Abstract of thesis entitled

"The Application of Generative Networks in MR Image Reconstruction"

Submitted by

Luo Guanxiong

for the degree of Master of Philosophy
at The University of Hong Kong
in October 2019

In this thesis, the application of generative networks in MR image reconstruction is demonstrated from the perspective of Euclidean loss to the perspective of likelihood loss. The method trained by Euclidean loss consists of a generator and a discriminator. The generator serves as the proximal operator in the ADMM loop, and the discriminator is to define an MR image space when training the generator. The other method utilizes the autoregressive network as a prior model. The reconstruction is modeled by Bayesian theorem and achieved by maximizing the posterior. These two proposed methods are all generalizable for difference reconstruction settings and reserve the conventional way to enforce data consistency instead of embedding that into the network, that's to say, the learned component used in the reconstruction is separated from the process of k-space data generation. These two methods are tested with knee and brain MRI database, and it shows considerable improvement over the former reconstruction methods such as parallel imaging and compressed sensing.

(200 words)

The Application of Generative Networks in MR Image Reconstruction

by

Luo Guanxiong

B.Eng. *X.J.T.U.*

A thesis submitted in partial fulfilment of the requirements for
the Degree of Master of Philosophy
at The University of Hong Kong.

October 2019

To my loved ones.

Declaration

I declare that this thesis represents my own work, except where due acknowledgement is made, and that it has not been previously included in a thesis, dissertation or report submitted to this University or to any other institution for a degree, diploma or other qualifications.

Signed . . . . . .

Luo Guanxiong

Acknowledgements

It is really an extraordinary experience for me to complete my MPhil program at HKU, which is way more than I expected. Firstly, I would like to express my sincere gratitude to my supervisor Dr. Cao for the continuous support of my MPhil study and related research, for his patience, inclusive mind, and immense knowledge. His guidance helped me in the time of research and writing of this thesis. I could not have imagined having a better supervisor and mentor for my MPhil study.

Besides my supervisor, I would like to thank the Lap-Chee college for granting me two years of residential right. That means a lot for me, especially in such a place where housing is insanely expensive, let alone the atmosphere, provided by the college, where the residents from different countries with the diverse cultural background can exchange thoughts without any limits. And also the culture of freedom advocated by HKU indeed diversified me, but also makes me realize that freedom does not mean that there is no restriction and it does mean that there should exist mutual respect and understanding.

What's more, this program did enrich me a lot not only in academic fields but also grows me in many aspects of my life. And I began to understand how to make balance and compromise. The academic performance is indeed essential and meaningful for a young person who wants to be influential but makes sure to remember life is more than your career when you are frustrated or depressed by what you are passionate about. Sometimes, your passion is not driven by your inner thoughts but pushed by others' recognition. It is too arbitrary to tell which one is superior. But what is supposed to keep in mind is any sort of emotion emerged in an extreme way will decay rapidly with only banality and vacuity left.

I want to end my acknowledgement with emotional words. Sometimes, when I feel lonely, I step out my dorm and follow the street leading to the seaside of Kennedy Town, enjoying the sunset splashed on the peaceful sea, listening to the in-depth thoughts from your mind. This is the feeling of serenity. To reserve the gratitude for beautiful things you meet and try to explore as more possibility as possible.

Contents

List of Figures

List of Tables

List of Abbreviations

ℓ_1-ESPIRiT an eigenvalue approach to autocalibrating parallel MRI 23

ADMM alternating direction method of multipliers 14

GAN generative adversarial network 15

GRAPPA generalized autocalibrating partially parallel acquisition 1

MAP maximum a posterior estimation 33

MODL model-based deep learning architecture for inverse problems 14

PSNR peak signal-to-noise ratio 23

QSM quantitative susceptibility mapping 47

SENSE sensitivity encoding for fast MRI 1

Chapter 1

Introduction

Magnetic Resonance Imaging (MRI) is commonly used imaging modality in clinical practice. The shortcoming of MRI is its inherent slow imaging speed. Therefore, a method how to shorten the long scanning time has been a hot research topic in MRI. The methods to accelerate imaging has evolved from the perspective of parallel imaging to the perspective of compressed sensing imaging. With the advent of deep learning, many studies are seeking to develop learning-base reconstruction with the existing MR image database. The establishment of generic image prior is a challenge problem and the focus of this thesis. A brief outline describing the focus of each chapter is provided below.

1.1 Thesis Outline

Chapter 1 gives a brief introduction to the principle of MR imaging from the perspective of signal processing. Parallel imaging techniques, SENSE and GRAPPA, are discussed with several pages. As the origin of learning-based, the application of compressed sensing to image reconstruction is introduced. Chapter 2 presents generalizable image reconstruction with a modified optimization algorithm whose proximal operation for ℓ_1 regularization has been replaced by a separated neural network. The network is trained with the existing knee MR images. Chapter 3 explores image reconstruction for the undersampled k-space data with Bayesian inference. In this method, a generative network serves as prior model which is learnt from image database and the reconstruction is achieved by maximizing posterior probability. Chapter 4 gives a brief summary of the development of image reconstruction and future research recommendation.

1.2 MR imaging principles

Magnetic resonance imaging is a non-invasive medical imaging modality, and effective for studying the anatomy and function of central nervous and musculoskeletal

systems, such as brain and knee. MR imaging is not only free of ionizing radiation but also full of means for characterizing the tissues and organs.

1.2.1 Signal generation and k-space

MR imaging is based on the physics of magnetic resonance, which is a phenomenon that the proton spins interact with the magnetic field. Different magnetic fields are utilized to manipulate the hydrogen spins resided in various tissues, and the radiofrequency coils are used to detect the precession of those spins after excitation. With the presence of external static filed B_0, the magnetic moment vector generated by a spinning proton precesses around the field direction, and the precession frequency (Larmor frequency) is given by

$$\omega_0 = \gamma B_0 \ , \tag{1.1}$$

where γ is gyromagnetic ratio. And a magnetic moment vector has either parallel alignment along with static field or anti-parallel alignment. The number of spins parallel to the magnetic field exceeds the number anti-parallel to that field, and the spin excess is the source of equilibrium magnetization, given by

$$M_0 = \frac{\rho_0 \gamma^2 \hbar}{4kT} B_0 \tag{1.2}$$

In order to get a detectable magnetization component in the transverse plane, the radiofrequency magnetic field (RF pulse) is switched on for a short time period to tip magnetization away from equilibrium. The transverse component is represented by a complex function defined as below:

$$M_{xy}(\vec{r},t) = M_x(\vec{r},t) + iM_y(\vec{r},t) \ , \tag{1.3}$$

where $\vec{r}$ is a vector in a 3-D coordinate, the component M_x and M_y are detected by the receiver coils. It should be noted that the relaxation of transverse magnetization and the regrowth of longitudinal magnetization would be triggered by the RF pulse excitation. The relaxation of transverse magnetization is referred to as the decay of transverse magnetization induced by the variation in the local field. The local field comes from the externally applied field and the field of the surrounding spins (i.e., spin-spin relaxation). Let the time origin $t = 0$ start at when the RF (excitation) pulse finishes, we have

$$M_{xy}(\vec{r},t) = M_{xy}(\vec{r},0)e^{-t/T_2(\vec{r})} \tag{1.4}$$

Here, we define $M(\vec{r},0)$ as the instantaneous transverse magnetization after excitation, T_2 as the relaxation time. In imaging reconstruction, two aspects of transverse magnetization are normally considered: the relaxation and the precession. Commonly in MR terminologies, T_2 only accounts for spin-spin relaxation, T_2^* accounts for both external field and spin-spin effects. In conjunction with the T_2 or T_2^* relaxation, the regrowth of longitudinal magnetization, i.e., magnetization returns back to the equilibrium, follows

$$M_z(\vec{r},t) = M_z(\vec{r},t)e^{-t/T_1(\vec{r})} + M_0(1-e^{-t/T_1(\vec{r})}) \tag{1.5}$$

Here, $M_z(t)$ denotes the longitudinal magnetization. Furthermore, different tissues have different T_1 and T_2 values, which contribute to the image contrast.

The precession frequency is used for Fourier encoding. Referring to as the Larmor frequency, and the precession frequency at each spatial location is given by

$$\omega(\vec{r},t) = \gamma B_z(\vec{r},t) \tag{1.6}$$

Here, $B_z(\vec{r},t)$ is comprised of three components

$$B_z(\vec{r},t) = B_0 + \Delta B_0(\vec{r}) + \vec{G}(t)\cdot\vec{r} \tag{1.7}$$

where $\Delta B_0(\vec{r})$ denotes the static field's inhomogeneity, which is called B0 field map. The spatial-varied field gradient $\vec{G}(t)\cdot\vec{r} = \vec{G}_x(t)\vec{i} + \vec{G}_y(t)\vec{j} + \vec{G}_z(t)\vec{k}$ is applied during the signal readout stage to provide spatial encoding with designed gradient waveform. Therefore, the transverse magnetization at a given spatial location is a time-varying function.

$$M_{xy}(\vec{r},t) = M_{xy}(\vec{r},0)e^{-t/T_2(\vec{r})}e^{-i\gamma\int_0^t B_z(\vec{r},t)dt} \tag{1.8}$$

$$= M_{xy}(\vec{r},0)e^{-t/T_2(\vec{r})}e^{-i\omega_0 t}e^{\phi_0(\vec{r})}e^{-i\gamma\int_0^t \vec{G}(\tau)\cdot\vec{r}d\tau} \tag{1.9}$$

For the simplicity, the relaxation term $e^{-t/T_2(\vec{r})}$ and the field map term $e^{\phi_0(\vec{r})}$ are ignored. On MR scanner, the receiver coil can detect the magnetic field flux changes induced by the precession of transverse magnetization. The received MR signal $s_r(t)$ is from all precessing transverse magnetization in the volume:

$$s_r(t) = \int_V M_{xy}(\vec{r},t)dv = \int_x\int_y\int_z M(x,y,z,t)dxdydz. \tag{1.10}$$

For two-dimensional (2D) imaging, we use RF pulse to excite spin in a slice-by-slice

manner. Hence, we write

$$m(x,y) = \int_{z_0-\Delta z/2}^{z_0+\Delta z/2} M(x,y,z)dz \tag{1.11}$$

and

$$s_r(t) = \int_x \int_y m(x,y) e^{-i\omega_0 t} e^{-i\gamma \int_0^t \vec{G}(\tau)\cdot\vec{r}d\tau} dxdy \tag{1.12}$$

where $m(x,y)$ is a composite representation of transverse magnetizations, which is modulated by the gradient and RF pulses applied. The term $e^{-i\omega_0 t}$ can be removed with signal demodulation during the image reconstruction. For 2D spatial encoding, only field gradients along $\vec{G}_x(t)$ and $\vec{G}_y(t)$ are needed. The Eq (1.12) is rewritten as

$$s(t) = \int_x \int_y m(x,y) e^{-i2\pi[k_x(t)+k_y(t)]} dxdy \tag{1.13}$$

where

$$\begin{aligned} k_x(t) &= \frac{\gamma}{2\pi} \int_0^t G_x(\tau)d\tau \\ k_y(t) &= \frac{\gamma}{2\pi} \int_0^t G_y(\tau)d\tau \end{aligned} \tag{1.14}$$

are time integrals of the gradient waveforms, which are referred to as the trajectory function in the k-space. At a given time t, $s(t)$ equals to the coefficient given by the 2D Fourier transform for $m(x,y)$ at the some spatial-frequency location. Commonly, the Fourier transform space is referred to as k-space $M(k_x,k_y)$. To fully represent $m(x,y)$, a set of $\{s(t)\}$ is required, which covers a sufficient part of k-space to allow the reconstruction of $m(x,y)$.

1.2.2 Sampling requirements

MR image formation depends on the appropriate coverage in k-space. In 2D imaging, the ”raster”-like k-space trajectory corresponds to rectangular coverage on a Cartesian grid, allowing convenient image reconstruction with a 2D DFT. By far, our analysis has been with continuous-time signals, but in practice, the k-space data is sampled at discrete points in k-space. For each repeated sampling in k-space, frequency encoding (or so-called readout) gradient pulse G_y is fixed and phase encoding gradient pulse G_x is incremental, covering the entire Cartesian k-space. The discrete sampling periods are Δk_x and Δk_y; while the highest spatial-frequencies sampled are k_{xmax} and k_{ymax}. With 2D imaging, an asymmetry exists in the k-space acquisition as samples along the k_x direction accumulate during a single readout

while samples along the k_y direction accumulate by repeating the excitation for different phase encodes. We can express the sampled k-space data $\hat{M}(k_x, k_y)$ as

$$\hat{M}(k_x, k_y) = M(k_x, k_y) \cdot \left(\frac{1}{\Delta k_x \Delta k_y} \right) \text{III} \left(\frac{k_x}{\Delta k_x}, \frac{k_y}{\Delta k_y} \right) \cdot \text{Rect} \left(\frac{k_x}{W_{k_x}}, \frac{k_y}{W_{k_y}} \right) \tag{1.15}$$

where

$$\begin{aligned} W_{k_x} &= \left(k_{xmax} + \frac{\Delta k_x}{2} \right) \\ W_{k_y} &= \left(k_{ymax} + \frac{\Delta k_y}{2} \right) \end{aligned} \tag{1.16}$$

are width of k-space or distance traversed in k-space for each direction. The resultant Fourier transform into the object domain relates the image $\hat{m}(x, y)$ to the theoretical image content $m(x, y)$, as given by

$$\hat{m}(x, y) = m(x, y) * \text{III}(\Delta_{k_x} x, \Delta_{k_y} y) * W_{k_x} W_{k_y} \, \text{sinc}(W_{k_x} x) \, \text{sinc}(W_{k_y} y) \tag{1.17}$$

where $*$ denotes convolution. Eq. 1.17 describes how the image field of view (FOV) and spatial resolution relate to the k-space sampling periods $(\Delta_{k_x}, \Delta_{k_y})$ and k-space widths (W_{k_x}, W_{k_y}).

According to the Nyquist sampling theorem, discrete sampling in frequency domain amounts to replications in the corresponding spatial domain. In MRI, sampling takes place in the k-space, i.e., the reciprocal space; hence, replications in the spatial domain is unavoidable and must be considered in designing the k-space trajectory. Assuming the k-space sampling has an infinite/ideal coverage, then from Eq. 1.17 the resultant object is

$$\hat{m}(x, y) = m(x, y) * \text{III}(\Delta_{k_x} x, \Delta_{k_y} y) \tag{1.18}$$

Convolution of $m(x, y)$ with $\text{III}(\Delta_{k_x} x, \Delta_{k_y} y)$ implies replication of $m(x, y)$ at intervals of $1/\Delta_{k_x}$ in the x direction and by $1/\Delta_{k_y}$ along the y direction. The separation of these replications has to be greater than the effective FOV in the image domain. From Eq. 1.18, we can write that

$$\begin{aligned} \text{FOV}_x &= \frac{1}{\Delta_{k_x}} = \text{sampling rate along } k_x \\ \text{FOV}_y &= \frac{1}{\Delta_{k_y}} = \text{sampling rate along } k_y \end{aligned} \tag{1.19}$$

Fig1.1 illustrates the relation between the FOV and $(\Delta_{k_x}, \Delta_{k_y})$ in k-space for 2D

imaging. Along the frequency encoding direction, the (Δ_{k_x} depends on the readout gradient amplitude G_x and the sampling interval Δt. Along the phase-encoding direction, the Δ_{k_y} is determined by the gradient amplitude G_y and the fixed phase-encoding duration τ_y. Their mathematical expressions are

$$\begin{aligned} \Delta_{k_x} &= \frac{\gamma}{2\pi} G_{xr} \Delta t \\ \Delta_{k_y} &= \frac{\gamma}{2\pi} G_{yi} \tau_y \end{aligned} \tag{1.20}$$

Therefore the expression for FOV can be written as

$$\begin{aligned} \mathrm{FOV}_x &= \frac{1}{\Delta_{k_x}} = \frac{1}{\frac{\gamma}{2\pi} G_{xr} \Delta_t} \\ \mathrm{FOV}_y &= \frac{1}{\Delta_{k_y}} = \frac{1}{\frac{\gamma}{2\pi} G_{yi} \tau_y} \end{aligned} \tag{1.21}$$

To summarize, the k-space sampling and Fourier transform in MRI reconstruction allow:

- recording N k-space lines at different k_y positions by changing the area under the G_y gradient for each measurement.
- reconstructing $m(x, y)$ by 2D Fourier transformation of the k-space data, which reside on a Cartesian grid.

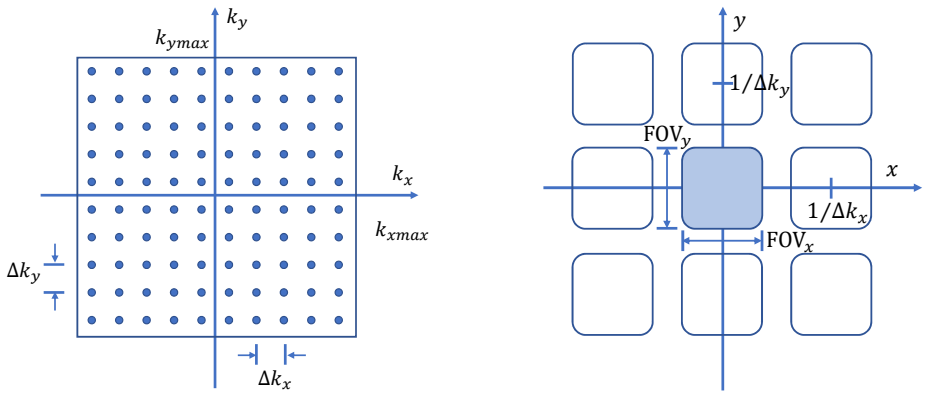

Figure 1.1 (a)Sampling in k-space. (b) Corresponding replication in object domain

In Nyquist sampling theorem, aliasing arises from that the actual sampling rate is below the Nyquist sampling rate. In MRI, insufficient sampling rate in k-space, i.e., in the reciprocal space, results in aliasing artifacts in the spatial domain. Because objects are spatially bounded, aliasing can be avoided in MR imaging

by increasing the FOV through reducing gradient areas, i.e., $G_{xr}\Delta t$ or $G_{yi}\tau_y$ in the k-space acquisition.

In the phase-encoding direction, the gradient area, $G_{yi}\tau_y$, can be reduced to increase the FOV. However, for fixed scan time, the number of phase-encoding lines remains fixed, and thus, the highest spatial frequency sampled along k_y is reduced. This reduces the spatial resolution. For maintaining the same resolution along y, additional phase encodes must be acquired, which increase the scan time. Hence an increase in scan time or a reduction in the spatial resolution are the trade-offs for avoiding aliasing in the phase-encoding direction.

In the readout direction, k-space samples are acquired rapidly through an A/D converter with a sampling rate far beyond the Nyquist sampling rate for a given FOV. Then the readout signal from A/D converter is processed by an anti-aliasing filter with bandwidth commensurate with the sampling frequency of the A/D converters and FOV. The residual aliasing effects along the readout direction could be caused by the tail in the filter's frequency response. But the aliasing artifact along readout direction should be sufficiently small and can be ignored in any MRI scans.Regarding the computation, we express the relationship between the image $m(x,y)$ and k-space (k_x,k_y) with a compact linear equation

$$(k_x,k_y) = \mathscr{F}m(x,y) \tag{1.22}$$

where $\mathscr{F}$ is DFT.

1.3 Methods to accelerate imaging

Along the phase encoding direction, the full traverse in k-space requires a relatively long time to finish, especially for high-resolution imaging. Therefore, in practice, it is preferable to accelerate imaging, utilizing reconstruction techniques such as parallel imaging and compressed sensing. Meanwhile, when applying those accelerated reconstruction techniques, there is a trade-off between acceleration factor and the image quality. In the following subsections, a brief introduction of acceleration methods is presented.

1.3.1 Parallel imaging

The long scanning time is a limiting factor in many clinical applications, but the speed of sufficient k-space acquisition is restricted by the gradient slew rate and nerve stimulation threshold, making high-quality images time-consuming. Parallel imaging is an effective approach for shortening the scanning time, using k-space data received with a phase-array coil, i.e., an array of coils.

The long MRI scan is caused by a large number of phase encodings, which are k-space lines at difference k_y positions. Usually, RF coils have inhomogeneous sensitivity maps over the image object, and parallel imaging utilizes this property to reduce the scanning time by exploiting the spatially varied information characterized by those sensitivity maps. With phase-array coil, k-space can be then undersampled by skipping lines in phase encoding, usually with a constant interval, which is required by the particular reconstruction method. Finally, the reconstruction method is employed to reconstruct a full k-space data or an aliasing-free image.

Various methods have been developed for the parallel imaging reconstruction in both spatial and frequency domains, or both[2]. The signal from a receiver RF coil array with N channels is given by

$$s_i(t) = \int_\Omega \mathrm{d}\vec{r} m(\vec{r}) c_j(\vec{r}) e^{-2\pi i \vec{k}(t)\cdot\vec{r}} \quad 1 \leq j \leq N \tag{1.23}$$

The image or volume is modulated by complex-valued sensitivity maps c_j of all receiver coils. The k-space signals $s_i(t)$ are then given by the Fourier transform of the coil images $c_j m$ sampled at discrete time points t_l along a given k-space trajectory $\vec{k}(t)$. Undersampling can be used to accelerate the measurement, but this can cause aliasing artifact with direct reconstruction, as discussed in Section 1.2.2. The objective of parallel imaging is to reconstruct artifact-free images from the undersampled k-space data by optimally exploiting the complementary information from multiple receiver coils. Due to the weak encoding power of the coil sensitivities, it can accelerate acquisition using a modest factor of 2 in each spatial dimension. In the following subsections, two commonly used parallel imaging methods: SENSE and GRAPPA are introduced. SENSE operates in the image domain, while GRAPPA operates on k-space data.

SENSE

The sensitivity encoding for fast MRI was proposed by Pruessmann [21]. With SENSE, the scan time is reduced by the acceleration factor of R, through skipping phase-encoded lines in k-space. According to Eq. (1.21), the FOV of the reconstructed image is reduced by the factor R. If the object extends outside the reduced FOV, some pixels will be aliased or wrapped. Aliased pixels are displaced by integer multiples of L/R, where L is the original phase-encoding FOV. Without loss of generality, we take the phase-encoded direction to be the y direction of the image. Let the total number of replicates be N_A. The value of N_A is pixel-dependent and determined by R and the size and shape of the object. For a square object with the same dimension in the phase-encoded direction as the original FOV, the relation is given as $N_A = R$. If the object dimension is smaller than FOV, then some pixels will have $N_A < R$. If the object diameter dimension is larger than L, some pixels will have $N_A > R$ and there will be phase wrap even without SENSE FOV reduction. The R-fold FOV reduction results in a N_A-fold aliased image. For each location y, we can write the image intensity $I_j(y)$ as s superposition of the original and the displaced replicates:

$$I_j(y) = \sum_{n=0}^{N_A-1} C_j(y+nL/R)M(y+nL/R) \quad j = 0,1,...,N_c-1 \tag{1.24}$$

where N_c is the number of receive coils. Depending on the specific location of y, the number of replicates can vary. Thus, the range of summation is changeable, as schematically shown in Figure. Here unknowns are the aliased image intensity values $M(y+nL/R)$. $I_j(y)$ are known because they are simply the reconstructed aliased images. It is convenient to generalize Eq. (1.24) to a matrix equation. With N_c coils, we define appropriate matrices I, C, and with dimensions $N_c \times 1, N_c \times N_A$, and $N_A \times 1$, respectively, and rewrite Eq. (1.24) as:

$$I = CM \tag{1.25}$$

where:

$$I = \begin{bmatrix} I_0(y) \\ I_1(y) \\ \vdots \\ I_{N_c-1}(y) \end{bmatrix}, M = \begin{bmatrix} M(y) \\ M(y+L/R) \\ \vdots \\ M(y+(N_A-1)L/R) \end{bmatrix} \tag{1.26}$$

and

$$C = \begin{bmatrix} C_0(y) & \dots & C_0(y+(N_A-1)L/R) \\ \vdots & \ddots & \vdots \\ C_{N_c-1}(y) & \dots & C_{N_c-1}(y+(N_A-1)L/R) \end{bmatrix} \tag{1.27}$$

If $N_c \geq N_A$, Eq. (1.25) can be inverted to find the estimated transverse magnetization $\hat{M}$. The most general solution that gives maximum image SNR is the pseudo-inverse

$$\hat{M} = \left[\left(C^{\dagger} \psi^{-1} C \right)^{-1} C^{\dagger} \psi^{-1} \right] I \tag{1.28}$$

where ψ is the $N_c \times N_c$ coil noise correlation matrix in which a diagonal element represents noise variance from a single coil and an off-diagonal element represents a noise cross-correlation between two coils. If $N_c > N_A$, the inversion problem of Eq (1.25) is overdetermined. The extra degrees of freedom are then used with the coil-noise correlation matrix ψ to improve the SNR. If $N_c = N_A$ there are no extra degrees of freedom available to improve the SNR. In the latter case, the coil-noise correlation matrix can be dropped in the solution. This can be derived from the Eq. 1.25 as given by

$$\hat{M} = C^{-1} I \tag{1.29}$$

where $N_c = N_A$, and C is a square matrix. Normally, the off-diagonal elements of ψ are negligible, and the diagonal elements are nearly equal. Therefore, ψ can be replaced by the identity matrix and Eq. (1.25) can be simplified as

$$\hat{M} = \left[(C^{\dagger} C)^{-1} C^{\dagger} \right] I \tag{1.30}$$

Usually, almost no SNR difference can be found in results those omitted the coil-noise matrix from the reconstruction algorithm when coils are well decoupled.

GRAPPA

GRAPPA represents the generalized implementation of parallel imaging in the k-space domain. Although GRAPPA shares the same acquisition scheme with VD-AUTO-SMASH, they differ significantly in the way of reconstructing missing k-space lines. One fundamental difference is that the individual coil signals $S_k(k_y)$ are fit to just a single component coil ACS signal $S_l^{ACS}(k_y + m\Delta_{k_y})$, not a composite signal, thereby deriving the linear weightings to reconstruct missing k-space lines of

each individual coil:

$$S_l^{ACS}(k_y + m\Delta_{k_y}) \cong \sum_{k=1}^{N_c} n_k^{(m)} S_k(k_y) \tag{1.31}$$

This procedure needs to be repeated for every individual coil, and since the coil sensitivities also change along the readout direction, the weights for the GRAPPA reconstruction are normally determined at multiple positions along readout direction. After Fourier transformation, images for each coil in the receiver array are obtained. Furthermore, GRAPPA uses multiple k-space lines from all coils to fit one single coil ACS line, resulting in a further increased accuracy of the fit procedure and therefore a better artifact suppression. The GRAPPA reconstruction can also be written in matrix equations. The vector $\vec{S}$ represents the collected signal in each coil at some position k and therefore has length N_c. Using GRAPPA in its simplest form, a set of weightings $\hat{n}^{(m)}$ has the dimension $N_c \times N_c$ and may shift the k-space data in each coil by $m\Delta K$.

$$\vec{S}^{(m)} = \hat{n}^{(m)} \vec{S} \tag{1.32}$$

Different to a SMASH or VD-AUTO-SMASH, the GRAPPA algorithm results in single-coil images, which can be combined using a magnitude reconstruction procedure (i.e., the sum of squares). This provides a significantly improved SNR performance, especially at low reduction factor. Furthermore, signal losses due to phase cancellations are essentially eliminated using a magnitude reconstruction procedure. Thus previous drawbacks on k-space-based techniques, namely, phase cancellation problems, low SNR, and poor reconstruction quality due to a suboptimal fit procedure, are eliminated. Furthermore, similar to SENSE, the GRAPPA algorithm works with arbitrary coil configurations. Finally, as an additional benefit, ACS lines used to derive the reconstruction coefficients can be integrated into the final image reconstruction.

Sensitivity estimation

As mentioned above, successful applications of SENSE and GRAPPA are strongly associated with accurate characterization of the coil sensitivity maps. Since the sensitivity varies with coil loading, the sensitivities must be accessed by an additional reference scan. This can be performed, for example, by a low-resolution 3D full k-space acquisition, which allows arbitrary slice positioning and orientation. Thus, sensitivity maps can be derived by either one of these methods

- dividing each component coil image by an additional body coil image.

- dividing each component coil image by a "sum of square." image, including phase modulation.
- dividing each component coil image by one component coil image (relative sensitivity maps).
- an adaptive sensitivity assessment based on the correlation between the component coil images.

In an additional numerical process, this raw-sensitivity maps need to be refined using smoothing (i.e., minimizing the propagation of additional noise from the calibration scan into the reconstructed image) and extrapolation algorithms (i.e., to provide coils sensitivity information from regions where MR signal is hard to obtain).

1.3.2 Compressed sensing MRI

Compressed sensing is an effective tool to recover alias-free images from undersampled k-space data, exploiting the sparsity in MR images. The successful application of compressed sensing requires: (a) the desired image have a sparse representation in a known transform domain, (b) the aliasing artifacts due to k-space undersampling be incoherence in that transform domain, and (c) a nonlinear reconstruction be used to enforce both sparsity of the image representation and consistency with the acquired data.

The transform sparsity of MR images can be demonstrated by applying a sparsifying transform to a fully sampled image and reconstructing an approximation to the image from a subset of the largest transform coefficients. The sparsity of the image is the percentage of the non-zero transform coefficients. For specific applications, it is possible to get an empirical sparsity estimate by performing a clinical trial and evaluating reconstructions of many images quantitatively or qualitatively.

The reconstruction is obtained by solving the following constrained optimization problem:

$$\begin{aligned} \hat{m} = \operatorname{argmin} \; & \|\Psi m\|_1 \\ \text{s.t. } & \|\mathscr{F}_\mu - y\|_2 < \varepsilon \end{aligned} \tag{1.33}$$

where m is the reconstructed image, y is the measure k-space data from the scanner, and ε controls the fidelity of the reconstruction to the measured data. The objective function is the ℓ_1-norm to enforce sparsity.

Chapter 2

A Generalized Deep Learning-based Reconstruction Method for Parallel Imaging and Compressed Sensing

2.1 Introduction

In recent years there has been increasing interest in the deep learning based magnetic resonance imaging (MRI) reconstruction. Specifically, the fully convolutional neural network (FCN) with theoretically high representation capacity has been shown effective in resolving the detailed anatomical structures from undersampled images [33, 28, 17, 35]. The applicability of deep learning reconstruction has been successfully demonstrated in several recent studies [33, 28, 17, 35]. Meanwhile, in these studies, experimental simplifications were employed, and they were fixed encoding matrix, due to the predetermined k-space undersampling scheme, coil sensitivity, and image size, and developing/evaluating the reconstruction solely based on magnitude images. There is a need for developing a flexible and practical deep learning reconstruction method for MRI, supporting changes of encoding matrix without re-training the deep learning model and capable of processing the complex-valued MR data. Before employing the deep learning in accelerated MRI reconstruction, conventional methods such as parallel imaging and compressed sensing were widely used, and they were based on the numerical pseudo-inversion of ill-posed MRI encoding matrix, which could be prone to reconstruction error at poor conditioning [21, 2, 15, 8]. The encoding matrix stemmed from the k-space undersampling scheme and the inverse Fourier transfer coupled with the sensitivity encoding from multichannel radiofrequency (RF) coils. The traditional reconstruction involved some gradient descent methods for minimizing the cost function of the k-space fidelity and the regularization [15, 31, 27]. The performance of reconstruction depended on the conditioning of the encoding matrix. There was a tradeoff between the image artifact level and the undersampling rate as limited by the encoding capacity of coil sensitivities, i.e., the number and the decoupling of coil elements. Nevertheless, the parallel imaging technique with acceleration factors from R = 2 to R = 4 were routinely used clinically [21, 2, 8, 29, 30, 20]. Based on the compressed sensing framework, using the prior knowledge provided by regularization function [15]. The

sparsity property of MR images in a specific transform domain, such as wavelet domain, in combination with the incoherent undersampling in k-space, enabled the successful application of compressed sensing to MRI reconstruction [15]. The reconstruction performance can be significantly improved, achieving the undersampling rate from R = 2 to R = 12.5 [11]. However, in practice, a fully randomized k-space sampling was unattainable, as the undersampling was governed by the feasibility of k-space trajectory designs and frequency/phase encoding choices, resulting in insufficient incoherence and compromising the reconstruction. It should be noted that an insufficient sparsity or incoherence would introduce artifacts such as image blurring, caused by the loss of small coefficients in the sparse transform domain, i.e., a filtering effect [11].

The deep neural network could learn a prior probability distribution for MRI, i.e., so-called manifold-learning, with application in MRI reconstruction [35]. Furthermore, the generative adversarial network (GAN) has been frequently used for learning the probability distribution of MRI (i.e., manifold) and making the prediction based on the learned probability distribution [17]. Meanwhile, due to the stochastic nature of the deep neural network, there was no guarantee for the prediction to fulfill the k-space fidelity that can be described as $Ax = y$, where A the encoding matrix, x the reconstructed/predicted image and y the sampled k-space data. State-of-the-art deep learning approaches typically employed an additional k-space fidelity layer to avoid unexpected errors in reconstruction [33, 28, 17, 1, 14]. These k-space-fidelity-encouraged methods generally fell into two categories: unrolling of an iterative compressed sensing and parallel imaging algorithm, e.g., alternative direction method of multipliers (ADMM)-Net [33], and interleaving an FCN and the data fidelity layer, e.g., deep cascaded network (Deep-Cascade) [28], model-based image reconstruction using deep learned priors (MODL) [1], and sampling-augmented neural network with incoherent structure (SANTIS) [14]. In Yan Yang et al., 2016 [33], ADMM-Net was proposed for mimicking the compressed sensing MRI reconstruction, which used a convolutional layer as sparsity term and learning all the trainable parameters based on the dataflow of 3 or 4 iterations of the unrolled ADMM. In another study, the Deep-Cascade was demonstrated to perform cascaded denoising and reconstruction operations, which eliminated the noise-like artifacts, together with the k-space fidelity enforcement [28]. Recently, a MODL method was developed, and authors recast the denoising operation as a regularization term [1]. In that study, a conjugate gradient algorithm was used as the fidelity layer and embedded in the network with coil sensitivity included in the encoding matrix [1]. This conjugate gradient algorithm and the use of coil sensitivity, i.e., applying parallel imaging, made the last-mentioned method outperformed

several other deep learning approaches [1]. More recently, SANTIS used a sampling-augmented training together with cycle-GAN method to improve the rodustness against sampling pattern discrepancy [14]. Nevertheless, in all methods mentioned above, MRI encoding matrices were fully integrated into the neural network models. These models were trained with predetermined encoding matrices and corresponding undersampling artifacts. After training, imaging configurations, including the k-space undersampling schemes and coil sensitivities, associated encoding matrices, must also be unchanged or changed only within predetermined sampling patterns, during the validation and application, which could be cumbersome or to some extent impractical for the potential clinical use.

To tackle this design challenge, we proposed a hybrid approach that combined the FCN-based GAN with the non-neural-network ADMM implementation. The proposed method was based on a recent deep learning image completion study, namely One-Net [23]. Meanwhile, our modified One-Net for MRI reconstruction can handle the complex-valued data. The objective of this study was to develop a flexible and practical deep learning-based MRI reconstruction method and to implement and validate the proposed method in an experimental setting, regarding changeable k-space undersampling schemes.

2.2 Methods

We recast the One-Net into an MRI reconstruction ADMM [23]. The encoding matrix is left out of the neural network, i.e., the GAN performs the proximal operator for regularization based on the "learned" MRI prior probability distribution and as one step within the ADMM loop, as shown in Figure 2.1. While the training data for the GAN consists of images with additive spatial distributed Gaussian noise, i.e., no undersampling artifacts are needed for training the GAN. This setting allows the training and the validation of neural network to be independent of the k-space fidelity term, as the k-space fidelity enforcement demands quadratic programming using the encoding matrix. Furthermore, the designed GAN takes the complex-valued data as input and output. Such a design allows flexibility in the MRI acquisition and reconstruction settings.

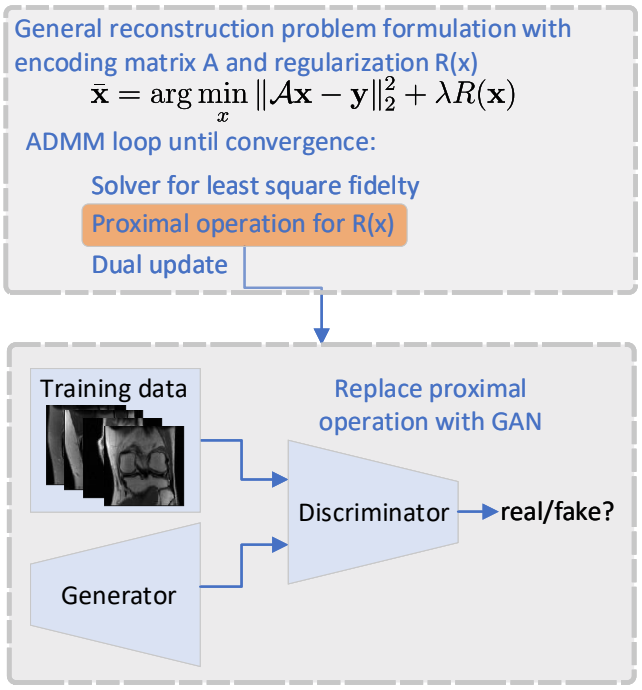

Figure 2.1 We recast the One-Net into an MRI reconstruction ADMM. The GAN performs the proximal operator for regularization based on the "learned" MRI prior probability distribution as one step within the ADMM loop.

2.2.1 Problem formulation

The image reconstruction for MR k-space data acquired with changeable k-space trajectories such as Cartesian, radial, spiral et al., and parallel imaging, can be considered as an inverse problem for the corresponding encoding matrix. We followed the MRI reconstruction problem formulation used in l_1-ESPIRiT [31]. Let an encoding matrix/operator $\mathscr{A}$ to map the MR image $\mathbf{x}$ to the sampled k-space data $\mathbf{y}$. The operator $\mathscr{A}$ consists of a Fourier transform operator $\mathscr{F}$, coil sensitivity $\mathscr{S}$, and a k-space sampling operator $\mathscr{P}$, i.e., $\mathscr{A} = \mathscr{P}\mathscr{F}\mathscr{S}$ [31]. The well-known solution to the inverse problem can be formulated as an optimization problem with a regularization term [31]:

$$\bar{\mathbf{x}} = \arg\min_x \|\mathscr{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda R(\mathbf{x}). \tag{2.1}$$

where $R(\mathbf{x})$ a regularization term that introduces an MRI prior projection for the reconstruction, and λ the regularization parameter.

The proximal operator for $R(x)$ in the proposed GAN approach, $\mathbf{prox}_R(x_0)$, is

defined as:

$$\hat{x} = \mathbf{prox}_{R,1/\lambda}(x_0) := \arg\min_x \|x - x_0\|_2^2 + \lambda R(x) \tag{2.2a}$$

$$\hat{x} \in \chi \sim P_{model} \approx P_{data} \tag{2.2b}$$

where x_0 the initial, $\hat{x}$ the estimate, χ the sub-space that contains MR images (including the ground truth), P_{data} the probability distribution of MR images, and P_{model} the probability distribution of the model prediction.

Intuitively, taking the sparse constraint in compressed sensing as an analogy, $R(x)$ for the compressed sensing prior is a ℓ_1-norm for coefficients from the wavelet transform or total variation function which can be solved by a soft-thresholding as the proximal operator [31]. For GAN prior definition, the $R(x)$ is square of the Euclidean distance between the estimate and the ground truth, i.e., $\| x - x_m \|_2^2$, $x, x_m \in \chi$, where x_m is the ground truth, and proximal operator for $R(x)$ projects the estimate onto the sub-space χ that contains MR images, approaching x_m. In addition, the $P_{model} \approx P_{data}$ is encouraged by the "min-max"-type loss function in GAN's training [7].

2.2.2 ADMM implementation for separating fidelity and MRI prior projection

The objective function in Eq. (2.1) contains a fidelity term and a regularization term. The regularization term facilitates the MRI prior projection. In an ADMM form, fidelity and regularization terms can be separated into two sub-problems: one quadratic programming concerning k-space fidelity and another proximal operator for the regularization exploiting the MRI prior probability distribution. We can rewrite the Eq. (2.1) as an ADMM problem [32]:

$$\begin{aligned} &\arg\min_{x,z} \frac{1}{2}\|Az - y\|_2^2 + \lambda R(x) \\ &s.t. \qquad x = z \end{aligned} \tag{2.3}$$

where λ is the regularization parameter. The augmented Lagrangian of ADMM problem can then be expressed as [32]:

$$L_\rho(x, z, u) = 1/2 \| Az - y \|_2^2 + \lambda R(x) + \rho/2 \| x - z + u \|_2^2 \tag{2.4}$$

where u the dual variable and ρ the penalty parameter for the Lagrangian multiplier. ADMM has the following iterative steps:

$$\begin{aligned} &S1: x^{(k+1)} \leftarrow \arg\min_{x} \lambda R(x) + \rho/2 \|x - z^{(k)} + u^{(k)}\|_2^2 \\ &S2: z^{(k+1)} \leftarrow \arg\min_{z} 1/2 \|Az - y\|_2^2 + \rho/2 \|x^{(k)} - z + u^{(k)}\|_2^2 \\ &S3: u^{(k+1)} \leftarrow u^{(k)} + x^{(k+1)} - z^{(k+1)} \end{aligned} \tag{2.5}$$

The $x^{(k+1)}$ from step S1 in Eq. (2.5) can be computed by using the generator of GAN since S1 is designed to be the loss function for training the generator. It can be also viewed as a proximal operator $\mathbf{prox}_{R,\frac{\rho}{\lambda}}(z^k - u^k)$ for the regularization function $R(x)$, which is given by substituting $z - u$ for x_0 in Eq. (2.2a). The step S2 is well-posed and can be solved by the non-linear conjugate algorithm for each $z^{(k+1)}$ update. It should be noted that the ADMM splits the MRI prior projection and the k-space fidelity term, which enabled the prior projection (facilitated by the GAN), learning a genetic MRI prior that is applicable for any encoding matrices $\mathscr{A}$ in Eq. (2.5) S2.

2.2.3 GAN implementation for MRI prior projection

The consideration for the MRI prior projection introduced by the GAN has three folds: (1) encouraging the probability distribution of data, P_{data}, to be equal to that of the generator prediction, P_{model} [7]; (2) using a proximal function loss as the generator loss for $R(x)$, i.e., letting the generator be the proximal operator for $R(x)$; and (3) using an additive spatial distributed Gaussian noise (as a random perturbation) with MRI data to train the generator, which determines the functionality of $R(x)$ and is also the key for the convergence of the ADMM iteration.

Intuitively, the proximal operator for $R(x)$ is realized by the generator of GAN, which projects the generator input, i.e., $x_0 = z^k - u^k$ in Eq. (2.2a) and Eq. (2.5) S1, onto a set of artifact-free MRI datasets, as given by the MRI prior, P_{data} in Eq. (2.2a). We also empirically find that the additive spatial distributed Gaussian noise for training is the key ingredient that leads to the convergence of ADMM. This finding could be explained by the following facts or observations when the ADMM converges: (a) ADMM must converge at when $z^k \approx x^k$; (b) when $z^k \approx x^k$, i.e., the $z^k - x^k$ map contains minimal structural information except for spatially distributed noise-like discrepancies; (c) in Eq. (2.5) S3, which is the dual ascent for ADMM, the u^k is the accumulative sum of the $z^k - x^k$ map, with the tendency of amplifying

discrepancies; and (d) discrepancies in u^k transfers back to Eq. (2.5) S1, i.e., it inputs u^k into the generator. Therefore, GAN learns fully sampled MR image datasets and estimates the underlying probability distribution of the data. Meanwhile, GAN eliminates discrepancies from image features, which is iteratively performed each time after the ADMM dual ascent step in Eq. (2.5) S3. In this way, we created a "trap" around the ground truth, x_m, implemented by training generator of GAN with the additive noise or randomized perturbation on data, $x_m + \varepsilon$, where ε denotes the noise/perturbation. During the ADMM iteration, any vector x_0, when close enough to the x_m, i.e., when $x_0 = x_m + \varepsilon$, will be pushed closer to x_m by the generator of GAN, which performs the prior projection function, leading to the convergence of ADMM, as illustrated in Figure 2.2.

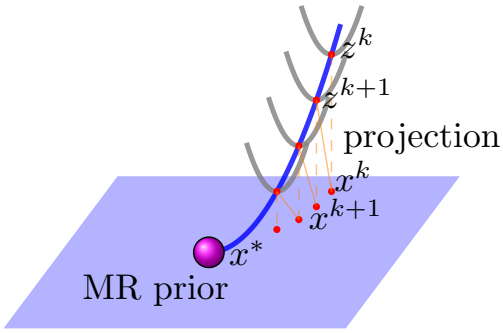

Figure 2.2 "MR prior" is a set or a manifold that contains MR images. The ADMM splits the MRI prior projection and the k-space fidelity enforcement into two iterative steps. The prior projection is performed by the generator of GAN. GAN learns a genetic MRI prior, so that the generator of GAN can project the z_k onto the closest MR image, x_k. Another step in ADMM is encouraging the k-space fidelity by projecting x_k onto z_{k+1}, on a convex feasible set that is determined by the fidelity term as a series of quadratic functions (gray curves), which can be implemented using convex optimization algrithm outside the neural network. By doing such alternative projections using ADMM, one can reach the intersection of two sets, x^*, which is the final solution for the ADMM reconstruction.

As shown in Figure 2.3, the GAN in this study is composed of a U-Net generator [24] and a Res-Net discriminator [9]. The U-Net has an encoder-decoder structure, which can preserve fine details of images [24]. The U-Net used in this study has 6 CNN blocks on both encoder and decoder sides. The CNN block consists of repeated application of two 3×3 convolutions (unpadded convolutions), each layer with a rectified linear unit (ReLU) activation function. On the encoder side, the CNN block is followed by a 2×2 average pooling operation implemented by a stride size of

2×2 in a convolution layer for downsampling. The CNN block of decoder side is applied after the up-convolution and a concatenation layer, which correspondingly merges the cropped feature map with the map that comes from the up-convolution path. The average-pooling is used instead of the commonly used max-pooling, as suggested by [25]. The generator input is normalized to the intensity range of $[-1,1]$ for both real and imaginary channels. The virtual batch normalization is also applied after each convolution layer to improve the convergence of the network training[25]. No activation function is used for the final layer of the generator to ease the design of generator loss in GAN. Also, as suggested by [19], a channel-wise fully connected layer accompanied by a convolution kernel is necessary for the generator to extract the image features at the first layer.

For training the U-Net generator, fully sampled image and image with the additive spatial distributed noise are both used as inputs and are projected onto the closest noise-free image, i.e., both x_m and $x_m + \varepsilon$ map to x_m. To deal with the complex-valued data, we separate the complex data into real and imaginary channels for both the generator$'$s input and output, as well as for discriminator$'$s two inputs. Also, the additive spatial distributed Gaussian noises are combined with both real and imaginary channels separately. The additive noise is generated independently for each training batch. The additive spatial distributed noise is computed as $\varepsilon = zp$, where p is a spatial distribution mask interpolated with bi-cubic method from a low-resolution noise whose maximum is 1.414, and z is a pixel-wise Gaussian noise with mean = 0 and standard deviation = 0.1.

The training of GAN encourages the output of the generator network and the ground truth to share the same probability distribution, i.e., $P_{model} \approx P_{data}$ in Eq. (2.2b). Res-Net is used as the discriminator because it is effective at the classification task, eliminating the "gradient vanishing" during the training [9]. In the Res-Net, the first layer is a convolutional layer followed by 40 bottleneck blocks, and the last layer is a fully connected layer. The details of the architecture can be found in [10].

We rewrite the proximal operation $x^{(k+1)} = \mathbf{prox}_R\left(z^k - u^k\right)$ using the project symbol: $x^{(k+1)} \leftarrow \mathscr{P}\left(z^k - u^k\right)$. Then we define the following combined loss function for training the prior projection network, i.e., the generator loss:

$$\min_{\theta} \lambda_1 \parallel v - \mathscr{P}(x) \parallel_2^2 + \lambda_2 \parallel x - \mathscr{P}(x) \parallel_2^2 + \lambda_3 \parallel v - \mathscr{P}(v) \parallel_2^2 + \lambda_4 \log((D(P(v)) \tag{2.6}$$

where θ trainable parameters in generator $\mathscr{P}(x)$, x the original image, $v := x + \varepsilon$ the image with spatial distributed Gaussian noise, $D(\cdot)$ the discriminator, and λ_1 to λ_4 regularization parameters for four loss terms. The $\parallel v - \mathscr{P}(x) \parallel_2^2$ term encourages

the generator to eliminate the noise, i.e., encourage the projection $x \leftarrow \mathscr{P}(x+\varepsilon)$, which can be considered as the regularization function $R(x)$ in Eq. (2.2a). The $\| x - \mathscr{P}(x) \|_2^2$ and $\| v - \mathscr{P}(v) \|_2^2$ stem from the proximal function for $R(x)$, corresponding to the first term in Eq. (2.2a), which is a necessary component to maintain stability of the ADMM algorithm in Eq. (2.5). Intuitively, they promote projects: $x \leftarrow \mathscr{P}(x)$ and $v \leftarrow \mathscr{P}(v)$, i.e., outputs are "close to" inputs for both x and v. The $\log(D(\mathscr{P}(v)))$ term partly contributes to the $P_{model} \approx P_{data}$ in Eq. (2.2b) when training the generator of GAN. Briefly, Eq. (2.6) as the generator loss for GAN is designed to perform the proximal operator for $R(x)$, as described in Eqs. 2a and 2b. Finally, a binary cross entropy loss is used in training the discriminator.

To summarize, the regularization has the proximal function defined as the generator loss of GAN, and the training processing of GAN also encourages learning the prior, $P_{data} \approx P_{model}$, i.e., the prediction from the generator should be within the set that follows the MRI prior probability distribution. In addition, the implemented GAN takes the complex-valued data, as input and output by separating the real and imaginary parts into two channels for the neural network.

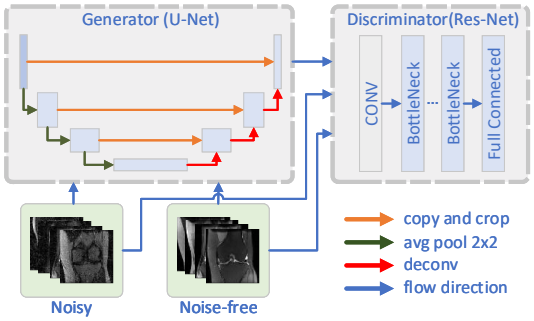

Figure 2.3 The GAN is composed of a U-Net generator and a Res-Net discriminator. The generator of GAN performs the prior projection in the ADMM. For training the U-Net generator, noise-free image and image with the additive spatial distributed Gaussian noise are both used as inputs and are projected onto the closest noise-free image.

2.2.4 MRI databases, pre-processing, and analysis

Data used in this study were obtained from the NYU fastMRI Initiative database (fastmri .med.nyu.edu) [34]. As such, NYU fastMRI investigators provided data but did not participate in analysis or writing of this report. A listing of NYU fastMRI investigators, subject to updates, can be found at:fastmri.med.nyu.edu.The primary goal of fastMRI is to test whether machine learning can aid in the reconstruction of

medical images. We downloaded multi-channel k-space data for all 973 scans from fastMRI reconstruction database [34]. The knee data have two contrast weightings, i.e., proton-density with and without fat suppression. Scan parameters include 15 channel knee coil and 2D multi-slice turbo spin echo (TSE) acquisition, two field strengthes (i.e., 1.5 T and 3.0 T), and other settings can be found in [34]. For the Cartesian sampling, we resized the kneel MR images into 128×128 in k-space. We used a relatively small image size for the proof of concept experiment because of the lack of computation power for handling the original full-size data. We used 6745 2D images in the training set and 1000 2D images in the testing set, respectively. Real and imaginary parts of all 2D images were separated into two channels when inputted into the neural network. We trained our network for 100 epochs on an NVIDIA P6000 graphic card with batch size = 50 and Adam optimizer, using tensorflow software (https://www.tensorflow.org/). It took 8 hours to train the neural network. The non-neural-network part of ADMM was implemented in python language using the MRIPY toolbox (https://github.com/peng-cao/mripy). The fidelty term in Eq. (2.5) S1 was computed using non-linear conjugate gradient function like how it was used in the ADMM algorithm in MRIPY. The number of ADMM iteration was set to 20 for all experiments using the proposed method.

For comparison, we also performed l_1-ESPIRIT [31], GRAPPA [8], and MODL, a deep learning reconstruction method [1].In the l_1-ESPIRiT reconstruction, we set the regularization parameter to be 0.01, using the default algorithm, i.e., the fast iterative thresholding method. Like our approach, MODL supported the complex-valued MRI data and utilized the parallel imaging in the reconstruction [1]. Therefore, MODL could serve as a representative deep learning model for comparison. We followed settings in Ref [1] when training MODL to reconstruct the undersampled kneel data. The only difference was the undersampling mask in Ref [1] was 2D, while we only used 1D undersampling in this study since the MR measurement was a 2D multi-slice sequence, i.e., the frequency encoding was fully sampling. During the experiment, the coil sensitivity map was used in l_1-ESPIRIT, MODL, and the proposed method. We applied ESPIRiT function in Berkeley MRI reconstruction toolbox to estimate the sensitivity map from a 20×20 calibration region in the central k-space. For parallel imaging, the GRAPPA reconstruction was performed, and 20 center lines were used as the auto-calibration region [8]. For all three methods, we calculated difference maps, which were discrepancies between the reconstructed image and the ground truth, as well as the peak signal to noise ratio (PSNR), which was given in dB unit as the ratio of max image intensity to the root mean square of the difference map.

2.3 Results

The proposed method was validated in two experiments with changeable k-space undersampling schemes, varied contrasts, due to different field strengths and sequences, and a large amount of clinical images (N=1000 for 2D slices) those with potentially different coil sensitivities due to variations in the subject positioning. We also demonstrated the convergence of the proposed hybrid GAN and non-neural-network ADMM to show their compatibility.

2.3.1 Persudo-random undersampling with compressed sensing and deep learning reconstructions

Figure 2.4 shows the comparison of knee MRI compressed sensing and deep learning reconstructions using 15% k-space frequency encoding lines. ℓ_1-ESPIRiT, a compressed sensing method, and MODL, a deep learning approach, showed apparent blurring effects on images. Residual aliasing artifacts were also observed in images from MODL. Meanwhile, images reconstructed by l_1-ESPIRiT were in qualitative agreement with the ground truth in Figure 2.4. However, l_1-ESPIRiT reconstruction confused the ligament with the bulge of the tibia. While the proposed reconstruction had successfully recovered the anterior cruciate ligament. The proposed reconstruction also preserved more tissue boundary information, compared with those of l_1-ESPIRiT and MODL reconstructions, as shown on difference maps in Figure 2.4. The proposed method generally resulted in the highest PSNR among the three methods, as summarized in Table 2.1. Also, for the proposed method, two sets of images in Figure 2.4, with different undersampling schemes, as shown in Supporting Figure S1 a and b, and two contrast weightings, were reconstructed by one trained neural network model. This result demonstrated that the proposed method could restore realistically undersampled MRI with both high acceleration factor (i.e., 15% samples) and high PSNR (i.e., $>$34 dB).

Table 2.1 Comparison of PSNRs (in dB, mean $\pm$ standard deviation, N = 1000) for three compressed sensing and deep learning reconstructions.

Undersampling rate	L1-ESPIRiT	MODL	Ours
15%+10%	32.57$\pm$1.78	29.49$\pm$3.08	34.81$\pm$1.79
20%+10%	33.74$\pm$1.73	30.13$\pm$3.04	35.60$\pm$1.87

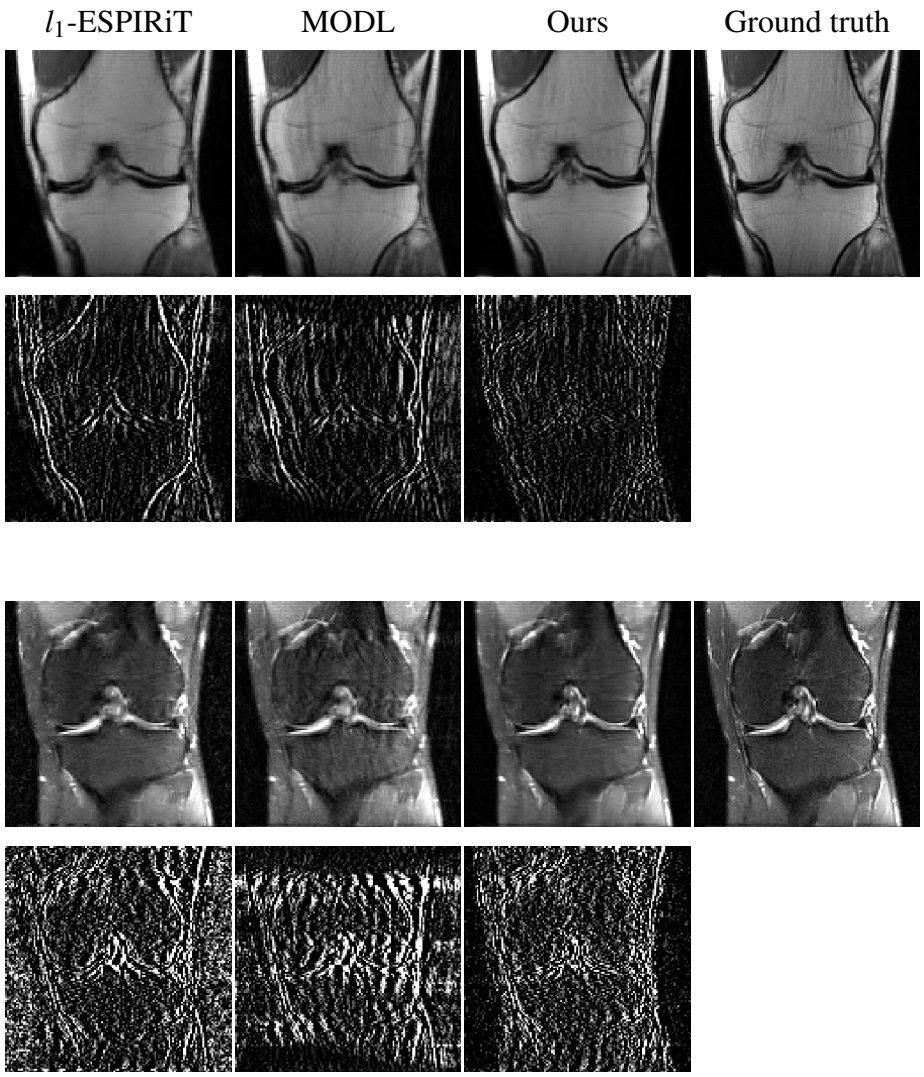

Figure 2.4 Comparison of knee MRI compressed sensing and deep learning reconstructions using 15% k-space frequency encoding lines and 1D undersampling along k_x. (first row) Proton density weighted TSE knee images reconstructed from l_1-ESPIRiT, MODL, and proposed method, and full k-space ground truth. (second row) Magnified difference maps ($\times 10$) for three reconstruction methods in the first row. The quantified PSNRs were 33.72 dB, 34.68 dB, and 37.37 dB for l_1-ESPIRiT, MODL, and proposed method, respectively. (third and fourth rows) The same comparison for proton density weighted TSE with fat suppression on a different volunteer. The quantified PSNRs were 32.75 dB, 25.72 dB, and 34.16 dB for l_1-ESPIRiT, MODL, and proposed method, respectively. l_1-ESPIRiT, a compressed sensing method, and MODL, a deep learning approach, showed apparent blurring effect on images such as in challenging cruciate ligament region, and increased error on edges of anatomical structures. Residual aliasing artifacts were also observed in reconstruction results from MODL. The proposed method generally resulted in the highest PSNR among three methods, as summarized in Table 1.

2.3.2 Parallel-imaging-type undersampling with GRAPPA and deep learning reconstructions

Figure 2.5 shows the comparison of knee MRI parallel imaging and deep learning reconstructions using simulated 3-fold acceleration, i.e., R = 3, as the k-space mask shown in Supporting Figure S1c. The proposed method had a better suppression of the aliasing artifacts, compared with that of GRAPPA reconstruction. The noise amplification was observed in GRAPPA reconstructed images, which may lead to a slight degradation of image quality around the condyle structure. Meanwhile, the proposed method generally provided higher PSNR compared with that of GRAPPA at high accelerations, as summarized in Table 2.2. With R=2, the GRAPPA and the proposed methods had similar PSNRs, and both achieved the PSNR $>$40 dB, i.e., $< 1\%$ root mean square artifact level, which means both can achieve near identical reconstruction. With the increase of acceleration factor, GRAPPA$'$s PSNR decreased drastically, i.e., from 41 dB to 29 dB; on the other hand, the PSNR of proposed method only reduced slightly, i.e., from 40 dB to 35 dB. The proposed method showed a significant improvement over GRAPPA reconstruction at high accelerations. Furthermore, in this experiment, we applied the same neural network model that was used in the previous compressed sensing experiment, but with a different k-space undersampling scheme, as shown in Supporting Figure S1c. This result demonstrated the proposed method generally applicable for different experimental settings, i.e., compressed sensing or parallel imaging or both, without the need of re-training the neural network model.

Table 2.2 Comparison of PSNRs (in dB, mean $\pm$ standard deviation, N = 1000) for parallel imaging and deep learning reconstructions.

Undersampling factor	GRAPPA	Ours
R=2	41.36$\pm$3.80	39.51$\pm$3.64
R=3	35.44$\pm$3.18	37.53$\pm$4.06
R=4	29.49$\pm$2.57	35.29$\pm$3.69

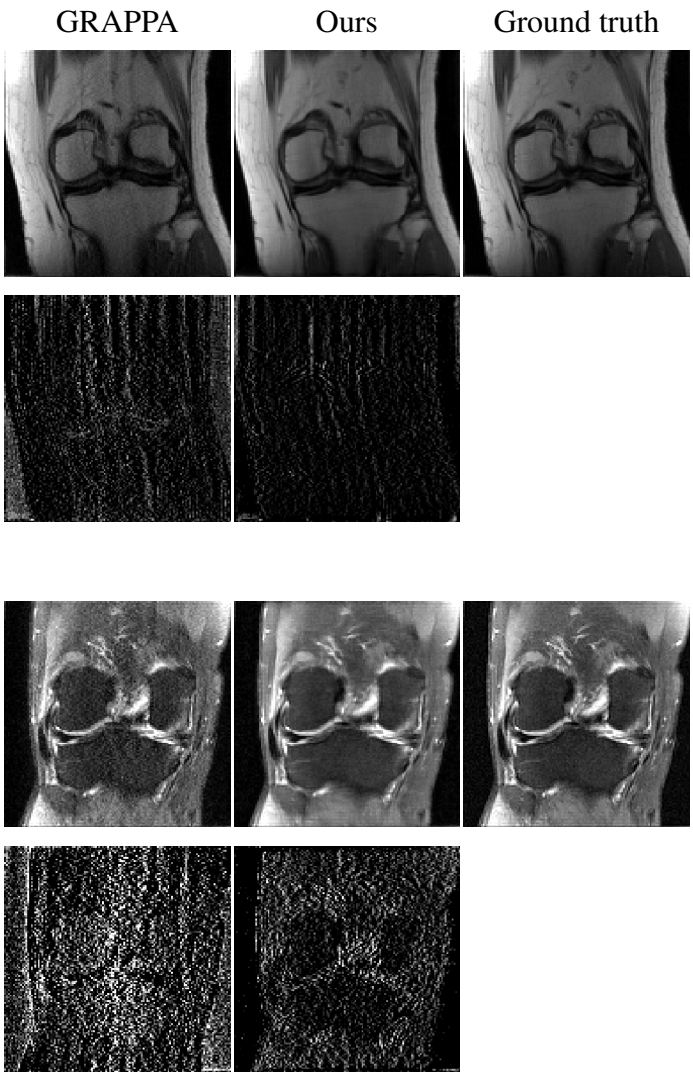

Figure 2.5 Comparison of knee MRI parallel imaging and deep learning reconstructions using simulated 3-fold acceleration, i.e., R = 3, and 1D undersampling along k_x. (first row) Proton density weighted TSE knee images reconstructed from GRAPPA and proposed method, as well as full k-space ground truth. (second row) Magnified difference maps ($\times 10$) for two reconstruction methods in the first row. The quantified PSNRs were 38.42 dB and 42.55 dB for GRAPPA and proposed method, respectively. (third and fourth rows) The same comparison for proton density weighted TSE with fat suppression on a different volunteer. The quantified PSNRs were 33.83 dB and 37.19 dB for GRAPPA and proposed method, respectively. GRAPPA caused noise amplification in the low sensitivity region of RF coils, i.e., the center of FOV. Meanwhile, the proposed method generally provided higher PSNR compared with that of GRAPPA at high accelerations, as summarized in Table 2.

2.3.3 Convergence of proposed hybrid of GAN and non-neural-network ADMM

In the proposed reconstruction scheme, the non-neural-network ADMM was used as the outer iteration in combination with the GAN implementation of MRI prior projection. Special consideration was made to ensure the compatibility and the convergence of the iteration. Figure 2.6 shows the $\|x-z\|_2^2$ and $\|y-Ax\|_2^2$ (defined in Eq. (2.5)) decreased with the number of iteration for a representative case. The $\|y-Ax\|_2^2$ measured the residual error of the fidelity term for reconstructing MR image, x, regarding the measured k-space, y. The monotonous decay of $\|y-Ax\|_2^2$ indicated the fidelity term in ADMM, i.e., the non-neural-network part of ADMM in Eq. (2.5) S2, was compatible with the neural network implementation of the proximal operation, i.e., in Eq. (2.5) S1. The $x-z$ was the gradient of the augmented Lagrangian equation for ADMM. The $\|x-z\|_2^2$ decayed nearly to 0, confirming the convergence of ADMM.

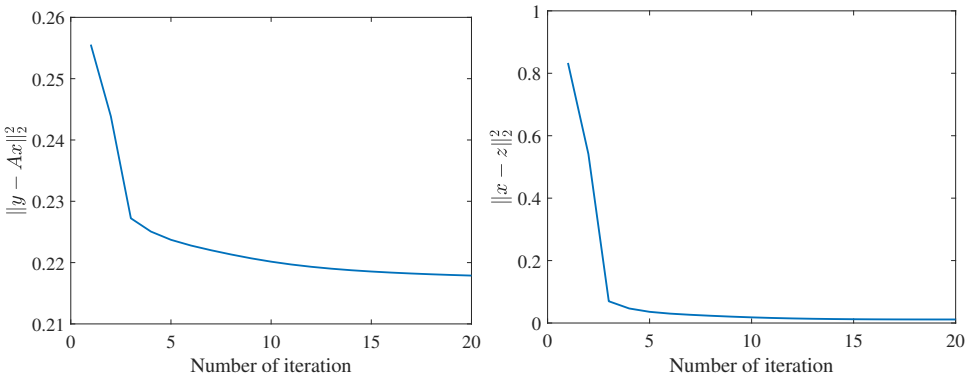

Figure 2.6 Convergence of the proposed method for a representative case. Both $\|x-z\|_2^2$ and $\|y-Ax\|_2^2$ (defined in Eq. (2.5)) decreased with the number of iteration. The $\|y-Ax\|_2^2$ measured the residual error of the fidelity term in ADMM. The monotonous decay of $\|y-Ax\|_2^2$ indicated the fidelity term was compatible with the neural network implementation of the proximal operation (as in Eq. (2.5) S1). The $x-z$ was the gradient of the augmented Lagrangian equation for ADMM. The $\|x-z\|_2^2$ decayed to 0 at the end of iteration, confirming the convergence of ADMM. The scale of y-axis was normalized.

2.4 Discussion

Previous deep learning reconstruction studies used to train the neural network with fixed encoding matrices [33, 28, 17, 35]. In the conventional paradigm, such

experimental simplification was unavoidable because most available deep neural networks were designed using the feedforward architecture, i.e., they are not recurrent nor interactive; however, MRI reconstruction required the iteration of regularization and k-space fidelity enforcement. Meanwhile, for a decade, the compressed sensing and parallel imaging researches have established a set of algorithms basing on the convex optimization for various MR acquisition settings and applications [15, 31, 27, 11]. These MR acquisition settings could be varied from Cartesian to non-Cartesian, from 2D to 3D, from static to dynamic acquisition et al., arguing the need of building a deep learning function that can be inserted in the existing MRI reconstruction algorithms, as an adaptive regularization function or an image prior. To our knowledge this is the first study that demonstrated the feasibility of building a compatible GAN-regularization for a non-neural-network ADMM, a widely used algorithm for MRI reconstruction. In this study, one trained neural network model can be applied to different reconstruction tasks, taking advantage of the separable equations of the MRI prior projection and the fidelity enforcement. Our result demonstrated the initial capacity of such hybrid GAN and non-neural-network ADMM approach for improving the reconstruction performance of both parallel imaging and compressed sensing, evaluated using realistic undersampling schemes and sufficient sample size, i.e., N=1000, two contrast weightings, and two field strengths.

Our method can be applied to MRI reconstruction with flexibility to change the number of RF coils or choosing different RF coils, vary the subject positioning that can affect the coil sensitivity, and scan with various undersampling schemes or types of readout trajectories, in contrast to fixed acquisition settings in existing deep learning methods [33, 28, 17, 35]. For example, SANTIS [14] used an embedded fidelity term in the neural network, which was trained with predetermined but augmented undersampling schemes. MODL [1] technically supported changeable sensitivity maps, but it used a conjugated gradient method inside the neural network. These previous deep learning methods all required MRI acquisition setting unchangeable, such as the number of coils and the type of MRI trajectory or the distribution of undersampling, due to encoding matrces were embedded in the neural network and undersampling artifacts were used for training the network. On the other hand, the proposed method as a regularization can be inserted into the existing non-neural-network ADMM algorithm for compressed sensing and parallel imaging reconstructions. Therefore, the proposed method can be directly applied to a broad range of accelerated MRI applications. Also, the proposed method is readily applicable to the non-Cartesian sampled data, for which MRIPY toolbox (https://github.com/peng-cao/mripy) with functions, such as ADMM and nonuniform fast Fourier transform, can be utilized.

In this initial experiment, we focused on demonstrating the utility of the proposed method in classic compressed sensing and parallel imaging reconstructions, and we used the knee MRI to verify the theory. However, the proposed method should be generally applicable for most MRI applications.

Our method showed a slight smoothing effect on the image texture at high accelerations; although, with 15% of k-space data, the reconstructed image had PSNR >34dB with the root mean square artifact level below 2.0%. This smoothing effect could be partly caused by the use of l_2 loss when training the generator of GAN and partly from the high acceleration. Recent deep learning study showed the benefit of using a perceptual loss for better resolving the texture of natural images [12]. The advantage of using perceptual loss could be investigated in the future. Also, ADMM had two tunable parameters in Eq. (2.5): λ, an adjustable parameter for regularization, and ρ, a setting empirically chosen for the robustness of ADMM when applied to non-convex regularization function. In this study, we found ρ should be adjusted carefully to reduce the smoothing effect while maintaining the fidelity and the iteration convergence. This additional tuning step for ρ was likely related to the non-convex nature of the neural-network-based regularization. As tested in the experiment, additive spatial-distributed Gaussian noise for training the GAN was an essential ingredient for the convergence of ADMM, reflecting the complexity of adapting the neural network to the conventional ADMM method. Further studies on preserving the image texture and straighten the implementation of optimization to be compatible with neural network regularization are currently undergoing in our lab. Finally, the proposed deep neural network could be modified as a patch based U-Net. By doing patch based rather than the current whole image approach, one could overcome the current GPU memory limitation and enable variable image size in the reconstruction. This will also be investigated in the future experiment.

2.5 Conclusion

In this study, we demonstrated a generalized deep-learning-based MRI reconstruction method that utilized the MRI prior probability distribution to improve the performance of compressed sensing and parallel imaging reconstructions. Like conventional iterative reconstruction algorithms, k-space fidelity in proposed hybrid approach was enforced by the least-square term and implemented outside the neural network, which allowed the high flexibility in MRI acquisition configuration, including k-space undersampling schemes and RF coil settings. This is the first study that confirmed the feasibility and applicability of the deep neural network as a

regularization function in a non-neural-network ADMM reconstruction algorithm. For quantitative comparison, our methods achieved highest PSNR, as well as 30% and 49% artifact reductions for compressed sensing and parallel imaging reconstructions, compared with l_1-ESPIRiT, GRAPPA, and MODL, showing great potential in clinical applications.

Chapter 3

Reconstruction Using Deep Bayesian Estimation

3.1 Introduction

Before the introduction of deep learning into MRI reconstruction, the convex regularization-based sparse or low-rank algorithm was commonly used under compressed sensing framework [15]. The convex regularization ensured the convergence of the iterative algorithm to achieve a globally optimal solution, i.e., an artifact-reduced MR image [15]. The convex regularization can impose a prior knowledge such as the sparsity in total variation or wavelet domain for MR image. The ℓ_1 regularization exploited the sparsity of MR images in the transform domain and served as an analytical prior in compressed sensing reconstruction [15]. The analytical prior due to the deduction of a convex regularization ensured the convergence of the algorithm and the generalization. Besides, the dictionary learning or the low-rank regularization was an extension of analytical prior, providing an improvement over the ℓ_1 regularization in certain compressed sensing application [22]. In the design of such analytical prior, the model selection mainly focused on the compression and representation capacity of the regularization and the optimization of parameters for eliminating the reconstruction artifacts [13]. Furthermore, deep learning studies, using models stemmed from the analytical prior, focused on unrolling iterative optimization algorithm, which was designed to solve the convex regularized reconstruction problem, and replacing the regularization function with the deep learning model [33, 28, 17, 1]. However, the convexity of the regularization in analytical prior may not be easily replicated by the deep learning model, posing a challenge in the generalization.

With the advent of the deep neural network, research started shifting the paradigm to structured feature representation of MRI, such as cascade, deep residual, and generative deep neural network [33, 28, 17, 1]. Especially method proposed in [33] recast the compressed sensing reconstruction into a specially designed generative deep neural network that still partly utilized the analytical data fidelity and regularization terms. However, the compromise did exist, such as the lack of flexibility, when adapting to a generative MRI reconstruction. Specifically, the lack of explicit statistical

assumption and induction (e.g., they mixed the image prior and the MRI encoding) caused those deep learning methods to performed poorly when generalized to a changeable MRI acquisition setting, e.g., under-sampling scheme, radio-frequency coil, and matrix size or spatial resolution; though they may show improved performance in some predetermined acquisition settings or pre-trained imaging tasks. Therefore, in previous studies, the objective was training a nonspecific deep neural network with data inconsistency loss in mimicking analytical optimization.

To tackle this design challenge in generalization, in this study, we applied a statistical representation of an existent MRI database, i.e., an image prior, to Bayesian inference. Compared with the analytical optimization approach, Bayesian inference was naturally compatible with deep learning-based prior in a statistical manner. This approach was based on a recent study on the deep learning-based image completion that demonstrated a pixel convolution neural network (pixelCNN++) applied to the Bayesian inference [26]. In [26], posterior distribution $p(x|y)$ was the probability of Bayesian model parameters in image prior x given the observation y. Following the well-known maximum a posteriori estimation (MAP), Bayesian inference estimated parameters of the image prior x through maximizing the likelihood of the prior. To make this inference tractable, in [26], authors proposed to use a generative deep neural network for modeling image prior, i.e., $\texttt{NET}(x) = p(x)$, through a pixel-wise predictive model, i.e., the pixel CNN. In this study, concerning MRI reconstruction, MR images were from the image prior and model parameters, x; while the measured k-space data were the observation, y. The MRI encoding equation can determine the k-space data likelihood, $p(y|x)$, which was a constant for a certain noise level in the measurement and was ignored in MAP. Because the relation between acquired k-space data and the image was deterministic, i.e., MRI acquisition setting could not affect the $p(y|x)$; thereby, the Bayesian inference in such MRI reconstruction would be solely based on the generic image prior. To summarize, the Bayesian inference could utilize the image prior in MRI reconstruction through a statistical model on pixel-wise dependencies of MR images, which was generic because of its decoupling with the changeable MRI acquisition.

This paper presented a generic and interpretable deep learning-based reconstruction framework, employing the MRI prior and Bayesian inference. The proposed framework was capable of exploiting the MRI database with a generative prior model, regardless of the changeable MRI acquisition settings. Also, the reconstruction was achieved by a series of inferences those employed the maximum likelihood of posterior with the image prior, i.e., iteratively applying the Bayesian inference. The reconstruction iterated over the data fidelity enforcement in k-space and the image

refinement with the Bayesian inference. During the iteration, algorithms used were gradient ascent and projected sub-gradient methods performed outside the neural network, which were compatible with changeable MRI acquisition settings. The method is theoretically described based on the methodology proposed by others [26] and then demonstrated in different MRI acquisition scenarios. The robustness and the reproducibility of the algorithm were also validated in experiments.

3.2 Theory

In this theory section, we discuss about how to utilize MR image dataset in a tractable way with the generative network.

3.2.1 Statistical approach to modelling MR image reconstruction

Let the vector space X represent the MR image space and the vector Y represent the MR raw k-space space. The model of image reconstruction is formulated with Bayes' theorem as following:

$$f(x|y) = \frac{f(y \mid x)g(x)}{f(y)}, x \in X, y \in Y \tag{3.1}$$

where $f(y \mid x)$ is given by the data model that models how the measured k-space data y is generated from a particular image x and $g(x)$ is the prior model of the image. It should be noted that the image x is the Bayes model's parameter conditional on the k-space data y. The prior model $g(x)$ is learned from a huge dataset of MR images which are samples from the vector space X. The data model is given by imaging principles in terms of signal generation, k-space encoding, and sampling. The image reconstruction is achieved by exploring the posterior $f(x \mid y)$ by a proper estimator. Here, we use maximum a posterior (MAP) estimator which maximizes the posterior probability. That is to say the estimator gives the 'most likely' image $\hat{x}$ given the measured k-space data y. The reconstruction image $\hat{x}$ is given by:

$$\hat{x}_{\mathrm{MAP}}(y) = \arg\max_{x \in X} f(x \mid y) = \arg\max_{x \in X} f(y \mid x)\, g(x). \tag{3.2}$$

In the following section, the deep learning based approach to designing the computationally tractable prior model $g(x)$, the data model of MR data acquisition $f(y \mid x)$ and the computationally feasible means for maximizing the posterior $f(x \mid y)$ are specified.

3.2.2 Scalable prior model of MR images

We utilize the deep autoregressive network as our prior model, which gives the predicted mixture distribution of the input MR image. In our model, we assume that, with respect to each voxel, there is a latent intensity ν with a continuous distribution which gives representation to real channel's signal intensity or imaginary channel's. Like in the VAE and pixel CNN++ [13, 26], the distribution of ν is a mixture of the logistic distribution, given by

$$\nu \sim \sum_{i=1}^{K} \pi_i \text{logistic}(\mu_i, s_i). \tag{3.3}$$

Here, π_i is the mixture indicator, μ_i and s_i are the mean and scale of logistic distribution separately. We could calculate the probability on each observed sub-voxel ν with following equation:

$$P(\nu;\pi,\mu,s) = \sum_{i=1}^{K} \pi_i [\sigma(\nu+0.5-\mu_i)/s_i - \sigma(\nu-0.5-\mu_i)/s_i] \tag{3.4}$$

modification needed

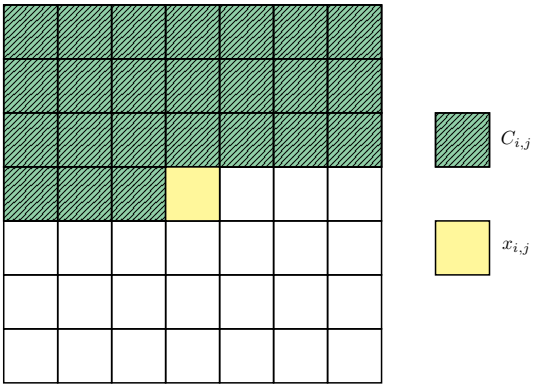

Figure 3.1 The conditional model. $x_{i,j}$ is conditional on all the voxels on its up and left.

Furthermore, following Ref [26, 18], the voxel is not only conditioned on whole voxels up and to the left in an image as showed in Fig '3.1, but also modelled with joint predictive distributions over real and imaginary channel. The very first voxel's real channel is predicted with a mixture of logistics as described in Eq (3.3). We assume that the means of mixture components of imaginary channel is linearly dependent on real channel. The conditional distribution of the subsequent voxel

$(\mathrm{Re}(x_{i,j}), \mathrm{Im}(x_{i,j}))$ at position $\vec{r} = (i, j)$ is given by

$$p(x_{i,j}|C_{i,j}) = p(\mathrm{Re}(x_{i,j}), \mathrm{Im}(x_{i,j}) \mid C_{i,j}) = P(\mathrm{Re}(x_{i,j}) \mid \mu_{\mathrm{Re}}(C_{i,j}), s_{\mathrm{Re}}(C_{ij})) \times \quad (3.5)$$
$$P(\mathrm{Im}(x_{i,j}) \mid \mu_{\mathrm{Im}}(C_{i,j}, \mathrm{Re}(x_{i,j})), s_{\mathrm{Im}}(C_{i,j}))$$
$$\mu_{\mathrm{Im}}(C_{i,j}, \mathrm{Re}(x_{i,j})) = \mu_{\mathrm{Im}}(C_{i,j}) + \alpha(C_{i,j})\,\mathrm{Re}(x_{i,j}) \ , \quad (3.6)$$

where the $C_{i,j}$ denotes the context information which is comprised of the mixture indicator and the previous voxels, α is the coefficient which is related to mixture indicator and previous voxels. We assume that mixture indicator is shared across all 2 channels. The joint distribution of whole pixels on an image vector $x = (x^{(1)}, ..., x^{(n^2)})$ can be expressed as following:

$$p(x; \pi, \mu, s) = p(x^{(1)}) \prod_{i=2}^{n^2} p(x^{(i)} \mid x^{(1)}, .., x^{(i-1)}) \quad (3.7)$$

The network $\mathscr{N}(x, \Theta)$ is trained by the maximizing of the likelihood, given by

$$\hat{\Theta} = \arg\max_{\Theta} \ p(x; \mathscr{N}(x, \Theta)), \quad (3.8)$$

where Θ is the network's parameters to be trained. It should be noted that the output of the network $\mathscr{N}(x, \Theta)$ is the predicted parameters (π, μ, s) which give representation to each voxel's intensity distribution. Here, we defined the prior model $g(x)$ as

$$g(x) = p(x; \mathscr{N}(x, \hat{\Theta})) \quad (3.9)$$

Thus far, we establish a prior model for the image x from MR image space X with the state-of-art autoregressive network, and we can view this prior model as data-driven tool which utilizes the existing dataset in a computationally tractable way. The network used in this paper shares the same architecture proposed in Ref [26] and the parameters are listed in appendix.

3.2.3 Data model of k-space acquisition

The image $x(\vec{r})$ is in the representation of the instant transverse magnetization after the excitation if the relaxation is not taken into consideration. The signal generated in the receiver coil is the k-space data under the Fourier encoding scheme. Because the acquisition of k-space is achieved by repeated experiments, it is assumed that the T_2^* weighted effect on each voxel is the same[4]. The measured k-space data

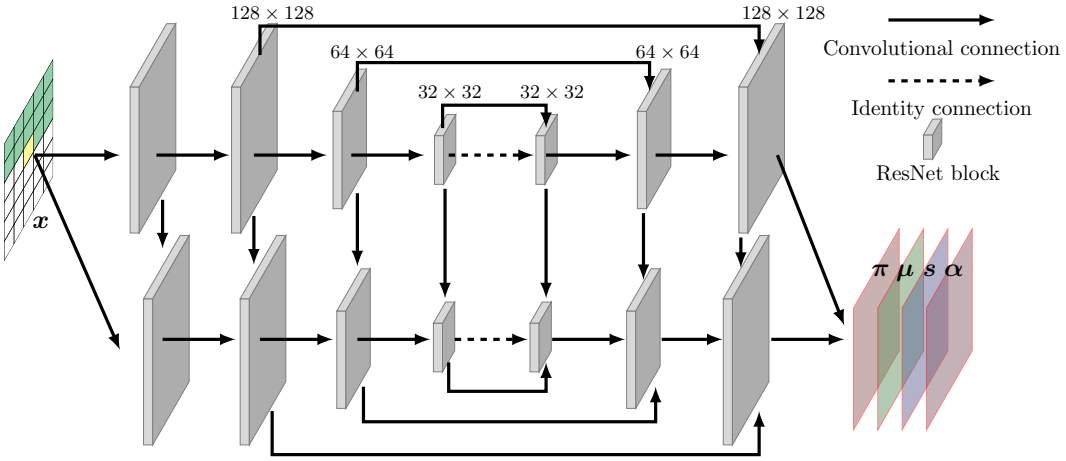

Figure 3.2 The architecture of the generative network.

is given by

$$y = \mathbf{A}x + \varepsilon, \tag{3.10}$$

where $\mathbf{A}$ is the system matrix and ε is the noise. The matrix $\mathbf{A}$ commonly consists of Fourier matrix, sampling trajectory and coil sensitivity(response pattern)[20], which serves as the bridge between the MR image space X and the raw k-space space Y. It should be noted that the encoding matrix is realized by corresponding operation such as FFT, NUFFT and Radon transform for the sake of high computational efficiency.

3.2.4 Reconstruction by maximizing the posterior

In the following, we give the derivation of the maximize the posterior probability. Substituting Eq. (3.9) into the log of the third equality of Eq. (3.2) yields

$$\hat{x}_{\text{MAP}}(y) = \arg\max_{x \in X} \log f(y \mid x) + \log p(x \mid \mathcal{N}(x, \hat{\Theta})) \tag{3.11}$$

From the data model, we know that the likelihood term $f(y \mid x)$ has less uncertainty with the presence of the imaging principles. Even though the noise and measurement error has potential to induce uncertainty, the likelihood is approximated to a constant with the dominant uncertainty of x. Hence, Eq (3.11) can be rewritten as

$$\hat{x}_{\text{MAP}}(y) = \arg\max_{x \in X} \log p(x \mid \mathcal{N}(x, \hat{\Theta})) \qquad \text{s.t.} \quad y = \mathbf{A}x + \varepsilon \tag{3.12}$$

The equality constraint to data consistency is the result of the elimination of the likelihood term. The projected subgradient method was used to solve the equality constrained problem [3]. Ref [3] gives the way to computing gradients through

random variables for deep generative models and refers this strategy as stochastic backpropagation. We can use tensorflow to do such stochastic backpropagation and get the subgradient $\nabla_x \log g(x)$. What's more, we notice that the dropout is compulsory when use the gradient to update the voxels x, which is investigated theoretically in Ref [6]. The dropout in the posterior approximation serves as the representation of model uncertainly which yields a considerable improvement in predictive log-likelihood. Therefore, the reconstruction from the maximization of the posterior $f(y \mid x)$ has follow steps:

1. Get the descent direction $\nabla_{x^{(k)}} \log g(x^{(k)})$
2. Pick a step size $\alpha_k = 1/k$
3. Update $z^{(k+1)} = x^{(k)} - \alpha_k \nabla_{x^{(k)}} \log g(x^{(k)})$
4. Projection $x^{(k+1)} = \arg\max_{x \in X} \frac{1}{2} \|x - z^{(k+1)}\|_2^2$

The projection of z onto $\{x \mid y = \mathbf{A}x + \varepsilon\}$ is given by

$$\mathscr{P}(z) = z - \mathbf{A}^*(\mathbf{A}\mathbf{A}^*)^{-1}(\mathbf{A}(x) - y). \tag{3.13}$$

So far, we have incorporated the tractable prior model $g(x)$ into the reconstruction of x through the maximization of the posterior. In the following, the superiority of the proposed reconstruction framework is demonstrated in several sorts of reconstruction modality.

3.3 Experiments

3.3.1 MRI data and pre-processing

For knee imaging, we downloaded multi-channel k-space data for all 973 scans from fastMRI reconstruction database [34]. As such, NYU fastMRI investigators provided data but did not participate in analysis or writing of this report. A listing of NYU fastMRI investigators, subject to updates, can be found at:fastmri.med.nyu.edu.The primary goal of fastMRI is to test whether machine learning can aid in the reconstruction of medical images. The knee data had two contrast weightings: proton-density with and without fat suppression (PDFS and PD). Scan parameters included 15 channel knee coil and 2D multi-slice turbo spin-echo (TSE) acquisition, and other settings could be found in Ref. [34].

For brain MRI, we collected 2D multi-slice images T1 weighted, T2 weighted, T2 weighted FLAIR, and T2* weighted brain images from 16 healthy volunteers examined with clinical standard-of-care protocols. All brain data were acquired using our 3T MRI (Philips, Achieva). An eight-channel brain RF coil was used. T1 weighted, T2 weighted, and T2 weighted FIAIR images were all acquired with TSE readout. Meanwhile, T2*-weighted images were obtained using a gradient-echo sequence. Brain MRI parameters for four contrast weightings were listed in Table 3.1.

Training images were reconstructed from 15 and 8 channels (for knee and brain MRI, respectively) k-space data without undersampling. Then, these image datasets after coil combination were scaled to a magnitude range of $[-1, 1]$ and resized to an image size of 256×256. The training of PixelCNN++ model required a considerable computational capacity when a large image size was used. In this study, the 128×128 was the largest size that our 4-GPUs server could handle. Hence, the original 256×256 images were resized into 128×128 low-resolution images by cropping in k-space for knee MRI. For brain MRI, we split each raw 256×256 image into four 128×128 image patches, before fed into the network for training. Real and imaginary parts of all 2D images were separated into two channels when inputted into the neural network. For knee MRI, 15541 images were used as the training dataset, and 1000 images were left for testing. For brain MRI, 1300 images were used as the training dataset, and 300 images were left for testing.

Table 3.1 The scan parameters of different weightings used in brain MRI experiments.

Type	Dimension	Voxel(mm)	TSE factor	TR/TE(ms)	TI
T1	256×256×24	0.9× 0.9× 4	7	2000/20	800
T2	256×256×24	0.9× 0.9× 4	13	3000/80	-
FLAIR	256×256×24	0.9× 0.9× 4	31	8000/135	135
T2*	256×256×28	0.9× 0.9× 4	-	770/16	-

3.3.2 Deep neural network

The PixelCNN++ was modified from the code in https://github.com/openai/pixelcnn. We implemented the reconstruction algorithm using Python, as explained in Eq. 3.13 and Appendix. With the trained prior model, we implemented the iterative

reconstruction algorithm for maximizing the posterior while enforcing the k-space data fidelity (as explained in Appendix and Fig. 3.3). Only two deep learning models were trained and utilized, one for knee MRI with two contrast weightings, and another for brain MRI with four contrast weightings. These two models can support all experiments performed in this study with variable undersampling patterns, coil sensitivity maps, channel numbers, image sizes, and trajectory types. Our networks training was performed in Tensorflow software, and on four NVIDIA RTX-2080Ti graphic cards. Other hyperparameters were 500 epochs, batch size = 4, and Adam optimizer. It took more than five days to train the network for knee dataset and two days for brain dataset under the above mentioned configuration.

3.3.3 Parallel imaging and ℓ_1 or ℓ_2 regularization driven reconstruction

The GRAPPA reconstruction was performed with block size of 4, and 20 central k-space lines as the auto-calibration area [8]. We simulated GRAPPA accelerations with undersampling factors from 2 to 4. The representative undersampling masks were shown in Supplementary Figure 1. We chose l_1-ESPIRIT and MODL [1] as baseline methods for comparison. They were analytical regularizations. The l_1-ESPIRIT exploited the sparsity of image, and the MODL was a deep learning method for compressed sensing reconstruction, trained via minimizing l_2 reconstruction error. In the l_1-ESPIRiT reconstruction, we set the l_1 regularization parameter to be 0.01, using BART software. One reason for choosing MODL was that it supported the coil sensitivity map for applying paralleling imaging. We followed settings in Ref [1] when training MODL to reconstruct the undersampled knee data. The only difference was the k-space mask in Ref [1] was 2D undersampled, while in the current study the 1d undersampling was applied. The central 20 k-space lines were sampled which account for 7% of the full k-space of 256×256 image. The others in the outer region were picked randomly with a sampling rate.

For the proposed method, MR images with 256×256 matrix size were reconstructed, using the prior model in Eq. (3.9) that was trained by 128×128 images or image patches. During inference, the 256×256 image was split into four 128×128 patches for applying the prior model, as shown in Figure 3.3. After updating $s^{(k+1)}$, four patches for one image were concatenated to form an image with the original size of 256×256, before it was projected onto $\{x \mid y = \mathbf{A}x + \varepsilon\}$ in Eq. (3.13). The detailed algorithm was presented in Appendix. In this method, we reconstructed images with 256×256 matrix size, using the prior model that was trained with 128×128 images $g(x)$, in Eq. 3.9. To reconcile this mismatch, we split one 256×256 image

into four 128×128 patches for applying the prior model. After updating $s^{(k+1)}$, four patches for one image were merged to form an image with the original size of 256×256. Then the merged image was projected onto $\{x \mid y = \mathbf{A}x + \varepsilon\}$ in Eq. 3.13. Further more, the random shift along phase encoding direction was applied to mitigate the stitching line in-between patches.

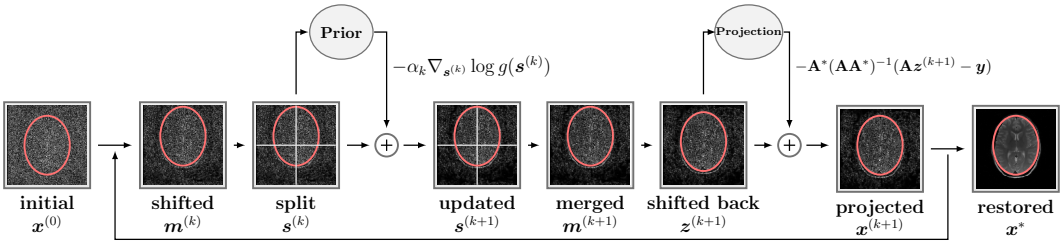

Figure 3.3 Flowchart illustrates the proposed reconstruction algorithm.

3.3.4 Non-Cartesian k-space acquisition

In this experiment, spiral sampled k-space from the acquired T2*-weighted k-space data was simulated. The method proposed in Ref [16] was used to design the spiral trajectory. The full k-space coverage required 24 spiral interleaves for the spatial resolution used in this study. The spiral trajectory was shown in Supplementary Figure 1. Besides, the implementation of non-uniform fast Fourier transform was based on the method in Ref [5]. For comparison, we used the iterative SENSE, i.e., conjugate gradient SENSE (CG SENSE), proposed in Ref [20], as a baseline method.

3.4 Results

3.4.1 Parallel imaging

Figures 3.4-3.7 show the comparison of knee and brain MRI reconstructed using GRAPPA and proposed method. The proposed method had an improved performance in recovering brain and knee image details and reducing the aliasing artifacts, compared with GRAPPA. As expected, parallel imaging amplified the noise in the low coil sensitivity region and along the undersampled dimension. On the other hand, error maps demonstrated in Figure 3.4-3.7 showed that the proposed method effectively eliminated the noise amplification. Table 3.2 presents the comparison

of GRAPPA reconstruction and proposed method for 100 knee MRI testing cases and 100 brain MRI testing cases. With the increase of the undersampling factor, the PSNR of the proposed method decreased less, compared with that of GRAPPA. Meanwhile, the GRAPPA achieved an almost "invisible-loss" (PSNR $>$ 40 dB) with acceleration factor R = 2 in brain MRI. The proposed method still showed 8dB improvement with the same acceleration factor.

Table 3.2 PSNR comparison (in dB, mean $\pm$ standard deviation, N = 100) for parallel imaging and the proposed method on knee and brain MRI.

Undersampling factor	organ	GRAPPA	Ours
R=2	knee	40.98$\pm$4.20	45.64$\pm$3.24
R=3	knee	34.87$\pm$3.38	41.71$\pm$3.42
R=4	knee	29.42$\pm$2.46	38.44$\pm$3.64
R=2	brain	37.81$\pm$4.7	48.40$\pm$2.18
R=3	brain	31.72$\pm$3.20	45.39$\pm$2.65
R=4	brain	28.85$\pm$2.87	43.58$\pm$2.66

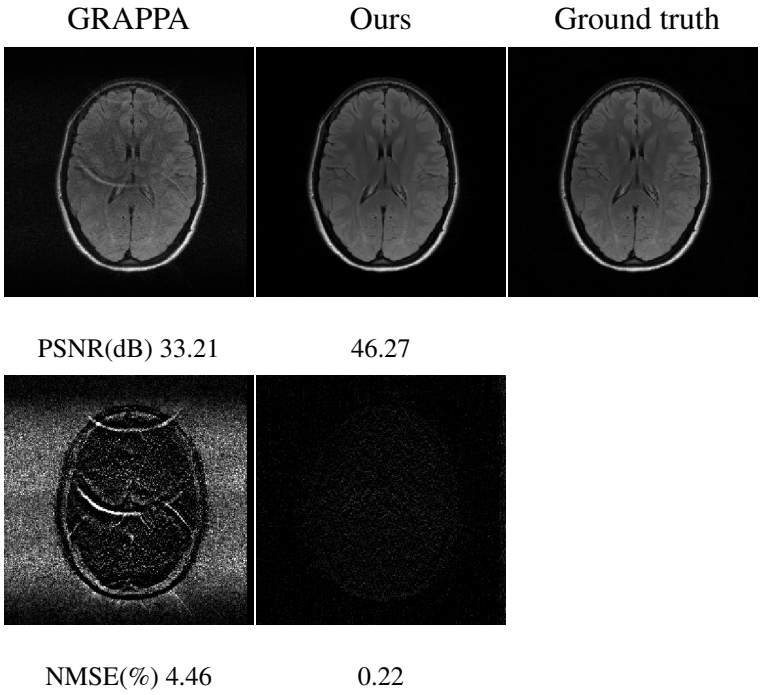

Figure 3.4 Comparisons on FLAIR-T2 weighted image reconstruction, using parallel imaging and the proposed reconstruction with R=3 acceleration and 256×256 matrix size.

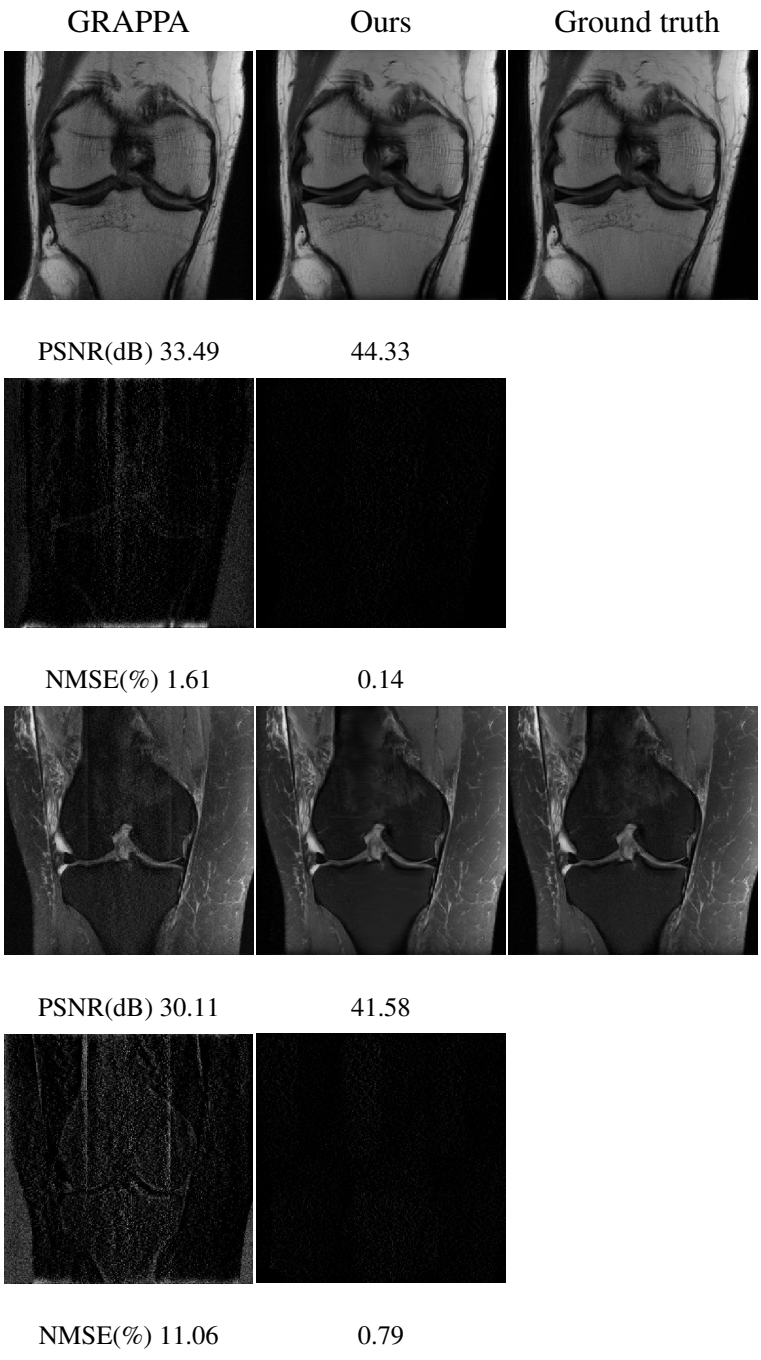

Figure 3.5 Comparisons on PD and PDFS contrasts using GRAPPA and the proposed reconstructions with R=3 acceleration and 256×256 matrix size. The intensity of error maps was five times magnified. The proposed method effectively eliminated noise amplification and aliasing artifact in GRAPPA reconstruction.

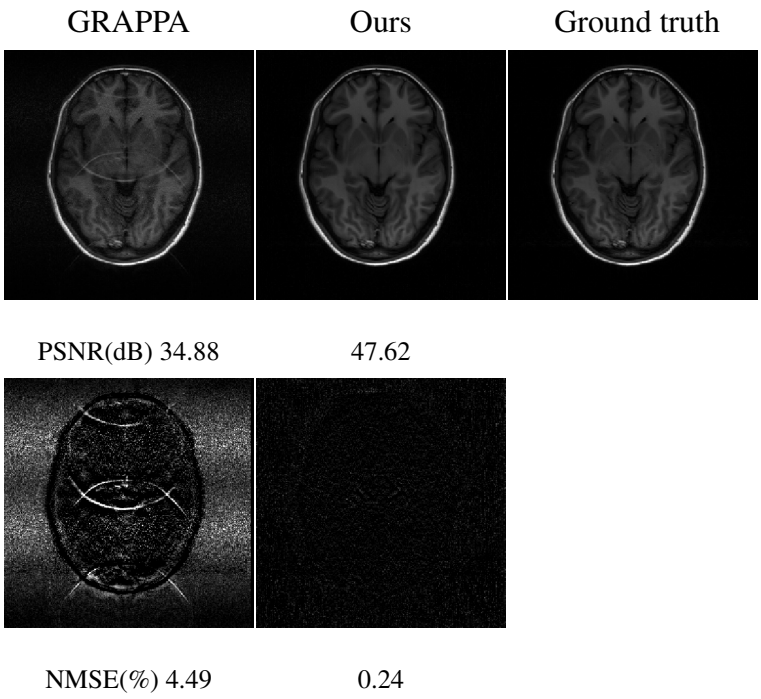

Figure 3.6 Comparisons on T1 weighted image reconstruction, using parallel imaging and the proposed reconstruction with R=3 acceleration and 256×256 matrix size.

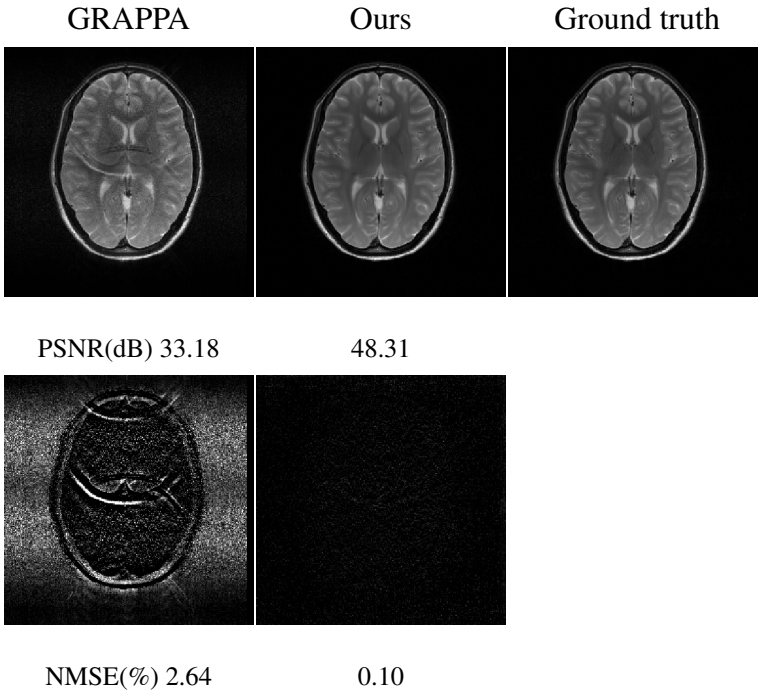

Figure 3.7 Comparisons on T2 weighted image reconstruction, using parallel imaging and the proposed reconstruction with R=3 acceleration and 256×256 matrix size.

3.4.2 Compressed sensing reconstruction

In Figures 3.8-3.11, the ℓ_1-ESPIRiT had caused the apparent image blurring for both knee and brain MRI results. Residual aliasing artifacts were also observed in reconstruction results from ℓ_1-ESPIRiT and MODL. Meanwhile, images reconstructed by ℓ_1-ESPIRiT were in qualitative agreement with the ground truth in Figures 3.8-3.11. While the proposed reconstruction recovered most anatomical structures and sharp boundaries in knee and brain MR images, compared with those from ℓ_1-ESPIRiT and MODL reconstructions, as shown on error maps in Figure 3.8-3.11. Tables 3.3 summarized ℓ_1 regularization reconstruction, MODL, and proposed reconstruction results. The proposed method generally showed more than 5 dB PSNR improvement compared with ℓ_1-ESPIRiT and MODL.

Table 3.3 PSNR comparison (in dB, mean $\pm$ standard deviation, N = 100) for compressed sensing and the proposed method on knee and brain MRI.

Undersampling rate	organ	ℓ_1-ESPIRiT	MODL	Ours
15% + 7%	knee	29.33$\pm$2.82	27.63$\pm$3.41	35.34$\pm$3.53
20% + 7%	knee	31.51$\pm$3.60	29.29$\pm$3.76	37.45$\pm$3.81
15% + 7%	brain	32.86$\pm$3.46	30.60$\pm$2.78	39.78$\pm$2.83
20% + 7%	brain	34.72$\pm$3.89	32.46$\pm$2.95	41.24$\pm$2.81

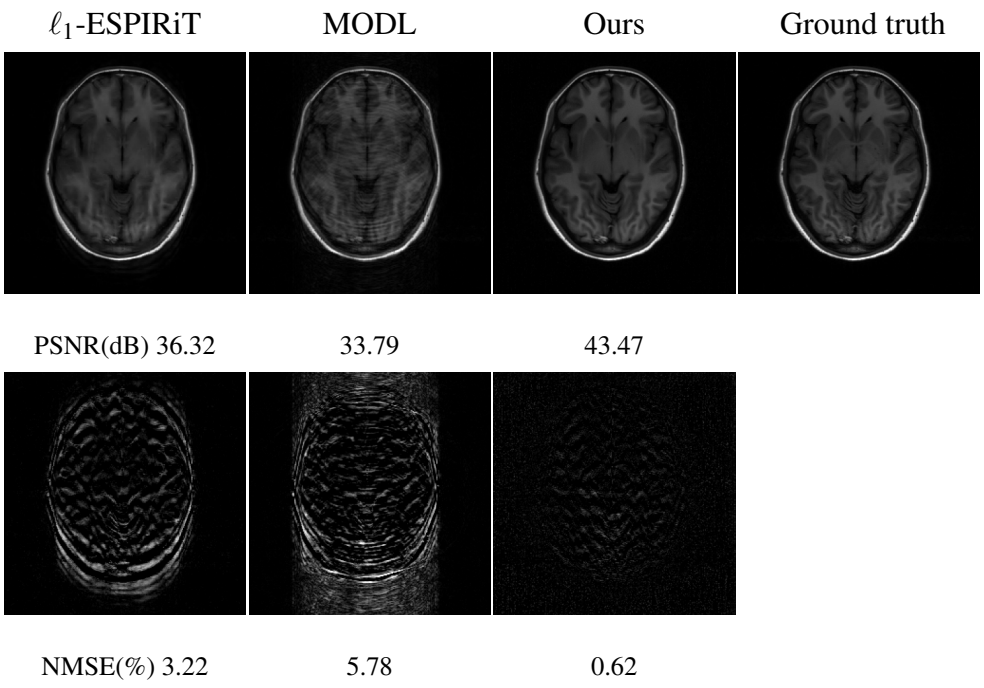

Figure 3.8 Comparison of compressed sensing and deep learning approaches for T1 weigted, using 22% 1D undersampled k-space and 256×256 matrix size.

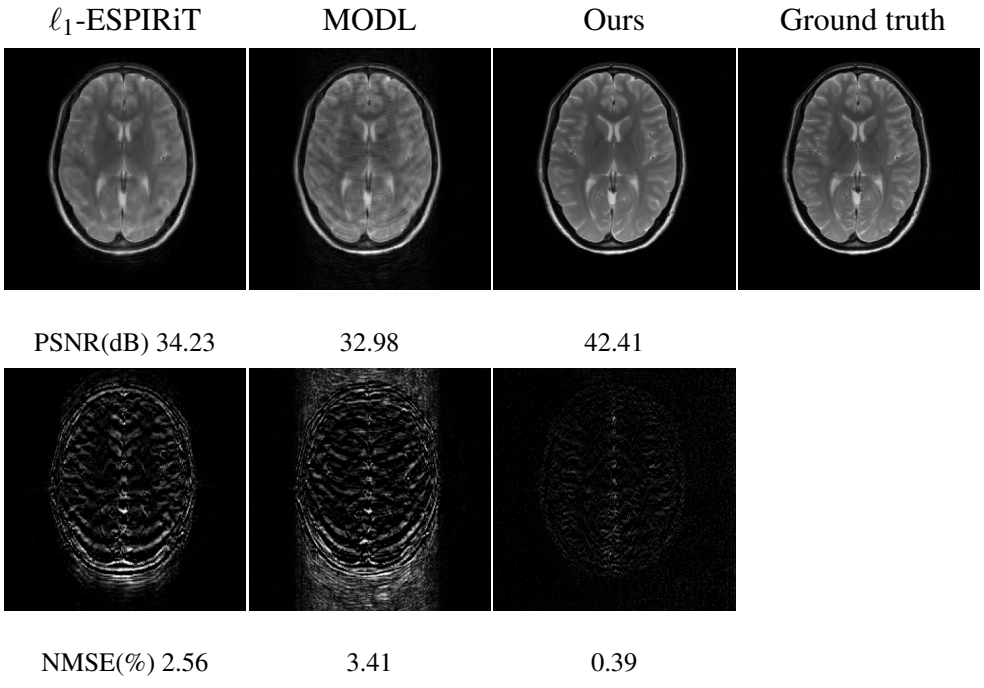

Figure 3.9 Comparison of compressed sensing and deep learning approaches for T2weighted image reconstructions, using 22% 1D undersampled k-space and 256×256 matrix size.

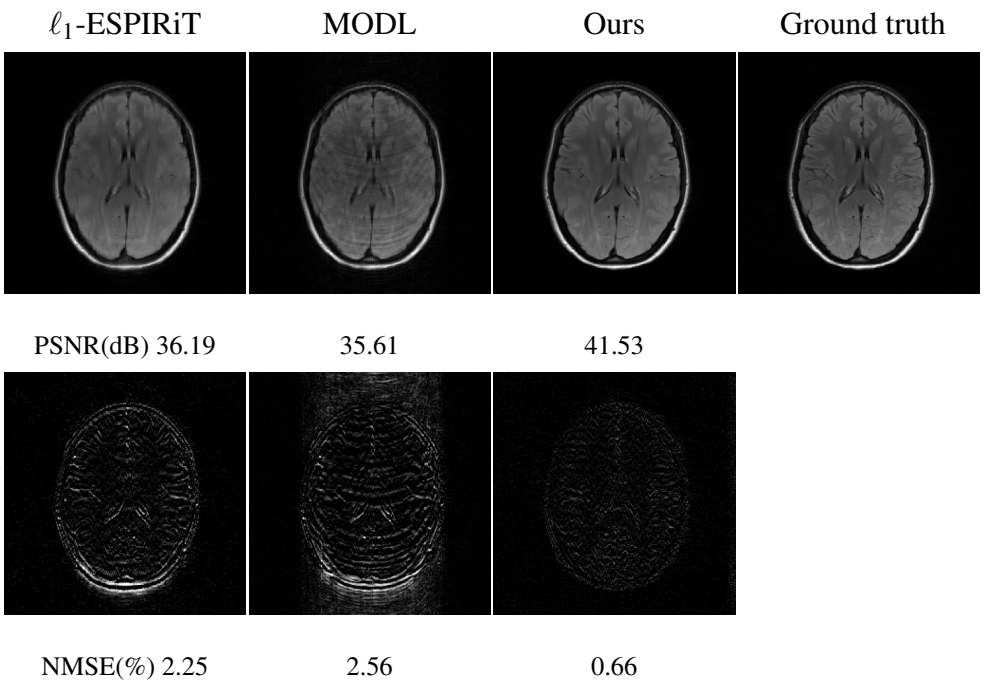

Figure 3.10 Comparison of compressed sensing and deep learning approaches for FLAIR-T2 weighted image reconstructions, using 22% 1D undersampled k-space and 256×256 matrix size.

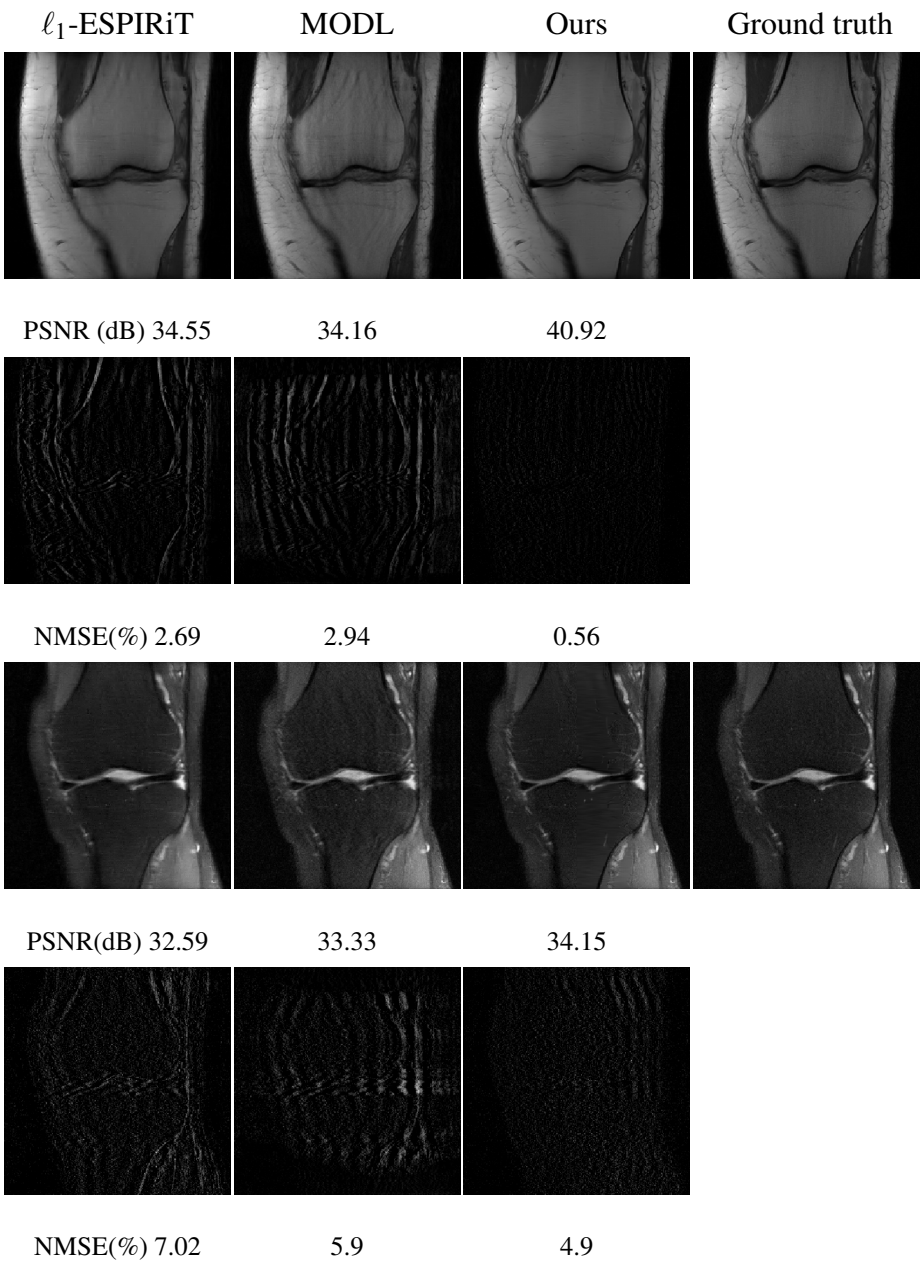

Figure 3.11 Comparison of different methods on PD and PDFS contrasts, using 27% 1D undersampled k-space and 256×256 matrix size. The intensity of error maps was five times magnified. The proposed method substantially reduced the aliasing artifact and preserved image details in compressed sensing reconstruction.

3.4.3 Preliminary result in non-Cartesian MRI reconstruction and quantitative susceptibility mapping (QSM)

In this study, we used T2* weighted gradient-echo images to simulate the spiral MRI data with 4-fold acceleration. The reconstructed images from the CG SENSE

and the proposed method were compared. The proposed method showed apparent improvement regarding the artifact reduction and the preservation of T2* contrast between gray matter and white matter. Meanwhile, the proposed method also showed a slight denoising effect on the reconstructed image compared with the ground truth. Noted that the same deep learning model used in the previous Cartesian k-space reconstruction experiments in Figures 3.4-3.7 and 3.8-3.10 was applied for spiral reconstruction, without the need of re-training the deep learning model. Figure 3.13 shows the preliminary result from the proposed accelerated reconstruction in QSM with 4-fold acceleration. Noted that the same deep learning model used in the previous brain experiments was applied to this experiment, with phase information preserved in all reconstructed images. The proposed deep learning method also showed an apparent de-noising effect on QSM maps, while still preserved the major phase contrast even with high acceleration.

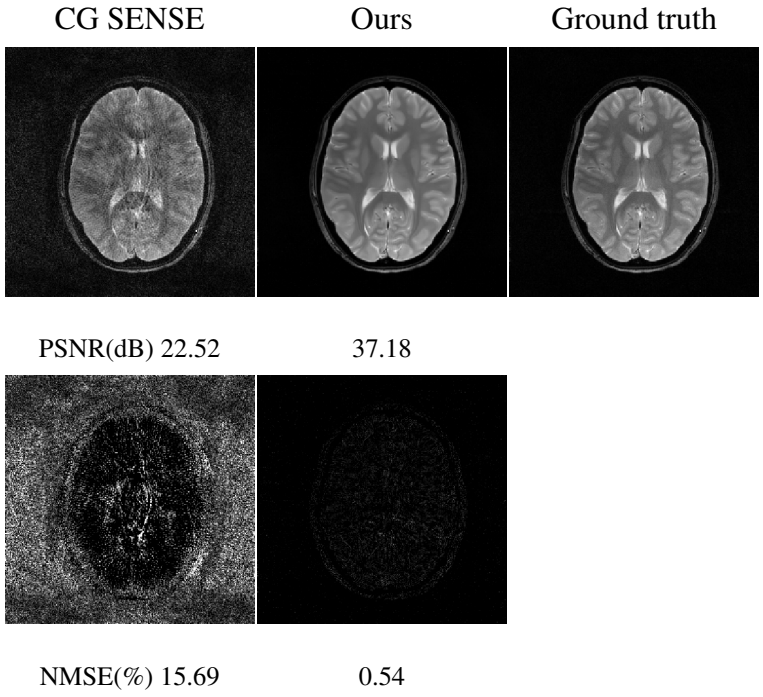

Figure 3.12 Comparison of the CG SENSE and proposed reconstruction for simulated spiral k-space with 4-fold acceleration (i.e., 6 out of 24 spiral interleaves), acquired by T2* weighted gradient echo sequence. The intensity of error maps was five times magnified. The proposed method substantially reduced the aliasing artifact in spiral reconstruction. Noted that the same deep learning model used in the previous Cartesian k-space reconstruction was applied to spiral reconstruction, without the need of re-training the deep learning model.

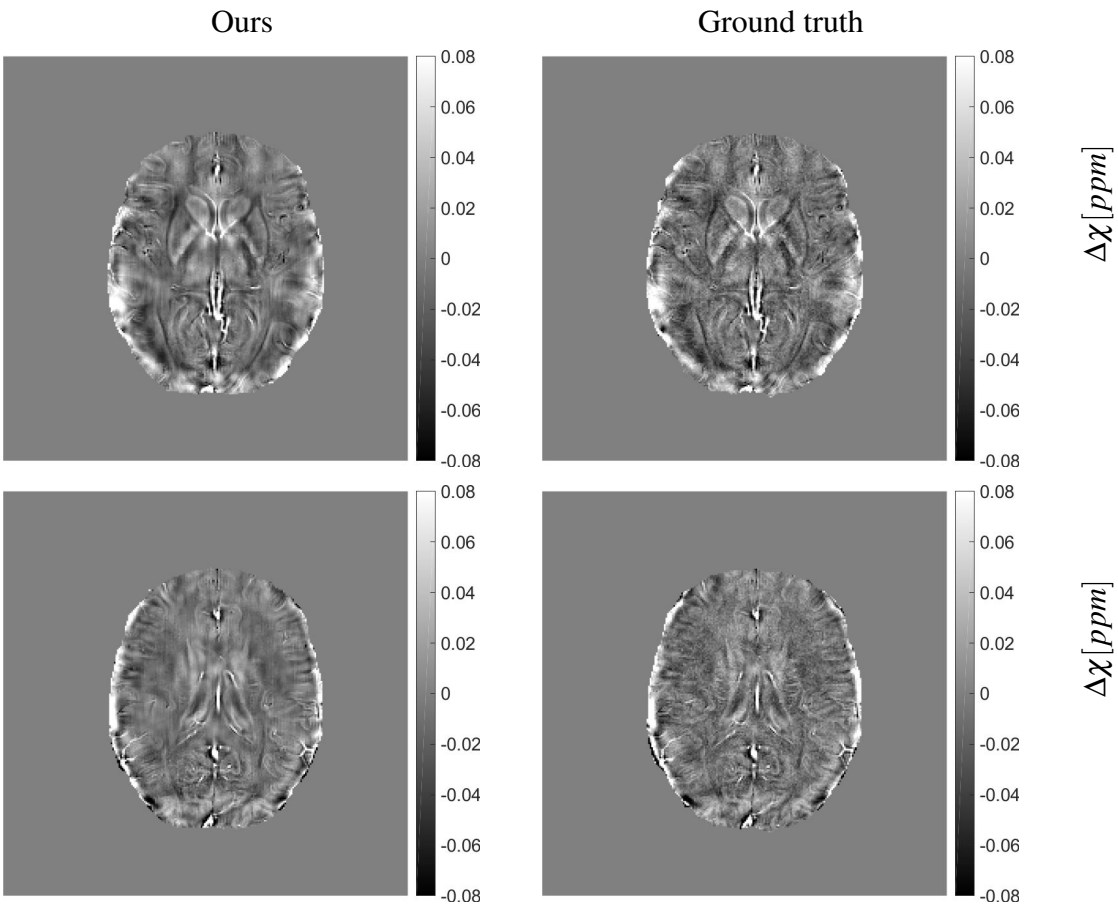

Figure 3.13 The preliminary result from the proposed accelerated reconstruction in quantitative susceptibility mapping (QSM), with R = 4 and GRAPPA type of 1D undersampling. The raw images were acquired by T2* weighted gradient echo sequence. Noted that the same deep learning model used in the previous experiments was applied to this experiment, with phase information preserved in all reconstructed images. The proposed deep learning method also showed an apparent de-noising effect on QSM maps, while still preserved the major phase contrast even with high acceleration, i.e., R = 4. Two rows show maps on different slices from one healthy volunteer.

3.5 Discussion

The proposed method can reliably and consistently recover the nearly aliased-free images with relatively high accelerations. Meanwhile, as expected, the increase of image smoothing with high acceleration was noticed, reflecting the loss of intrinsic resolution. The estimated image from the maximum of the posterior can not guarantee the full recovery of the image details, i.e., PSNR $>$ 40 dB for a full recovery.

However, at modest acceleration, the reconstruction from a maximum of posterior showed the successful reconstruction of the detailed anatomical structures, such as vessels, cartilage, and membranes in-between muscle bundles.

In this study, the results demonstrated the reconstruction of high-resolution image (i.e., 256 $\times$ 256 matrix) with low-resolution prior (i.e., trained with 128 $\times$ 128 matrix), confirming the feasibility of reconstructing different size images without the need for retraining the prior model. Even the prior model was trained by 128x128 images; it was still valid and applicable for the reconstruction of a high-resolution image. The proposed methods provided more than 8 dB improvement over the conventional GRAPPA reconstruction at the 4-fold acceleration in knee MRI. Besides, in contrast with other deep learning-based methods, those focused on the ℓ_2 loss, the likelihood that was conditioned by pixel-wise dependencies of the whole image showed an improved representation capacity, leading to a higher reconstruction accuracy. The applicability of the proposed method in the patch-based reconstruction also suggested its high representation capacity and flexibility. Even when the inputs were image patches, the prior model could still recover the whole image.

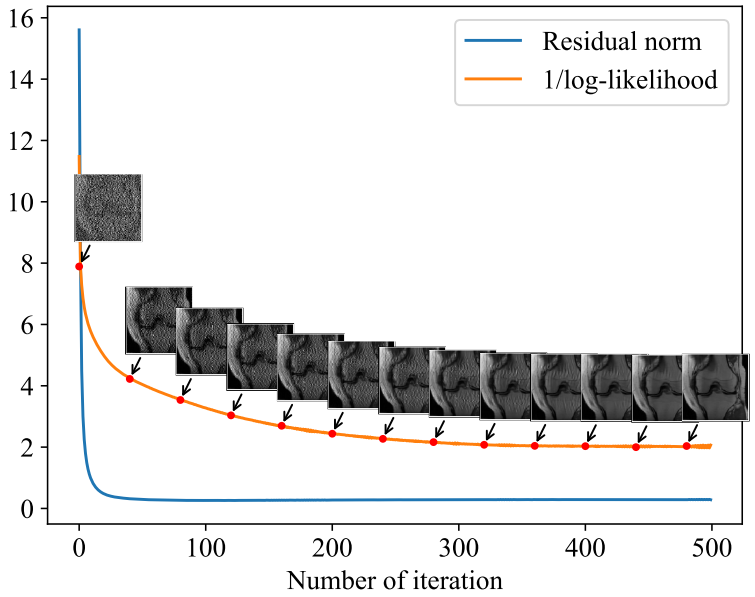

Figure 3.14 Convergence curves reflected stabilities of iterative steps: 1) maximizing the posterior, which effectively minimized the log-likelihood of MRI prior model and 2) k-space fidelity enforcement, which reduced the residual norm on k-space fidelity. The 22% sampling rate and 1D undersampling scheme were used in this simulation. The residual norm was written as $\|y - \mathbf{A}x\|_2^2$ in Eq. 3.13, and the reciprocal of log-likelihood for MRI prior model was given in Eq. 3.9.

The projected subgradient approach to solving Eq. (3.12) was computationally inexpensive but converged slowly, as shown in Figure 3.14. For a random initialization, the algorithm needed about 500 iterations to converge with a fixed step size. Meanwhile, we noticed that if the zero-filled-reconstructed image was used for initialization, the number of iterations required could be reduced to 100. Besides, the decay of residual norm stopped earlier than that of the log-likelihood, i.e., when the residual norm stopped decaying, the likelihood can still penalize the error. This evidence indicated that the residual norm as the ℓ_2 fidelity alone was sub-optimal, and the deep learning-based statistical regularization can lead to a better reconstruction result compared with the ℓ_2 fidelity. The use of deep learning-based statistical regularization was key improvement led the proposed method outperformed other conventional methods trained by image-level ℓ_2 loss. ℓ_2 loss did not give an explicit description of the relationship amid all pixel in the image, while the likelihood used in conjunction with the proposed image prior model was conditioned by the pixel-wise relationship and demonstrated superior performance compared with the conventional methods, under the current experimental setting.

Furthermore, the demonstrated image prior can be extended to a more elaborate form with clinical information, such as organs and contrast types, as the model inputs. For example, one could input the image prior with labels such as brain or knee. Then hypothetically, the image prior can be designed as a conditional probability for the given image label. In other words, the posterior would be dependent on both the k-space data and image labels. Moreover, the MR images acquired with pulse sequence parameters could serve as image labels for the prior, such as echo time and repetition time. In short, the prior model can be used to describe clinical information or acquisition parameters. This setting opens up a future direction on a more elaborated image prior to clinical information and MR sequence parameters for more intelligent image representation and pattern detection.

In this study, the generative network solely served as an image prior model, in contrast with how neural network was used in other deep learning-based reconstructions [33, 28, 17, 1]. Specifically, in previous studies [33, 28, 17, 1], embedding k-space fidelity term into the network made the algorithm inflexible because image prior and undersampling artifacts were mixed during the training. The proposed method used the standard analytical term for fidelity enforcement; therefore, its flexibility was comparable to the traditional optimization algorithm, such as ℓ_1 regularization. Due to unavoidable changes of the encoding scheme, e.g., the image size and the RF coils, during MRI experiment in practice, it was essentially needed to separate the learned component (the image prior) from the encoding matrix used

in the fidelity term in reconstruction. Besides, the proposed method showed the feasibility of incorporating the coil sensitivity information in the fidelity term, which enabled the changeable RF coils without the need of retraining the model [20, 4]. In summary, the separation of the image prior and the encoding matrix embedded in the fidelity term made the proposed method more flexibility and generalizable compared with conventional deep learning approaches.

3.6 Conclusion

In summary, this study presented the application of Bayesian inference in MR imaging reconstruction with the deep learning-based prior model. We demonstrated that the deep MRI prior model was a computationally tractable and effective tool for MR image reconstruction. The Bayesian inference significantly improved the reconstruction performance over that of conventional ℓ_1 sparsity prior in compressed sensing. More importantly, the proposed reconstruction framework was generalizable for most reconstruction scenarios.

Chapter 4

Summary

The acceleration of MR imaging has been the central topic of MR research since it was introduced into clinical practice. However, after years of the development of sophisticated hardware system, there is less potential to shorten the scanning time through the improvement of hardware due to the restriction of physiology. Many studies sought to obtain the clinically acceptable image from undersampled k-space for the sake of faster acquisition, utilizing the information provided by some prior knowledge such as sensitivities maps and the sparsity in a particular transformed domain. With these ideas, parallel imaging and compressed sensing MR emerged and indeed accelerated the imaging process without a considerable loss of image quality under the moderate acceleration factor. After that, the reconstruction started to be achieved by some model-based methods in order to employ the prior knowledge, and the definition of the encoding matrix began to appear in some reference. From this perspective, the compressed sensing MRI is a method of taking advantage of sparsity, which is evaluated by ℓ_1-norm. The development of deep learning provides an effective tool to exploit the existing image database when reconstructing the image from undersampled k-space data. In the beginning, the convolutional layers were trained to replace the sparsifying transform, such as wavelet transform, to achieve a much sparser representation in the trained transform domain. However, these sort of methods was still under the bound of compressed sensing without uncovering the intrinsic distribution of the existing database, even though it showed improvement over the conventional sparse kernels. What's more, those methods were difficult to be generalized for various reconstruction settings, when it comes to the different number of coils, image size and sampling trajectory and so on.

Therefore, extracting a generalizable prior model from the existing image database is the key to applying deep learning to MR image reconstruction. And in this thesis, two types of prior implemented with generative networks were elaborately presented. In Chapter 2, the discriminator of GAN was used to confine a manifold for the operation of ℓ_2 projection, and the generator was used to replace the proximal operation which is to enforce sparsifying transform. The other steps of ADMM for MR image reconstruction remained the same except the modification of proximal. In Chapter 3, the prior model of existing image database was defined with an autoregressive net-

work, Pixel-CNN++. And the reconstruction was described by Bayesian theorem and achieved by maximizing the posterior distribution. The reconstruction via Bayesian inference has shifted from the constraint of sparsity to the constraint of likelihood. The sparsity is scalable in Euclidean space and the distance is measured pixel-by-pixel without considering the relationship between them, while the likelihood has statistical assumption of the whole image. The prior model trained by existing image database was evaluated by likelihood which makes it more appropriate to serve as generic prior knowledge for various reconstruction settings.

In the future work, the investigation is supposed to be pushed into a deeper level that the prior model is trained to describe the distribution of T_1, T_2, and proton density. By doing that, the prior model is enforced to be closer and more accurate to the formation of MR images, because the MR image is a specific representation of those three parameters under a certain imaging experiment. With this prior model, the potential to obtain different weighted images from only one imaging experiment is guaranteed.

Appendix I

Pseudo-code

I.1 Reconstruction for the fixed image size with deep prior model

Algorithm 1 Reconstruction algorithm with deep prior model

Input:

y - k-space data

$\mathbf{A}$ - encoding matrix

λ - maximum iteration

Output:

x - the restored image

Give a random initial point $x^{(0)}$ ▷ Initialization

while $\|g^{(k)}\|_2^2 > \varepsilon$ and $k < \lambda$ **do** ▷ Iteration

    Generate a random shifting offset d

    Shift $x^{(k)}$ d pixels away from the center circularly

    Split $x^{(k)}$ into pieces $s^{(k)}$ for feeding to network

    Get subgradient $\nabla_x \log(g(x))$ at $x^{(k)}$

    Pick a step size $\alpha_k = 1/k$

    Update $s^{(k+1)} = s^{(k)} - \alpha_k \nabla_{s^{(k)}} \log g(s^{(k)})$

    Merge pieces $s^{(k+1)}$ into $z^{(k+1)}$ for projection

    Shift $z^{(k+1)}$ $-d$ pixels away from the center circularly

    Projection $x^{(k+1)} = \arg\max_{x \in X} \frac{1}{2}\|x - z^{(k+1)}\|_2^2$

return $x^{(k+1)}$

References

[1] Hemant K Aggarwal, Merry P Mani, and Mathews Jacob. "Modl: Model-based deep learning architecture for inverse problems". In: *IEEE transactions on medical imaging* 38.2 (2018), pp. 394–405.

[2] Martin Blaimer et al. "SMASH, SENSE, PILS, GRAPPA: how to choose the optimal method". In: *Topics in Magnetic Resonance Imaging* 15.4 (2004), pp. 223–236.

[3] Stephen Boyd, Lin Xiao, and Almir Mutapcic. "Subgradient methods". In: *lecture notes of EE392o, Stanford University, Autumn Quarter* 2004 (2003), pp. 2004–2005.

[4] Jeffrey A Fessler. "Model-based image reconstruction for MRI". In: *IEEE Signal Processing Magazine* 27.4 (2010), pp. 81–89.

[5] Jeffrey A Fessler and Bradley P Sutton. "Nonuniform fast Fourier transforms using min-max interpolation". In: *IEEE transactions on signal processing* 51.2 (2003), pp. 560–574.

[6] Yarin Gal and Zoubin Ghahramani. "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning". In: *Proceedings of The 33rd International Conference on Machine Learning*. Ed. by Maria Florina Balcan and Kilian Q. Weinberger. Vol. 48. Proceedings of Machine Learning Research. New York, New York, USA: PMLR, 2016, pp. 1050–1059. URL: http://proceedings.mlr.press/v48/gal16.html.

[7] Ian Goodfellow et al. "Generative adversarial nets". In: *Advances in neural information processing systems*. 2014, pp. 2672–2680.

[8] Mark A Griswold et al. "Generalized autocalibrating partially parallel acquisitions (GRAPPA)". In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 47.6 (2002), pp. 1202–1210.

[9] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *CoRR* abs/1512.03385 (2015). arXiv: 1512.03385. URL: http://arxiv.org/abs/1512.03385.

[10] Kaiming He et al. “Identity mappings in deep residual networks”. In: *European conference on computer vision*. Springer. 2016, pp. 630–645.

[11] Oren N Jaspan, 1 Roman Fleysher, and Michael L Lipton. “Compressed sensing MRI : a review of the clinical literature”. In: *Br J Radiol.* 1056.August (2015), pp. 1–12. DOI: 10.1259/bjr.20150487.

[12] Justin Johnson, Alexandre Alahi, and Li Fei-fei. “Perceptual Losses for Real-Time Style Transfer and Super-Resolution”. In: *arXiv* (2016). arXiv: 1603.08155v1.

[13] Diederik P Kingma et al. “Improving variational inference with inverse autoregressive flow.(nips), 2016”. In: *URL http://arxiv. org/abs/1606.04934* ().

[14] Fang Liu et al. “SANTIS: Sampling-Augmented Neural neTwork with Incoherent Structure for MR image reconstruction”. In: *Magnetic Resonance in Medicine* (Apr. 2019). DOI: 10.1002/mrm.27827.

[15] Michael Lustig, David Donoho, and John M Pauly. “Sparse MRI: The application of compressed sensing for rapid MR imaging”. In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 58.6 (2007), pp. 1182–1195.

[16] Michael Lustig, Seung-Jean Kim, and John M Pauly. “A fast method for designing time-optimal gradient waveforms for arbitrary k-space trajectories”. In: *IEEE transactions on medical imaging* 27.6 (2008), pp. 866–873.

[17] Morteza Mardani et al. “Deep generative adversarial neural networks for compressive sensing MRI”. In: *IEEE transactions on medical imaging* 38.1 (2018), pp. 167–179.

[18] Aaron van den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. “Pixel recurrent neural networks”. In: *arXiv preprint arXiv:1601.06759* (2016).

[19] Deepak Pathak et al. “Context Encoders: Feature Learning by Inpainting”. In: *CoRR* abs/1604.07379 (2016). arXiv: 1604.07379. URL: http://arxiv.org/abs/1604.07379.

[20] Klaas P Pruessmann et al. “Advances in sensitivity encoding with arbitrary k-space trajectories”. In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 46.4 (2001), pp. 638–651.

[21] Klaas P Pruessmann et al. “SENSE: sensitivity encoding for fast MRI”. In: *Magnetic resonance in medicine* 42.5 (1999), pp. 952–962.

[22] Saiprasad Ravishankar and Yoram Bresler. “MR image reconstruction from highly undersampled k-space data by dictionary learning”. In: *IEEE transactions on medical imaging* 30.5 (2010), pp. 1028–1041.

[23] JH Rick Chang et al. “One Network to Solve Them All–Solving Linear Inverse Problems Using Deep Projection Models”. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2017, pp. 5888–5897.

[24] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *CoRR* abs/1505.04597 (2015). arXiv: 1505.04597. URL: http://arxiv.org/abs/1505.04597.

[25] Tim Salimans et al. “Improved Techniques for Training GANs”. In: *CoRR* abs/1606.03498 (2016). arXiv: 1606.03498. URL: http://arxiv.org/abs/1606.03498.

[26] Tim Salimans et al. “PixelCNN++: Improving the PixelCNN with Discretized Logistic Mixture Likelihood and Other Modifications”. In: *CoRR* abs/1701.05517 (2017). arXiv: 1701.05517. URL: http://arxiv.org/abs/1701.05517.

[27] Alexei A. Samsonov et al. “POCSENSE: POCS-based reconstruction for sensitivity encoded magnetic resonance imaging”. In: *Magnetic Resonance in Medicine* 52.6 (2004), pp. 1397–1406. ISSN: 07403194. DOI: 10.1002/mrm.20285.

[28] Jo Schlemper et al. “A deep cascade of convolutional neural networks for dynamic MR image reconstruction”. In: *IEEE transactions on Medical Imaging* 37.2 (2017), pp. 491–503.

[29] Daniel K Sodickson and Warren J Manning. “Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays”. In: *Magnetic resonance in medicine* 38.4 (1997), pp. 591–603.

[30] Jonathan I Tamir et al. “Generalized magnetic resonance image reconstruction using the Berkeley Advanced Reconstruction Toolbox”. In: *ISMRM Workshop on Data Sampling & Image Reconstruction, Sedona, AZ*. 2016.

[31] Martin Uecker et al. “ESPIRiT an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA”. In: *Magnetic resonance in medicine* 71.3 (2014), pp. 990–1001.

[32] Bo Wahlberg et al. “An ADMM algorithm for a class of total variation regularized estimation problems”. In: *IFAC Proceedings Volumes* 45.16 (2012), pp. 83–88.

[33] Yan Yang et al. “Deep ADMM-Net for Compressive Sensing MRI”. In: *Advances in Neural Information Processing Systems 29*. Ed. by D. D. Lee et al. Curran Associates, Inc., 2016, pp. 10–18. URL: http://papers.nips.cc/paper/6406-deep-admm-net-for-compressive-sensing-mri.pdf.

[34] Jure Zbontar et al. “fastMRI: An Open Dataset and Benchmarks for Accelerated MRI”. In: *CoRR* abs/1811.08839 (2018). arXiv: 1811.08839. URL: http://arxiv.org/abs/1811.08839.

[35] Bo Zhu et al. “Image reconstruction by domain-transform manifold learning”. In: *Nature* 555.7697 (2018), p. 487.