A MACHINE LEARNING BASED METHOD FOR SENSITIVITY
ESTIMATION FOR ACCELERATED MAGNETIC RESONANCE
SPECTROSCOPY IMAGING USING PHASED ARRAY COILS

BY

KEVIN PERKINS

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2018

Urbana, Illinois

Adviser:

Professor Zhi-Pei Liang

ABSTRACT

Magnetic resonance spectroscopic imaging (MRSI) enables in-vivo analysis of the spatial distribution of chemicals within the human body. Through MRSI, one can infer the concentration of various metabolites in different regions throughout the body. While the medical implications of such an imaging paradigm are remarkable, a poor trade-off between imaging speed and image resolution has stunted development of MRSI applications.

A combination of many technological advancements is necessary to bring MRSI to its full potential; one advancement is an accelerated imaging technique known as parallel imaging. Parallel imaging exploits differences in receiver sensitivities in phased array coils to recover additional location information. Accurate estimation of the sensitivity profiles is necessary to prevent parallel imaging induced artifacts. However, accurate sensitivity profile estimations require fully sampled high-resolution images which adds an excessive data acquisition burden.

A novel sensitivity profile estimation strategy which relies on deep learning is presented. It is shown how prior information in the form of learned image feature representations may be combined with noisy imaging data to produce high-resolution, artifact-free sensitivity profiles. An in-vivo experiment demonstrates the effectiveness of the proposed method. The relative SENSE reconstruction error for the proposed method is 1.96% compared to a signal processing baseline of 2.52%.

To Kate.

ACKNOWLEDGMENTS

I'm very grateful for the opportunities that have been given to me and the support which has enabled me to realize my potential. My parents have always motivated me and encouraged me to always ask questions and look for answers.

I would like to especially thank my advisor Professor Zhi-Pei Liang for all of the time and energy he has taken to give me the support, advice and direction I needed to make it to this point. I'm especially grateful for the kind but candid mentoring that has helped me to grow. Early on, Zhi-Pei made it clear that his ultimate goal was to provide me with a learning experience that would be a positive influence in my career development, and he has stayed true to this goal.

I am very grateful to Professor Brad Sutton for lending me the computational as well as human resources needed to take on this large-scale learning task. I also appreciate all my labmates who have selflessly given time to help me whenever I was in need. Special thanks to Steve Peng, Yang Chen, Brian Clifford, Ronny Guo and Yudu Li. Each of them went to great lengths to help me with data acquisition and preparation, comparative experiments, software setup and anything else I needed.

Finally, I would like to thank my loving wife Katelyn for following me across the country to Champaign, IL and agreeing to start our family here in our new home. I couldn't have done any of this without her.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Background

1.1.1 Magnetic Resonance Spectroscopic Imaging

In-vivo applications of Nuclear Magnetic Resonance (NMR) spectroscopy
roughly consist of Magnetic Resonance Imaging (MRI) and Magnetic Reso-
nance Spectroscopy (MRS). In both settings the interaction of atomic spins
with an external magnetic field is observed in order to infer properties about
the subject in a non-invasive way. Both MRI and MRS rely on proton res-
onant frequency which is proportional to the strength of the magnetic field
in which it resides, although they exploit this information in different ways.
MRI employs spatially varying magnetic fields known as gradients which en-
code location information in the frequency (or phase) of each nuclear spin.
MRS infers the molecular structure in which the protons reside by relying on
a phenomenon known as chemical shielding.

MRI and MRS can be combined into what is known as Magnetic Reso-
nance Spectroscopic Imaging or MRSI. In an MRSI image, a spectrum is
acquired for every pixel. These data are particularly interesting to neurolo-
gists, for example, because they enable the determination of concentrations
of various metabolites in different regions throughout the brain. However,
the inclusion of spectroscopy to MRI adds another dimension to the acquired

data. Unfortunately, the time to acquire the data scales exponentially with the dimensionality of the data to acquire. This problem is known as the *curse of dimensionality*, and explains why many otherwise promising MRSI applications have been considered infeasible over the last four decades.

Much effort has been spent in attempt to reduce data acquisition time with the goal of realizing the full potential of MRSI. High field scanners and fast pulse sequences have each contributed to faster scans. A breakthrough in the 1990s called parallel imaging has been widely applied to accelerate scans in both clinical and research settings. These speedups, however, come with a cost. Parallel imaging requires a sensitivity profile which adds an additional data acquisition burden [1].

1.1.2 Machine Learning

Machine learning has a long history originating with the invention of the perceptron by Frank Rosenblatt [2] in the 1950s. Machine learning can be roughly divided into two tasks: classification and regression. In classification tasks data points are given in some geometric space with accompanying labels. The objective of the classifier is to determine a suitable separatrix which can divide the data according to the labels. Regression tasks are similar; however, the objective is to predict a numerical rather than categorical value.

Deep learning is a class of machine learning techniques which is capable of performing both regression and classification. Recently deep learning has been widely popularized due to astounding success in image classification and segmentation as well as voice recognition and machine translation tasks.

1.2   Problem Formulation

In this thesis we propose the use of a deep learning framework to alleviate some of the acquisition burden required to produce a sensitivity profile. We assume that the sensitivity profile generated from a set of anatomical images is equivalent to that which is derived from MRSI data. We leverage high-resolution MR image data generated from a flash sequence to generate the sensitivity profile. These images are acquired in a very short time, and are very noisy. We propose the use of deep learning to denoise the anatomical images prior to the generation of the sensitivity profile.

1.3   Motivation

This work addresses the challenge of long acquisition time; however, the focus of this thesis is placed on an MRSI imaging framework known as SPICE. The SPICE framework has been used to produce high-resolution, high-SNR 3D metabolite maps in about 25 minutes [3] - a significant advancement over state-of-the-art MRSI. Much of the speedup enjoyed by SPICE comes from a novel subspace imaging approach which is implemented using custom pulse sequences and a hybrid sparse sampling strategy. Further, SPICE employs a number of well-known accelerated imaging strategies such as parallel imaging.

Prior studies suggest that the SPICE framework could further increase imaging acceleration, while maintaining image integrity, through the use of parallel imaging. However, this finding comes with a caveat. Accurate, high-resolution sensitivity profiles are necessary to ensure that image quality is not jeopardized by increased acceleration.

The two most common methods for estimating sensitivity profiles were given in the original paper [1]. They are:

- Sum-of-squares normalization of full-fov image

- Prior image of uniform phantom

The first method produces very accurate estimations of the sensitivity profile, however it requires the acquisition of a full-FOV (field of view) image for every scan, which adds an imaging burden that negates the use of parallel imaging. The second method eliminates the additional scan time, but produces inaccurate results. This is because the sensitivity profile is altered by loading effects which change with every new subject in the scanner. While it is theoretically possible to calculate the sensitivity profile, it is not practical because it requires detailed knowledge of loading effects and receiver coil calibrations.

1.4 Summary of Results

An in-vivo experiment demonstrates the effectiveness of the proposed method of sensitivity profile estimation. Three cases were compared in this experiment.

- Deep denoised SENSE

- Total variation denoised SENSE

- No denoising

This thesis considers the deep denoised SENSE method, and compares it to the total variation method which exploits sparsity in the MR images as a method of denoising. In terms of visual quality and relative error, the deep denoised SENSE method outperforms the total variation as well as no denoising methods.

1.5   Thesis Outline

Chapter 2 provides the necessary literature review. A brief overview of the fundamentals of MRSI are presented. Parallel imaging is introduced with an emphasis on the SENSE algorithm given in [1]. A review of the history of machine learning is given, justifying the use of deep convolutional neural networks.

Chapter 3 gives the details of the problem formulation as well as the proposed solution. The methods used to accomplish the stated solution are presented. A detailed description of the modifications made to the training procedure is discussed.

Chapter 4 presents the experiment results and contains a quantitative analysis of the predicted sensitivity profile. For comparison, SENSE reconstructions are given using total variation denoised and noisy sensitivity profiles. Finally, results from suboptimal deep learning strategies are shown which reinforce the methods as set forth in chapter 3.

Chapter 5 concludes the thesis with a summary of findings and direction for future work.

CHAPTER 2

LITERATURE REVIEW

2.1 Magnetic Resonance Spectroscopic Imaging

2.1.1 MRI Fundamentals

Upon placement in a high magnetic field (denoted $\vec{B_0}$) nuclei with spin $\frac{1}{2}$ will split into two energy states according to the theory of quantum mechanics. These states can be interpreted as parallel and anti-parallel with the main magnetic field. Since the observable signal comes from a collection of many spins, a classical framework may be used to describe the dynamics of the vector sum of magnetic moments known as the *bulk magnetization*. Using the Boltzmann relationship, the magnitude of the bulk magnetization in equilibrium is derived in [4] and is given as

$$M_z^0 = \frac{\gamma^2 \hbar^2 B_0 N}{4KT_s} \tag{2.1}$$

where N is the total number of nuclei, γ is the gyromagnetic ratio, $\hbar$ is the plank constant, K is the Boltzmann constant and T_s is the absolute temperature.

In equilibrium the bulk magnetization is directly aligned with the main external magnetic field, conventionally known as the $\vec{z}$ direction. A properly tuned RF field called $\vec{B_1}(t)$ can be used to excite the spin population which induces a coherent spin known as *precession*. The spin population precesses

6

at a frequency known as the *Larmor frequency* which is given by

$$\omega_0 = \gamma B_0 \tag{2.2}$$

This excitation as well as a subsequent relaxation is described by a phenomenological equation known as the *Bloch equation*.

$$\frac{d\vec{M}}{dt} = \gamma \vec{M} \times \vec{B} - \frac{\vec{M}_\perp}{T_2} - \frac{(M_z - M_z^0)\vec{z}}{T_1} \tag{2.3}$$

Here $\vec{M}$ represents the bulk magnetization vector, $\vec{B}$ represents $\vec{B}_0 + \vec{B}_1(t)$, $\vec{z}$ represents the longitudinal direction vector, and $\vec{M}_\perp$ represents the transverse component of the bulk magnetization. The second two terms on the right-hand side of the equation describe the relaxation process and include relaxation constants T_1 and T_2 which are material specific.

Using the Faraday law of induction as well as the reciprocity principle [5] the *emf* at a receive coil can be derived as

$$emf = -\frac{d}{dt} \int_{sample} \vec{M}(\vec{r}, t) \cdot \vec{\mathcal{B}}_{receive}(\vec{r}) d\vec{r} \tag{2.4}$$

In this derivation $\vec{M}(\vec{r}, t)$ represents the 3D bulk magnetization which varies both spatially (over $\vec{r}$) and temporally (over t), and $\vec{\mathcal{B}}_{receive}(\vec{r})$ represents the sensitivity profile of the receive coil, which is given as $\vec{\mathcal{B}}(\vec{r})$ for brevity. This thesis considers the problem of estimating the sensitivity profile for each receive coil in the phased array. Accurate estimation of the sense profile is a vital step toward achieving imaging acceleration through parallel imaging while maintaining image quality. Note that as the data are collected in the complex domain, the sensitivity profile contains a magnitude and a phase component. The terms sensitivity map, magnitude map and phase map all

refer to the sensitivity profile or its complex components.

Modern MRI scanners typically employ quadrature detection, and the common mathematical signal model uses the complex $\mathcal{C}^1$ domain to represent the two-dimensional transverse plane. In this model, the received signal is given as

$$s(t) = \omega_0 \int_{sample} \mathcal{B}^*_{xy}(\vec{r}) M_{xy}(r) e^{-\frac{t}{T_2(r)}} e^{-i\gamma \int_0^t \Delta B(\vec{r},\tau)d\tau} d\vec{r} \qquad (2.5)$$

where

$$\mathcal{B}_{xy}(\vec{r}) = \mathcal{B}_x(\vec{r}) + i\mathcal{B}_y(\vec{r})$$

$$M_{xy}(\vec{r}) = M_x(\vec{r}) + iM_y(\vec{r})$$

$$M_{xy}(\vec{r}) = M_{xy}(\vec{r}, 0)$$

In equation 2.5 $\mathcal{B}^*_{xy}(\vec{r})$ refers to the complex conjugate of $\mathcal{B}_{xy}(\vec{r})$ and $\Delta B(\vec{r}, t)$ refers to field inhomogeneities. The signal received immediately following excitation is called a *free induction decay* (FID) and is characterized by Larmor frequency oscillations inside a decaying exponential envelope.

2.1.2 Spatial Encoding

MRI employs gradients, which are spatially varying magnetic fields, to encode location into the magnetic spins. More specifically, we will consider a linear gradient $\vec{G}(t) = G_x \vec{x} + G_y \vec{y} + G_z \vec{z}$ which induces a shift in magnetic field given by

$$\Delta B(\vec{r}, t) = \vec{G}(t) \cdot \vec{r} \qquad (2.6)$$

With the inclusion of linear gradients, and ignoring relaxation terms and constants for simplicity, the signal model becomes

$$s(t) = \int_{sample} \mathcal{B}_{xy}^*(\vec{r})M_{xy}(r)e^{-i\gamma \int_0^t \vec{G}(\tau)\cdot\vec{r}\,d\tau}d\vec{r} \qquad (2.7)$$

Take

$$\vec{k} = \frac{\gamma}{2\pi}\int_0^t \vec{G}(\tau)d\tau$$

Equation 2.7 can be rewritten as

$$s(t) = \int_{sample} \mathcal{B}_{xy}^*(\vec{r})M_{xy}(r)e^{-i2\pi\vec{k}(t)\cdot\vec{r}}d\vec{r} \qquad (2.8)$$

It is common to replace $\mathcal{B}_{xy}^*(\vec{r})M_{xy}(r)$ with $\rho(\vec{r})$ which is the *spin density*, defined as the concentration of nuclei precessing at the Larmor frequency. The final form of the equation is given as

$$s(\vec{k}) = \int_{sample} \rho(\vec{r})e^{-i2\pi\vec{k}\cdot\vec{r}}d\vec{r} \qquad (2.9)$$

Equation 2.9 gives the Fourier transform as the relation between the underlying image $\rho(\vec{r})$ and the k-space measurements $s(\vec{k})$. Consequently a simple inverse Fourier transform is all that is required to recover an image from the acquired data.

2.1.3 Spectral Encoding

NMR spectroscopy exploits a property known as chemical shift to infer properties about the chemical structure in which nuclei reside. This method has been successfully employed as a non-destructive chemical analysis tool for several decades. Chemical shielding refers to local perturbations of magnetic
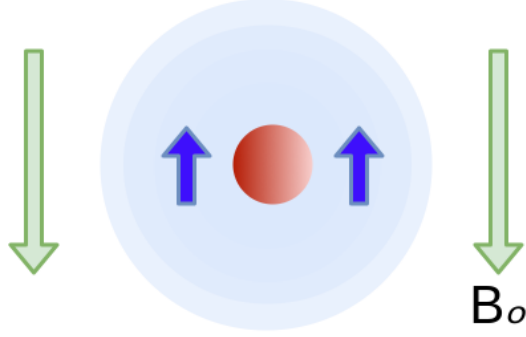
Figure 2.1: Illustration of the chemical shielding effect. The arrows indicate that the local magnetic field perturbation from the electron cloud opposes the main magnetic field B1 causing the proton to 'feel' a smaller B_0.

field, which are caused by electron cloud currents as seen in figure 2.1. These local perturbations cause slight shifts in the resonant frequency of neighboring nuclei. The chemical shielding effect is dependent on the geometric distribution of the electron cloud, meaning the effect of chemical shift varies across different molecules. Therefore, every chemical structure has its own spectral marker from which it can be identified.

A spectrum can be acquired simply by sampling the FID and performing a Fourier transform on the acquired data. Unfortunately, with regard to in-vivo studies, a single spectrum is often insufficient to meet the desired application objective. This is because the human body does not have a homogeneous distribution of metabolites, and the spatial distribution of metabolites is lost without spatial encoding as described in section 2.1.2. The solution is to combine MRI and MRS into what is known as Magnetic Resonance Spectroscopic Imaging (MRSI).

In MRSI a spectrum is acquired for every pixel (or voxel) location. This is given as a simple modification to equation 2.9, namely

$$s(\vec{k}, t) = \int_{sample} \rho(\vec{r}, t) e^{-i2\pi \vec{k} \cdot \vec{r}} e^{-i2\pi t \Delta f_0(\vec{r})} d\vec{r} \qquad (2.10)$$
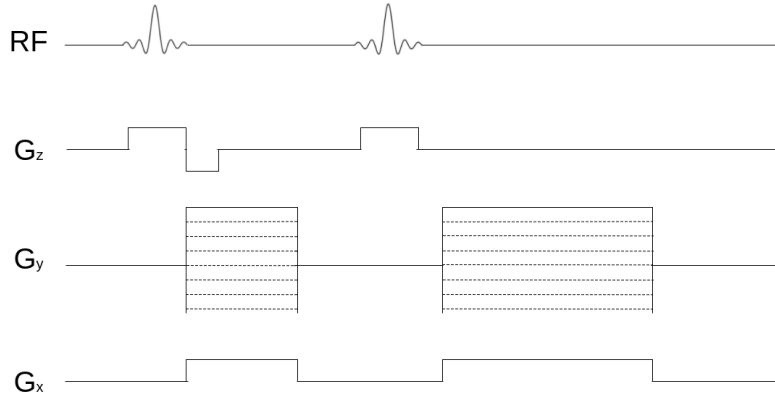
Figure 2.2: Example 2D-imaging pulse sequence. The G_y gradient phase encodes the signal and the G_x gradient is applied during readout and frequency encodes the signal.

Here Δf_0 refers to field inhomogeneities in B_0. The acquired data resides in a space which includes a temporal dimension commonly referred to as kt space. The desired spectrograms can be recovered using a Fourier transform.

2.1.4 Data Acquisition

RF and gradient activation are controlled through pulse sequence. Figure 2.2 shows an example pulse sequence. The RF pulse in conjunction with the G_z gradient excites a two-dimensional slice. The G_y gradient phase encodes the signal over n excitation cycles. The G_x, or readout gradient, frequency encodes the signal. Sub-sampling in the phase encode direction directly translates into fewer excitation cycles which can significantly reduce scan time.

2.2 Parallel Imaging

Parallel imaging provides one opportunity to reduce scan times without sacrificing image quality. Parallel imaging techniques are designed with algo-
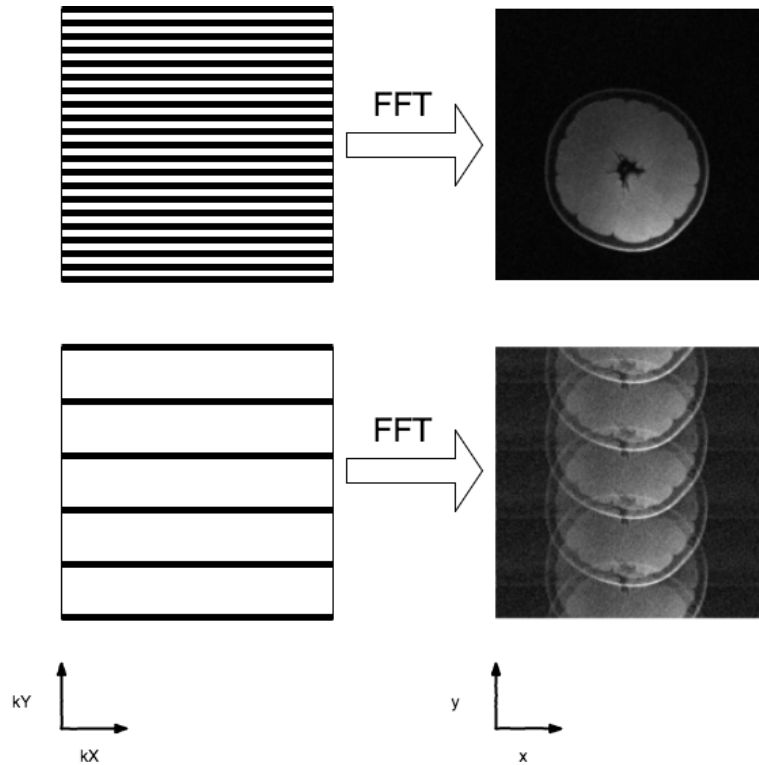
Figure 2.3: Aliasing artifact produced by subsampling in the phase encoding direction. The top image comes from fully sampled k-space data. In the bottom image, one in four lines is acquired which is considered a reduction factor of four (R = 4). The result in image space is four shifted identical copies which are superimposed.
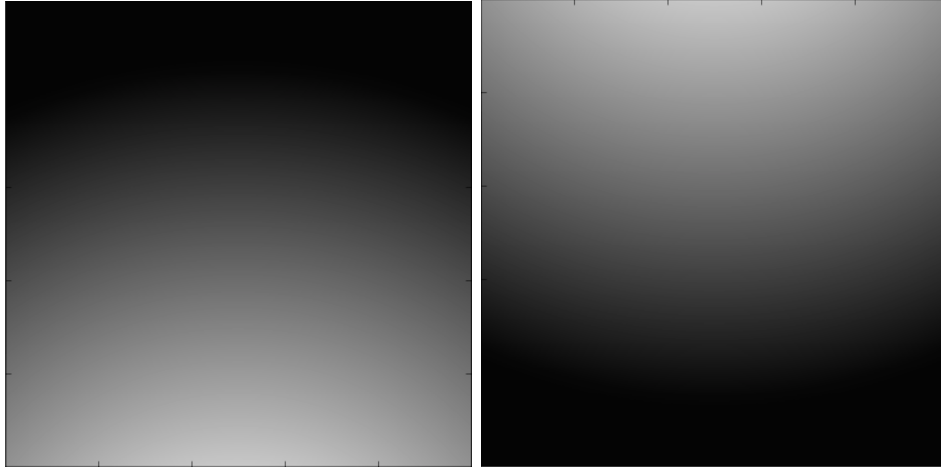
Figure 2.4: Example sensitivity profiles. Each coil has its own unique profile. For maximum parallel imaging acceleration profiles should be designed to be as orthogonal as possible.

rithms that operate in either k-space or the image domain. Examples of k-space-based parallel imaging algorithms are Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA) [6] and iterative self-consistent parallel imaging reconstruction from arbitrary k-space (SPIRiT) [7]. On the other hand, sensitivity encoding for fast MRI (SENSE) [1] is an image domain parallel imaging technique. See figure 2.3 for an example of how SENSE operates.

Like all parallel imaging algorithms, SENSE exploits sensitivity profile variations across receiver coils, such as those seen in figure 2.4, to recover additional spatial information to resolve spatial ambiguity. SENSE relies on high-quality sensitivity profile maps which must be obtained in addition to the routine data acquisition. Because the sensitivity profile is dependent on loading effects which change with every subject scanned, it is not possible to obtain a satisfactory sensitivity profile using a homogeneous phantom. Accurate sensitivity maps may be acquired by obtaining high-SNR, fully sampled MR data for every coil and normalizing by the sum-of-squares of

all coils. The acquisition of such high-SNR, fully sampled data, however, is a time-consuming process which negates the benefit of applying parallel imaging in the first place.

SENSE provides a general method for accelerating 2D and 3D Cartesian image scans, with well-understood SNR trade off. For this reason, SENSE has been widely adopted in research as well as clinical settings. SENSE resolves spatial ambiguity which is caused by aliasing in the image domain according to the Poisson summation formula. Consider the following example of a 2D imaging sequence with an acceleration factor of 2 and two independent receive coils.

$$
\begin{aligned}
s_1(x,y) &= w_1(x,y)\rho(x,y) + w_1\left(x, y + \frac{1}{2}FOV\right)\rho\left(x, y + \frac{1}{2}FOV\right) \quad (2.11) \\
s_2(x,y) &= w_2(x,y)\rho(x,y) + w_2\left(x, y + \frac{1}{2}FOV\right)\rho\left(x, y + \frac{1}{2}FOV\right)
\end{aligned}
$$

Here w_1 and w_2 refer to the sensitivity profiles of coils 1 and 2, respectively. The acquired signals s_1 and s_2 are given as linear combinations of spin density ρ from two locations. Equation 2.11 can be written in matrix form as follows:

$$
\vec{s} = W\rho \quad (2.12)
$$

W is a sensitivity matrix with a row for each coil and a column for each superimposed pixel. An unfolding matrix can be created which optimally separates superimposed pixels in a least squares sense and is given by

$$
U = (W^H \Psi^{-1} W)^{-1} W^H \Psi^{-1} \quad (2.13)
$$

In equation 2.13 W^H denotes the transposed complex conjugate of W and Ψ

represents the receiver noise matrix which gives noise levels and correlations for each of the coils. Note that in the absence of noise the unfolding matrix is simply the Moore-Penrose pseudo-inverse of W. Assuming i.i.d. Gaussian noise, this method produces the maximum likelihood estimation of the true image. The separated pixels can be recovered through the following:

$$\hat{\rho} = U\vec{s} \tag{2.14}$$

Here $\hat{\rho}$ represents the separated pixels, and $\vec{s}$ represents the measurements. This process can be repeated for all pixels in the image. Note that to correctly solve a system of linear equations there must be more equations than unknowns. This means that the acceleration factor cannot exceed the number of coils. In practice, the acceleration factor is often several factors lower than the number of coils due to sub-optimal, non-orthogonal coil sensitivities.

2.2.1   Condition Number

SENSE seeks an inverse solution of a linear system of equations, such as $Ax = b$. An important property of any such systems of equations is called the condition number (κ) which is defined in the following way:

$$\kappa(A) = ||A||_2||A^+||_2 \tag{2.15}$$

where $|| \cdot ||_2$ denotes the spectral norm which is equal to the largest singular value of A, and A^+ denotes the pseudo-inverse of A. The condition number is the ratio of the relative error in the solution (A^+b) to the relative error in the input (b). It is used to provide a bound on output errors provided that a bound on the input errors is known.

Linear systems of equations with high condition number are described as

15

ill-conditioned. The inverse solution of such a system suffers from magnified errors. Such errors enforce a strict limit on the amount of acceleration which can be achieved through SENSE, because the condition number increases as the acceleration factor increases. Errors in the sensitivity profile are similarly magnified by a large condition number, therefore the quality of the sensitivity profile is one determining factor for the amount of acceleration which can be achieved.

2.3   Machine Learning

Most machine learning formulations pose learning as an optimization problem, over a set of model parameters, whose objective is a task-dependent loss function. Empirical risk minimization is the process by which the objective function is minimized over a training set which is carefully constructed to represent the real world. A well-crafted model trained on a sufficient training dataset will avoid the problem of over fitting; in other words it will generalize to new data.

An artificial neural network (ANN) is a type of parametric machine learning model which bears resemblance to the original perceptron and has been compared to the biological brain. The parameters of an ANN are a set of weights and biases. Each layer in a neural network is made up of many nodes called neurons. An example neuron is shown in figure 2.5a. To enable the network to learn arbitrary, nonlinear mappings every neuron ends in a possibly nonlinear function known as an activation function. Neural networks can have an arbitrary number of layers. The intermediate layers are trained to discover the latent structure in the data and are therefore named hidden layers. A graphical representation of a neural network is shown in figure
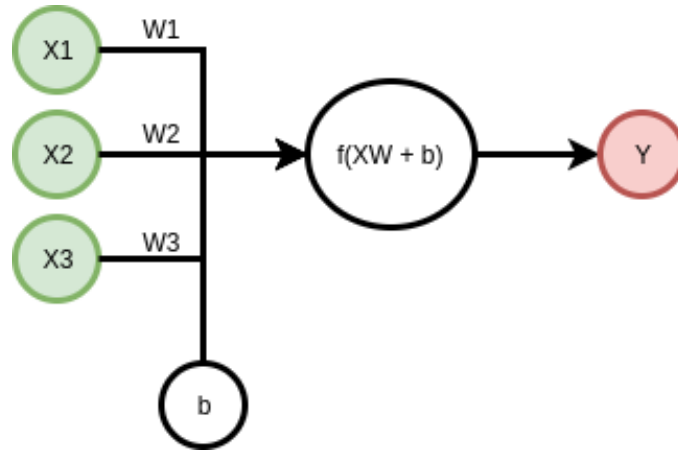
(a) Single Neuron.
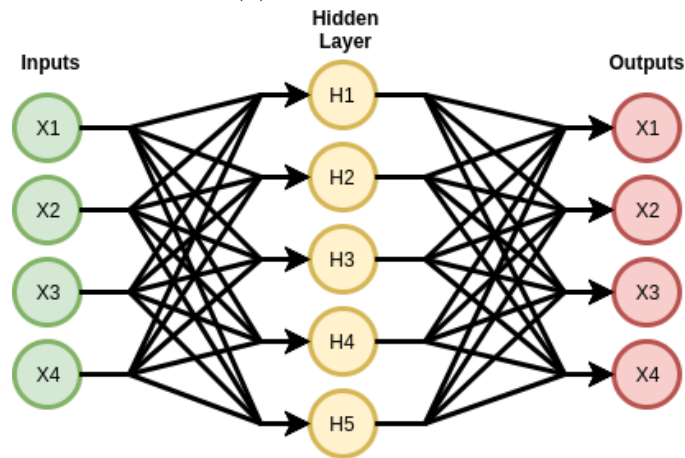
(b) Example neural network.

Figure 2.5: Graphical representation of an ANN. Each neuron contains a set of weights which connect with the previous layer and a bias. Layers are composed of an arbitrary number of neurons.

2.5b. A neural network with a single hidden layer can approximate continuous functions on compact subsets of $\mathcal{R}^n$ using a finite number of neurons [8, 9].

While single-layer ANNs can approximate arbitrary functions, Mhaskar et al. [10] showed that deep neural networks can accurately represent a class of compositional functions using exponentially fewer parameters than their shallow counterparts, and similar findings were reported in [11, 12]. Neural networks are commonly trained using gradient descent. Gradients

are propagated through deep neural networks according to the calculus chain rule in a process known as *backpropagation.* The use of deep neural networks, or deep learning, has recently advanced many areas of research such as image classification, segmentation, speech recognition and machine translation. For a review of modern deep learning, see [13].

2.3.1 Convolutional Neural Networks

A fully connected neural network layer refers to a layer in which the input of all of the neurons is connected to all previous neuron outputs. A network consisting of only fully connected layers is highly suboptimal for image processing tasks for two reasons:

- Lack of shift invariance

- Unscalable to even moderate image sizes

A *convolutional neural network* (CNN) addresses these two problems by modifying the standard neural network architecture through massive weight pruning and weight sharing. The effect is that hidden layer representations, called feature maps, are produced by convolving prior feature maps with a small kernel whose entries are the parameters of the network. Figure 2.6 demonstrates the main concept of a CNN.

2.4 Related Work

To the best of our knowledge deep learning has not been applied to the problem of estimating sensitivity maps for improved parallel imaging. Deep learning has been applied to MRI image reconstruction, see [14, 15, 16, 17]. However, these works have given deep neural networks the task of directly
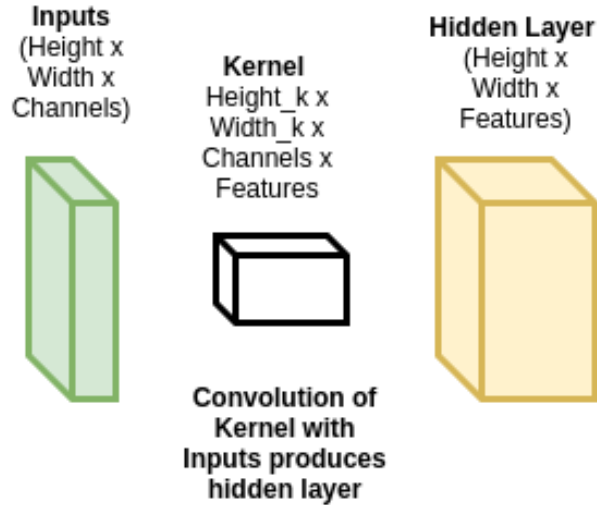
Figure 2.6: Base unit CNN. Data is represented by feature maps which are essentially images with an arbitrary number of channels which encode latent information.

predicting the final image reconstruction given under-sampled data. While some positive results have emerged, little theory or justification is given to demonstrate that deep learning reconstruction can work in a general case. Furthermore, a general deep learning reconstruction strategy would replace a well-understood reconstruction pipeline with a black box.

In the context of producing denoised sensitivity maps, learning is justified. The image denoising task uses only low-level image features and thereby does not require a large training dataset. Furthermore, SENSE reconstruction is a simple linear operation. This has a number of advantageous properties such as the Gaussian-noise in Gaussian-noise out relationship.

CHAPTER 3

METHODS

3.1 Formulation

It is assumed that a sensitivity profile derived from anatomical MR data is equivalent to a sensitivity profile derived from MRSI data. Therefore, the goal of the thesis is to generate high-resolution denoised sensitivity profiles from noisy MR images. More precisely, the objective is to find a function $f : x \mapsto \hat{y}$ which maps noisy MR images (x) to denoised MR images $(\hat{y})$. It is assumed that the training images contain iid white Gaussian, additive noise. Under this assumption, an appropriate learning objective which maps noisy images to denoised images takes the form:

$$w_{optimal} = arg \min_{w} \sum_{n=1}^{N} ||f_w(x_n) - y_n||^2 \tag{3.1}$$

Inspired by residual CNN networks [18] and based on empirical findings, it was determined that the residual error provided a better target for training. If $x_n = y_n + v$, where v is iid-Gaussian noise, then a residual learning network is trained to approximate $\hat{f}(x) = v$. Therefore, the objective to optimize is

$$w_{optimal} = arg \min_{w} \sum_{n=1}^{N} ||\hat{f}_w(x_n) - (x_n - y_n)||^2 \tag{3.2}$$

It is proposed that a deep neural network learn the mapping from corrupted images to residual error. However, the stated problem is ill-posed which
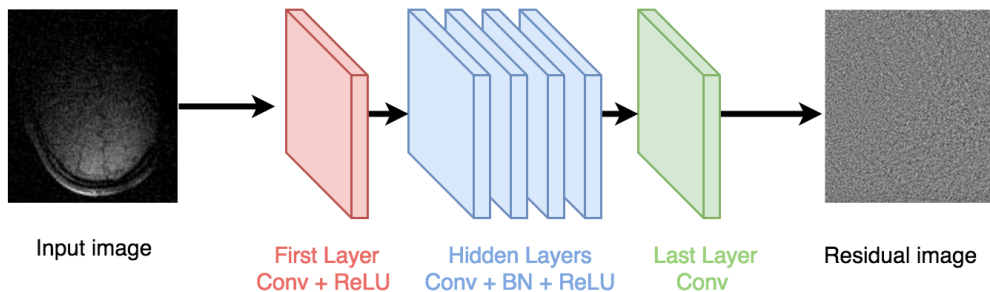
Figure 3.1: DnCNN architecture. Each hidden layer consists of a convolution batch-normalization and ReLU nonlinearity. The input and output images contain two channels, and the hidden layers each have 64 feature maps.

means is impossible to find a solution without relying on prior knowledge.

Prior knowledge comes in the form of image feature representations which are learned from a training set. It is assumed that high-dimensional MR images reside on a low-dimensional manifold, which means that each image can be represented using fewer parameters than those necessary to specify all pixel intensities.

3.2 Approach

3.2.1 DnCNN

The Deep Denoising Convolutional Neural Network (DnCNN) architecture [19] was developed for the task of Gaussian noise removal. The DnCNN architecture is shown in figure 3.1. The DnCNN is capable of outperforming standard signal processing methods because it utilizes image features in the denoising processes rather than simple signal characteristics such as sparsity. The DnCNN excels at learning good representations of image features because it is a deep, convolutional neural network with a large receptive field.

The depth of the network is a crucial feature, as it was shown indepen-

dently by Mhaskar and Telgarsky [12, 11] that deep neural networks have more representational power than a similarly sized shallow counterpart. The DnCNN is convolutional, which means that many small kernels, size 3 x 3, are applied at every layer to produce the next layer or set of feature maps. Convolutional layers have far fewer parameters than fully connected layers which gives the network scalability and reduces the likelihood of overfitting. The receptive field refers to the spatial connectivity extent within a network. A large receptive field is important for neural networks because input values which fall outside of a receptive field cannot influence the corresponding output. The receptive field in a CNN grows exponentially with the number of pooling layers in the network [20] and linearly with the number of convolutional layers. The DnCNN was designed without pooling layers to mitigate blurring effects, and contains a large number of convolutional layers to produce a large receptive field.

3.2.2   Residual Learning and Batch Normalization

Inspired by residual CNNs [18] the DnCNN is trained to predict the residual error rather than predicting the denoised image directly. In this way, the denoised image can be obtained by simply subtracting the output of the network from the input. This has many advantages in practice such as faster training, and mitigated effects of the vanishing gradient problem which enables effective training of very deep networks.

The DnCNN network follows the batch normalization principle from Ioffe and Szegedy [21] which accelerates training by reducing internal covariate shift. Batch normalization is applied on every hidden layer which means that all feature maps are normalized by the means and variances computed across

the batch. It was demonstrated by [19] that residual learning and batch normalization have complimentary effects, and that the best performance and fastest training times come when both methods are applied.

3.2.3 Proposed Solution

The proposed solution is given in graphical form in figure 3.2. A DnCNN is trained to predict high-resolution, denoised MR images given the high-resolution noisy images. As the input images are very noisy, it was desired that the network employ a large receptive field, therefore the depth of the network was set as 20 layers. The sensitivity profile is created by simple sum-of-squares normalization across images from all coils according to the following equation

$$w_j = \frac{s_j}{\sqrt{\sum_{k \in numCoils} s_k^2}} \tag{3.3}$$

where w_j refers to the sensitivity profile of coil j and s_j refers to the image of coil j. Following the sum-of-squares normalization, the phase of each coil is computed relative to that of the first coil as follows

$$w_j = w_j * exp(-i * \angle w_1) \tag{3.4}$$

where $\angle w_1$ is the phase of the sensitivity of the first coil.

As usual for the SENSE algorithm, imaging data is sub-sampled by acquiring only one in n lines in k-space, where n is the acceleration factor. Upon prediction of sensitivity profiles, the SENSE algorithm proceeds as normal to reconstruct the desired images.
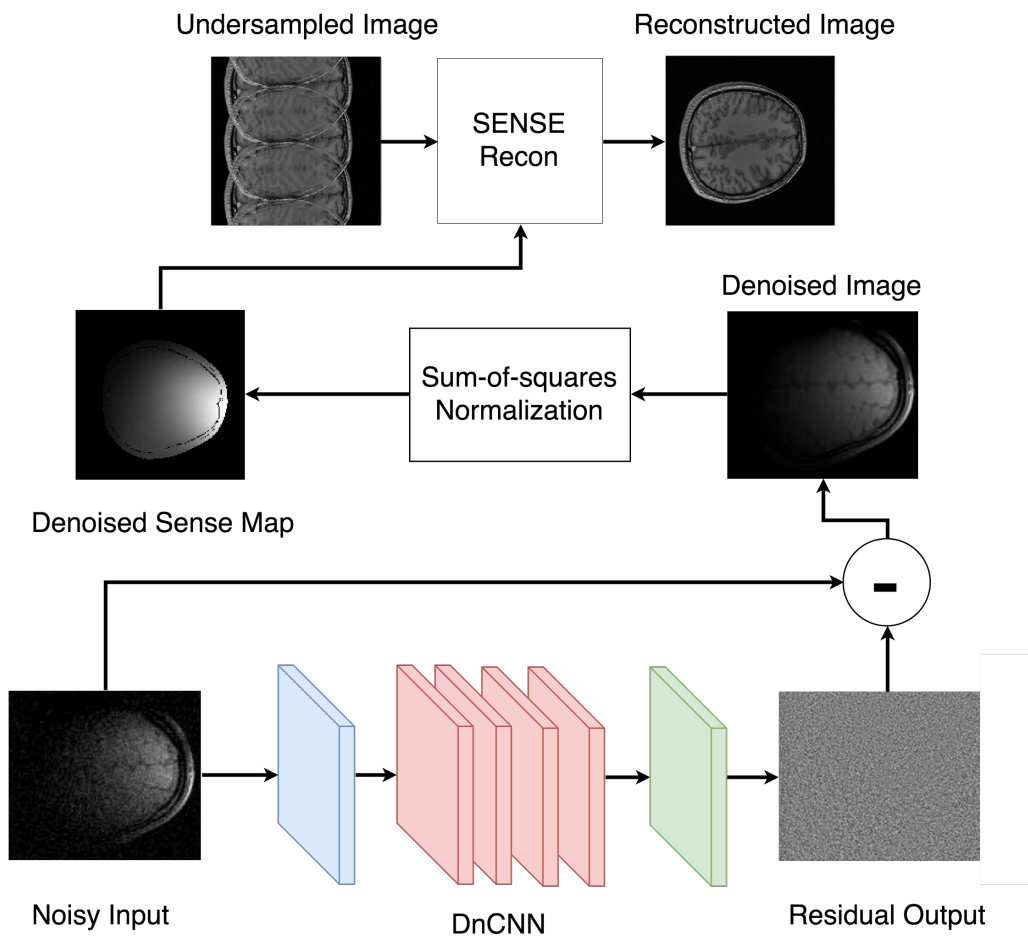
Figure 3.2: Graphical representation of the solution outline.

3.3   Implementation

3.3.1   Architecture

The DnCNN architecture is designed according to the guidelines in [19]. As discussed in section 2.3, MSE loss is statistically optimal when dealing with Gaussian noise.

The ground truth training images are created by acquiring and averaging eight flash images. The noisy images simply include additive Gaussian noise with a fixed noise variance. The test set is generated in a similar manner; however, it is derived from scans of a different subject in order to eliminate the effect of overfitting.

Care must be taken to ensure that phase wrapping does not harm the learning procedure. Phase wrapping causes the mean squared error loss function to be highly suboptimal in the presence of Gaussian noise. Consider the following: Let Y be a true image with values $\pi + N_{ij}$. N_{ij} is drawn iid from a zero mean Gaussian with variance ϵ for every pixel in the phase map. Whenever $N_{ij} > 0$, the value of Y_{ij} wraps to $-\pi$, therefore any pixel Y_{ij} will be close to π with probability $\frac{1}{2}$ and close to $-\pi$ with probability $\frac{1}{2}$. Under these conditions the predicted output $\hat{Y}$ that minimizes MSE is 0, which is exactly out of phase from π.

A simple way to correct for this type of wrapping is to train a network to predict the real and imaginary components of the signal simultaneously. In this training setup, the network does not predict circular outputs.

3.3.2 Training Procedure

In contrast to the original DnCNN paper [19], this work denoises entire images (180 x 180) simultaneously rather than in patches (40 x 40). This simplifies the testing process, and eliminates the possibility of boundary artifacts between patches. Due to memory constraints, the batch size used in this project was 16 as compared to 128 in [19]. However, due to the larger size of the inputs, the amount of information processed in a single batch is actually larger, thereby providing stable gradients for training.

The proposed DnCNN was developed in Tensorflow 1.3.0. Code was taken from [22] and is used to generate training batches. Data augmentation was done in the form of random flips in the horizontal and vertical directions as well as random rotations. Input preprocessing consists of scale and shift normalization. Batch normalization (batch size 16) is employed at every layer, and the network is trained to predict the residual image function. The network is trained using stochastic gradient descent (SGD) with an Adam optimizer initialized with the following parameters: learning_rate = 0.005, beta1 = 0.9, beta = 0.999, epsilon = 1e-08. Training requires approximately 7 hours on an Nvidia Titan Xp GPU with a memory clock rate of 1.582 (GHz).

3.4 Evaluation

3.4.1 Comparison of Methods

The various methods are compared by separate SENSE reconstructions. High-SNR data was retrospectively undersampled according to the given acceleration factors and used as input to the SENSE reconstructions. Sen-

sitivity profiles were obtained from anatomical images using the following sum-of-squares approach given in equation 3.4.

3.4.2   Simulation

A simulation was run which shows the effects of sensitivity profile errors in SENSE reconstructions with the following acceleration factors: 4, 6, 8. Three cases were studied: no errors, added Gaussian noise, Gibbs ringing. High-resolution sensitivity maps were constructed using fully sampled k-space data. These maps were corrupted with zero-mean, Gaussian noise and under-sampled in the Fourier domain in a way that simulates the acquisition of only 28 ACS lines. Fully sampled imaging data were acquired and used as ground truth for comparison with the reconstruction. The imaging data were retrospectively sub-sampled according to the specified acceleration factors with the inclusion of 28 ACS lines.

Figures 3.3a, 3.3b, and 3.3c display the results of the SENSE reconstruction using the error-free, Gaussian noise, and Gibbs ringing sensitivity profiles respectively. The first, second and third columns refer to SENSE reconstructions with acceleration factors of 4, 6 and 8 respectively. Similarly, figures 3.4a, 3.4b, and 3.4c show the difference images with ground truth. To improve visualization, the difference images were scaled by 5.

Table 3.1 gives the relative error for each of the SENSE reconstruction which is calculated using equation 3.5:

$$error = \frac{||(y - \hat{y}) \cdot m||_2}{||y \cdot m||_2} \tag{3.5}$$

Here y refers to the ground truth image and $\hat{y}$ refers to the predicted image. The binary mask used in the weighted MSE calculation is given as m.

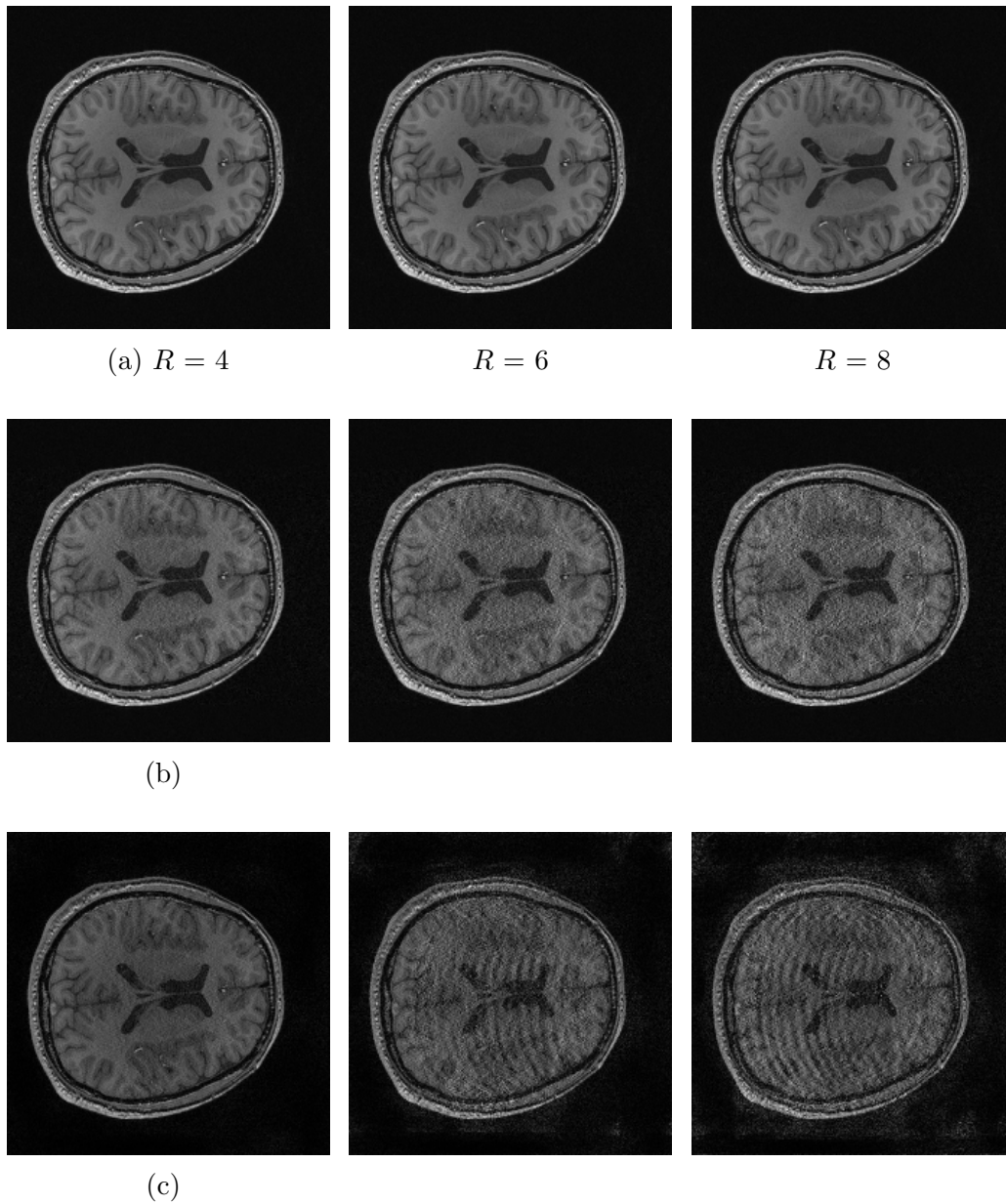|		
(a) $R = 4$	$R = 6$	$R = 8$

(b)

(c)

Figure 3.3: SENSE simulation - reconstruction using (a) error-free sensitivity profile, (b) Gaussian noise sensitivity profile, and (c) Gibbs ringing sensitivity profile. Acceleration factors R = 4, 6, 8. Notice, SENSE reconstructions are exact given an error free sensitivity profile.
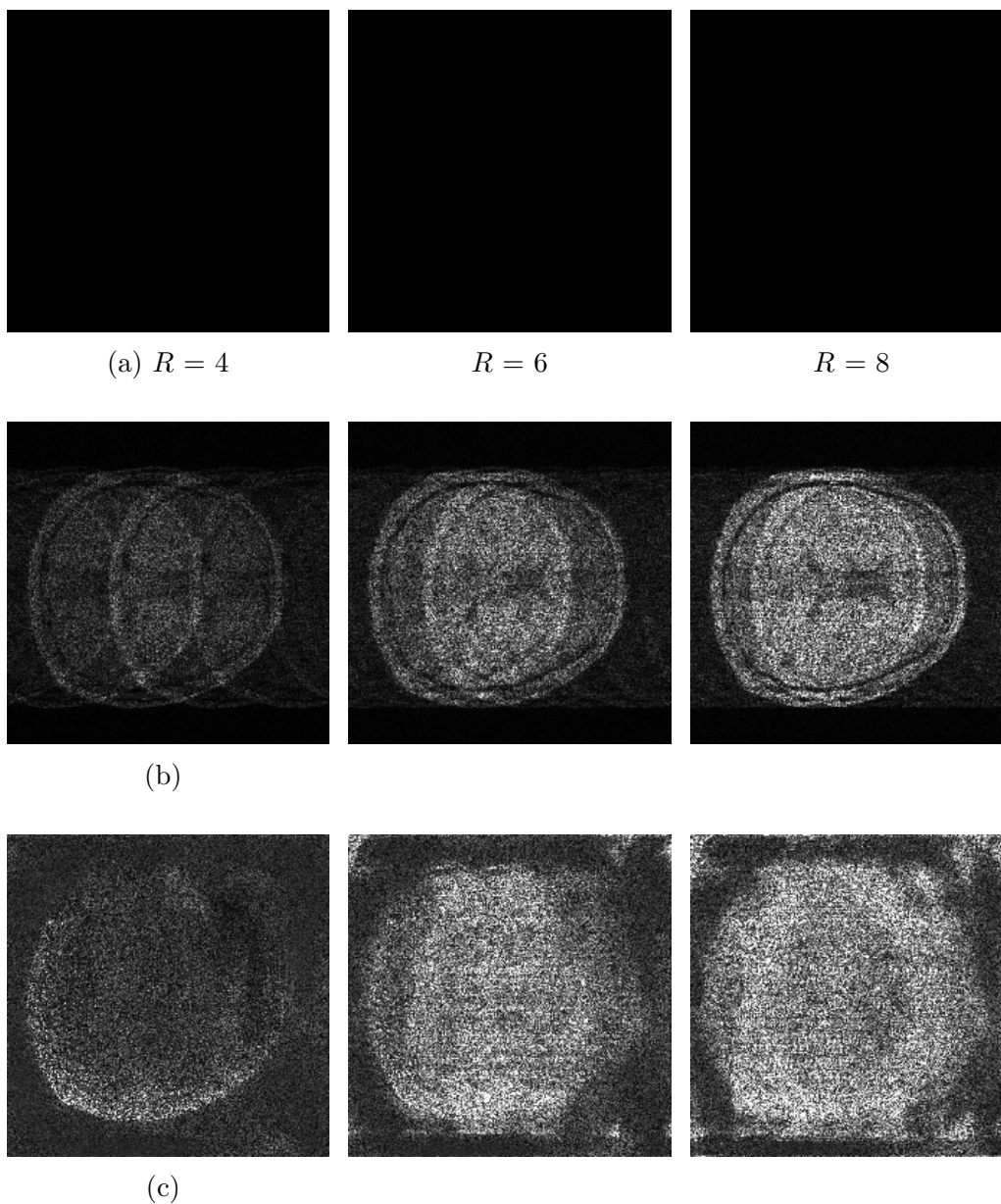
| (a) $R = 4$ | $R = 6$ | $R = 8$ |

(b)

(c)

Figure 3.4: SENSE simulation - difference from ground truth. Reconstructions use (a) error-free sensitivity profile, (b) Gaussian noise sensitivity profile, and (c) Gibbs ringing sensitivity profile. Acceleration factors R = 4, 6, 8.

Table 3.1: SENSE reconstruction percent errors using (1) error-free sensitivity profile, (2) sensitivity profile with Gaussian noise, (3) sensitivity profile with Gibbs ringing. Final column shows percent error of sensitivity profile.

Acceleration Factor:	4	6	8	Percent Error Map
Error-Free Sensitivity	6.2e-4%	4.8e-3%	.44%	0
Added Gaussian Noise	11%	15%	27%	17%
Gibbs Ringing	20%	39%	44%	33%

3.5 Experiments

To assess the effectiveness of the proposed solution an in-vivo experiment was conducted. Data were acquired on a 3T Siemens Trio scanner with a 16-channel head coil receiver. A 3D image volume of size 160 x 160 x 36 was acquired using a 3D flash sequence. As this project considers the task of two-dimensional image denoising, the 3D volume was separated into 16 x 36 = 576 separate, complex images which were each considered independent training samples. This is a reasonable sized training dataset for the low-level task of denoising, as it was demonstrated by [19] that only marginal gains are achieved by increasing the training set size above 400.

A training set was created which consists of 576 noisy images. During training random noise was added to the images in both the real and imaginary domains. The DnCNN was trained for 1500 Epochs. A single slice was selected as the test set. As a comparative benchmark, a total variation denoising algorithm was implemented to denoise the MR images prior to the generation of the SENSE map. The test slice comes from a subject that was not included in the training set. Identical noisy versions of the test slice were used to test both the total variation and the DnCNN.

The results which are given show the final images after SENSE reconstruction. Fully sampled, average images were taken as ground truth. Ret-

rospectively under-sampled images were taken as the input to the SENSE reconstruction with an acceleration factor of 2 in the x direction and accelerations of 2, 3, and 4 in the y direction.

Results from this thesis are given in two forms. In each case, sensitivity profiles generated from a variety of method are compared through SENSE simulations. In the first case, a comparison of the deep denoising method with baselines is given, comparing the following sensitivity profiles:

- DnCNN deep denoised

- Total variation denoised

- Noisy

The second case compares sensitivity profiles which were generated using deep learning in a non-ideal network configuration such as:

- No batch normalization or residual learning

- U-Net structure

- Direct estimation of sensitivity profiles

- Joint estimation of magnitude and phase

3.5.1 Other Learning Configurations

U-Net

The U-Net neural network architecture was developed for the purpose of biomedical segmentation [23], however it is a very general image-to-image architecture which has been applied to the tasks of MRI super-resolution [24] as well as image denoising [25]. The key features of a U-Net are five pooling layers which produce small but semantically rich feature maps.

Direct SENSE map predictions

For best results, learning is applied in the image domain rather than in predicting SENSE maps directly. This is because magnitude and phase sensitivity maps have sporadic values in regions of low signal intensity. These sporadic values must not be included in the training of the network. When training on sensitivity profiles directly, a simply way to account for sporadic values is to use weighted MSE and take the weights to be a binary mask function created by thresholding the original high-resolution images.

As the network was trained to make accurate predictions only where the binary mask is equal to one, network predictions in zero regions of the mask are unusable. This leads to aliasing artifacts on the image boundaries. This trade-off between training accuracy and boundary artifacts does not exist when training the network to directly predict images.

Magnitude and phase

In the magnitude and phase network, the DnCNN outputs the magnitude and phase of the MR images as separate channels. In like manner the input to the network was in magnitude-phase form. Appropriate scaling was used in preprocessing to ensure that each channel was given equal weight in training.

# CHAPTER 4

RESULTS

4.1   Result Description

Table 4.1 gives the relative error of SENSE reconstructions using (1) the proposed DnCNN method, (2) a total variation denoising approach and (3) no denoising. In addition, deep denoised results with sub-optimal learning configurations are given using (4) no batch normalization or residual learning, (5) U-Net neural network architecture, (6) direct sensitivity profile learning and (7) magnitude and phase estimates. The relative error was calculated using equation 3.6 with an averaged, fully sampled image taken as ground truth. In addition, the last column of table 4.1 shows the relative error of the predicted SENSE maps. It can be seen that the DnCNN method significantly outperforms the case where no denoising is applied. Additionally, the DnCNN method outperforms the total variation method across all acceleration factors with error reductions up to 28%.

Figures 4.1a, 4.1b and 4.1c show SENSE reconstructions generated using (a) the proposed DnCNN method, (b) a total variation denoising approach and (c) no denoising. Similary, figures 4.2a, 4.2b and 4.2c show ground-truth difference images from (a) the proposed DnCNN method, (b) a total variation denoising approach and (c) no denoising. It can be seen that the DnCNN method is less noisy and contains diminished aliasing artifact when compared to the total variation denoised reconstruction. SENSE reconstructions taken

Table 4.1: SENSE reconstruction percent errors using (1) deep denoised sensitivity profile, (2) total variation denoised sensitivity profile, (3) noisy sensitivity profile, (4) deep denoised sensitivty profile without batch normalization or residual learning, (5) deep denoised profile using pooling layers, (6) deep denoised profile with direct profile estimation and (7) magnitude and phase estimated image data. The final column shows the percent error of sensitivity profile.

| Acceleration Factor:	Rx2 Ry2	Rx2 Ry3	Rx2 Ry4	Percent Error Map
Deep Denoised	1.96%	3.58%	6.4%	20%
TV Denoised	2.52%	4.55%	8.3%	24%
Noisy	11.1%	17.5%	22.5%	52%
No Batch No Res	2.13%	3.78%	7.14%	21%
U-Net Denoised	2.27%	4.09%	7.53%	24%
Direct Sensitivity	4.17%	5.41%	8.63%	19%
Magnitude and Phase	8.40%	13.03%	22.33%	37%

from unaltered sensitivity profiles show severe aliasing artifact even at low accelerations.
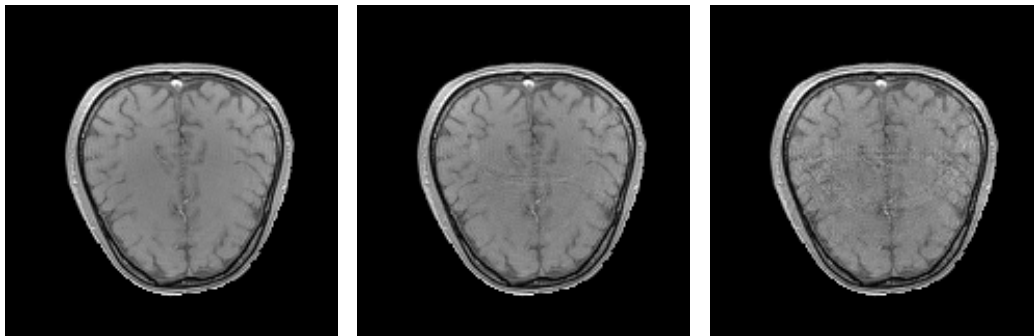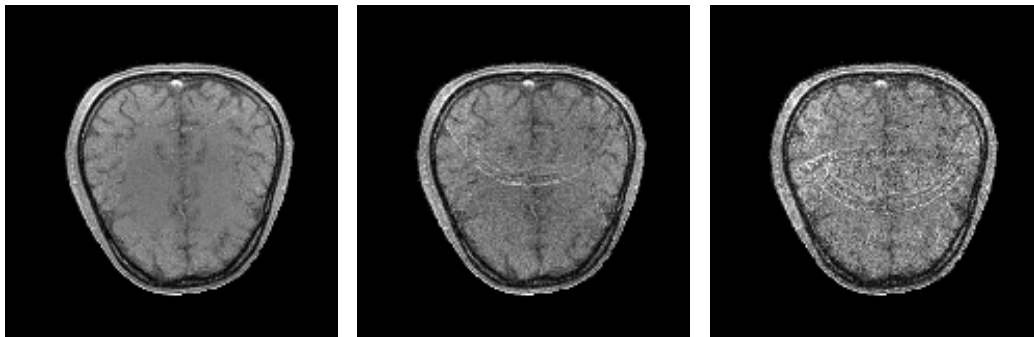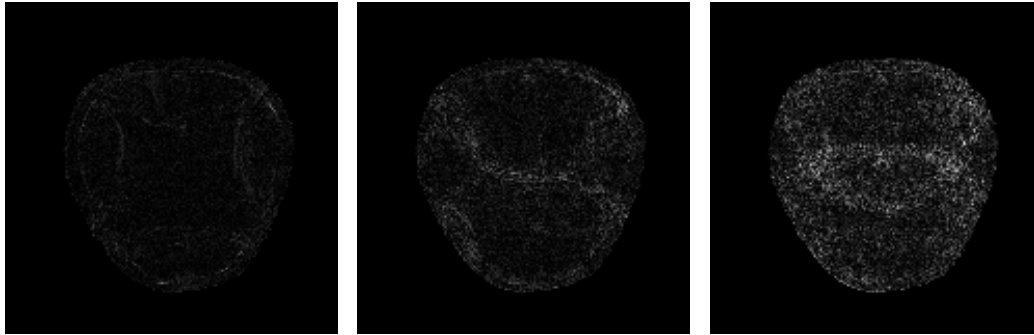
(a) $Ry = 2$ $Ry = 3$ $Ry = 4$
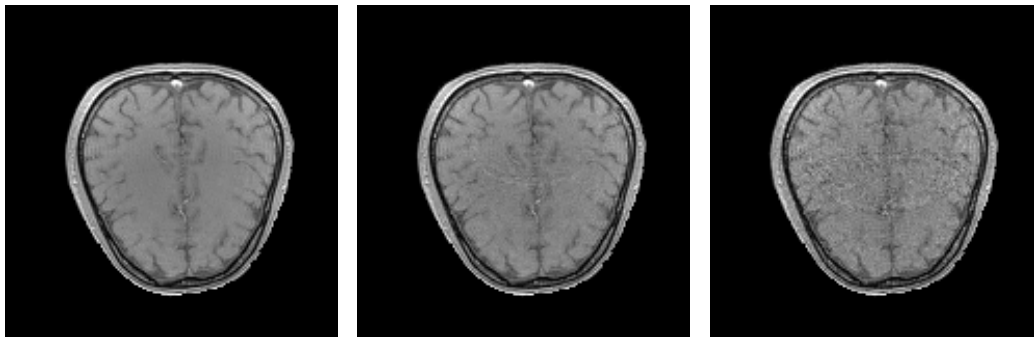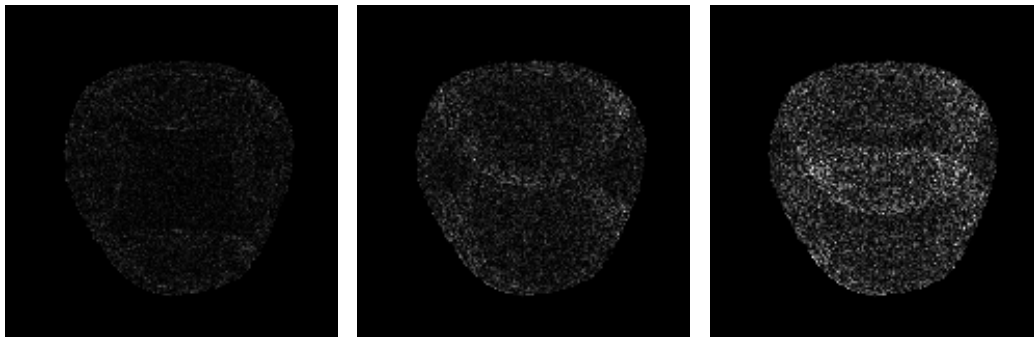
(b)

(c)

Figure 4.1: SENSE reconstruction results using (a) deep denoised sensitivity profiles, (b) total variation denoised sensitivity profiles, (c) noisy sensitivity profiles. Acceleration factors Rx = 2 on all images Ry = 2, 3, 4 across each of the columns.
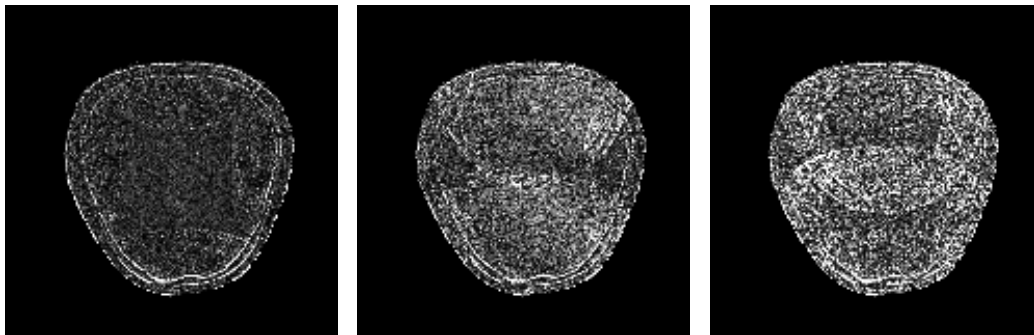
(a) $Ry = 2$ $Ry = 3$ $Ry = 4$

(b)

(c)

Figure 4.2: SENSE reconstruction difference images using (a) deep denoised sensitivity profiles, (b) total variation denoised sensitivity profiles, (c) noisy sensitivity profiles. Acceleration factors Rx = 2 on all images Ry = 2, 3, 4 across each of the columns.

Figures 4.3a, 4.3b and 4.3c show suboptimal deep denoised SENSE reconstructions generated using (a) direct sensitivity profile learning, (b) deep learning without batch normalization or residual learning and (c) deep learn-

ing with U-Net architecture. Similary, figures 4.4a, 4.4b and 4.4c show ground-truth difference images using (a) direct sensitivity profile learning, (b) deep learning without batch normalization or residual learning and (c) deep learning with U-Net architecture. It can be seen that SENSE maps which were directly learned by the neural network contain significant errors on the boarders of the subject. In like manner, the U-Net architecture or removal of residual learning leads to suboptimal results. Figure 4.5 shows the reconstructions and difference images from the denoised SENSE maps from the network trained to jointly predict the magnitude and phase of the noisy image data.

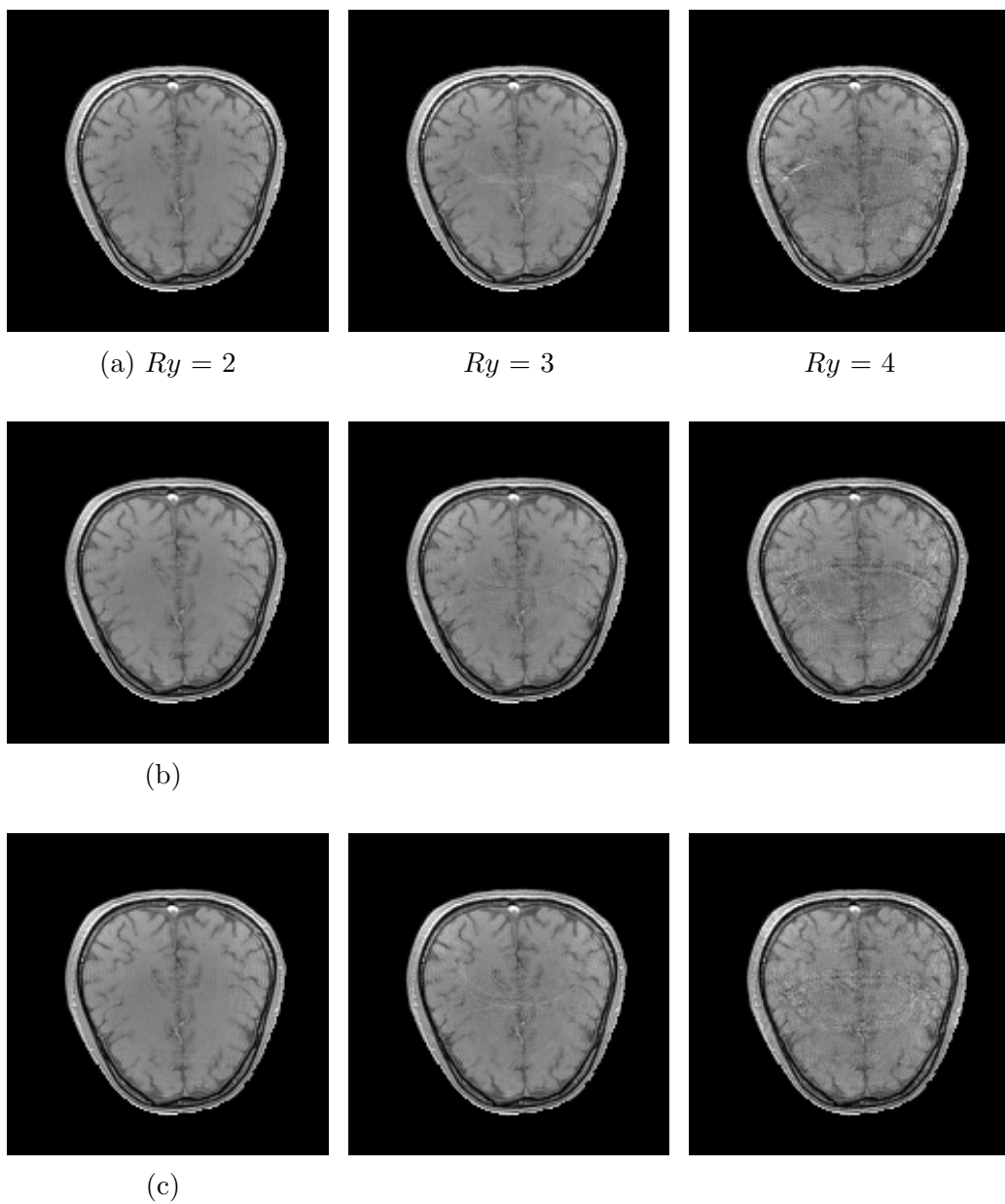(a) $Ry = 2$ $Ry = 3$ $Ry = 4$

(b)

(c)

Figure 4.3: SENSE reconstruction results using (a) deep directly learned sensitivity profiles, (b) deep profiles without batch normalization or residual learning, (c) deep profiles using U-Net architecture. Acceleration factors Rx = 2 on all images Ry = 2, 3, 4 across each of the columns.

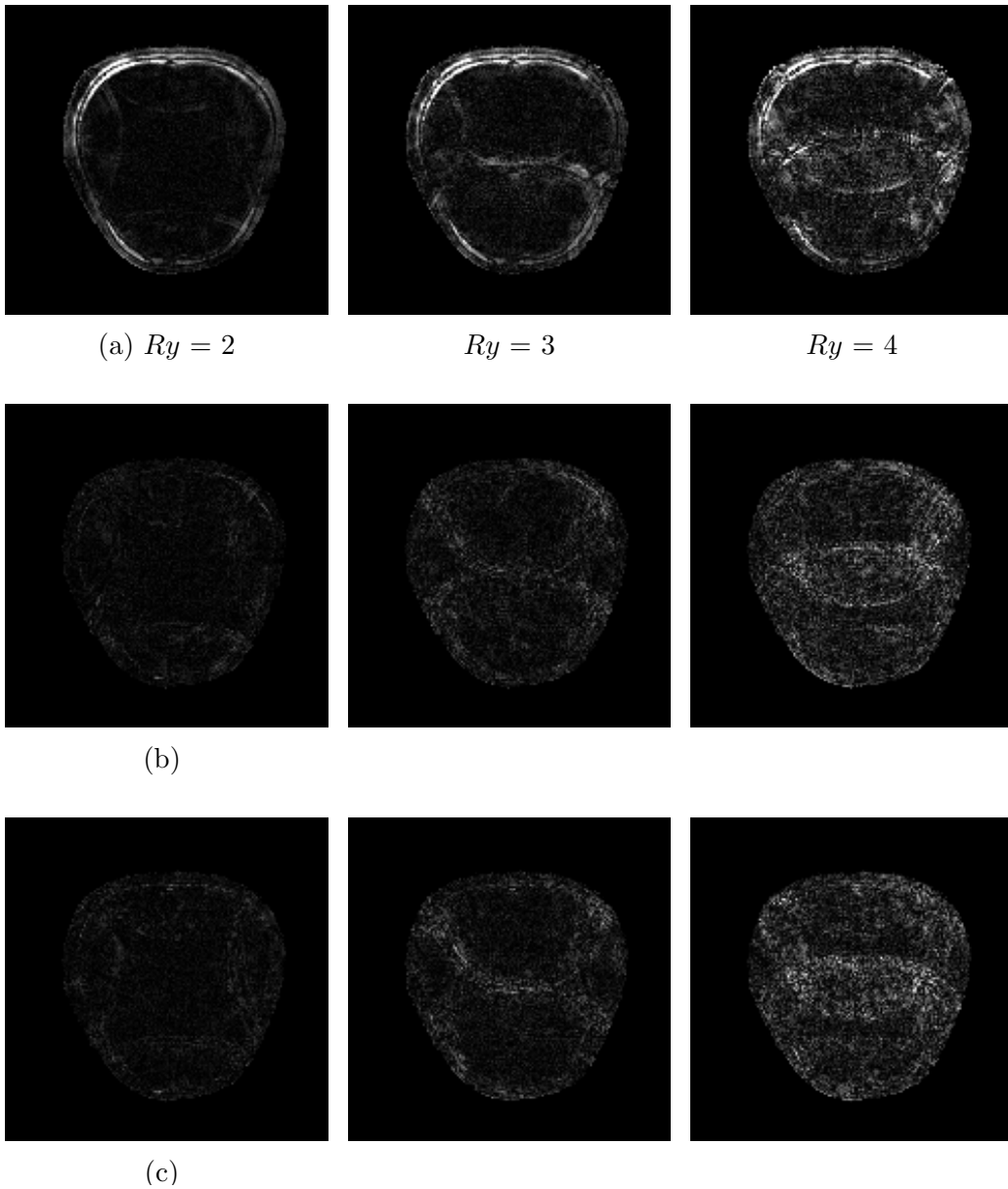(a) $Ry = 2$ $Ry = 3$ $Ry = 4$

(b)

(c)

Figure 4.4: SENSE reconstruction difference images using (a) deep directly learned sensitivity profiles, (b) deep sensitivity profiles without batch normalization or residual learning, (c) deep sensitivity profiles using U-Net architecture. Acceleration factors Rx = 2 on all images Ry = 2, 3, 4 across each of the columns.

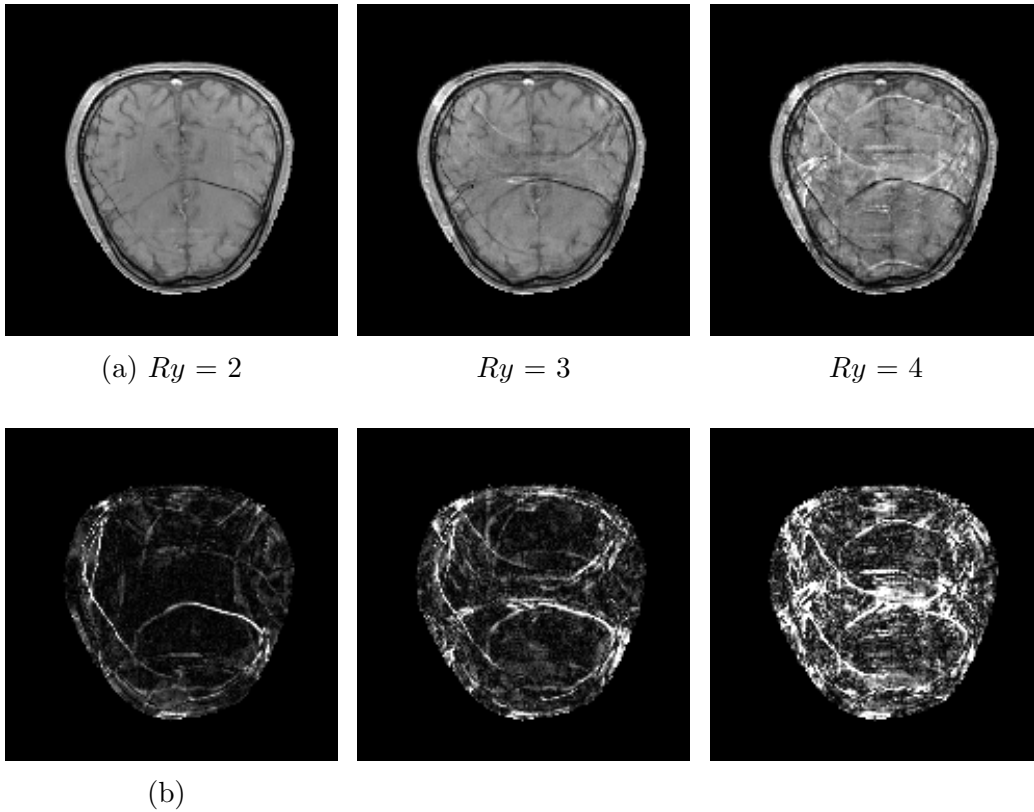(a) $Ry = 2$ $Ry = 3$ $Ry = 4$

(b)

Figure 4.5: SENSE reconstructions (a) and difference images (b) using network predictions where magnitude and phase are predicted. Acceleration factors Rx = 2 on all images Ry = 2, 3, 4 across each of the columns.

CHAPTER 5

CONCLUSION

5.1   Direction for Future Work

As sensitivity profiles are three-dimensional, there is additional information that can be exploited by making three-dimensional estimations. This can easily be achieved by modifying the network architecture to allow three-dimensional inputs and outputs. Due to memory constraints on the hardware used in this work, a 3D network would be trained on patches no greater than 80 x 80 x 80, however this is satisfactory for the task of denoising.

Transfer learning and prior information can be included in the learning process in many ways, such as:

- Baseline sensitivity profiles taken from a uniform phantom previously may serve as prior information.

- Low-resolution, high-SNR images can be rapidly acquired and used to solve a joint image-denoising and super-resolution problem.

- Images from different coils can be jointly estimated to enable transfer learning.

While this project estimated the MR data which was subsequently normalized, it is possible for the network to directly estimate the sensitivity profile that optimizes SENSE reconstructions. Work would need to be done to im-

plement the SENSE reconstruction in Tensorflow in a manner that allows efficient backpropagation of gradients.

5.2   Summary of Findings

It has been shown that deep convolutional neural networks outperform total variation on the task of denoising MRI images. Furthmore, these gains directly lead to improved SENSE reconstructions. Improvements in image reconstruction come in the form of reduced aliasing artifact and reduced random noise. Sensitivity estimation with deep-denoising enables a factor-of-two increase in image acceleration with image quality improvements as compared to the case of no noise removal.

5.2.1   Network Characteristics

Pooling layers

The success of deep neural networks is heavily dependent of the neural network architecture which is chosen. In the task of denoising MR images, it can be seen that the DnCNN architecture outperforms the U-Net architecture. This is likely because the U-Net has a significant number of pooling layers which shrink the intermediate feature map sizes and cause blurring in the final output. The removal of pooling layers does not have a significant negative impact on training, because only low-level image features are required for the denoising task.

SAME convolutions

This network utilizes "SAME" convolutions which means that feature maps are zero padded prior to convolution layers to avoid a change in feature map size. It has been supposed that such zero padding can lead to boundary artifacts; however, no such artifacts were discovered in the results.

Phase wrapping

While some authors have had success in directly learning images corrupted by phase wrapping [14], the results show that the combination of noisy images, phase wrapping and MSE loss causes significant errors. This problem is exacerbated by noise in the training image which causes the network to predict smooth transitions from -pi to pi rather than a sharp discontinuity. This causes severe artifacts in the location of the discontinuity.

Direct sensitivity estimation

The neural network which directly predicts sensitivity profiles shows severe boundary artifacts. This is due to the masks which are used in the weighted MSE function. While it is possible to train the network without masks, the high-variance background noise of the sensitivity profile dominates the training error and prevents convergence.

Residual learning

This network uses residual learning and batch normalization. As can be seen from the results, this combination leads to improved image reconstructions. Furthermore, convergence happens faster when batch normalization and residual learning are used.

REFERENCES

[1] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, "SENSE: Sensitivity encoding for fast MRI," *Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 952–962, 1999. [Online]. Available: http://dx.doi.org/10.1002/(SICI)1522-2594(199911)42:5¡952::AID-MRM16¿3.0.CO;2-S

[2] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychological Review*, pp. 65–386, 1958.

[3] F. Lam, C. Ma, B. Clifford, C. L. Johnson, and Z.-P. Liang, "High-resolution 1H-MRSI of the brain using SPICE: Data acquisition and image reconstruction," *Magnetic Resonance in Medicine*, vol. 76, no. 4, pp. 1059–1070, 2016. [Online]. Available: http://dx.doi.org/10.1002/mrm.26019

[4] Z.-P. Liang and P. C. Lauterbur, *Principles of Magnetic Resonance Imaging*. IEEE Press, 2000.

[5] R. W. Brown and Y.-C. Cheng, *Magnetic Resonance Imaging: Physical Principles and Sequence Design*, 2nd ed. John Wiley & Sons Inc., 2018.

[6] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magnetic Resonance in Medicine*, vol. 47, no. 6, pp. 1202–1210, 2002. [Online]. Available: http://dx.doi.org/10.1002/mrm.10171

[7] M. Lustig and J. M. Pauly, "Spirit: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space," *Magnetic Resonance in Medicine*, vol. 64, no. 2, pp. 457–471, 2010. [Online]. Available: http://dx.doi.org/10.1002/mrm.22428

[8] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals and Systems*, vol. 2, no. 4, pp. 303–314, Dec 1989. [Online]. Available: https://doi.org/10.1007/BF02551274

[9] K. Hornik, "Approximation capabilities of multilayer feedforward network. neural networks, 251-257," vol. 4, 01 1991.

[10] H. Mhaskar, Q. Liao, and T. A. Poggio, "Learning Real and Boolean Functions: When Is Deep Better Than Shallow," *CoRR*, vol. abs/1603.00988, 2016. [Online]. Available: http://arxiv.org/abs/1603.00988

[11] M. Telgarsky, "Benefits of depth in neural networks," *CoRR*, vol. abs/1602.04485, 2016. [Online]. Available: http://arxiv.org/abs/1602.04485

[12] M. Bianchini and F. Scarselli, "On the complexity of neural network classifiers: A comparison between shallow and deep architectures," vol. 25, pp. 1553–1565, 08 2014.

[13] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 5 2015.

[14] D. Lee, J. J. Yoo, and J. C. Ye, "Deep artifact learning for compressed sensing and parallel MRI," *CoRR*, vol. abs/1703.01120, 2017. [Online]. Available: http://arxiv.org/abs/1703.01120

[15] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, and D. Liang, "Accelerating magnetic resonance imaging via deep learning," in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, April 2016, pp. 514–517.

[16] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, "Learning a variational network for reconstruction of accelerated MRI data," *CoRR*, vol. abs/1704.00447, 2017. [Online]. Available: http://arxiv.org/abs/1704.00447

[17] C. M. Hyun, H. Pyung Kim, S. M. Lee, S. Lee, and J. K. Seo, "Deep learning for undersampled MRI reconstruction," *ArXiv e-prints*, Sep. 2017.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[19] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *CoRR*, vol. abs/1608.03981, 2016. [Online]. Available: http://arxiv.org/abs/1608.03981

[20] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning." *CoRR*, vol. abs/1603.07285, 2016. [Online]. Available: http://dblp.uni-trier.de/db/journals/corr/corr1603.html

[21] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CoRR*, vol. abs/1502.03167, 2015. [Online]. Available: http://arxiv.org/abs/1502.03167

[22] M. Kweon, "U-Net implementation in tensorflow," Jan 2018. [Online]. Available: https://github.com/kkweon/UNet-in-Tensorflow

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015. [Online]. Available: http://arxiv.org/abs/1505.04597

[24] Z. Iqbal, D. Nguyen, and S. Jiang, "Super-Resolution 1H Magnetic Resonance Spectroscopic Imaging utilizing Deep Learning," *ArXiv e-prints*, Feb. 2018.

[25] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *CoRR*, vol. abs/1611.03679, 2016. [Online]. Available: http://arxiv.org/abs/1611.03679