since MR recovery aims at eliminating noise caused by undersampling, we believe that an RDB-based framework can be effective.

During learning, each raw k-space data F_s first gets undersampled through a randomly generated mask \mathcal{M}_s . The results are then transformed from k-space to image space, and concatenated before fed to the recovery network, which outputs $I_{1:S}^R$. The loss function is defined as follows:

$$\mathcal{L} = \|I_{1:S}^R - I_{1:S}\|_1. \quad (4.11)$$

Chapter 5: Experimentation

5.1 Settings

5.1.1 Implementation Details

We implement the proposed frameworks using PyTorch¹. The RDN [21] architecture with two RDBs are used as the basic unit for our networks. For fusion, refinement, and baseline 2D CNN models, where the inputs and outputs have the same image size, we replace the upsampling network in RDN with one convolutional layer. The input to the MSR network has $s = 3$. Note that due to memory constraints, 3D CNN only uses one RDB. We train the models with Adam optimization, with a momentum of 0.5 and a learning rate of 0.0001, until they reach convergence.

5.2 Deep Slice Interpolation

5.2.1 Data sets

We employ 120 T1 MR brain scans from the publicly available Alzheimer’s Disease Neuroimaging Initiative (ADNI) data sets. The MR scans are isotropically

¹<https://pytorch.org>

sampled at $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$, and zero-padded to $256 \times 256 \times 256$ pixels, ending up with 30,720 slices in each of sagittal, coronal, and axial directions. We further down-sample the isotropic volumes by factors of $k = 4$ and $k = 8$, yielding $I_{\downarrow k}(x, y, z)$ of sizes $256 \times 256 \times 64$ and $256 \times 256 \times 32$, respectively. The data is split into training/validation/testing sets with 95/5/20 samples. Note that during test time, we only select slices that contain mostly brain tissue; the number of samples for each sparsity are presented in Table 5.3.

5.2.2 Evaluation metrics

We compare different slice interpolation approaches using two types of quantitative metrics. First, we use Peak Signal-to-Noise Ratio (PSNR) and Structured Similarity Index (SSIM) to measure low-level image quality. Second, we evaluate the quality of the interpolated slices through gray/white-matter segmentation. The segmentation network has a U-Net architecture, which is one of the winning models in MRBrainS challenge [28], and is trained on the OASIS data sets [29]. Dice Coefficient (DICE) and Hausdorff Distance (HD)² between the segmentation maps of ground truth slices and generated slices are calculated. Due to the memory limitation of 3D CNNs, we can at most super-resolve a limited region of $144 \times 144 \times 256$ pixels during evaluation. For fair comparisons, the evaluation metrics are calculated over the same region across all methods.

²To reduce the effect of outliers, HD is calculated on the 90th percentile displacement.

5.2.3 Quantitative Evaluations

In this section, we evaluate the performance of our method and baseline approaches. Quantitative comparisons are presented in Table 5.3. We observe that all the three CNN-based methods have higher PSNR and SSIM than the widely used linear interpolation. The 3D CNN-based method slightly outperforms 2D CNN in 4x sparsity, but performs worse in 8x sparsity. Among the three CNN methods, our method consistently outperforms 2D CNN and 3D CNN baselines.

The performance gain in accurately segmenting gray and white matters is large from linear interpolation to baseline CNN-based methods. However, at 8x sparsity, the HD scores of linear interpolation are comparable with 2D CNN and 3D CNN, while our method outperforms these approaches by at least 10%. This demonstrates the robustness of our method even at very high sparsity.

5.2.4 Visual Comparisons

In Fig. 5.1, we present the observed slices $I_{\downarrow k}^z$ and $I_{\downarrow k}^{z+1}$ along with the interpolated slices produced by different methods. Specifically we demonstrate the second of three interpolated MR slices for 4x sparsity, and the third of seven interpolated slices for 8x sparsity. We highlight the region where the anatomical structures significantly change compared to the observed slices $I_{\downarrow k}^z$ and $I_{\downarrow k}^{z+1}$. We observe that although 2D CNN has comparable performance in terms of PSNR and SSIM, it tends to produce false anatomical structures in the zoomed regions. 3D CNN is able to resolve more accurate details. However, the improvement is quite limited,

Sparsity	Method	PSNR(dB)	SSIM	DICE	HD(90th pct.)
				GM/WM	GM/WM
4	LI	26.39	0.8317	0.7716/0.7296	3.607/7.965
	2D CNN	31.24	<u>0.9313</u>	<u>0.8813/0.8334</u>	3.176/12.36
	3D CNN	<u>31.34</u>	0.9292	0.8536/0.8265	<u>2.898/7.373</u>
	Ours	32.22	0.9441	0.9021/0.8593	2.494/6.240
8	LI	23.45	0.7165	0.6611/0.6105	4.487/10.59
	2D CNN	<u>27.88</u>	<u>0.8444</u>	<u>0.7783/0.7425</u>	<u>4.322/12.84</u>
	3D CNN	27.38	0.8390	0.7684/ <u>0.7468</u>	4.583/ <u>9.017</u>
	Ours	28.87	0.8808	0.8189/0.7828	3.960/8.127

Table 5.1: Quantitative evaluations for different slice interpolation approaches. For DICE and HD performance metrics, we present results on gray matter (GM)/white matter (WM) segmentation. The best results are in **bold** and the second best are underlined.

which we attribute to the fact that 3D CNN requires more training MR volumes in order to generalize and has smaller receptive field due to patch-based training. Our method benefits from the large receptive field of 2D CNN and two-view fusion, which not only produces sharper images, but also correctly estimates the brain anatomy. The sharp and accurate estimation is crucial in clinical applications such as diagnosing Alzheimer’s Disease by brain volume estimation.

In Fig. 5.2, we demonstrate the advantage of the proposed method in brain

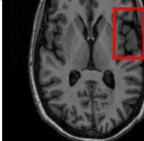
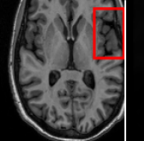
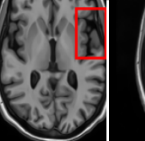
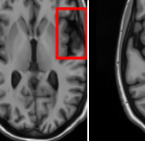
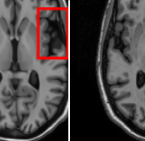
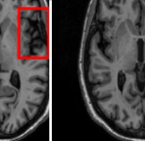
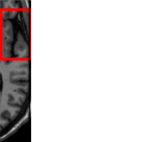
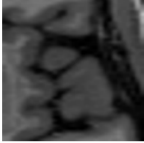
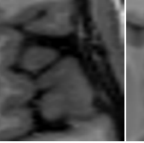
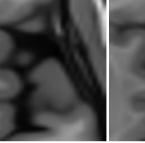
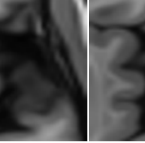
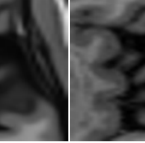
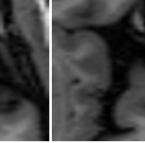

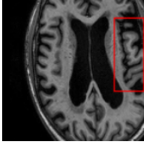
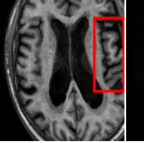
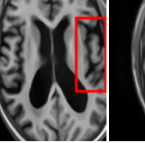
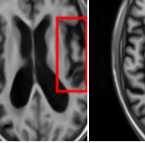
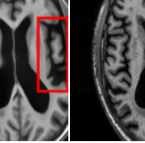
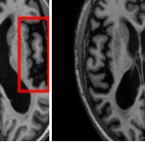
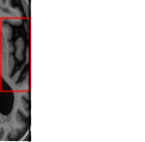

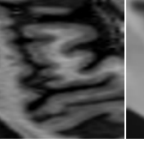


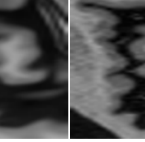
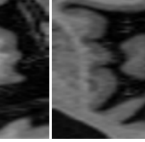

Sparsity	$I_{\downarrow k}^z$	LI	2D CNN	3D CNN	Ours	GT	$I_{\downarrow k}^{z+1}$
4							
							
	27.37/0.8465 32.34/0.9441 32.72/0.9436 34.11/0.9607 PSNR(dB)/SSIM						
8							
							
	25.51/0.7681 28.29/0.8205 29.51/0.8824 31.87/0.9249 PSNR(dB)/SSIM						

Figure 5.1: Visual comparisons of slice interpolation approaches. For 4x sparsity, the second of three interpolated MR slices is presented. For 8x sparsity, the third of seven interpolated slices is presented.

matter segmentation. It is clear that although 2D and 3D CNN-based methods generate visually plausible interpolation as presented in Fig. 5.1, the brain matters are easily misclassified due to incorrect anatomical structures and blurred details.

5.2.5 Ablation study

In this section, based on 4x sparsity, we evaluate the effectiveness of each proposed components. The following settings are considered:

- MSR_{sag}^n : Slice interpolation based on only sagittal view MSR. We consider number of input slices $n = 1, 3$.

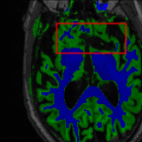
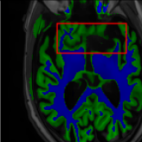
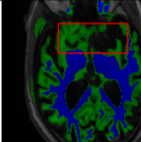
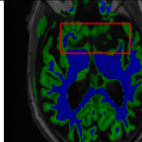
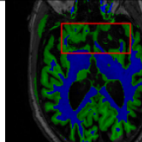
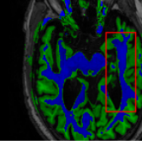
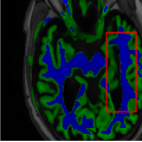
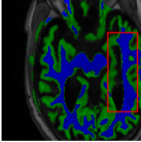
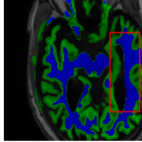
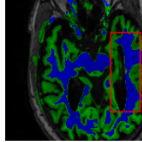
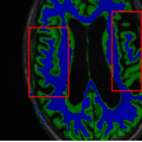
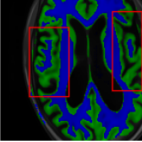
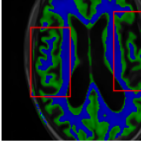
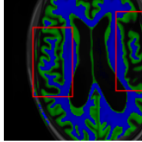
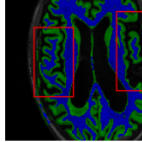
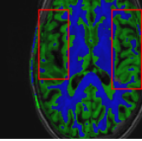
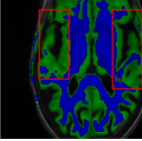
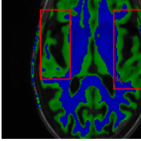
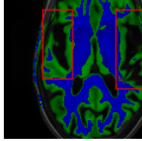
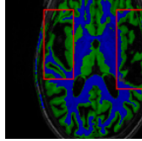
Sparsity	LI	2D CNN	3D CNN	Ours	GT
4	 0.6787/0.7972	 0.8143/0.8776	 0.8190/0.8714	 0.8664/0.9085	 GM/WM
	 0.6808/0.7161	 0.8103/0.8631	 0.7950/0.8606	 0.8598/0.9115	 GM/WM
8	 0.5139/0.7240	 0.6619/0.8224	 0.6878/0.8584	 0.7798/0.8853	 GM/WM
	 0.5910/0.6947	 0.6516/0.8021	 0.6507/0.8186	 0.7471/0.8540	 GM/WM

Figure 5.2: Visual comparison of gray matter (Green)/white matter (Blue) segmentation over different methods, with respective DICE scores listed under the images.

- MSR_{cor}^n : Slice interpolation based on only coronal view MSR. We consider number of input slices $n = 1, 3$.
- Fused: Slice interpolation with fusion network. Inputs to the network are MSR_{sag}^3 and MSR_{cor}^3 .
- Refined: The proposed full pipeline.

From Table 5.2, it is clear that each proposed component improves the quality of slice interpolation. Notice that even without fusion and refinement, the axial slices interpolated by MSR_{sag}^3 and MSR_{cor}^3 are already better than the baseline 2D/3D CNNs.

Stage	PSNR (dB)	SSIM
baseline 2D CNN	31.24	0.9313
baseline 3D CNN	31.34	0.9292
MSR_{sag}^1	30.28	0.9129
MSR_{cor}^1	30.56	0.9178
MSR_{sag}^3	31.43	0.9314
MSR_{cor}^3	31.61	0.9339
Fused	<u>32.02</u>	<u>0.9413</u>
Refined	32.22	0.9441

Table 5.2: Quantitative ablation study. Baseline numbers are also included for comparison. The best results are in **bold** and the second best are underlined.

Visual comparisons are shown in Fig. 5.3, where we select a challenging slice with abundant anatomical details. From Fig. 5.3, it is clear that marginally super-resolving axial slices from coronal and sagittal views leads to noticeable horizontal (MSR_{sag}^n) and vertical (MSR_{cor}^n) artifacts. Furthermore, some small details are better resolved by MSR_{sag}^3 , while others are better resolved by MSR_{cor}^3 . The fusion network combines the features from MSR_{sag}^3 and MSR_{cor}^3 , which effectively reduces inconsistency. With additional axial information, the fused slice is then further improved by the refinement network.

In addition to L_1 loss, we also experiment on GAN loss at refinement stage. However, we find that GAN tends to generate fake anatomical details, which is

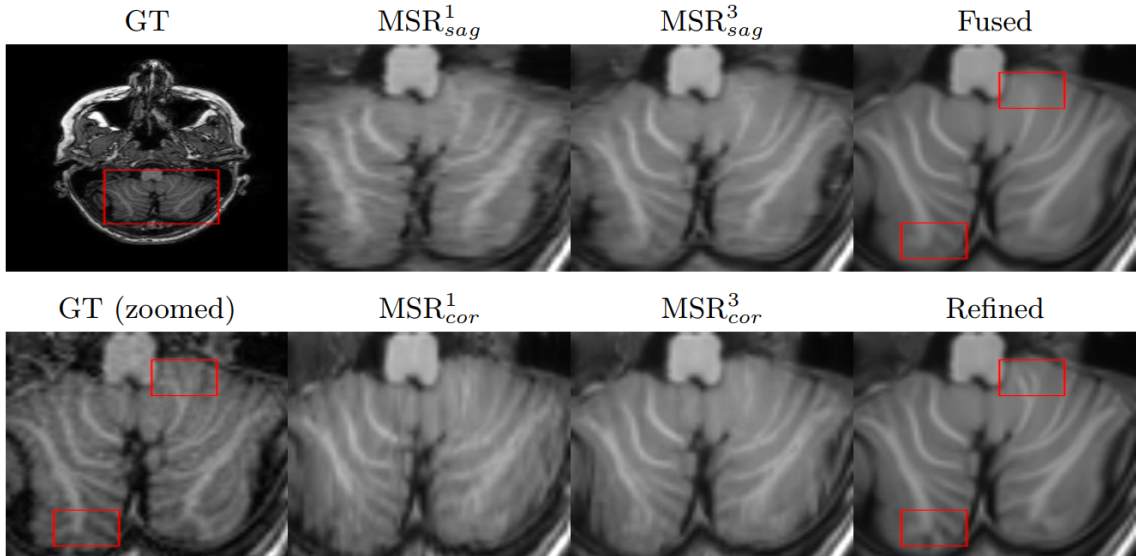


Figure 5.3: Visual comparison for the proposed components.

undesirable in medical applications.

5.3 Multi-Sequence Recovery

5.3.1 Data sets

We employ two datasets. The first one is a privately collected, k-space raw data of three sequences (T_1 , T_2 , FLAIR) from 20 patients, with each sequence containing 18 slices. The sequences are co-registered and taken with an MRI machine with 8 channels; in order to augment training, we treat each channel as an individual image to result in a total of 2,880 three-sequence images, which are divided into a ratio of 17:1:2 for training, validation, and testing. We refer to this dataset as “real data”. In order to further validate our research, we also employ the Brain Tumor Image Segmentation (BraTS) dataset [30, 31], which contains T_1 , T_2 , and FLAIR.

The sequence are co-registered to the same anatomical template, skull-stripped, and interpolated to the same resolution. We divide the selected 167 cases into a ratio of 140:10:17 for training, validation, and testing. From every case, we select the middle 60 slices that contain most of the anatomical details. Because BraTS does not provide raw k-space data, we follow common practices [8,27] to simulate k-space data. We refer to this dataset as “simulated data”. *Below, our insights are first demonstrated with experiments on real data and are further validated on simulated data.*

5.3.2 Acquisition time and undersampling settings

In general, T_2 and FLAIR have longer repetition time (TR) than T_1 ; however, the acquisition time of each sequence also depends on the number of excitations. A larger number of excitations helps better resolve sequences but take a longer time; therefore, the acquisition time of each sequence is rather machine-dependent. Here we consider three experimental settings: $t_{T_1}:t_{T_2}:t_{flair} =$ (1) 1:1:1, (2) 1:4:6, and (3) 2:3:6.

We experiment on both low-pass sampling [27] and random sampling [8]. We found that random sampling works better on real data but worse on simulated data. As our approach is agnostic of sampling strategy, we choose the better performing sampling strategy for each dataset.

It is worth noting that during training of BRM, the masks $\mathcal{M}_{1:S}$ are generated based on a random $\lambda_s \in [1, k]$, where k is the maximum undersampling factor that we

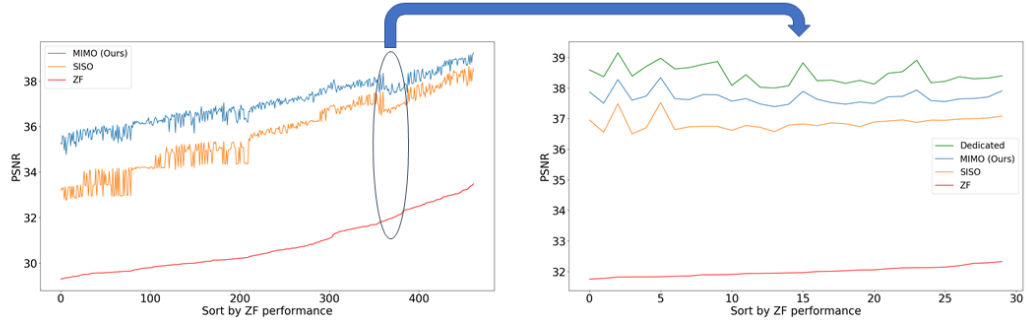


Figure 5.4: Quantitative recovery performance comparison. The Pearson correlation coefficient between Dedicated and MIMO vs between Dedicated and ZF is 0.85 vs -0.33 in the selected range

empirically set to 8. This means that BRM, after training, can handle a continuous set of undersampling factors on every sequence.

5.3.3 Evaluation metrics

We utilize two metrics to gauge image quality: PSNR (peak signal-to-noise ratio) and SSIM (structural similarity). Since we mainly focus on three sequences, calculation of these metrics on three-sequence outputs is the same as on RGB images. This is easily extensible with a larger number of sequences. MRI images do not have a fixed dynamic range, as a consequence PSNR values should be regarded in terms of relative improvements. For example, a T_2 image tends to have a lower PSNR as it has the highest peak out of all three sequences.

5.3.4 Results

We evaluate the effectiveness of BRM in order to empirically demonstrate that a properly trained network f_θ performs well regardless of the choices of $\mathcal{M}_{1:S}$, and serves as a good estimator of best sampling strategy. Furthermore, we want to show that MIMO BRM performs better than SISO BRM.

The study is done by training (i) one MIMO BRM, (ii) three SISO BRM for three sequences, and (iii) many models that are dedicated for specific sampling ratios. All the models follow the same structure as shown in Fig. 4.4. The proposed training scheme for continuous $\lambda_s \in [1, k]$ allows us to efficiently investigate the performance of different undersampling strategies. For each acquisition time setting $\{t_s\}_{s=1}^S$, we search through possible $\{\lambda_s\}_{s=1}^S$ on the following simplex: $\sum_{s=1}^S \frac{t_s}{\lambda_s} = T_{max}$, which maximally utilizes the budgeted time T_{max} . We select hundreds of $\{\lambda_s\}_{s=1}^S$ under the 1:1:1 time setting, and set $T_{max} = \frac{T}{4}$, or 75% reduction in time. We run the trained models on the test set, and plot the reconstruction performances in Fig. 5.4. The top-three performing sampling strategies for different acquisition time setting are shown in Table 5.3.

Fig. 5.4 shows a clear performance gap between MIMO and SISO. Overall, the reconstruction performance of ZF images is the good indicator of the performances of BRMs; however, the correlation fluctuates often, and two sets of ZF that are similar in PSNR can swing for more than 1dB after going through BRM. To limit the number of dedicated models we need to train, we select a range of sampling factors of which ZF performance does not correlate well with MIMO/SISO performance,

and train 30 dedicated models to see how well BRM predicts the performance of dedicated models. As we observe from the right image in Fig. 5.4, our BRM, both

$t_{T_1}:t_{T_2}:t_{flair}$	$\lambda_{T_1}, \lambda_{T_2}, \lambda_{flair}$	ZF	SISO	MIMO	MIMO (tuned)
1 : 1 : 1	6.6, 2.1, 8.0	33.48/0.918	38.57/0.980	39.24/0.984	40.00/0.987
Real	8.00, 2.11, 6.63	33.43/0.920	38.36/0.979	39.16/0.984	40.07/0.987
	7.25, 2.11, 7.25	33.39/0.918	38.50/0.980	39.15/0.984	40.07/0.986
1 : 4 : 6	2.90, 2.44, 7.82	33.81/0.926	38.85/0.983	39.33/0.985	40.28/0.988
Real	3.01, 2.44, 7.69	33.60/0.924	38.83/0.983	39.32/0.985	40.37/0.987
	3.93, 2.44, 6.99	33.58/0.925	38.81/0.983	39.31/0.986	40.13/0.987
1 : 1 : 1	5.66, 3.14, 3.93	32.21/0.887	37.69/0.974	38.32/0.978	38.99/0.980
Simulated	5.27, 3.41, 3.74	32.31/0.889	37.88/0.975	38.31/0.979	38.98/0.980
	6.10, 3.14, 3.74	32.21/0.887	37.51/0.973	38.31/0.978	38.99/0.980
2 : 3 : 6	2.61, 3.74, 5.16	32.87/0.899	38.01/0.976	38.67/0.980	39.37/0.982
Simulated	2.44, 3.74, 5.40	32.84/0.899	37.87/0.975	38.66/0.980	39.35/0.982
	2.61, 3.41, 5.66	32.82/0.899	37.80/0.975	38.65/0.980	39.33/0.982

Table 5.3: Quantitative evaluations for the top performing $\lambda_{1:S}$ under different acquisition time assumption. The performance numbers presented here are PSNR (dB) and SSIM.

from MIMO and SISO settings, predicts the performance of dedicated models with a high correlation. We further choose the best three $\{\lambda_s\}_{s=1}^S$, and perform the last stage of fine-tuning accordingly to (4.7). A visual evaluation on real data is shown

in Fig. 5.5, 5.6, 5.7, 5.8, 5.9.

Base on the best performing $\{\lambda_s\}_{s=1}^S$, we perceive that among T_1 , T_2 , and FLAIR, the results are best when T_2 is sampled the most. We suggest that this makes intuitive sense as T_2 images provide the best contrast out of the three sequences, which can compensate for the details lost in other images. The same observation can be made on the simulated data, where both T_2 and FLAIR show good contrast. When the time setting is changed to non-uniformity, we can see that our search for the best sampling strategy reflects the change. T_1 is sampled more as a result of faster acquisition time, while T_2 is still sufficiently sampled.

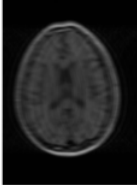
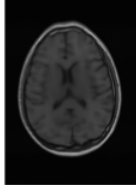
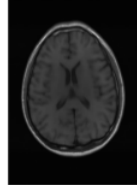
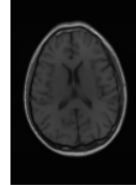
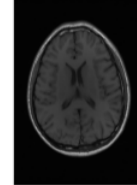
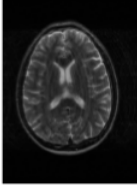
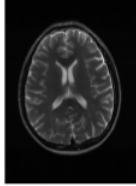
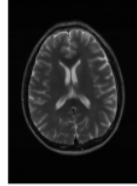
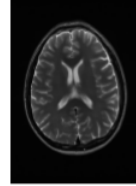
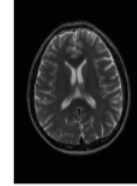
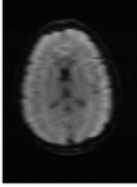
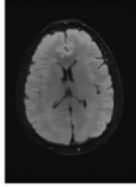
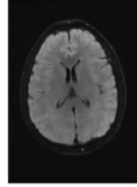
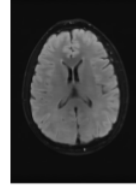
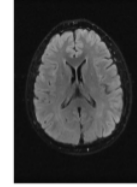
Sequence	LR	SISO	MIMO	MIMO tuned	GT
$\lambda_{T_1} = 6.63$					
	34.38/0.9371	42.42/0.9883	44.60/0.9920	45.50/0.9940	PSNR/SSIM
$\lambda_{T_2} = 2.11$					
	29.74/0.8903	36.25/0.9734	36.42/0.9752	37.70/0.9832	PSNR/SSIM
$\lambda_{flair} = 8.00$					
	39.89/0.9311	43.94/0.9864	44.74/0.9883	45.49/0.9894	PSNR/SSIM

Figure 5.5: Visual comparison of different recovery methods on real data, with PSNR and SSIM values listed under the images. After successful recovery, the images become sharper with more visible details.

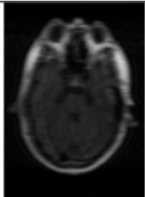
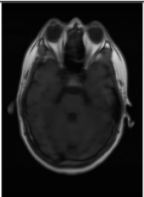
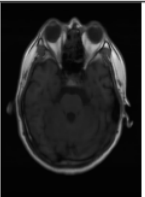
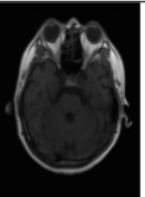
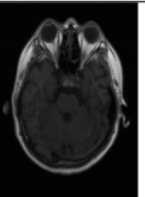
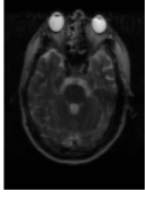
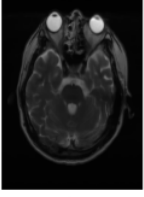
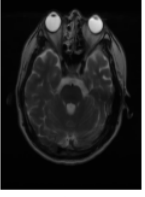
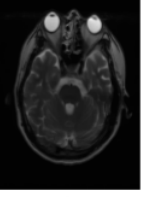
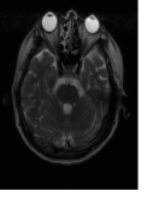
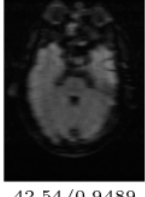
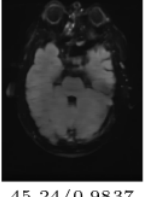
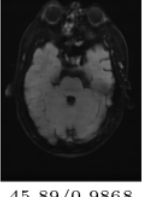
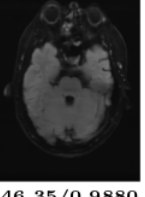
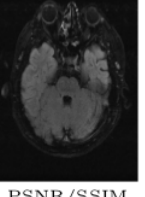
Sequence	LR	SISO	MIMO	MIMO tuned	GT
$\lambda_{T_1} = 6.63$	 35.46/0.9431	 40.82/0.9826	 41.84/0.9857	 42.12/0.9867	 PSNR/SSIM
$\lambda_{T_2} = 2.11$	 32.34/0.9254	 36.19/0.9699	 36.42/0.9711	 36.78/0.9695	 PSNR/SSIM
$\lambda_{flair} = 8.00$	 42.54/0.9489	 45.24/0.9837	 45.89/0.9868	 46.35/0.9880	 PSNR/SSIM

Figure 5.6: Additional visual comparison of different recovery methods on real data.

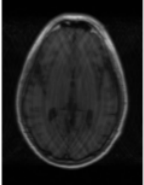
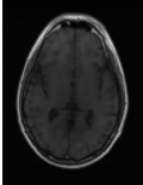
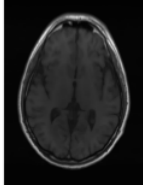
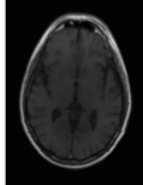
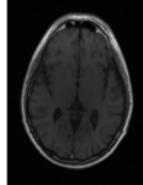
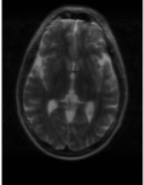
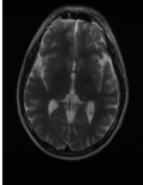
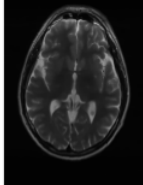
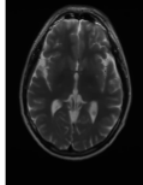
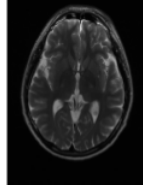
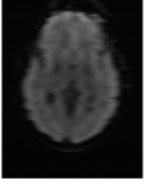
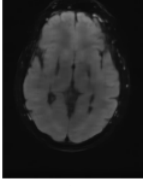
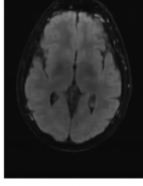
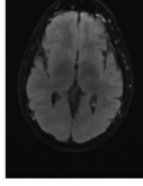
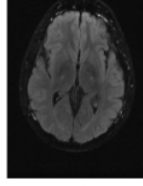
Sequence	LR	SISO	MIMO	MIMO tuned	GT
$\lambda_{T_1} = 2.90$	 38.30/0.9484	 45.03/0.9920	 45.35/0.9926	 46.70/0.9951	 PSNR/SSIM
$\lambda_{T_2} = 2.44$	 29.78/0.8990	 35.43/0.9720	 35.93/0.9752	 37.05/0.9809	 PSNR/SSIM
$\lambda_{flair} = 7.82$	 41.24/0.9412	 44.54/0.9850	 45.66/0.9885	 46.15/0.9891	 PSNR/SSIM

Figure 5.7: Additional visual comparison of different recovery methods on real data, with a different time setting.

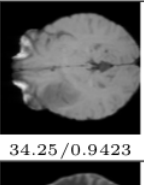
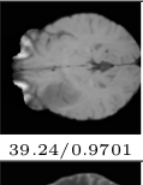
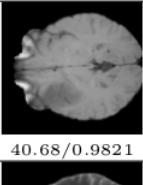
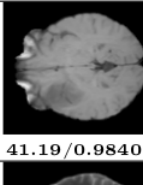
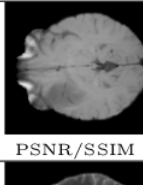
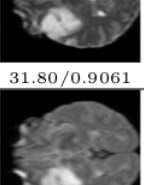
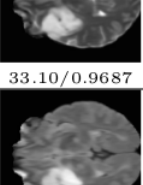
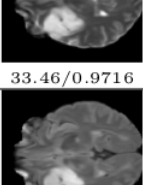
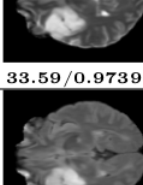
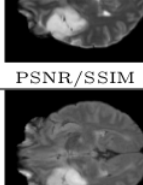
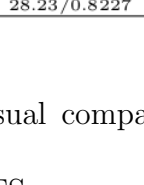
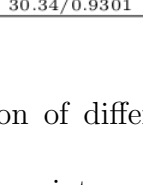
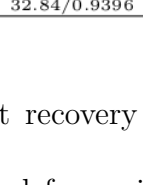
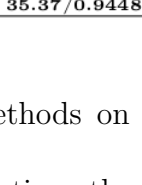
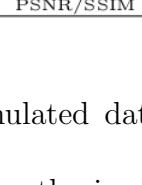
Sequence	LR	SISO	MIMO	MIMO tuned	GT
$\lambda_{T_1} = 2.61$	 34.25/0.9423	 39.24/0.9701	 40.68/0.9821	 41.19/0.9840	 PSNR/SSIM
$\lambda_{T_2} = 3.74$	 31.80/0.9061	 33.10/0.9687	 33.46/0.9716	 33.59/0.9739	 PSNR/SSIM
$\lambda_{flair} = 5.16$	 28.23/0.8227	 30.34/0.9301	 32.84/0.9396	 35.37/0.9448	 PSNR/SSIM

Figure 5.8: Visual comparison of different recovery methods on simulated data. Note that BraTS sequences are interpolated for registration; therefore the image quality is not as good as the real data.

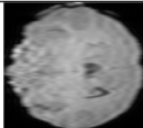
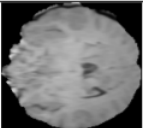
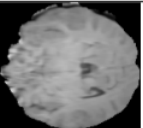
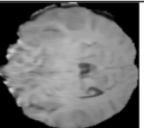
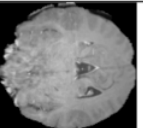
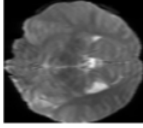
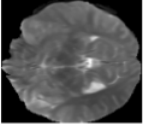
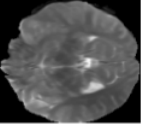
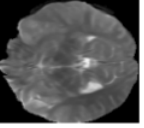
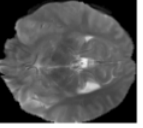
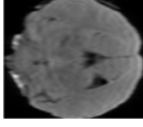
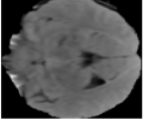
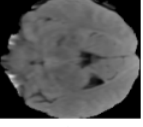
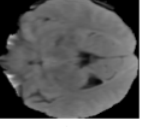
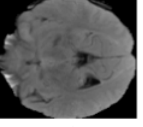
Sequence	LR	SISO	MIMO	MIMO tuned	GT
$\lambda_{T_1} = 5.66$	 25.06/0.7024	 29.79/0.8501	 29.86/0.8437	 30.41/0.8514	 PSNR/SSIM
$\lambda_{T_2} = 3.14$	 30.01/0.8725	 36.02/0.9500	 35.65/0.9459	 36.49/0.9523	 PSNR/SSIM
$\lambda_{flair} = 3.39$	 28.53/0.8360	 35.25/0.9270	 35.44/0.9311	 36.90/0.9396	 PSNR/SSIM

Figure 5.9: Visual comparison of different recovery methods on simulated data, with a different time setting.

Chapter 6: Conclusion

6.1 Deep Slice Interpolation

We proposed a multi-stage 2D CNN approach called deep slice interpolation which allows us to recover missing slices with high quality, even when the distance between observed slices are sparsely sampled. We evaluated our approach on a large ADNI data sets, demonstrating that our method outperforms possible 2D/3D CNN baselines both visually and quantitatively. Furthermore, we have illustrated that the MR slices estimated by the proposed method have superior segmentation accuracy. In the future, we plan to investigate the potential application of the proposed framework on real screening MRI images which often have a very low slice density.

6.2 Multi-sequence Recovery

We formulated multi-sequence MR recovery as a constrained optimization problem, and explored possible methods to solve such a problem. We proposed a CNN-based approach that has been experimentally proven to be degradation-agnostic, and an optimization scheme that helps us to find the best combination

of sampling strategy and recovery model on input sequences. We evaluated our approach on the BraTS data sets, demonstrating that our method finds the sampling strategy that agrees with the k-space energy distribution of each sequences, and that it outperforms single sequence recovery methods in recovery quality as well as in time and space complexity. While this work is a preliminary study on a complex problem, we believe that it builds the foundation for further researches in multi-sequence MR recovery. In the future, we plan to investigate questions that have arisen in this work, including finding the optimal sampling pattern for all sequences and the performance of a blind recovery model against noise introduced by randomized sampling pattern.

6.3 Future Work

In the thesis, we introduced two novel methods in two different, but practical scenarios often encountered in accelerating MRI acquisition. Deep Slice Interpolation provides a great framework in generating realistic slices; however, in its current form, it still relatively inhibited by the input size and cannot upsample to arbitrary output size without repetitive retraining or lower output quality. Further research, whether it is in knowledge distilling or network memorization, is needed to make this possible. Multi-sequence Recovery, on the other hand, is inhibited by the fact that there exist little public multi-sequence data sets to be tested with this method. Furthermore, current multi-sequence data sets do not take consistent number of slices and spatial locations into account, which creates difficulty in applying our method.

As a result, a very practical extension of this work will be to create a method that can create spatially registered, slice-wise consistent volumes. From this perspective, Deep Slice Interpolation can be thought of as the part that helps reach slice-wise consistency.

Bibliography

- [1] F. Bloch. Nuclear induction. *Phys. Rev.*, 70:460–474, Oct 1946.
- [2] E. M. Purcell, H. C. Torrey, and R. V. Pound. Resonance absorption by nuclear magnetic moments in a solid. *Phys. Rev.*, 69:37–38, Jan 1946.
- [3] G. A. Wright. Magnetic resonance imaging. *IEEE Signal Processing Magazine*, 14(1):56–66, Jan 1997.
- [4] Daniel K. Sodickson and Warren J. Manning. Simultaneous acquisition of spatial harmonics (smash): Fast imaging with radiofrequency coil arrays. *Magnetic Resonance in Medicine*, 38(4):591–603.
- [5] Michael Lustig, David Donoho, and John M Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.
- [6] Saiprasad Ravishankar and Yoram Bresler. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Trans. Med. Imaging*, 30(5):1028–1041, 2011.
- [7] Jo Schlemper, Jose Caballero, Joseph V. Hajnal, Anthony N. Price, and Daniel Rueckert. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans. Med. Imaging*, 37(2):491–503, 2018.
- [8] Guang Yang, Simiao Yu, Hao Dong, Gregory G. Slabaugh, Pier Luigi Dragotti, Xujiong Ye, Fangde Liu, Simon R. Arridge, Jennifer Keegan, Yike Guo, and David N. Firmin. DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Trans. Med. Imaging*, 37(6):1310–1321, 2018.
- [9] Tran Minh Quan, Thanh Nguyen-Duc, and Won-Ki Jeong. Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. *IEEE Trans. Med. Imaging*, 37(6):1488–1497, 2018.

- [10] A. Ardeshir Goshtasby, David A. Turner, and Laurens V. Ackerman. Matching of tomographic slices for interpolation. *IEEE Trans. Med. Imaging*, 11(4):507–516, 1992.
- [11] George J. Grevera and Jayaram K. Udupa. Shape-based interpolation of multidimensional grey-level images. *IEEE Trans. Med. Imaging*, 15(6):881–892, 1996.
- [12] Tong-Yee Lee and Wen-Hsiu Wang. Morphology-based three-dimensional interpolation. *IEEE Trans. Med. Imaging*, 19(7):711–721, 2000.
- [13] Graeme P. Penney, Julia A. Schnabel, Daniel Rueckert, Max A. Viergever, and Wiro J. Niessen. Registration-based interpolation. *IEEE Trans. Med. Imaging*, 23(7):922–926, 2004.
- [14] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *CoRR*, abs/1501.00092, 2015.
- [15] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. *CoRR*, abs/1511.04587, 2015.
- [16] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep CNN denoiser prior for image restoration. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 2808–2817, 2017.
- [17] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 1637–1645, 2016.
- [18] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. *CoRR*, abs/1608.00367, 2016.
- [19] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. *CoRR*, abs/1707.02921, 2017.
- [20] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 105–114, 2017.
- [21] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. *CoRR*, abs/1802.08797, 2018.

- [22] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. Densely connected convolutional networks. *CoRR*, abs/1608.06993, 2016.
- [23] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. Generative adversarial networks. *CoRR*, abs/1406.2661, 2014.
- [24] Yuhua Chen, Feng Shi, Anthony G. Christodoulou, Zhengwei Zhou, Yibin Xie, and Debiao Li. Efficient and accurate MRI super-resolution using a generative adversarial network and 3d multi-level densely connected network. *CoRR*, abs/1803.01417, 2018.
- [25] Junzhou Huang, Chen Chen, and Leon Axel. Fast multi-contrast MRI reconstruction. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2012 - 15th International Conference, Nice, France, October 1-5, 2012, Proceedings, Part I*, pages 281–288, 2012.
- [26] Akira Hirabayashi, Norihito Inamuro, K Mimura, T Kurihara, and T Homma. Compressed sensing mri using sparsity induced from adjacent slice similarity. In *Sampling Theory and Applications (SampTA), 2015 International Conference on*, pages 287–291. IEEE, 2015.
- [27] Lei Xiang, Yong Chen, Weitang Chang, Yiqiang Zhan, Weili Lin, Qian Wang, and Dinggang Shen. Ultra-fast t2-weighted MR reconstruction using complementary t1-weighted information. In *Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*, pages 215–223, 2018.
- [28] Zeynettin Akkus, Alfiya Galimzianova, Assaf Hoogi, Daniel L. Rubin, and Bradley J. Erickson. Deep learning for brain MRI segmentation: State of the art and future directions. *J. Digital Imaging*, 30(4):449–459, 2017.
- [29] Daniel S. Marcus, Tracy H. Wang, Jamie Parker, John G. Csernansky, John C. Morris, and Randy L. Buckner. Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cognitive Neuroscience*, 19(9):1498–1507, 2007.
- [30] Bjoern H. Menze, András Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin S. Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, Levente Lenczi, Elizabeth R. Gerstner, Marc-André Weber, Tal Arbel, Brian B. Avants, Nicholas Ayache, Patricia Buendia, D. Louis Collins, Nicolas Cordier, Jason J. Corso, Antonio Criminisi, Tilak Das, Herve Delingette, Çagatay Demiralp, Christopher R. Durst, Michel Dojat, Senan Doyle, Joana Festa, Florence Forbes, Ezequiel Geremia, Ben Glocker, Polina Golland, Xiaotao Guo, Andac Hamamci, Khan M. Iftekharuddin, Raj Jena, Nigel M. John, Ender Konukoglu, Danial Lashkari, José Antonio Mariz,

Raphael Meier, Sérgio Pereira, Doina Precup, Stephen J. Price, Tammy Riklin Raviv, Syed M. S. Reza, Michael T. Ryan, Duygu Sarikaya, Lawrence H. Schwartz, Hoo-Chang Shin, Jamie Shotton, Carlos A. Silva, Nuno Sousa, Nagesh K. Subbanna, Gábor Székely, Thomas J. Taylor, Owen M. Thomas, Nicholas J. Tustison, Gözde B. Ünal, Flor Vasseur, Max Wintermark, Dong Hye Ye, Liang Zhao, Binsheng Zhao, Darko Zikic, Marcel Prastawa, Mauricio Reyes, and Koen Van Leemput. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging*, 34(10):1993–2024, 2015.

- [31] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4:170117, 2017.