
A

Presented to
the faculty of the School of Engineering and Applied Science
University of Virginia

in partial fulfillment
of the requirements for the degree

by

APPROVAL SHEET

This

is submitted in partial fulfillment of the requirements
for the degree of

Author:

Advisor:

Advisor:

Committee Member:

Committee Member:

Committee Member:

Committee Member:

Committee Member:

Committee Member:

Accepted for the School of Engineering and Applied Science:

A handwritten signature in black ink that reads "Jennifer L. West". The signature is written in a cursive style with a large initial 'J' and 'W'.

Jennifer L. West, School of Engineering and Applied Science

Abstract

Magnetic resonance imaging (MRI) is a non-invasive medical imaging modality. It offers the capability to produce high-quality multi-contrast diagnostic images without the use of ionizing radiation. The resulting images not only provide detailed anatomical information but also capture functional process, making MRI invaluable for clinical diagnosis, treatment planning, and biomedical research. However, the full potential of MRI remains partially obscured by persistent challenges, notably the presence of imaging artifacts that compromise the quality of acquired images. The advent of potent graphics processing unit (GPU)-based computational platforms and the availability of open-access datasets provide previously unachievable opportunities to address these issues through deep learning methods.

For the signal-to-noise ratio (SNR) issues, a complex-valued convolutional network (CNN) incorporating the noise level map (non-blind \mathbb{C} DnCNN) was trained with ground truth and simulated noise-corrupted image pairs. The network was validated using both simulated and in vivo data collected from low-field scanners. The non-blind \mathbb{C} DnCNN showed superior quantitative metrics and significantly improved the SNR and visual quality of the image. By incorporating the noise level map, the method showed better performance on dealing with spatially varying parallel imaging noise.

For the motion artifacts, a multi-task conditional generative adversarial network (MT-cGAN) was developed for simultaneous motion detection and compensation. The training images were generated with a realistic artifact simulation process, incorporating comprehensive rigid motion profiles, noise addition, and parallel imaging acquisitions. Performance was evaluated using both simulated and real in-vivo data. For the motion detection task, MT-cGAN achieved the

best classification accuracy on simulated and real in-vivo dataset. For the motion compensation task, the outputs of MT-cGAN showed less visual blurring, fewer residual artifacts, and better preservation of fine structures compared to other models.

For off-resonance artifacts, a deep-learning-based method (AutofocusNet) was developed to correct both field inhomogeneities and concomitant fields in spiral MRI. By training the network using images with simulated field inhomogeneity and concomitant field effect, AutofocusNet showed superior performance compared to the conventional autofocus method. It offers a practical and effective solution for off-resonance correction in spiral MRI without using the field map or computing the concomitant fields during the reconstruction.

For undersampled cardiac MRI reconstruction, a novel complex-valued cascading cross-domain CNN was proposed, named C3-Net, for improved balance between computation demands and image quality for accelerated CMR. C³-Net outperformed other comparison methods, especially at high acceleration rates (> 8). The short-axis results from C³-Net showed reduced residual artifact and improved temporal fidelity of cardiac motion.

Acknowledgement

The past years of study and research at the University of Virginia, along with the peaceful city of Charlottesville, have been exciting, enriching, and fulfilling. I am deeply grateful for the guidance, support, and opportunities I have received from the people around me during this journey.

First and foremost, I would like to express my sincere gratitude to my advisor, Dr. Craig Meyer, who introduced me to the world of MRI. His profound knowledge in MRI has shaped my understanding of this complex field. His kindness, patience, and support have made my research not only possible but enjoyable. I am incredibly grateful for his mentorship.

I am also deeply thankful to my committee member, Dr. Xue Feng, who generously shared his expertise in deep learning with me. Beyond research, he provided me with insightful advice for my career development, and his guidance has been invaluable in shaping my future path.

I would like to extend my gratitude to my committee chair, Dr. John Hossack, and committee members Dr. John Mugler, Dr. Wilson Miller, and Dr. Michael Salerno. Their extensive knowledge and deep insights in the field of medical imaging have been tremendously helpful.

Moreover, I am grateful to my lab mates in Meyer Lab: Dr. Steven Allen, Dr. Helen Sporkin, Dr. Zhixing Wang, Yekaterina Gilbo, Sheng Chen, and Kang Yan. Our discussions have always been enlightening. I would also like to acknowledge researchers and graduate students from other labs who have inspired and guided me: Dr. Yang Yang, Dr. Changyu Sun, Dr. Ruixi Zhou, Dr. Junyu Wang, Dr. Yu Wang, and Yi Huang.

I especially want to thank my roommate, Dr. Yanjun Xie, for teaching me to play basketball and for all the memorable moments of friendship. I am also grateful to my longtime friends from high school: Meng Ba, Yexing Wang, Haofu Limiao, and Yixian Chen, whose friendship has been a constant source of encouragement.

Lastly, and most importantly, I wish to express my deepest appreciation to my wife, Xindi Zhang, for her understanding and support throughout my studies, and to my parents, whose love has encouraged me to keep going. None of this would have been possible without them.

Table of Contents

Abstract.....	I
Acknowledgement.....	III
Table of Contents	V
List of Figures.....	VIII
List of Tables.....	XV
Chapter 1: Introduction.....	1
1.1 MRI Overview	1
1.2 MRI Artifacts	3
1.3 Deep Learning Overview	7
1.4 Dissertation Outline	9
Chapter 2: MRI Denoising with a Non-Blind Deep Complex-Valued Convolutional Neural Network.....	11
2.1 Introduction.....	11
2.2 Methods.....	13
2.2.1 Training datasets	13
2.2.2 Network architecture.....	14
2.2.3 Noise level map.....	17
2.2.4 Evaluation	17
2.3 Results.....	21
2.4 Discussion	29
2.5 Conclusion	34

Chapter 3: Multi-Task Learning for Simultaneous Motion Detection and Compensation in Brain Imaging	35
3.1 Introduction.....	35
3.2 Methods.....	37
3.2.1 Motion simulation.....	37
3.2.2 Network architecture.....	40
3.2.3 Evaluation	43
3.3 Results.....	46
3.4 Discussion.....	49
3.5 Conclusion	52
Chapter 4: Convolutional Neural Network for Field Inhomogeneity and Concomitant Gradient Field Correction in Spiral Imaging.....	53
4.1 Introduction.....	53
4.2 Methods.....	55
4.2.1 Training data generation	55
4.2.2 Network architecture.....	57
4.2.3 Evaluation	59
4.3 Results.....	59
4.4 Discussion.....	61
4.5 Conclusion	62
Chapter 5: Complex Valued Cascading Cross-Domain Convolutional Neural Network for Reconstructing Undersampled CMR Images.....	64
5.1 Introduction.....	64

5.2 Methods.....	67
5.2.1 Network architecture.....	67
5.2.2 Training.....	69
5.2.3 Evaluation	70
5.3 Results.....	71
5.4 Discussion.....	74
5.5 Conclusion	75
Chapter 6: Conclusions and Future Directions	77
6.1 Conclusions.....	77
6.2 Future Directions	78
6.2.1 Inline implementations.....	78
6.2.2 Nonrigid motion compensation.....	78
6.2.3 Novel network architectures	79
Appendix – List of Publications.....	81
First Author Manuscripts	81
Other Manuscripts.....	81
Conference Abstracts	81
Invention Disclosures.....	83
Bibliography	84

List of Figures

Figure 1-1: Commonly used trajectories in MRI..... 2

Figure 1-2: Improvement of SNR by increasing NA..... 3

Figure 1-3: Motion artifacts in MRI. 4

Figure 1-4: SENSE reconstruction results with different undersampling ratios. Adapted from Questions and Answers in MRI Website..... 5

Figure 1-5: Blurred phantom image reconstructed with MFI..... 7

Figure 2-1: Architecture of the non-blind \mathbb{C} DnCNN for MRI denoising. The input is the complex-valued noisy image concatenated with the complex-valued noise level map, and the output is the complex-valued denoised image. The network consists of twelve 3×3 \mathbb{C} Conv. Each \mathbb{C} Conv is followed by a radial BN and a \mathbb{C} ReLU except for the first and last..... 16

Figure 2-2: (A) Performance of different denoising methods on the simulated testing dataset at different noise levels. The average NRMSE, PSNR, and SSIM were evaluated for each method. (B) Representative magnitude image results of different denoising methods. Top left is the ground truth image, and bottom left is the simulated noise-corrupted image with a noise standard deviation of 0.06. The output of \mathbb{C} DnCNN and non-blind \mathbb{C} DnCNN showed reduced noise and less visual blurring compared to other methods..... 20

Figure 2-3: (A) Phase difference of DnCNN, \mathbb{C} DnCNN, and non-blind \mathbb{C} DnCNN on the simulated testing dataset at different noise levels. The average phase ABSD was evaluated for each network. (B) Procedures for generating the brain mask to eliminate the impact of background random phase in metric calculation. (C) Representative phase image results of different methods. Top left is the ground truth image, and top center is the simulated noise-corrupted image with a

noise standard deviation of 0.06. Bottom row is the output of DnCNN, CDnCNN, and non-blind CDnCNN, from left to right. The complex-valued models preserved the phase information better than the real-valued model..... 21

Figure 2-4: Example of non-blind CDnCNN on spatially varying parallel imaging artifacts. Top left is the ground truth image. Top center is the simulated noise-corrupted image reconstructed by SENSE with a subsampling ratio of 4. Top right is the g -factor map from the SENSE reconstruction. Bottom left is the output of CDnCNN without noise level map. Bottom center is the output of non-blind CDnCNN with a uniform noise level map. Bottom right is the output of non-blind CDnCNN with the scaled g -factor map as the noise level map. The non-blind CDnCNN with the scaled g -factor map successfully reduced the noise at the center regions with large g -factor. 22

Figure 2-5: Performance of non-blind CDnCNN on modalities or anatomies out of the training distribution. The top row is a T1-weighted brain image, the middle row is a FLAIR brain image, and the bottom row is a PD-weighted knee image. From left to right, each column is the ground truth image, the simulated noise-corrupted image, and the output of non-blind CDnCNN, respectively. 23

Figure 2-6: Performance of non-blind CDnCNN on images with pathology. The top row is a FLAIR brain image with a white matter lesion (yellow arrow), and the bottom row is a knee image with bone lesion (yellow arrow). From left to right, each column is the ground truth image, the simulated noise-corrupted image, and the output of non-blind CDnCNN, respectively..... 24

Figure 2-7: (A) Performance of different denoising methods on the low-field M4Raw dataset. The average WM and GM SNRs of each testing volume were evaluated for each method. (B) Representative image results of different denoising methods. The output of CDnCNN and non-

blind CDnCNN showed clear structures and better noise reduction compared to other methods. The asterisks indicate statistically significant differences between the methods ($p < 0.05$). 25

Figure 2-8: Example from the M4Raw dataset of non-blind CDnCNN on spatially varying parallel imaging artifacts. Top left is the fully-sampled image. Top center is the undersampled image reconstructed by SENSE with a subsampling ratio of 2. Top right is the g -factor map from the SENSE reconstruction. Bottom left is the output of CDnCNN without noise level map. Bottom center is the output of non-blind CDnCNN with a uniform noise level map. Bottom right is the output of non-blind CDnCNN with the scaled g -factor map as the noise level map. The non-blind CDnCNN with scaled g -factor map showed the best denoising performance at regions with large g -factor. 26

Figure 2-9: Example of the manually defined ROIs for SNR calculation (white rectangles for air regions, cyan rectangles for GM regions, and yellow rectangles for WM regions) and the mean intensity and standard deviation of each ROI. 27

Figure 2-10: (A) Performance of different denoising methods on in vivo images collected from a low-field scanner with different NAs. The average WM and GM SNRs were evaluated for each method. (B) Representative image results of different denoising methods. Top left is the original noisy image acquired with NA of 3. The output of CDnCNN and non-blind CDnCNN showed sharper structures and less noise compared to other methods. 28

Figure 2-11: Performance of non-blind CDnCNN at different noise levels. Top row is the original noisy image, and each column was acquired with NA of 1, 2, 3, 4, and 5, from left to right. Bottom row is the output of CDnCNN. The small structures (yellow and cyan arrows) became sharp and visible as NA increased. 29

Figure 2-12: Performance of DnCNN, CDnCNN, and non-blind CDnCNN as a function of network depth on the simulated testing dataset at different noise levels. The average NRMSE, PSNR, and SSIM were evaluated for each method. The performance of all networks improved with an increase in network depth. 31

Figure 2-13: Performance of U-Net, CU-Net, and non-blind CU-Net on the simulated testing dataset at different noise levels. The average NRMSE, PSNR, and SSIM were evaluated for each method..... 32

Figure 3-1: Training data simulation process. (A) Motion profiles were generated to simulate realistic movement artifacts commonly observed in brain MRI, including periodic, linear, sudden motions, and random perturbations. (B) Spatial transformations were applied to the motion-free image according to the generated motion profile. The phase encoding lines corresponding to different motion states were combined to form the motion-corrupted k -space data. The motion-corrupted image was obtained by an inverse FFT. (C) To simulate real clinical MRI conditions, complex-valued AWGN was added, and k -space undersampling and parallel imaging reconstruction were applied. The ground truth image was obtained by the sum-of-squares reconstruction..... 38

Figure 3-2: Network Architecture of MT-cGAN. The network consists of a generator based on a modified U-Net and a discriminator. The generator includes an encoder-decoder structure to reconstruct the output image, and a classification branch at the bottleneck to predict whether the input image is motion-corrupted. The discriminator is a simple CNN..... 41

Figure 3-3: Confusion matrices and ROC curves for the motion detection task on the simulated test dataset. The AUC for XGBoost, ST-Net, and MT-cGAN is 0.8375, 0.9114, and 0.9247, respectively. The MT-cGAN outperformed the other two models. 46

Figure 3-4: Quantitative evaluation results and representative images for the motion compensation task on the simulated test dataset. (A) MT-cGAN achieved the best NRMSE, PSNR, and SSIM on motion-compensated images. Asterisks indicate a significant difference ($p < 0.05$). (B) MT-cGAN showed fewer residual artifacts and better fine structural preservation (marked by red arrows). 47

Figure 3-5: Confusion matrices and ROC curves for the motion detection task on the MR-ART dataset. The AUC for XGBoost, ST-Net, and MT-cGAN is 0.7761, 0.8578, and 0.9090, respectively. The MT-cGAN achieved the highest performance. 48

Figure 3-6: Quantitative evaluation results and representative images for the motion compensation task on the MR-ART dataset. (A) For compensation of images with motion artifacts, MT-cGAN achieved the lowest EFC and CJV, while the baseline model attained the highest SNR. Asterisks indicate a significant difference ($p < 0.05$). (B) The output of MT-cGAN exhibit minimal residual artifacts as indicated by red arrows. 49

Figure 3-7: Quantitative evaluation results and representative images for the motion compensation task on the local in-vivo data. (A) MT-cGAN achieved the best NRMSE, PSNR, and SSIM on compensating motion-corrupted images. Asterisks indicate a significant difference ($p < 0.05$). (B) Compared to the baseline model and ST-cGAN, the outputs of MT-cGAN showed reduced visual blurring and superior preservation of fine structures (marked by red arrows). 50

Figure 4-1: Off-resonance artifacts simulation process. The field map was simulated by combining random polynomial and Gaussian functions, and the concomitant field map was calculated based on the scan parameters. The k -space trajectory and corresponding gradient waveforms were generated and used for computing the time maps. Least-squares fitting was

adopted to estimate the parameters in mapping the time maps to a quadratic function. The final blurred image with off-resonance artifacts was obtained by MFI. 55

Figure 4-2: Architecture for AutofocusNet. The input blurred image is demodulated at multiple frequencies to produce a series of images, which are concatenated and fed into the network. The network consists of three residual blocks. The final layer is a 1×1 CConv to generate the deblurred output image. 58

Figure 4-3: Quantitative evaluation results and representative images on the simulated test dataset. (A) AutofocusNet achieved the better NRMSE, PSNR, and SSIM. Asterisks indicate a significant difference ($p < 0.05$). (B) The output of the autofocus method still showed blurring and artifacts in regions with high off-resonance frequencies, while the output of AutofocusNet showed sharper structures with reduced artifacts (marked by red arrows). 60

Figure 4-4: Comparison of model performance on phantom and in vivo data. Top-to-bottom rows: phantom image, in vivo head image with tagging lines to aid visualization of blurring, T1-weighted head image, T2-weighted head image. The autofocus method showed residual blurring and artifacts (marked by red arrows). The AutofocusNet achieved comparable results to the semiautomatic method. 61

Figure 5-1: Illustration of the proposed framework for undersampled CMR image reconstruction. The C^3 -Net alternates between the restoration step and DC step. The restoration part consists of k -space subnetwork and image subnetwork, which use complex U-Net as the backbone network architecture. 69

Figure 5-2: Comparison of one short-axis slice (end-diastole frame) generated using five reconstruction methods, from datasets retrospectively undersampled with accelerated ratios of 4 and 10. From left to right are images reconstructed from fully-sampled k -space as the reference,

zero-filling, L1-ESPIRiT, U-Net, C²-Net and C³-Net, respectively. From top to bottom are magnitude image (top), absolute difference image relative to the fully sampled reference (middle), and $x-t$ plots below the white lines (bottom). The error maps are windowed by scaling the image intensity by a factor of 5. The red arrows point to structures that show fine papillary muscles in ventricles and the blue arrows indicate preserved temporal fidelity. 73

Figure 5-3: Comparison of long-axis cardiac images from 2-chamber (top), 3-chamber (middle), and 4-chamber views (bottom), reconstructed using fully-sampled k -space as the reference, zero-filling, L1-ESPIRiT, U-Net, C²-Net, and C³-Net, respectively. Absolute difference images relative to the fully sampled reference are shown for performance comparison, as well. The error maps are windowed by scaling the image intensity by a factor of 5. 75

Figure 6-1: Nonrigid motion simulation with global affine transformation and local deformation. 79

List of Tables

Table 2-1: Computational cost of different algorithms for denoising one slice with size 320 × 320.	30
Table 3-1: Description of the parameters used in the motion simulation.	40
Table 3-2: MR acquisition parameters of local in-vivo T1-weighted and T2-weighted images.	45
Table 4-1: Description of the parameters used in the off-resonance simulation.	57
Table 5-1: Quantitative assessment of reconstruction performance on the CMRxRecon dataset.	72

Chapter 1: Introduction

1.1 MRI Overview

Magnetic Resonance Imaging (MRI) is a non-invasive imaging modality that enables visualization of anatomical information or monitoring of physiological functions without ionizing radiation. This attribute makes MRI a widely utilized technology for disease detection, diagnosis, and treatment planning. By adjusting the scan parameters, MRI provides the ability to produce multi-contrast images that highlight different tissue properties. Compared to other imaging modalities, such as X-ray or computed tomography (CT), MRI offers superior contrast for soft tissues. Since its first clinical application in 1980, MRI has become a powerful clinical tool in brain, musculoskeletal, abdominal, and cardiac imaging.

MRI scan involves a complex assembly of components, each serving an essential role in generating and manipulating the signal. The MRI scanner generates a strong magnetic field (B_0), typically through a superconducting magnet. This static field aligns the magnetic moments of protons (spins) in the body, creating a net magnetic moment along the direction of B_0 . The radiofrequency (RF) pulse is applied to tip the aligned spins away from the B_0 field direction and generate a detectable MR signal. Spatial localization is achieved by applying controlled gradient magnetic fields along the x , y , and z axes, which enable the slice selection, phase encoding, and frequency encoding during the image acquisition.

For 2D imaging, the total received MRI signal, ignoring relaxation terms, can be written as:

$$s(t) = \int_x \int_y m(x, y) e^{-i\gamma \left[\left(\int_0^t G_x(\tau) d\tau \right) x + \left(\int_0^t G_y(\tau) d\tau \right) y \right]} dx dy \quad (1-1)$$

where $m(x, y)$ is the object distribution, $G_x(t)$ and $G_y(t)$ are the applied gradients, and γ is the gyromagnetic ratio. From a frequency domain perspective, the signal can be rewritten as:

$$s(k_x, k_y) = \int_x \int_y m(x, y) e^{-i2\pi(k_x x + k_y y)} dx dy \quad (1-2)$$

where $k_x(t) = \frac{\gamma}{2\pi} \int_0^t G_x(\tau) d\tau$ and $k_y(t) = \frac{\gamma}{2\pi} \int_0^t G_y(\tau) d\tau$ are spatial frequencies in k -space. The k -space can be filled using different trajectories by controlling the encoding gradients. Common trajectories used in MRI include Cartesian, echo planar imaging (EPI), radial, and spiral, as shown in Figure 1-1. Once the k -space data is acquired, the MR image can be reconstructed by a fast Fourier transform (FFT) for Cartesian sampling, or a nonuniform FFT (NUFFT)¹ for non-Cartesian sampling.

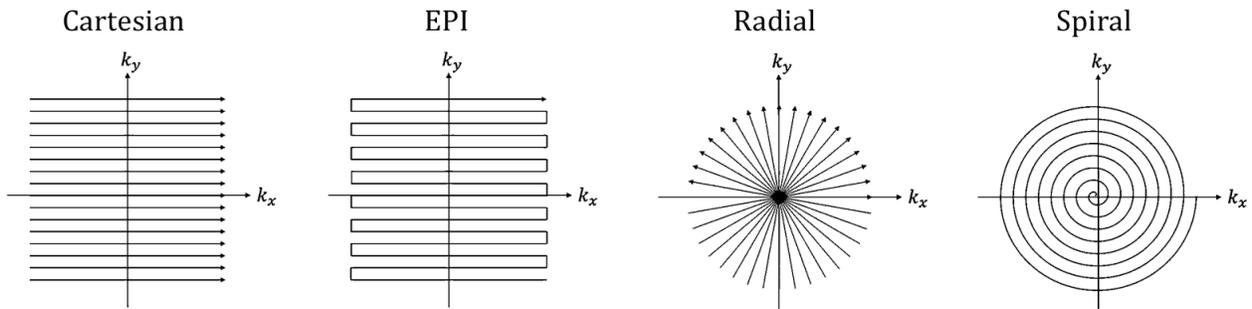


Figure 1-1: Commonly used trajectories in MRI.

One major disadvantage of MRI is the relatively slow imaging speed, since it acquires data in the frequency domain, rather than the image domain like optical imaging. Recent development of advanced acquisition and reconstruction methods dramatically accelerate the MRI scan. Parallel imaging techniques, such as SENSE² and GRAPPA³, reduce the imaging time by undersampling k -space data and leveraging information from multiple coils to remove the aliasing artifacts. On the other hand, compressed sensing techniques^{4,5} exploit mathematical sparsity, incoherent sampling, and nonlinear reconstruction to shorten the scan time.

1.2 MRI Artifacts

Like other imaging modalities, MRI is vulnerable to artifacts, affecting the image quality and diagnostic value. These artifacts can arise from various sources, including hardware limitations, imaging techniques, physiological motion, and tissue characteristics. Here, several common artifacts were introduced, focusing on signal-to-noise ratio (SNR) issues, motion artifacts, residual artifacts from parallel imaging, and off-resonance artifacts.

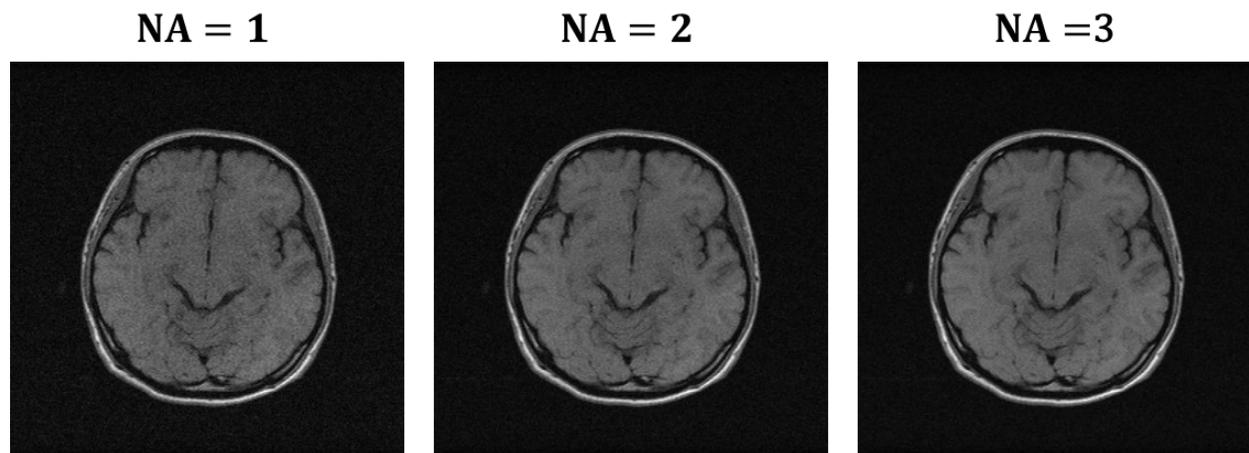


Figure 1-2: Improvement of SNR by increasing NA.

SNR is an important metric for image quality evaluation. SNR is proportional to spin polarization, which varies linearly with B_0 . At a fixed B_0 , the relationship between imaging parameters and SNR can be expressed as:

$$\text{SNR} \propto \Delta x \Delta y \Delta z \sqrt{\text{total readout interval}} \quad (1-3)$$

where Δx , Δy , and Δz is the spatial resolution along x , y , and z direction. This equation explains the trade-off between SNR, spatial resolution, and scan time. It is impossible to achieve high SNR while also maintain the resolution and scan time. To assess the SNR, region-of-interest (ROI) measurements are commonly performed on magnitude images by calculating the mean within a signal ROI, calculating the standard deviation in a background ROI, and applying a correction for

the Rayleigh distribution in the background⁶. In clinical practice, SNR can be improved by increasing the number of averages (NA), as shown in Figure 1-2. However, this results in a prolonged scan time, which may cause increased patient discomfort and higher burden on radiologists.

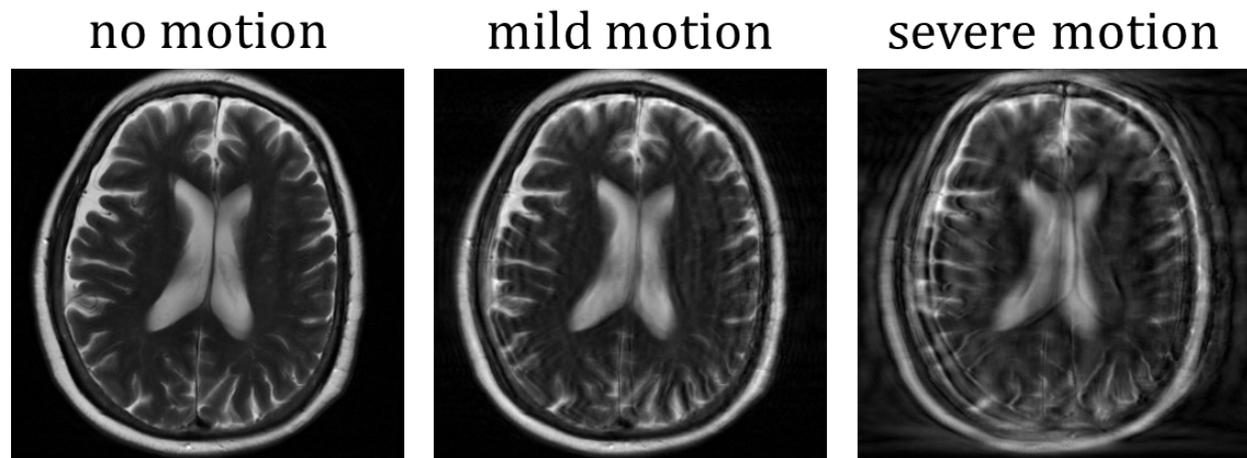


Figure 1-3: Motion artifacts in MRI.

Compared to other imaging modalities, such as CT or ultrasound, subject motion is a more challenging issue in MRI due to its relatively long scan time. During the data acquisition, subject movement causes inconsistencies in k -space data, leading to motion artifacts on the reconstructed image. Many different types of motion can degrade the image quality, including involuntary motion, such as cardiac or respiratory motion and vessel pulsation, or voluntary motion, such as swallowing or yawing. The appearance of motion artifacts is affected by the type of motion, the k -space trajectory, and the specifics of MR sequence. Common manifestations of motion artifacts include ghosting, blurring, and signal loss, as shown in Figure 1-3. Various strategies have been developed for motion prevention and compensation. Motion prevention methods include patient training, the use of foam restraints, and in some cases, sedation, which help to minimize voluntary motion during scanning⁷⁻¹⁰. Triggering and gating techniques are commonly used for cardiac and abdominal imaging, allowing to only collect data in specified low-motion intervals^{11,12}. During the

scan, the subject motion can be tracked by external sensors, such as optical cameras^{13,14}, or be estimated by the application of MR navigators¹⁵⁻²⁰. Motion information obtained from sensors or navigators can then be applied for either prospectively to adjust the scan²¹ or retrospectively to correct the motion artifacts²².

In clinical settings, parallel imaging is a powerful technique that enables faster data acquisition and reduced examination time. Accelerated scans are achieved by undersampling the k -space data and utilizing the spatial sensitivity information from multiple coils for image reconstruction. However, acquiring fewer data points in k -space results in a SNR loss by a factor of the square root of the undersampling ratio (R). The noise is further amplified by the coil geometry factor (g -factor). The g -factor is spatially variant and relates to geometry of surface coils. Figure 1-4 shows an example of SENSE reconstruction results with different undersampling ratios. Various methods to estimate or quantify the noise for parallel imaging have been developed to better evaluate different reconstruction strategies²³⁻²⁵.

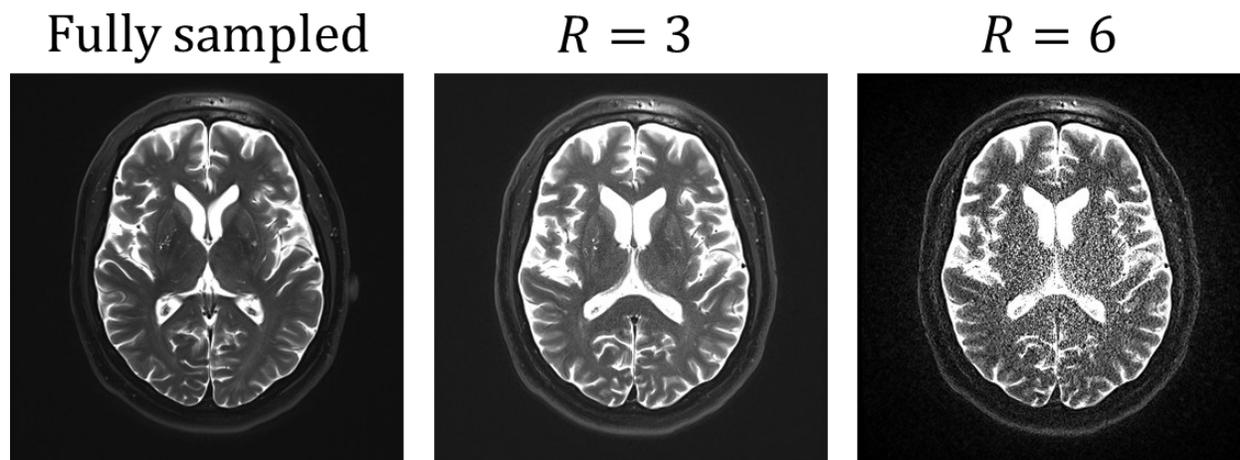


Figure 1-4: SENSE reconstruction results with different undersampling ratios. Adapted from [Questions and Answers in MRI Website](#).

Off-resonance artifacts in MRI primarily result from inhomogeneities in the B_0 field, causing different precession frequencies at different locations. The field inhomogeneities can arise from various sources, including magnetic susceptibility, chemical shift, and metal implants²⁶. Off-resonance artifacts typically manifest as signal loss, geometric distortions, and blurring in the resulted images, and are particularly problematic in sequences with long readouts, such as EPI and spiral. For most of existing off-resonance correction methods, it is necessary to acquire a field map. A common way measuring the field map is to acquire two images at different echo times and calculate the phase difference between them. This phase difference is used to estimate the off-resonance frequency at each voxel:

$$\Delta\omega_0(x, y) = \frac{\angle\{m_1^*(x, y)m_2(x, y)\}}{\Delta TE} \quad (1-4)$$

where the image $m_1(x, y)$ is acquired at TE_1 , image $m_2(x, y)$ is acquired at TE_2 , and $\Delta TE = TE_2 - TE_1$. After getting the field map, the conjugate phase reconstruction (CPR) can be applied to correct the off-resonance artifacts. However, exact CPR is extremely slow, and multiple efficient implementations have been developed to approximate the exponential term:

$$e^{i\Delta\omega_0(x, y)t} \approx \sum_l^L b_l(t)c_l(x, y) \quad (1-5)$$

where b_l and c_l are the basis functions in k -space and image domain. According to the choice of b_l and c_l , these algorithms can be roughly categorized into time-segmented approximation^{27,28} and frequency-segmented approximation²⁹. Figure 1-5 shows an example of a blurred phantom image in spiral imaging reconstructed with multifrequency interpolation (MFI). One major limitation of CPR and its approximations is the assumption of a smooth field map. Iterative conjugate gradient reconstruction^{28,30} shows better performance in dealing with abrupt variations of $\Delta\omega_0$ in the field map. The automatic deblurring method (autofocus)^{31,32} offers an alternative for off-resonance

correction without requiring a field map. Autofocus can estimate a field map based on the blurred image by examining some image features, such as the imaginary component of the image.

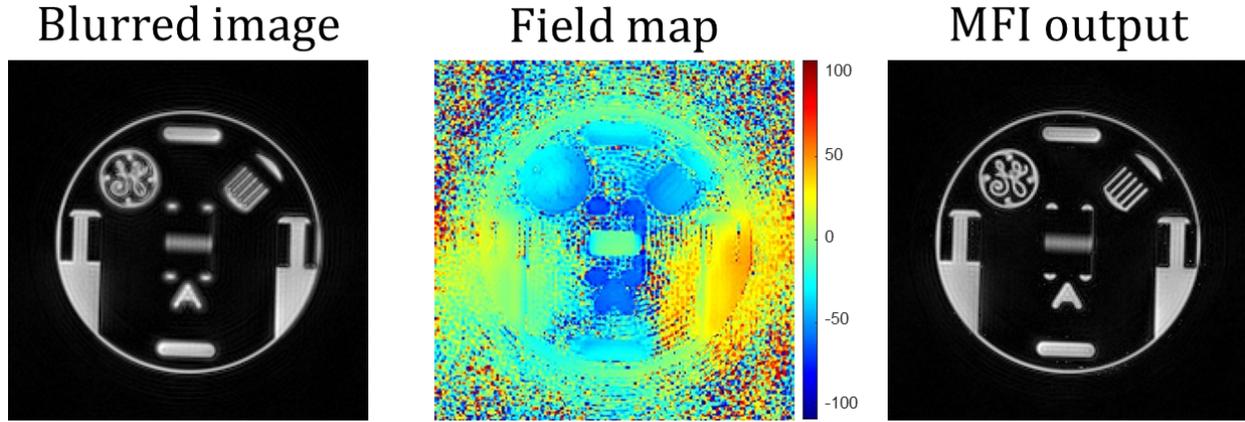


Figure 1-5: Blurred phantom image reconstructed with MFI.

Concomitant fields are another source of off-resonance artifacts. As implied by Maxwell's equations, any applied linear gradients during the scan induces additional spatially varying magnetic fields. Failing to correct the phase accumulation caused by concomitant fields can lead to inadequate off-resonance correction. As described by Bernstein et al³³, the concomitant field to the lowest order can be calculated by:

$$B_c \approx \left(\frac{G_z^2}{8B_0} \right) (X^2 + Y^2) + \left(\frac{G_x^2 + G_y^2}{2B_0} \right) Z^2 - \left(\frac{G_x G_z}{2B_0} \right) XZ - \left(\frac{G_y G_z}{2B_0} \right) YZ \quad (1-6)$$

where G_x , G_y , and G_z are the time-dependent gradients, and X , Y , and Z are the scanner spatial coordinates. After calculating the concomitant fields, the B_0 inhomogeneity correction algorithms can also be extended for correcting concomitant fields.

1.3 Deep Learning Overview

Deep learning is a branch of artificial intelligence and machine learning. Deep learning focuses on training artificial neural networks with multiple layers to extract features and generate output based

on large datasets. The neural networks are typically trained through backpropagation, where the model iteratively adjusts its parameters to minimize some objective function. In recent years, different network architectures have been proposed and optimized for various data types and applications. Convolutional neural networks (CNNs) are widely used in computer vision and image processing due to their ability to learn spatial hierarchies of features in data. Unlike traditional fully-connected networks, CNNs are highly efficient, as they reduce the number of parameters by sharing weights through the convolution operation. A CNN typically consists of dozens or hundreds of convolutional layers, depending on the complexity of the intended purpose. By combining convolutional layers with other building blocks, such as nonlinear activation functions, pooling layers, and fully-connected layers, the model is able to both detect low-level image features, such as edges and shapes, and extract more abstract high-level features that are more abstract and task-specific.

Generative adversarial networks (GANs)³⁴ are a class of deep learning models that are originally proposed for generating new data that closely resemble training data. GANs consist of two neural networks, where the generator network aims to produce realistic high-quality data that can “fool” the discriminator, and the discriminator network is trained to maximize its ability to distinguish between real and synthetic data. The training of GANs involves alternating between training the generator and discriminator in an adversarial setup. Conditional GAN (cGAN)³⁵ is a variant of GAN that takes advantage of an additional conditioning input. By incorporating the conditional information, cGANs provide controlled data generation, which is particularly useful in image-to-image translation tasks, like super-resolution and image restoration³⁶.

Recently, Transformers have become one of the most influential architectures in deep learning, significantly advancing natural language process. The core idea behind Transformers is

the self-attention mechanism³⁷, which allows the model to capture long-distance contextual relationships. Transformers have also been extended to the computer vision field by dividing the images into patches and treating all patches as a sequence³⁸.

The deep learning techniques have also gained significant traction in MRI image processing, bringing advancements across a wide range of tasks. The deep-learning-based image reconstruction methods drastically reduce the reconstruction time leveraging the implicit priors learned from the training data³⁹⁻⁴¹. The deep learning models have also been applied to MRI super-resolutions⁴²⁻⁴⁴, which reconstruct high-resolution images from low-resolution inputs. By training neural networks with multi-contrast MRI images, the models show impressive performance on synthesizing missing contrasts without extra acquisitions⁴⁵⁻⁴⁷. Additionally, deep learning offers efficient and effective solutions to reduce or correct the artifacts in MR images^{26,48-50}.

1.4 Dissertation Outline

This dissertation includes six chapters. Chapter 1 presents an overview of MRI and deep learning. It first describes the principles of MRI and its acquisition and reconstruction, leading to a discussion of common artifacts encountered in MRI. A brief introduction to deep learning and its applications in MRI is then followed.

Chapter 2 introduces a non-blind deep complex-valued convolutional neural network (CDnCNN) for denoising MRI images. The network is trained to boost the SNR of MRI images and address spatially varying noise, especially significant for low-field MRI applications, where SNR is a primary limitation. This chapter covers the complex-valued operations, network architecture, training strategies, and performance evaluation on simulated and in vivo datasets, showing that CDnCNN provides superior denoising performance over other comparison methods.

Chapter 3 presents a multi-task conditional GAN (MT-cGAN) designed for simultaneous motion detection and correction. The model uses multi-task learning to leverage shared representations for motion detection and compensation tasks, enhancing its robustness and performance compared to single-task models. This chapter covers the comprehensive simulation process to emulate motion artifacts under realistic clinical settings, multi-task framework, combined objective function, and evaluation on simulated and in vivo datasets.

Chapter 4 covers a deep-learning-based method for correcting both field inhomogeneity and concomitant field artifacts without requiring the field map acquisition or concomitant field computation. The network is trained using synthetic data with simulated off-resonance artifacts. The evaluation results demonstrate that the proposed method improve the image quality and reduce the off-resonance artifacts on both simulated and in vivo data.

Chapter 5 is focused on developing a complex-valued cascading cross-domain network (C³-Net) for reconstructing undersampled cardiac MRI. C³-Net utilizes the unrolled network structure, alternating between cross-domain learning and data consistency steps. Detailed descriptions of network architecture, training strategies, and experimental results are covered.

Chapter 6 summarizes the work of this dissertation and provides some potential directions for future research, including integrating these models within MRI reconstruction systems for real-time processing, non-rigid motion compensation, exploring novel deep learning architectures for enhanced performance, and developing explainable and robust models.

Chapter 2: MRI Denoising with a Non-Blind Deep Complex-Valued Convolutional Neural Network

2.1 Introduction

SNR is crucial for MR image analysis. High-SNR images enable better visualization of small structures, which not only facilitates the human interpretation, but also benefits subsequent processing techniques such as registration and segmentation. The SNR is intricately linked to various imaging conditions, including field strength, image resolution, and number of averages (NA). The application of parallel imaging and innovative reconstruction techniques can also impact the spatial distribution of noise. Increasing NA is a common method to boost the SNR, but it results in a longer scan time. On the other hand, low-field MRI has grown in popularity in recent years⁵¹⁻⁵⁴ due to advancements in hardware techniques and acquisition strategies. Because low-field MRI scanners are substantially less expensive to purchase and install, they enable MRI to reach underserved populations worldwide. However, images acquired on low-field scanners inherently have low SNR due to the low Boltzmann polarization, which hinders their clinical application. Therefore, efficient and effective denoising is important for increasing MRI availability in the clinical setting.

The noise in a complex MR image is typically modeled as a complex additive white Gaussian noise (AWGN), with zero mean and equal variance for real and imaginary parts. Thus, the pixel intensity in the noise-corrupted magnitude image follows the Rician distribution^{55,56}. Numerous MRI denoising methods have been proposed based on decades of research. The non-local means (NLM) algorithm relies on the non-local similarity to remove Gaussian noise⁵⁷⁻⁵⁹.

Wiest-Daesslé et al⁶⁰ adapted the NLM algorithm for Rician noise and applied it to diffusion tensor MRI. Coupé et al⁶¹ reduced the time complexity of NLM through blockwise implementation and parallel computation. The block-matching and 3D filtering (BM3D) takes advantage of the enhanced sparsity in the transform domain and uses collaborative filtering to remove noise⁶²⁻⁶⁴. In recent years, deep-learning-based approaches have shown great success in image denoising. A convolutional neural network (CNN) learns to restore the clean image by training it with a large number of noise-corrupted and ground-truth image pairs. Zhang et al⁶⁵ combined residual learning and batch normalization in their DnCNN model for Gaussian denoising with unknown noise level (i.e., blind denoising). Later, Zhang et al⁶⁶ demonstrated that incorporating the noise level information into the network (i.e., non-blind denoising) increased its generalizability. Quan et al⁶⁷ investigated the potentials of complex-valued CNNs for natural image denoising. Manjón et al⁶⁸ proposed a two-stage method that combines blind CNN denoising with the NLM algorithm. Tripathi et al⁶⁹ employed encoder-decoder structure and residual learning scheme for removing Rician noise from magnitude MR images. Li et al⁷⁰ used a progressive learning strategy, cascading two sub-networks for crude and refinement noise estimation, respectively. Tian et al⁷¹ developed an MRI denoising method based on the conditional generative adversarial network (GAN). Koonjoo et al⁷² presented an end-to-end DL-based noise-robust reconstruction method for low-field MRI data.

However, the vast majority of existing MRI denoising methods do not fully exploit the complex-valued feature of MRI data. The input and output of conventional model-based approaches, such as NLM and BM3D, are both magnitude images, and phase information is ignored. CNN-based approaches typically treat the real and imaginary parts as two separate real-valued channels, similar to the RGB channels of a color image, which may limit the network's

ability to extract features from complex-valued data. Recently, complex-valued CNNs have received increased attention. Complex-valued CNNs have several advantages over real-valued CNNs, including easier optimization, faster learning, and richer representational capacity⁷³⁻⁷⁶. Wang et al⁷⁷ used the residual learning strategy to accelerate the convergence of their complex-valued CNN. El-Rewaidy et al⁷⁸ applied a complex-valued CNN to reconstruct highly under-sampled cardiac MRI data. Cole et al⁷⁹ evaluated the performance of complex-valued CNNs on phase-related MRI applications by systematically analyzing the impact of different model design choices. All of these works concentrated on the MRI reconstruction task, and to the best of our knowledge, no previous work attempted to use complex-valued CNN in the MRI denoising task.

Therefore, in this chapter, we designed and implemented non-blind CDnCNN, a complex-valued CNN for non-blind MRI denoising. The noise level map was estimated from the noise-corrupted image and fed into the network. Complex-valued building blocks were used throughout the network. We trained the network on simulated data and tested it on both simulated and in vivo data. We compared our method to several other denoising algorithms, both quantitatively and qualitatively. In the comparison, both magnitude and phase performance were evaluated. We also investigated the role of the noise level map in dealing with parallel imaging noise that varies spatially.

2.2 Methods

2.2.1 Training datasets

For supervised neural network training, ground truth images are required. Assuming a noise-free image $m(\mathbf{x})$, we generated the noise-corrupted image $m'(\mathbf{x})$ by applying random, complex AWGN to $m(\mathbf{x})$:

$$m'(\mathbf{x}) = m(\mathbf{x}) + n(\mathbf{x}; \sigma^2) \quad (2-1)$$

where $n(\mathbf{x}; \sigma^2) = n_r(\mathbf{x}; \sigma^2) + j \cdot n_i(\mathbf{x}; \sigma^2)$ is random, complex AWGN with zero mean and variance σ^2 . However, ideal noise-free images do not exist. The training and validation datasets used in this work were built from the fastMRI brain dataset (<https://fastmri.med.nyu.edu/>)⁸⁰. The raw fastMRI dataset contains nearly 7000 fully sampled multi-coil brain MRIs obtained on 1.5T or 3T scanners, comprising axial T1-weighted, post-contrast T1-weighted, T2-weighted and FLAIR images. Detailed descriptions can be found on the project website. A reference study has indicated that the SNR at 0.55T low-field scanners is approximately 70% of that at 1.5T scanners⁸¹. We randomly selected 2000 T2-weighted imaging volumes for training the model and selecting model hyperparameters. We strategically reversed the other three modalities for testing the denoising performance across a diverse range of imaging conditions and contrasts. Single-coil data was reconstructed from multi-coil data through the use of an adaptive combination method⁸², in which the complex-valued coil sensitivities were estimated from a local matched filter. The adaptive combination method achieved near optimal SNR while retaining the phase information. The reconstructed images were then center cropped to have matrix size 320×320 . During the noise simulation, each image was normalized to have its magnitude between 0 to 1 and its phase unchanged. The noise standard deviation σ was sampled from a uniform distribution between 0 to 0.1.

2.2.2 Network architecture

A DnCNN⁶⁵ was used as our backbone network structure. The original DnCNN model was proposed for blind Gaussian denoising of natural images and achieved state-of-the-art performance. The network was designed to use complex-valued operations rather than splitting the real and imaginary components into two separate channels. Figure 2-1 shows the proposed non-blind

CDnCNN for MRI denoising. The input to the network is a 2D complex-valued MR image concatenated with a tunable complex-valued noise level map. The non-blind CDnCNN consists of a series of complex-valued convolution blocks. Three types of operations were adopted in each block: complex-valued convolution (CConv), radial batch normalization (BN), and complex-valued rectified linear unit (CReLU). The CConv operation between the input $\mathbf{d} = \mathbf{a} + j \cdot \mathbf{b}$ and filter $\mathbf{w} = \mathbf{x} + j \cdot \mathbf{y}$ can be accomplished by four real-valued convolutions³⁰:

$$\mathbf{w} \circledast \mathbf{d} = (\mathbf{x} \circledast \mathbf{a} - \mathbf{y} \circledast \mathbf{b}) + j \cdot (\mathbf{y} \circledast \mathbf{a} + \mathbf{x} \circledast \mathbf{b}) \quad (2-2)$$

where \circledast represents the convolution operation. Batch normalization is an important operation to expedite training and stabilize model performance⁸³. We adopted radial BN, which maintains the phase information while scaling the magnitude⁷⁸:

$$\mathbf{m}_{\text{BN}} = \left(\frac{\mathbf{m} - \mu_{\mathbf{m}}}{\sqrt{\sigma_{\mathbf{m}}^2}} \right) \gamma + \beta + \tau \quad (2-3)$$

$$\mathbf{d}_{\text{BN}} = \mathbf{m}_{\text{BN}} e^{j\theta} \quad (2-4)$$

where $\mathbf{d} = \mathbf{m} e^{j\theta}$ is the input expressed in its polar form, $\mu_{\mathbf{m}}$ and $\sigma_{\mathbf{m}}^2$ are the mean and variance of \mathbf{m} , β and γ are trainable parameters, and τ is a constant to ensure that the normalized \mathbf{m}_{BN} is positive (empirically set to 1). The CReLU function separately activates the real and imaginary components of the input⁷⁶:

$$\text{CReLU}(\mathbf{d}) = \text{ReLU}(\mathbf{a}) + j \cdot \text{ReLU}(\mathbf{b}) \quad (2-5)$$

The first block of the network is composed of a CConv and a CReLU, the middle blocks are composed of a CConv, a radial BN and a CReLU, and the last block is composed of a CConv to produce the output image. In order to balance the denoising performance and computational efficiency, all of the CConv kernels have a size of 3×3 and a number of channels of 64, and the

network has 20 convolution blocks. The effect of network depth on denoising performance was explored.

The network was implemented in the open-source machine learning library PyTorch⁸⁴ and trained with an L1 loss:

$$\mathcal{L}(\Theta) = \frac{1}{N} \sum_{i=1}^N |\mathbb{C}\text{DnCNN}(m'(\mathbf{x}_i); \Theta) - m(\mathbf{x}_i)| \quad (2-6)$$

where $m'(\mathbf{x}_i)$ is the noise-corrupted image, $m(\mathbf{x}_i)$ is the ground truth image, N is the total number of training pairs, and Θ represents trainable parameters in the network. The optimization was carried out by an Adam optimizer⁸⁵ with an initial learning rate of 0.0001 and momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The training batch size was fixed to 32. Random flips and random cropping were employed as training augmentation to reduce the possibility of overfitting and improve the model robustness.

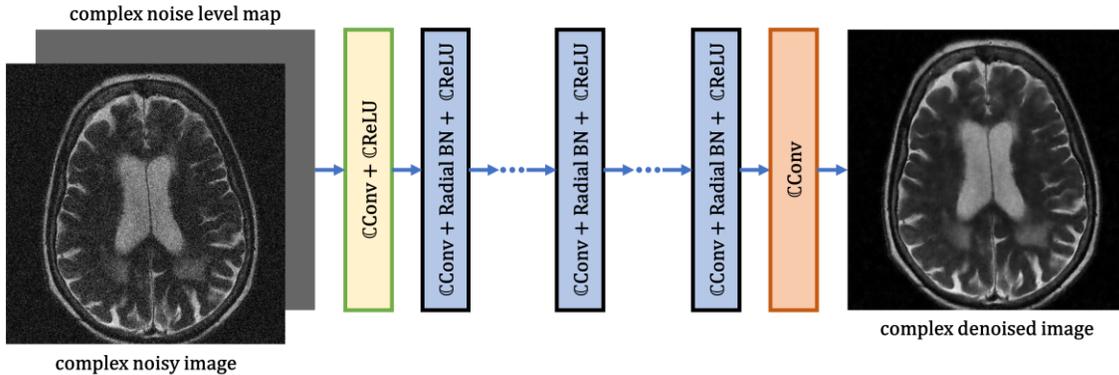


Figure 2-1: Architecture of the non-blind $\mathbb{C}\text{DnCNN}$ for MRI denoising. The input is the complex-valued noisy image concatenated with the complex-valued noise level map, and the output is the complex-valued denoised image. The network consists of twelve 3×3 $\mathbb{C}\text{Conv}$. Each $\mathbb{C}\text{Conv}$ is followed by a radial BN and a $\mathbb{C}\text{ReLU}$ except for the first and last.

2.2.3 Noise level map

To cope with images at different noise levels, specifying the noise standard deviation σ is required for most conventional model-based denoising techniques, such as NLM and BM3D. In practice, the noise standard deviation $\hat{\sigma}$ can be estimated from the k -space or image data. We adopted a commonly-used wavelet-based approach⁸⁶ in this work. Specifically, we used the `estimate_sigma` function from the `scikit-image` Python package⁸⁷. In order to incorporate this information into our CDnCNN for MRI denoising, we built a complex-valued noise level map with its size matching the input MR image and concatenated it with the input image. All pixels in the noise level map were set to $\hat{\sigma}_{\text{avg}} + j \cdot \hat{\sigma}_{\text{avg}}$, where $\hat{\sigma}_{\text{avg}}$ was the average of $\hat{\sigma}$'s estimated from the real and imaginary parts of the input image:

$$\hat{\sigma}_{\text{avg}} = \frac{\text{estimate_sigma}(\text{Real}\{m'(\mathbf{x})\}) + \text{estimate_sigma}(\text{Imag}\{m'(\mathbf{x})\})}{2} \quad (2-7)$$

Since CDnCNN is fully convolutional, it inherently provides the flexibility to deal with spatially non-uniform noise. For parallel MRI, such as sensitivity encoding (SENSE)² and generalized autocalibrating partially parallel acquisitions (GRAPPA)³, the noise is amplified by the geometry factor (g -factor). The g -factor is determined by the coil geometry and changes across the image. By weighting the uniform noise level map with a g -factor map, the network is capable of handling the spatially varying noise.

2.2.4 Evaluation

We first created a simulated testing dataset from the fastMRI brain dataset to evaluate the performance of the proposed denoising method. To avoid overlap between the training and testing subsets, another 200 T2-weighted imaging volumes were chosen at random. Simulated complex

AWGN with zero mean and standard deviation between 0 to 0.1 was added to the testing data. We compared our method to other denoising algorithms including NLM, BM3D (<https://webpages.tuni.fi/foi/GCF-BM3D/>), real-valued DnCNN, CDnCNN without noise level map (blind). The NLM and BM3D algorithms operated on the magnitude images, while the CDnCNN and non-blind CDnCNN operated on the complex-valued images, and the real-valued DnCNN treated the real and imaginary components as two separate channels. For quantitative assessment, the normalized root-mean-square error (NRMSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM)⁸⁸ of the magnitude images were calculated. In the following definitions, x denotes the output image with size $m \times n$, y denotes the reference image with the same size, and $|\cdot|$ means taking the magnitude.

$$\text{NRMSE}(|x|, |y|) = \sqrt{\frac{\text{MSE}(|x|, |y|)}{\text{MSE}(|y|, 0)}} \quad (2-8)$$

$$\text{PSNR}(|x|, |y|) = 20 \log_{10} \left(\frac{\max(|x|)}{\sqrt{\text{MSE}(|x|, |y|)}} \right) \quad (2-9)$$

$$\text{SSIM}(|x|, |y|) = \frac{(2\mu_{|x|}\mu_{|y|} + c_1)(2\sigma_{|x||y|} + c_2)}{(\mu_{|x|}^2 + \mu_{|y|}^2 + c_1)(\sigma_{|x|}^2 + \sigma_{|y|}^2 + c_2)} \quad (2-10)$$

where $\text{MSE}(|x|, |y|) = \frac{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (|x_{ij}| - |y_{ij}|)^2}{mn}$ is the mean-square error between $|x|$ and $|y|$, $\mu_{|x|}$ and $\mu_{|y|}$ are the means of $|x|$ and $|y|$, $\sigma_{|x|}^2$ and $\sigma_{|y|}^2$ are the variances of $|x|$ and $|y|$, $\sigma_{|x||y|}^2$ is the cross-correlation of $|x|$ and $|y|$, and $c_1 = 0.01$ and $c_2 = 0.03$ are regularization constants. Both $|x|$ and $|y|$ were normalized with respect to $|y|$ before metric calculation. For three CNN-based methods generating complex-valued output images, the absolute difference (ABSD) of the phase images was also calculated to test whether the phase information was altered:

$$\text{ABSD}(\angle x, \angle y) = \frac{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |\angle x_{ij} - \angle y_{ij}|}{mn} \quad (2-11)$$

where $\angle \cdot$ means taking the phase. For a fair comparison between different CNN-based methods, all CNNs had the same number of trainable parameters. Manual tuning was performed to determine the optimal combination of hyperparameters for each model.

The role of the noise level map in dealing with spatially non-uniform parallel imaging noise was also explored. We first utilized the fully sampled multi-coil k -space data from the raw fastMRI dataset to generate the coil sensitivity map for g -factor calculation. Complex AWGN was then added to each coil. We assume, for simplicity, that the standard deviation of Gaussian noise at each coil has the same value and there is no correlation across coils²⁴. The noise-corrupted k -space data was subsampled and reconstructed using SENSE. To address the spatially varying noise, the uniform noise level map was weighted by the g -factor map. In practice, the coil sensitivity map and g -factor map can be acquired during the preparation phase with no additional scan time cost.

To test the generalizability of non-blind CDnCNN, its performance on different contrasts and anatomies was also assessed. T1-weighted and FLAIR brain images and proton density (PD) weighted knee images from the fastMRI dataset were randomly selected to form a testing dataset out of the training distribution. Additionally, the network was applied to the low-field M4Raw dataset⁸⁹ and local in vivo data collected on a prototype 0.55 T MR scanner with high-performance gradients (ramped-down MAGNETOM Aera, Siemens Healthcare, Erlangen, Germany). The M4Raw dataset contains multi-contrast and multi-coil MRI data collected using a 0.3 T MR system. Detailed description can be found on the project website (<https://github.com/mylyu/M4Raw>). We randomly selected 5 T2-weighted volumes from the dataset for testing the network performance. We also applied retrospective undersampling and SENSE reconstruction on the multi-coil data to test the network performance on spatially varying noise. The local in vivo data was acquired with a SPRING-RIO TSE sequence⁹⁰ with TR/TE_{eff} ,

3000/105 ms; echo train length, 9; FOV, $230 \times 230 \text{ mm}^2$; voxel size, $0.70 \times 0.70 \times 4 \text{ mm}^3$. Written informed consent from all subjects were given. To evaluate the image quality without reference, the SNR for white matter (WM) and gray matter (GM) were computed based on manually defined regions of interest (ROIs):

$$\text{SNR} = 0.66 \times \frac{\mu_{\text{ROI}}}{\sigma_{\text{air}}} \quad (2-12)$$

where μ_{ROI} is the mean intensity of the ROI, σ_{air} is the standard deviation of the air region, and 0.66 is the correction factor for the Rayleigh distribution of the noise in the magnitude image⁶.

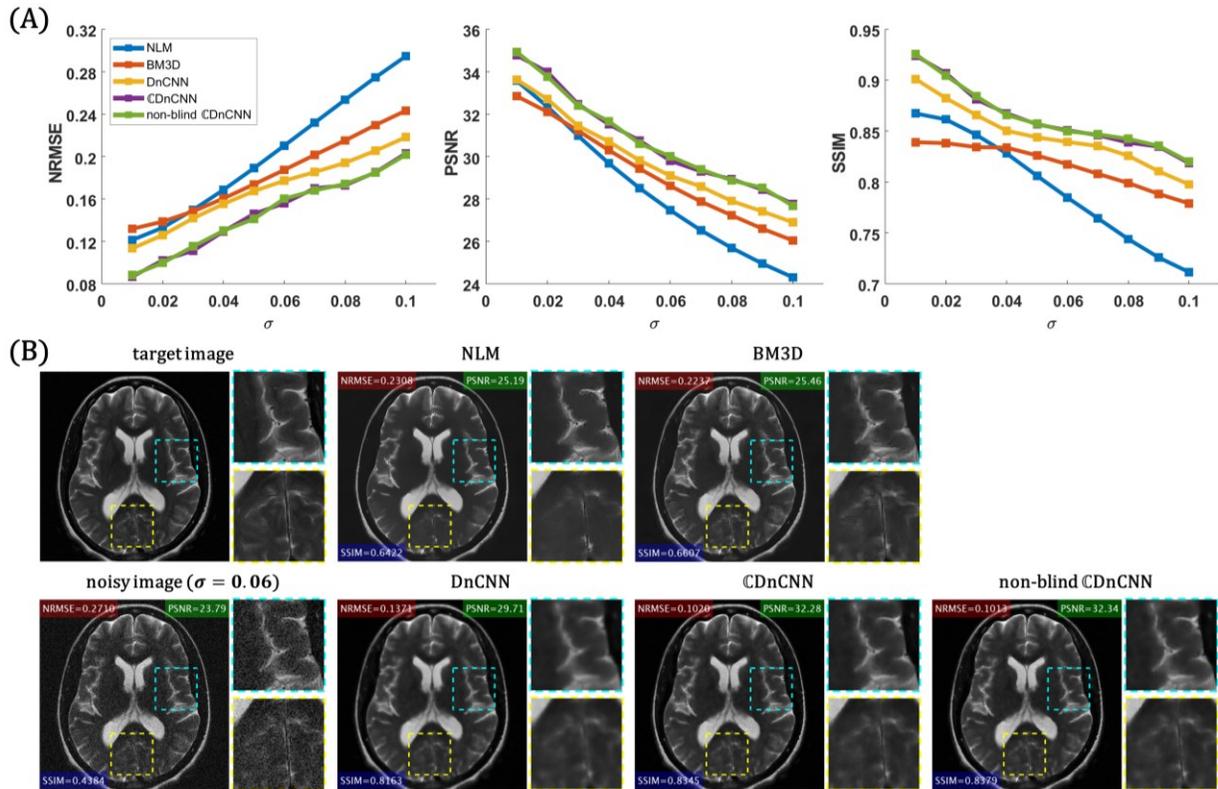


Figure 2-2: (A) Performance of different denoising methods on the simulated testing dataset at different noise levels. The average NRMSE, PSNR, and SSIM were evaluated for each method. (B) Representative magnitude image results of different denoising methods. Top left is the ground truth image, and bottom left is the simulated noise-corrupted image with a noise standard deviation of 0.06. The output of CDnCNN and non-blind CDnCNN showed reduced noise and less visual blurring compared to other methods.

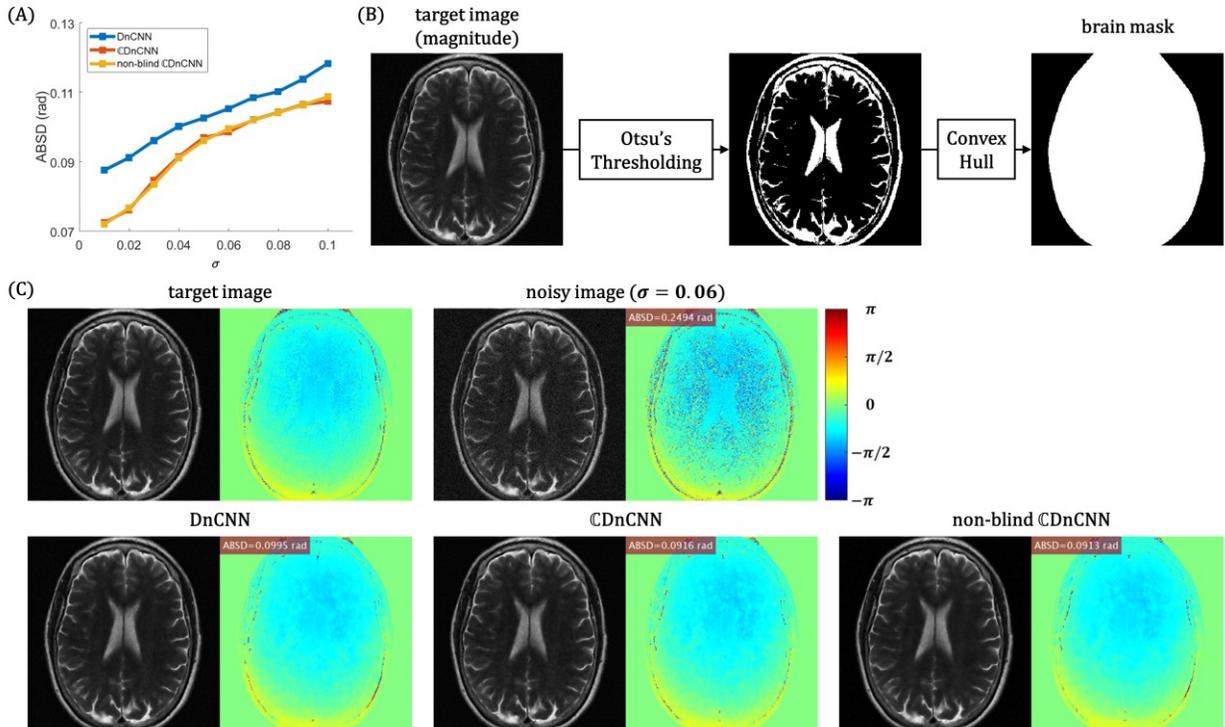


Figure 2-3: (A) Phase difference of DnCNN, CDnCNN, and non-blind CDnCNN on the simulated testing dataset at different noise levels. The average phase ABSD was evaluated for each network. (B) Procedures for generating the brain mask to eliminate the impact of background random phase in metric calculation. (C) Representative phase image results of different methods. Top left is the ground truth image, and top center is the simulated noise-corrupted image with a noise standard deviation of 0.06. Bottom row is the output of DnCNN, CDnCNN, and non-blind CDnCNN, from left to right. The complex-valued models preserved the phase information better than the real-valued model.

2.3 Results

Figure 2-2A shows the performance of different denoising algorithms on the simulated noise-corrupted dataset at different noise levels. When the noise standard deviation σ was larger than 0.04, three CNN-based algorithms outperformed NLM and BM3D. Compared to real-valued DnCNN with two-channel input, the output of CDnCNN and non-blind CDnCNN showed superior NRMSE, PSNR and SSIM over the entire range of 0 to 0.1. Representative images are displayed in Figure 2-2B. Compared to other methods, the output of CDnCNN and non-blind CDnCNN

showed reduced noise and less visual blurring. Figure 2-3A shows the phase difference for CNN-based methods. To eliminate the impact of random phase in the background, a mask covering the brain region was generated from the reference magnitude image using Otsu's thresholding⁹¹ and convex hull operation, as shown in Figure 2-3B. Representative phase images are displayed in Figure 2-3C. The phase of the output images did not deviate significantly from the reference phase. \mathbb{C} DnCNN and non-blind \mathbb{C} DnCNN showed comparable performance and outperformed the real-valued DnCNN in terms of preserving the phase information.

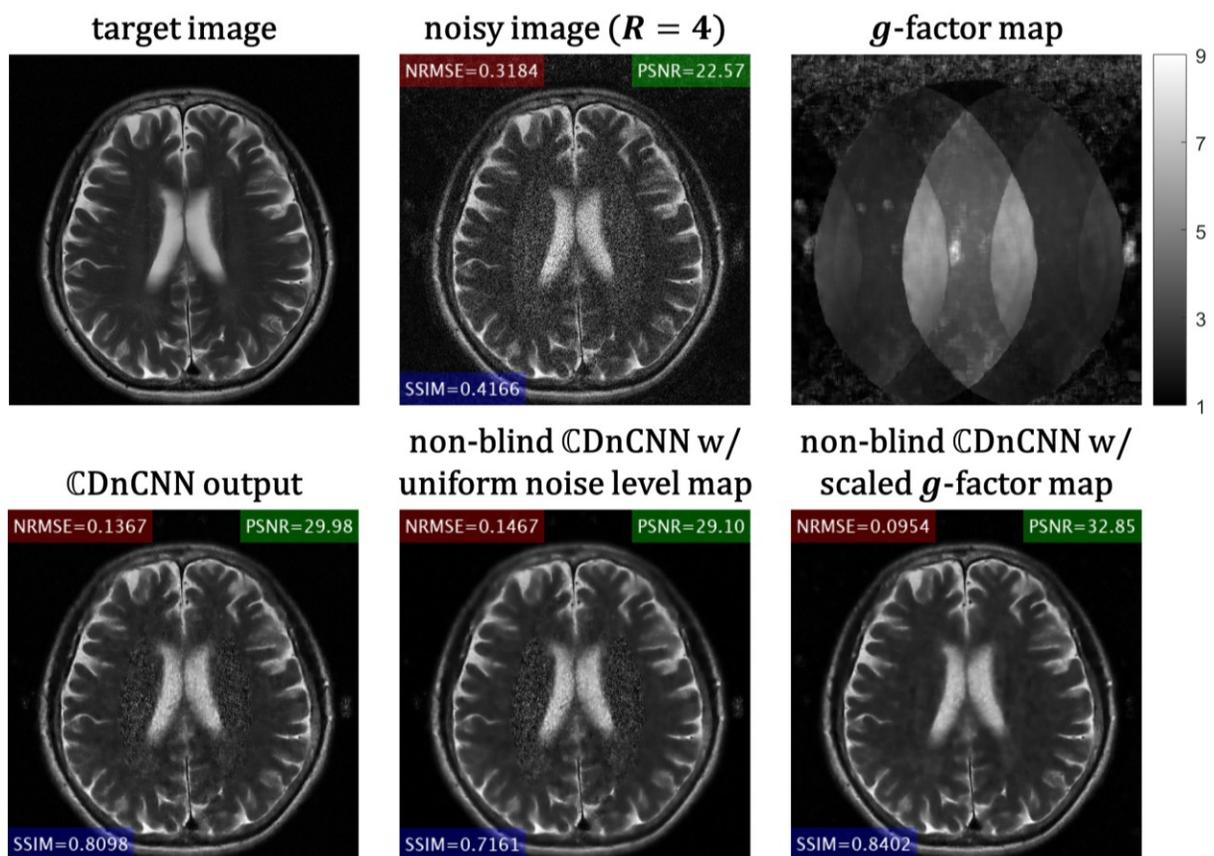


Figure 2-4: Example of non-blind \mathbb{C} DnCNN on spatially varying parallel imaging artifacts. Top left is the ground truth image. Top center is the simulated noise-corrupted image reconstructed by SENSE with a subsampling ratio of 4. Top right is the g -factor map from the SENSE reconstruction. Bottom left is the output of \mathbb{C} DnCNN without noise level map. Bottom center is the output of non-blind \mathbb{C} DnCNN with a uniform noise level map. Bottom right is the output of non-blind \mathbb{C} DnCNN with the scaled g -factor map as the noise level map. The non-blind \mathbb{C} DnCNN with the scaled g -factor map successfully reduced the noise at the center regions with large g -factor.

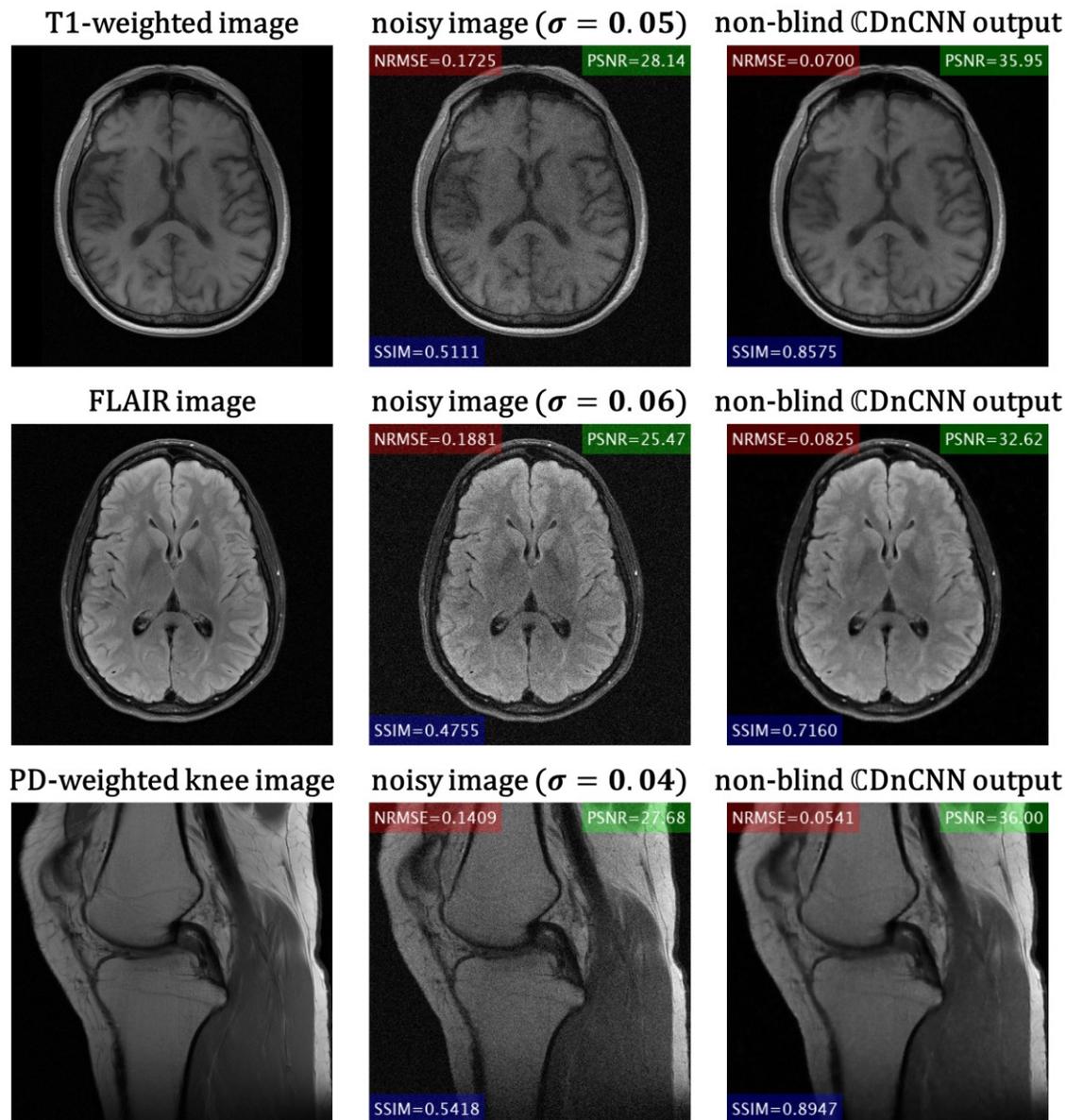


Figure 2-5: Performance of non-blind CDnCNN on modalities or anatomies out of the training distribution. The top row is a T1-weighted brain image, the middle row is a FLAIR brain image, and the bottom row is a PD-weighted knee image. From left to right, each column is the ground truth image, the simulated noise-corrupted image, and the output of non-blind CDnCNN, respectively.

Figure 2-4 gives an example showing the network performance on spatially varying parallel imaging artifacts. The raw k -space data from 16 coils was retrospectively undersampled by a factor of 4. The noise in the SENSE reconstructed image was amplified by the g -factor due

to the coil geometry. With scaled g -factor map as the noise level map, non-blind CDnCNN successfully reduced the noise and showed the greatest image quality, whereas other methods failed at the center regions with large g -factor.

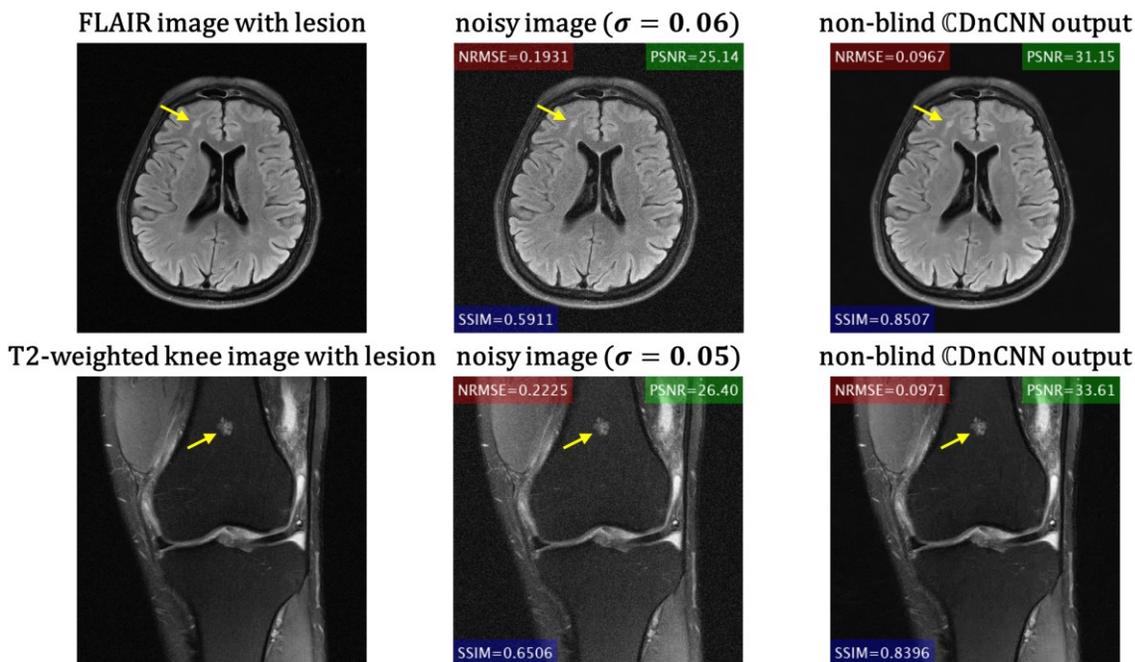


Figure 2-6: Performance of non-blind CDnCNN on images with pathology. The top row is a FLAIR brain image with a white matter lesion (yellow arrow), and the bottom row is a knee image with bone lesion (yellow arrow). From left to right, each column is the ground truth image, the simulated noise-corrupted image, and the output of non-blind CDnCNN, respectively.

Figure 2-5 shows the generalizability of non-blind CDnCNN on data out of the training distribution. The testing images had contrasts and anatomies that were different from the training dataset. The output images were less noisy and showed superior metrics, demonstrating that the model was able to generalize under these circumstances and did not overfit to the training data. To further test the generalizability of non-blind CDnCNN, we used the pathology annotations from the fastMRI+ dataset. Figure 2-6 shows the network performance on images with lesion. The proposed method effectively reduced the noise without compromising the structural details in

small pathological regions. The output images showed superior metrics compared to the input images.

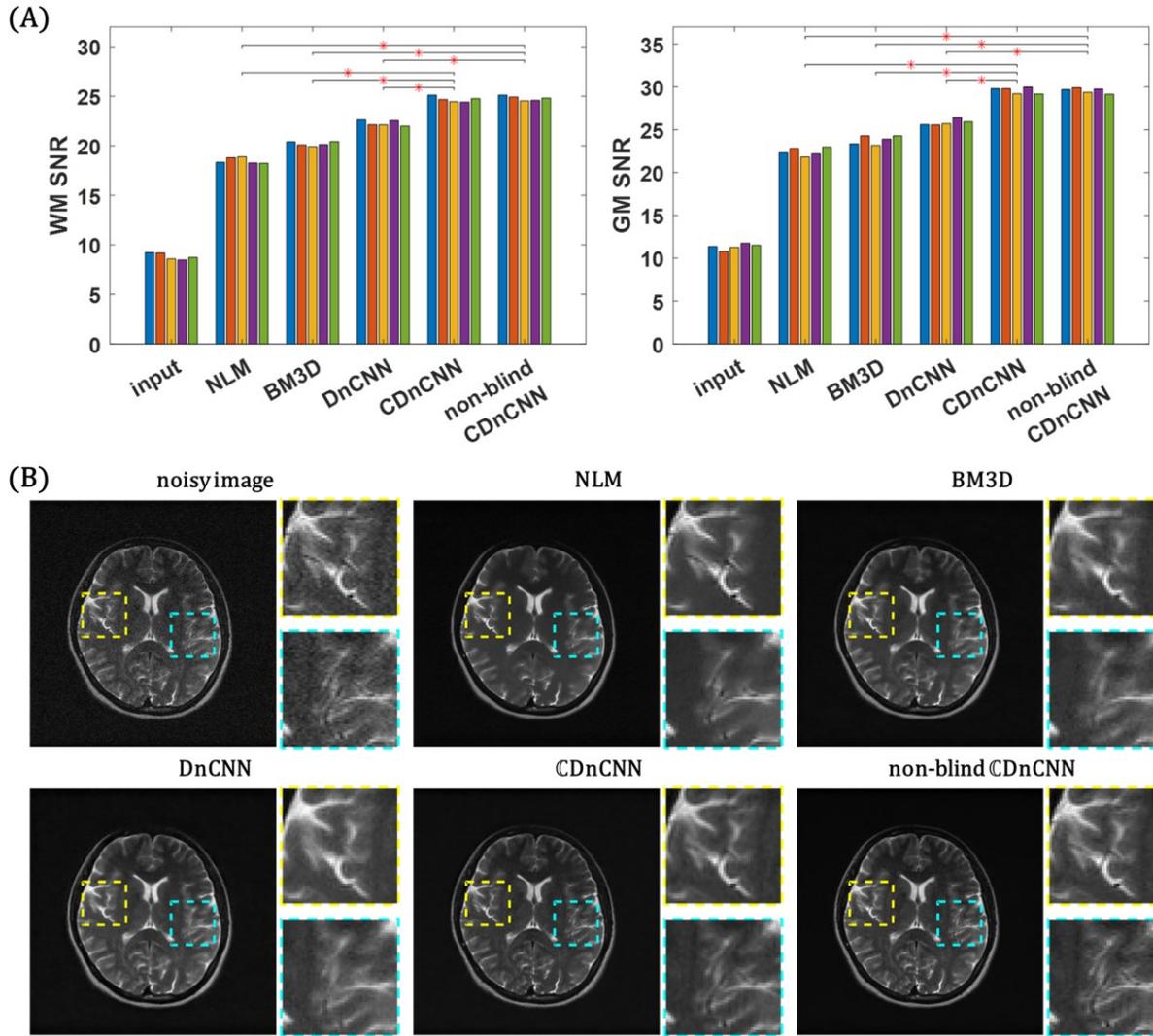


Figure 2-7: (A) Performance of different denoising methods on the low-field M4Raw dataset. The average WM and GM SNRs of each testing volume were evaluated for each method. (B) Representative image results of different denoising methods. The output of CDnCNN and non-blind CDnCNN showed clear structures and better noise reduction compared to other methods. The asterisks indicate statistically significant differences between the methods ($p < 0.05$).

Figure 2-7 shows the performance of different denoising methods on the low-field dataset M4Raw. The blind CDnCNN and non-blind CDnCNN showed better SNRs for WM and GM compared to other methods. Figure 2-8 gives an example from the M4Raw dataset showing the

network performance on spatially varying parallel imaging artifacts. The 4-coil data was undersampled by a factor of 2 and reconstructed by SENSE. The noise in the SENSE reconstructed image was amplified compared to the fully-sampled reconstruction. The non-blind CDnCNN with the scaled g -factor map showed the best denoising performance at regions with large g -factor, demonstrating its generalizability in different coil geometry.

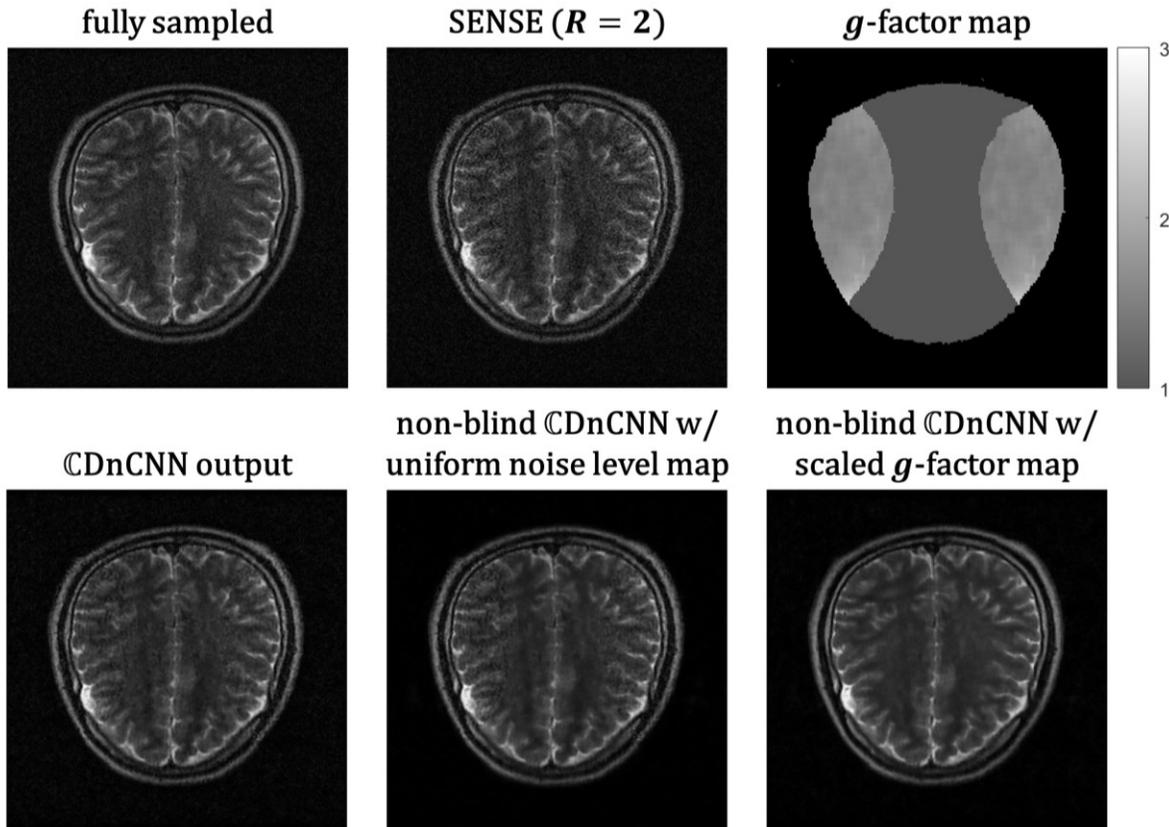
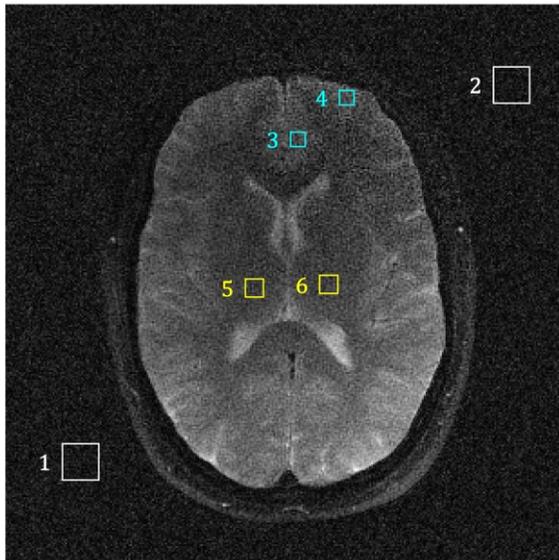


Figure 2-8: Example from the M4Raw dataset of non-blind CDnCNN on spatially varying parallel imaging artifacts. Top left is the fully-sampled image. Top center is the undersampled image reconstructed by SENSE with a subsampling ratio of 2. Top right is the g -factor map from the SENSE reconstruction. Bottom left is the output of CDnCNN without noise level map. Bottom center is the output of non-blind CDnCNN with a uniform noise level map. Bottom right is the output of non-blind CDnCNN with the scaled g -factor map as the noise level map. The non-blind CDnCNN with scaled g -factor map showed the best denoising performance at regions with large g -factor.

To evaluate the denoising performance on in vivo data without reference, we calculated the SNR for WM and GM. Figure 2-9 shows an example of the manually defined ROIs. We listed the

mean intensity and standard deviation of each ROI. Figure 2-10A shows the performance of different denoising methods on local in vivo data collected from a 0.55 T low-field scanner. It can be observed that all methods reduced the noise and increased the SNRs for WM and GM. The blind CDnCNN and non-blind CDnCNN showed superior performance compared to other methods. Representative images are displayed in Figure 2-10B. The output of CDnCNN and non-blind CDnCNN showed sharper structures and less noise. Figure 2-11 shows the output of non-blind CDnCNN with different NAs. The network showed its generalizability as it was able to enhance the overall image quality at different noise levels. When NA increased, small brain structures became more observable in the output image.



ROI	mean	standard deviation
1	0.0989	0.0528
2	0.0941	0.0515
3	0.2741	0.0765
4	0.3045	0.0862
5	0.2306	0.0503
6	0.2497	0.0598

Figure 2-9: Example of the manually defined ROIs for SNR calculation (white rectangles for air regions, cyan rectangles for GM regions, and yellow rectangles for WM regions) and the mean intensity and standard deviation of each ROI.

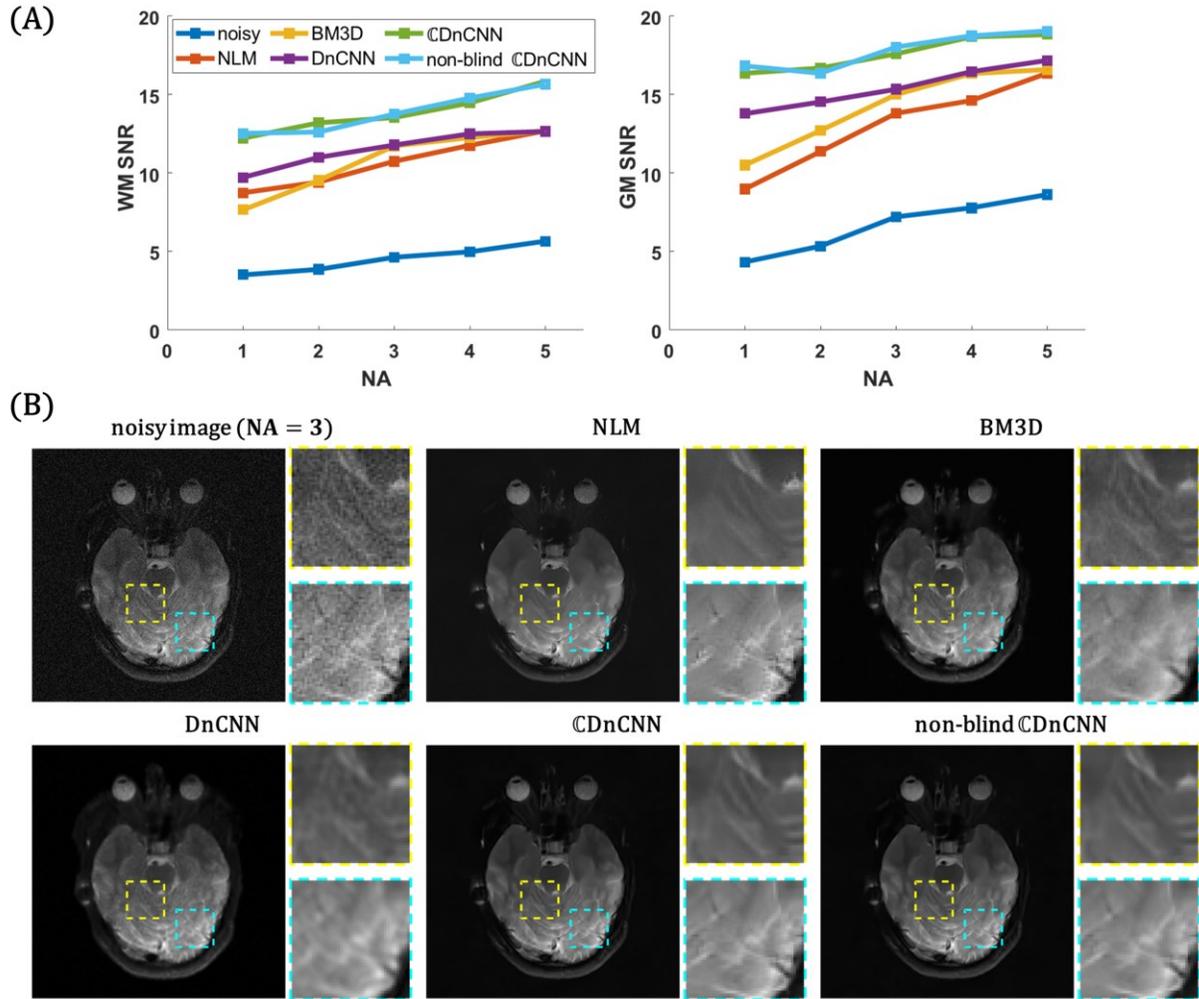


Figure 2-10: (A) Performance of different denoising methods on in vivo images collected from a low-field scanner with different NAs. The average WM and GM SNRs were evaluated for each method. (B) Representative image results of different denoising methods. Top left is the original noisy image acquired with NA of 3. The output of CDnCNN and non-blind CDnCNN showed sharper structures and less noise compared to other methods.

The total training time for non-blind CDnCNN on a system with a NVIDIA Titan Xp GPU, an Intel Xeon 3.3 GHz CPU, and 128 GB RAM was roughly 42 hours. We measured the computational cost of different algorithms on the same system. Table 2-1 summarizes the inference time and memory required for denoising one slice with size 320×320 . Note that the time for estimating the noise standard deviation was also counted, and NLM and BM3D were only tested

on CPU. The non-blind \mathbb{C} DnCNN was able to process one slice in less than 1.5 seconds on CPU. The inference time could be significantly shortened with GPU acceleration.

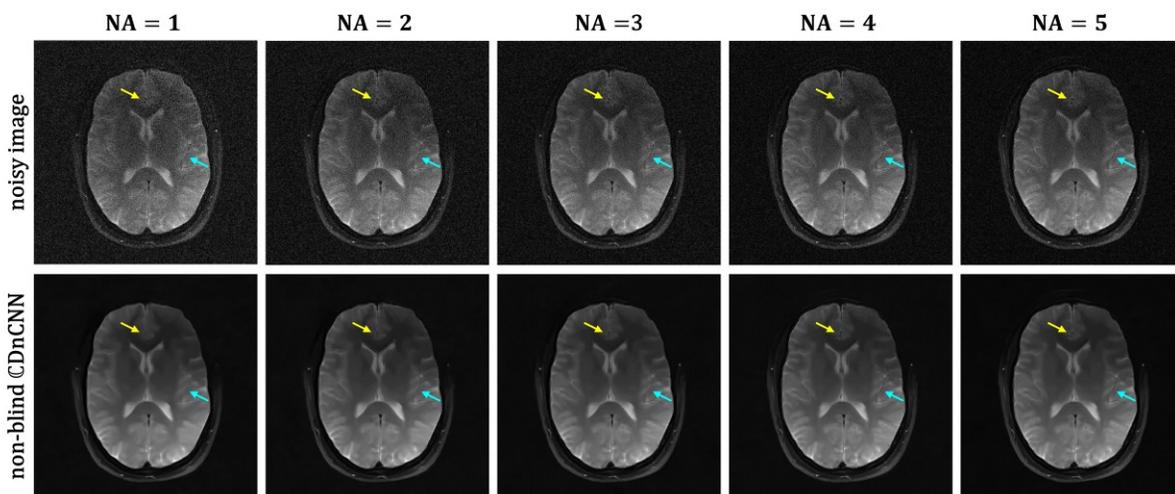


Figure 2-11: Performance of non-blind \mathbb{C} DnCNN at different noise levels. Top row is the original noisy image, and each column was acquired with NA of 1, 2, 3, 4, and 5, from left to right. Bottom row is the output of \mathbb{C} DnCNN. The small structures (yellow and cyan arrows) became sharp and visible as NA increased.

2.4 Discussion

In this chapter, we presented non-blind \mathbb{C} DnCNN, a network for MRI denoising that leverages complex-valued building blocks and noise level information to improve denoising performance in various settings. The proposed method achieved superior performance on both simulated and in vivo testing data compared to other algorithms.

The utilization of complex-valued operations allows the network to better exploit the complex-valued MRI data and preserve the phase information. For NLM and BM3D, the denoising is directly performed on the magnitude image. The phase information is lost and cannot be recovered after denoising. Thus, the phase performance for these approaches was not examined. For real-valued CNNs, the input image is split into real and imaginary channels, and real-valued operations are then applied on these channels. The output image is obtained by combining the two

separate output channels, and the reconstructed phase may be changed. For complex-valued CNNs, the input/output, learned convolutional kernels, and latent features are all in complex-valued representations, enabling the network to make use of the valuable information contained in the phase map. The superior metrics achieved by complex-valued CNNs demonstrate that integrating the phase information is beneficial for the denoising process.

	inference time / slice (GPU)	inference time / slice (CPU)	memory usage
NLM	-	456 ms	85 MB
BM3D	-	2562 ms	188 MB
DnCNN	27 ms	303 ms	288 MB
CDnCNN	79 ms	1297 ms	302 MB
non-blind CDnCNN	154 ms	1472 ms	305 MB

Table 2-1: Computational cost of different algorithms for denoising one slice with size 320×320 .

The DnCNN network architecture is used in this work, since the original DnCNN model attained remarkable performance on Gaussian denoising. We defined network depth as the number of convolution blocks in a DnCNN model. We varied the number of trainable parameters by training real-valued DnCNN, blind CDnCNN, and non-blind CDnCNN with different network depths. At each network depth level, we ensured that each model has approximately the same number of trainable parameters. Our goal was to investigate whether the difference in denoising performance of real-valued, blind complex-valued, and non-blind complex-valued models narrowed as the number of parameters increased. Figure 2-12 shows the performance of the models as a function of network depth on the simulated testing dataset. The output of CDnCNN and non-blind CDnCNN showed superior NRMSE, PSNR, and SSIM. With an increase in network depth, all models exhibited enhanced performance. The difference in denoising performance among different models remained relatively consistent as the network depth increased.

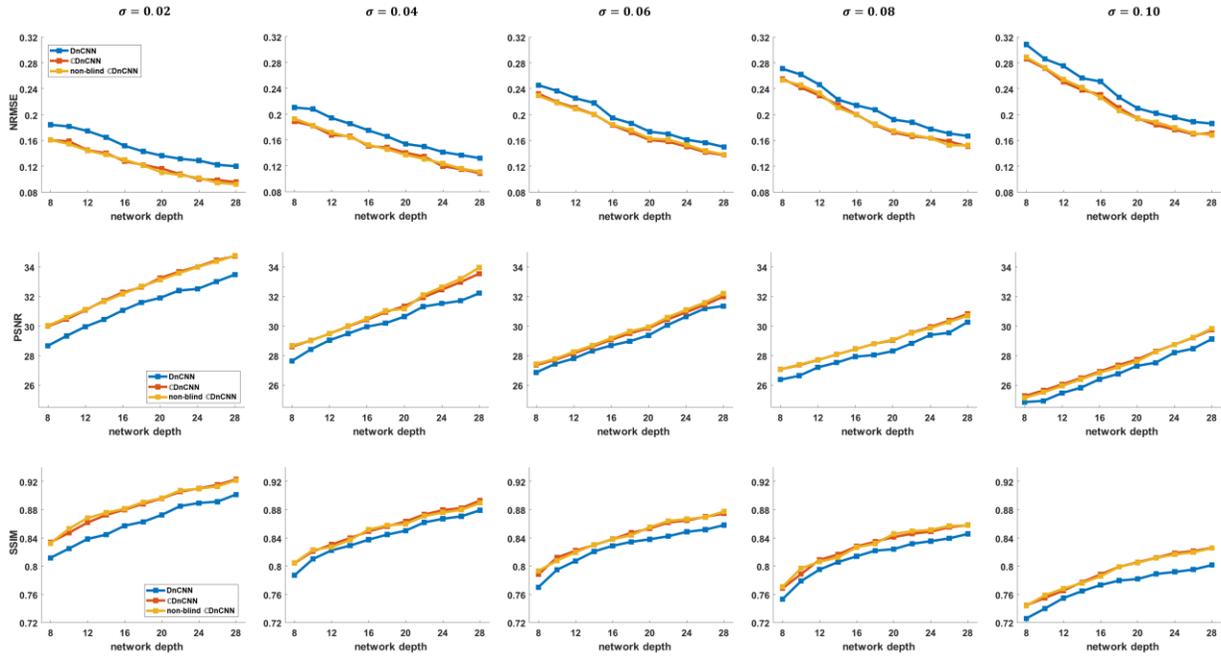


Figure 2-12: Performance of DnCNN, CDnCNN, and non-blind CDnCNN as a function of network depth on the simulated testing dataset at different noise levels. The average NRMSE, PSNR, and SSIM were evaluated for each method. The performance of all networks improved with an increase in network depth.

Recently, the U-Net architecture⁹² is of growing interest in solving problems like medical image segmentation and reconstruction^{78,79,93–95}. The downscaling/upscaling blocks in the U-Net-based model effectively increase the network receptive field and allow the network to utilize both global and regional features. We extended our investigations beyond DnCNN architecture to include U-Net-based networks. Specifically, we trained and evaluated the denoising performance of real-valued U-Net, blind complex-valued U-Net (CU-Net), and non-blind CU-Net. Our goal was to investigate whether the utilization of complex-valued operations can improve the denoising performance on other network architectures. As shown in Figure 2-13, the denoising performance from blind CU-Net and non-blind CU-Net improved compared to the real-valued U-Net.

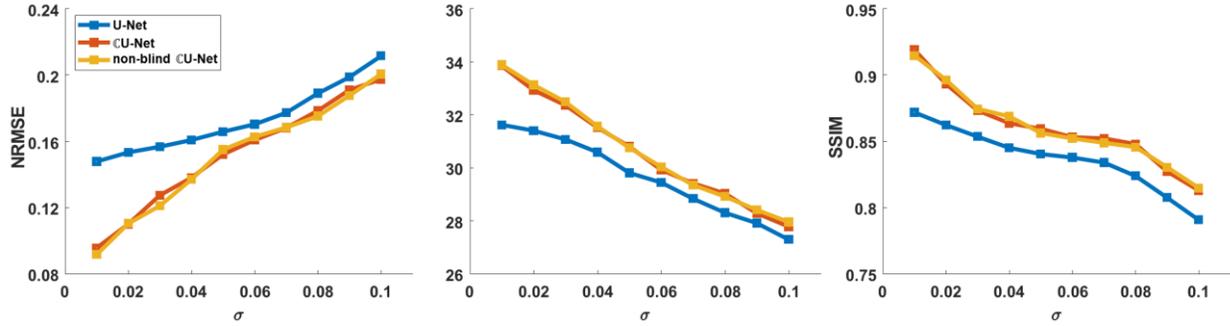


Figure 2-13: Performance of U-Net, CU-Net, and non-blind CU-Net on the simulated testing dataset at different noise levels. The average NRMSE, PSNR, and SSIM were evaluated for each method.

One major concern about CNN-based denoising techniques is the blurring effect introduced by the network. To mitigate this drawback, we chose L1 loss instead of L2 loss, because previous studies have shown that using L2 loss is prone to produce over-smoothed restored images^{96–98}. Incorporating the noise level into the network provides the possibility to control the balance between noise suppression and detail retention. However, in our experience, it is still challenging to balance the trade-off by simply adjust the estimated noise standard deviation. In the future work, more advanced methods, such as using attention mechanisms⁹⁹ to better integrate the noise level map, could be explored to achieve this goal. The spatially non-uniform noise level map allows the network to remove the spatially dependent parallel imaging noise. This advantage improves its clinical feasibility since parallel imaging is frequently used to accelerate scans. In our experiment, we used a wavelet-based method to estimate the noise level for real and imaginary parts separately, and adopted naïve averaging to get the final estimation of $\hat{\sigma}$. To address parallel imaging noise, we weighted the uniform noise level map with the g -factor map. There exist numerous MRI noise estimation schemes relying on wavelet domain analysis, local mutual information, or median absolute deviation estimator^{100–102}. We expect that applying these techniques will give a more accurate noise level map and further improve the network performance. Training the non-blind

CDnCNN with GAN^{103,104} is an alternative way to reduce blurring in the network output. However, a GAN scheme for complex-valued networks remains to be investigated.

Overfitting is another issue for supervised learning models. Due to the limited size of the training set, the model might fit too closely to the training data. In such situation, the network begins to memorize irrelevant information, e.g., the brain anatomy or image contrast, instead of finding a general strategy for denoising. This is a critical problem for medical image processing since the spurious structures or subtle artifacts created by the network can severely affect the diagnosis. To alleviate overfitting, image augmentations were employed to increase the diversity of the training set. The validation loss was also monitored after each epoch during the training stage. Additional tests on data out of the training distribution showed that the method is generalizable. The network also showed promising results on in vivo data acquired on a low-field scanner, demonstrating its ability to boost the SNR of low-field MR images and its potential to reduce the acquisition time of low-field MRI. However, more in vivo experiments on the low-field scanner are needed to validate the network robustness on images of different body regions with different scan parameters.

One potential limitation in our approach is that the current methodology focuses on employing deep learning solely for denoising after reconstructing the MR images. Several studies have tried to mitigate noise during the reconstruction to improve the reconstructed image quality^{105,106}. These methods naturally embed the phase modulation and may better exploit the raw complex-valued imaging data. The integration of denoising within the reconstruction process could be investigated to further enhance the overall imaging pipeline for future work.

2.5 Conclusion

We have shown that the proposed CDnCNN has superior denoising performance compared to the real-valued DnCNN and several other algorithms. The addition of noise level map provides it the ability to remove spatially varying parallel imaging noise. It offers rapid and significant SNR improvements, which is useful for low-field MRI.

Chapter 3: Multi-Task Learning for Simultaneous Motion Detection and Compensation in Brain Imaging

3.1 Introduction

As introduced in Chapter 1, motion artifacts caused by patient movement during the scan remain a significant challenge, particularly for brain MRI, which often requires long acquisition times. Motion artifacts can severely degrade image quality by introducing ghosting and blurring, leading to non-diagnostic images. Such artifacts are especially problematic in pediatric, elderly, and neurologically impaired patients who may have difficulty remaining still during the acquisition¹⁰⁷. A recent study reported that approximately 20% of clinical MR examinations are affected by patient motion and require a rescan¹⁰⁸.

Various methods have been developed to address motion artifacts in brain MRI. These approaches can be categorized broadly into prospective and retrospective techniques. Prospective methods involve motion tracking systems, such as optical cameras^{13,14}, or navigator sequences^{15–20}, that monitor patient movement during the scan and adjust the acquisition in real-time. However, most of these techniques require extra monitoring devices or modifications to pulse sequences, limiting their clinical applicability. On the other hand, retrospective methods attempt to compensate motion artifacts after image acquisition. Traditional retrospective methods often rely on model-based techniques and iterative reconstruction^{109,110}, which are computation-demanding and time-consuming.

In recent years, deep-learning-based methods have emerged as a promising solution for motion compensation in brain MRI. These methods leverage convolutional neural networks

(CNNs) or generative adversarial networks (GANs) to transform motion-corrupted images into motion-free images. Johnson and Drangova¹¹¹ implemented a pix2pix-based network to correct for 3D rigid-body motion. Küstner et al¹¹² investigated the influence of network architecture and introduced their MedGAN architecture with style transfer loss. Lee et al¹¹³ combined a registration network and a multi-input, multi-output network to perform motion compensation for multi-contrast MRI. Pawar et al¹¹⁴ used a multi-resolution network architecture for motion artifact suppression. Hewlett et al¹¹⁵ investigated the incorporation of multi-coil MRI data in deep-learning-based motion correction approaches. These deep learning models provide several advantages, including the ability to generalize across different motion types and imaging contrasts. Additionally, they are typically applied retrospectively without the need for hardware or sequence modifications, making them suitable for routine clinical practice. In addition to motion compensation, deep-learning-based motion detection techniques are receiving increasing attention. Vakli et al¹¹⁶ developed an effective end-to-end lightweight network for motion detection. Fantini et al¹¹⁷ utilized transfer learning to rate motion artifacts in neuroimaging. By learning to classify images as usable or requiring motion compensation, these methods help optimize the quality control process in clinical settings and reduce the burden on radiologists.

Multi-task learning (MTL) has gained attention in the field of medical imaging as a way to simultaneously learn multiple related tasks^{118,119}. Compared to single-task learning (STL), where models are trained independently for each task, MTL provides several advantages. First, MTL enables more efficient data utilization by leveraging shared features across multiple tasks, reducing the need for large task-specific datasets. Additionally, MTL models are less prone to overfitting, as they learn from a more diverse set of tasks, improving their generalizability on different types of data. Finally, MTL can accelerate the learning process by leveraging auxiliary tasks that improve

the model's ability to learn relationships between tasks^{120,121}. For example, the motion compensation task can benefit from the motion detection task by focusing on relevant features associated with motion artifacts.

In this chapter, we proposed a novel multi-task conditional GAN (MT-cGAN) for simultaneous motion detection and compensation in brain MRI. Rigid motion artifacts under realistic clinical settings were simulated on brain images. The network was trained to detect whether the image is motion-corrupted and learn the mapping between motion-corrupted and motion-free images. The proposed method was evaluated with both simulated and in vivo testing data sets and compared with STL models.

3.2 Methods

3.2.1 Motion simulation

It is difficult to get a large number of matched motion-free and motion-corrupted image pairs for supervised network training. Therefore, images with intershot rigid motion artifacts were simulated from motion-free images. The fastMRI dataset is an extensive open-source collection of multi-coil MRI raw data designed to advance MRI reconstruction techniques⁸⁰. It includes multi-coil k -space data for T1-weighted, T2-weighted, FLAIR, and post-contrast T1-weighted images. These modalities are commonly used in clinical brain MRI. Detailed imaging parameters can be found on the project website (<https://fastmri.med.nyu.edu/>). From the fastMRI dataset, we randomly selected 250 volumes from each of the four modalities, yielding a total of 1000 volumes for model training and validation. In addition, 200 more volumes (50 from each modality) were set aside for testing the model on simulated data. Prior to motion simulation, any volumes that exhibit apparent motion artifacts were manually excluded to ensure the quality of the data.

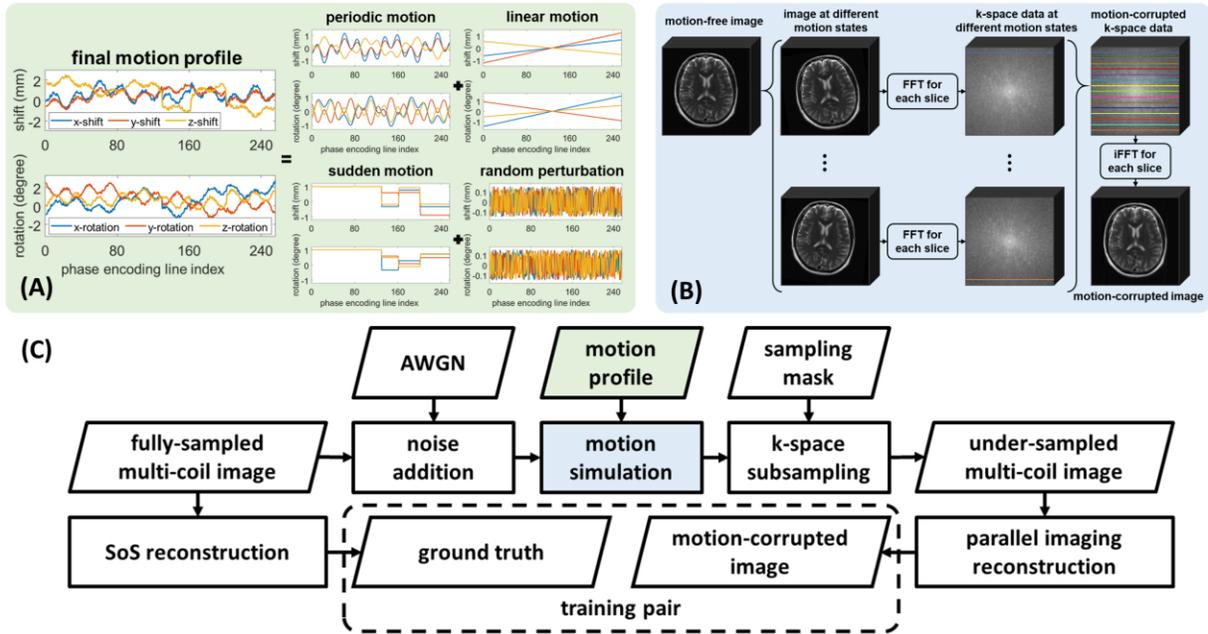


Figure 3-1: Training data simulation process. (A) Motion profiles were generated to simulate realistic movement artifacts commonly observed in brain MRI, including periodic, linear, sudden motions, and random perturbations. (B) Spatial transformations were applied to the motion-free image according to the generated motion profile. The phase encoding lines corresponding to different motion states were combined to form the motion-corrupted k -space data. The motion-corrupted image was obtained by an inverse FFT. (C) To simulate real clinical MRI conditions, complex-valued AWGN was added, and k -space undersampling and parallel imaging reconstruction were applied. The ground truth image was obtained by the sum-of-squares reconstruction.

To generate realistic motion artifacts, we first implemented a motion profile generation process. Our motion simulation model combines several types of motion commonly observed during brain MRI scans¹²². These include periodic motion, which typically corresponds to physiological movements such as respiration or cardiac cycles, linear motion, which simulates gradual shifts or rotations due to unintentional movements, and sudden motion, which represents abrupt actions like yawing or sneezing, mimicking the irregular and unpredictable nature of patient movement. Additionally, random perturbations were added to introduce variability. As shown in Figure 3-1A, the final motion profile was constructed by combining these components. The motion was simulated as 3D translations and rotations, allowing for a comprehensive representation of

head motion. Detailed parameters for generating motion profiles are listed in Table 3-1. This approach enabled us to create realistic and diverse motion-corrupted images for training and testing the network, ensuring the model's robustness in handling a wide range of motion artifacts commonly encountered in clinical brain MRI.

Assume an ideal motion-free image \mathbf{x}_0 , then the motion-corrupted k -space data \mathbf{y} can be written as:

$$\mathbf{y} = \sum_{n=1}^N \mathbf{S}_n \mathbf{F} \mathbf{U}_n \mathbf{x}_0 \quad (3-1)$$

where \mathbf{U}_n is the spatial transformation matrix for phase encoding line n , which operates in the image domain to move pixel values from one location to another¹²³, \mathbf{F} represents the fast Fourier transform (FFT), \mathbf{S}_n extracts the corresponding phase encoding line. As shown in Figure 3-1B, we can reconstruct the motion-corrupted image by applying an inverse FFT to the motion-corrupted k -space data.

In addition to reconstruction from fully sampled k -space data, in clinical settings, brain MRI scans are often accelerated using parallel imaging techniques such as Sensitivity Encoding (SENSE)² or Generalized Autocalibrating Partial Parallel Acquisition (GRAPPA)³. These acceleration methods, while reducing the scan time by undersampling the k -space lines, can lower the signal-to-noise ratio (SNR). To simulate such real-world scenarios and improve the robustness of our model, we incorporated these factors into our simulation pipeline. Specifically, we first added complex-valued additive white Gaussian noise (AWGN) to the multi-coil raw k -space data to simulate the typical SNR limitations encountered in MRI acquisitions. After simulating the motion artifacts, the multi-coil k -space data was then undersampled with an equispaced sampling mask, where the central 24 phase-encoding lines were kept. Detailed parameters of the added noise

and undersampling ratio used in the training are also listed in Table 3-1. Finally, the undersampled k -space data was reconstructed through SENSE or GRAPPA. For SENSE reconstruction, the coil sensitivity maps were estimated from the central 24 phase-encoding lines using the ESPIRiT algorithm¹²⁴. For GRAPPA reconstruction, the central lines constituted the autocalibration signal (ACS) region, allowing the algorithm to estimate missing k -space points from adjacent lines. This process, shown in Figure 3-1C, ensured that the simulation reflected realistic MRI acquisition conditions, enhancing the model's ability to generalize to real clinical data.

Parameter	Sampling range
x/y-shift	-3 mm ~ 3 mm
z-shift	-1.5 mm ~ 1.5 mm
x/y/z-rotation	-2.5° ~ 2.5°
periodic motion frequency	0.1 Hz ~ 2 Hz
No. of sudden motion events	1, 2, 3
standard deviation of AWGN	0 ~ 0.06
undersampling ratio	1 ~ 4
reconstruction method	SENSE, GRAPPA

Table 3-1: Description of the parameters used in the motion simulation.

3.2.2 Network architecture

We implemented a MT-cGAN for the simultaneous detection and compensation of motion artifacts in brain MRI scans. The architecture consists of two primary components: a generator and a discriminator. The generator is based on an adaptation of the U-Net architecture⁹², while the discriminator is designed as a simple CNN aimed at distinguishing between real motion-free images and generated motion-compensated images.

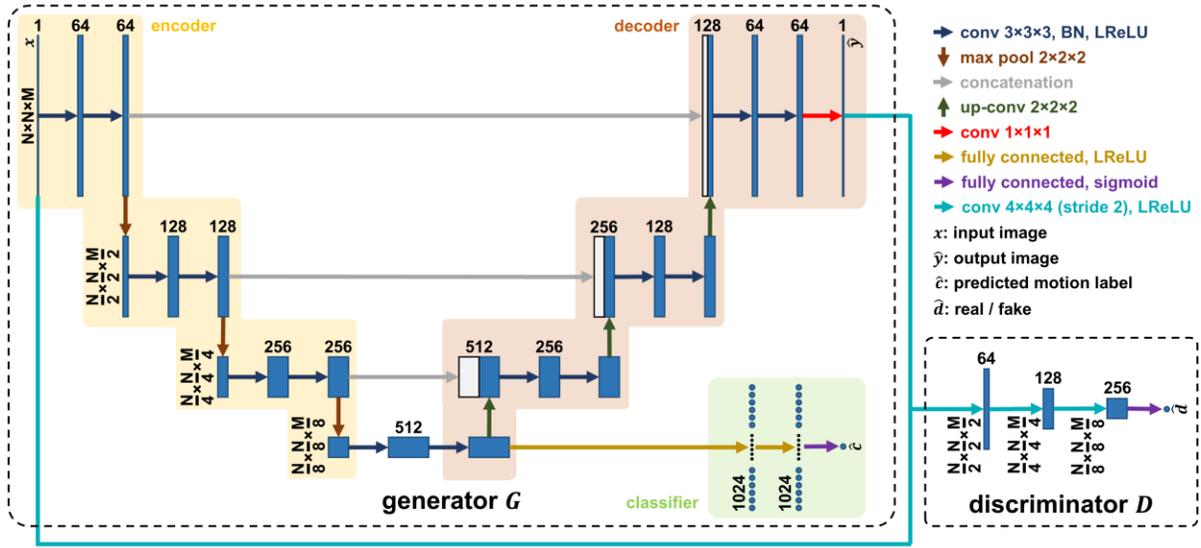


Figure 3-2: Network Architecture of MT-cGAN. The network consists of a generator based on a modified U-Net and a discriminator. The generator includes an encoder-decoder structure to reconstruct the output image, and a classification branch at the bottleneck to predict whether the input image is motion-corrupted. The discriminator is a simple CNN.

The generator follows a U-Net structure, which is well-suited for image restoration by utilizing skip connections between the encoding and decoding blocks. The encoder comprises three encoding blocks, each of which includes two convolutional layers with $3 \times 3 \times 3$ kernels, batch normalization (BN)⁸³, and leaky rectified linear units (LReLU) for non-linearity¹²⁵. Each encoding block is followed by a max-pooling layer with a $2 \times 2 \times 2$ stride that reduces the spatial dimensions of the feature maps, while the number of feature channels is increased from 64 to 256. The decoder mirrors the encoder, consisting of three decoding blocks. Each decoding block includes an up-sampling layer (implemented with a transposed convolution with a stride of $2 \times 2 \times 2$) and two convolutional layers followed by BN and LReLU. Skip connections are employed between corresponding encoding and decoding blocks, allowing original feature maps from the encoder to be merged with up-sampled feature maps in the decoder. The output of the decoder is passed through a final $1 \times 1 \times 1$ convolutional layer to generate the corrected image. In addition to the

main encoder-decoder structure, a classification branch was added at the bottleneck of the U-Net. At this point, the feature maps are flattened and passed through two fully connected layers with LReLU activations, producing a binary output representing whether the input image is motion-free or motion-corrupted. This multi-task approach allows the generator to perform both motion detection and compensation tasks simultaneously, as shown in Figure 3-2.

The discriminator is designed to classify images as either real (truly motion-free) or fake (motion-compensated). It consists of three convolutional layers, each with a stride of 2, followed by LReLU activations. The number of feature channels increases from 64 to 256 across these layers. After the final convolutional layer, the feature maps are flattened, and a fully connected layer is used to produce a single output, which indicates the probability that the input image is either a real motion-free image or a motion-compensated image generated by the generator.

To effectively train the MT-cGAN model, we incorporated three different loss functions: an image restoration loss, a classification loss, and a GAN loss. For the image restoration task, we used a combined L1 loss and multi-scale structural similarity (MS-SSIM) loss. The L1 loss encourages pixel-level accuracy by minimizing the absolute differences between the corrected and ground-truth images. Meanwhile, the MS-SSIM loss captures structural information at different scales, enhancing the model’s ability to preserve textural details and improve perceptual quality. This approach follows the work of Zhao et al⁹⁶ and Pezzotti et al¹²⁶, who demonstrated that the combined loss yields superior performance for image restoration. Mathematically, the image restoration loss \mathcal{L}_{img} is defined as:

$$\mathcal{L}_{\text{img}} = \|\mathbf{x}_{\text{GT}} - \hat{\mathbf{x}}\|_1 + \alpha(1 - \text{MS-SSIM}(\mathbf{x}_{\text{GT}}, \hat{\mathbf{x}})) \quad (3-2)$$

where \mathbf{x}_{GT} represents the ground truth motion-free image, $\hat{\mathbf{x}}$ represents the output image from the generator, and α is a hyper-parameter empirically set to 0.84. For the classification branch, the

widely used binary cross-entropy loss is employed to penalize misclassifications. The classification loss \mathcal{L}_{cls} is given by:

$$\mathcal{L}_{\text{cls}} = -c_{\text{GT}} \log(\hat{c}) - (1 - c_{\text{GT}}) \log(1 - \hat{c}) \quad (3-3)$$

where c_{GT} is the ground truth label of the input image (0 for motion-free, and 1 for motion-corrupted), and \hat{c} is the predicted probability from the generator. To train a generator producing realistic motion-compensated images, we also used a GAN loss. The discriminator aims to distinguish between real motion-free and generated motion-compensated images, while the generator seeks to produce images that the discriminator cannot differentiate from real ones. This adversarial learning process is characterized by:

$$\mathcal{L}_{\text{GAN}} = \mathbb{E}[\log(D(\mathbf{x}, \mathbf{x}_{\text{GT}}))] + \mathbb{E}[\log(1 - D(\mathbf{x}, \hat{\mathbf{x}}))] \quad (3-4)$$

where G denotes the generator, D denotes the discriminator, \mathbf{x} denotes the input image, and \mathbb{E} denotes the expectation operator. These components are combined in the final weighted objective function as follows:

$$\mathcal{L} = \lambda_{\text{img}} \mathcal{L}_{\text{img}} + \lambda_{\text{cls}} \mathcal{L}_{\text{cls}} + \lambda_{\text{GAN}} \mathcal{L}_{\text{GAN}} \quad (3-5)$$

where λ_{img} , λ_{cls} , and λ_{GAN} are regularization parameters optimized through grid search.

3.2.3 Evaluation

For the motion detection task, we compared our MT-cGAN with two other classifiers: a conventional classifier and a CNN. The first conventional classifier utilized an extreme gradient boosting (XGBoost) approach¹²⁷ trained on image quality metrics calculated using the MRIQC toolbox¹²⁸. This classifier leverages pre-calculated metrics, such as SNR, contrast, and blurriness, to assess the presence of motion artifacts. The second model was a single-task CNN (ST-Net) adapted from our generator architecture. Specifically, the ST-Net retained only the encoder and

classification components from the MT-cGAN generator, excluding the decoder branch responsible for motion compensation. For each model, performance was evaluated through various classification metrics, including confusion matrix, classification accuracy, sensitivity, specificity, receiver operating characteristic (ROC) curve, and the area under the ROC curve (AUC).

For the motion compensation task, our MT-cGAN model was compared with two other models. Both of these models follow the same network structure as the MT-cGAN, but without the classification branch dedicated to the motion detection task. The first comparison model, referred to as the baseline model, was trained on a dataset containing only simulated motion artifacts, excluding additional elements such as AWGN, k -space undersampling, and parallel imaging reconstruction. The second comparison model, named the ST-cGAN model, was trained on data that underwent the full simulation process. By comparing these two models, we can determine the impact of including AWGN and parallel imaging simulation on the model's robustness, especially when applied to real clinical data. For a quantitative evaluation of motion compensation performance, we used three metrics: normalized root-mean-square error (NRMSE), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM).

For the XGBoost classifier, we used the Python `xgboost` package¹²⁷. All networks were implemented in PyTorch⁸⁴ and trained using the Adam optimizer⁸⁵ with a learning rate of 0.0001 and momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Due to the limit of GPU memory, a batch size of 1 was used throughout training. The evaluation was first performed on a test dataset generated using the same simulation process as the training data, but without AWGN and with an undersampling ratio set to 2. Half of the test images contained simulated motion artifacts, while the remaining half were kept as motion-free, allowing the assessment of the model's motion detection accuracy.

	T1-weighted	T2-weighted
Sequence	3D MPRAGE	2D TSE
Voxel size [mm]	$1.0 \times 1.0 \times 1.0$	$0.5 \times 0.5 \times 5.0$
Repetition time [ms]	2100	4500
Echo time [ms]	3	100
Flip angle [°]	10	90
Orientation	Sagittal	Sagittal
Phase encoding direction	Anterior-posterior	Anterior-posterior
Acceleration factor	2	2
Acquisition time [min:s]	5:23	3:17

Table 3-2: MR acquisition parameters of local in-vivo T1-weighted and T2-weighted images.

To further assess the model's generalizability to real clinical data, we used the MR-ART dataset¹²⁹. This dataset comprises T1-weighted brain MRI scans from 148 healthy adults, collected using the magnetization-prepared rapid gradient echo (MPRAGE) sequence and 2-fold GRAPPA acceleration. The images are scored by expert neuroradiologists on a three-point scale, where images scored as 1 are considered motion-free, and those scored as 2 or 3 are labeled as motion-corrupted. Although the MR-ART dataset offers matched motion-free and motion-corrupted scans per participant, there is minor misalignment between the image pairs, preventing the use of reference-based metrics like NRMSE, PSNR, and SSIM. Therefore, we evaluated motion compensation performance on this dataset using three reference-free image quality metrics computed with the MRIQC toolbox: total SNR, entropy focus criterion (EFC)¹⁰⁹, and coefficient of joint variation (CJV)¹³⁰. These metrics correlate with head motion effects and allow assessment of motion compensation performance without a ground truth image.

Lastly, we validated the models on local in vivo T1-weighted and T2-weighted scans collected from six subjects. The scans were performed on a 3T scanner (MAGNETOM Prisma;

Siemens Healthcare; Erlangen, Germany). Imaging studies was performed under institutional review board-approved protocols, and written informed consent from all subjects were given. For each subject, motion-free images were first acquired while the participant remained still in the scanner. Subsequently, motion-corrupted images were obtained by instructing the participant to nod their head during the scan. Detailed imaging parameters for these acquisitions were listed in Table 3-2.

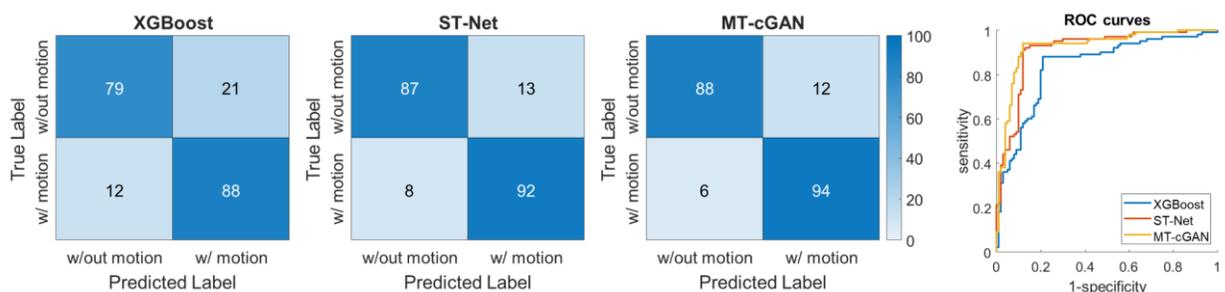


Figure 3-3: Confusion matrices and ROC curves for the motion detection task on the simulated test dataset. The AUC for XGBoost, ST-Net, and MT-cGAN is 0.8375, 0.9114, and 0.9247, respectively. The MT-cGAN outperformed the other two models.

3.3 Results

On the simulated test dataset, for the motion detection task, the XGBoost classifier achieved an accuracy of 83.00%, with a sensitivity of 88.00% and specificity of 79.00%. The AUC for XGBoost is 0.8375. The ST-Net outperformed XGBoost, reaching an accuracy of 90.00%, with a sensitivity of 92.00% and specificity of 87.00%. The AUC for ST-Net was 0.9114. Overall, MT-cGAN achieved the highest performance among the three models, with an accuracy of 91.00%, a sensitivity of 94.00%, a specificity of 88.00%, and an AUC of 0.9247, demonstrating that the shared features between motion detection and compensation tasks enhance the model’s robustness in detecting motion artifacts. The confusion matrices and ROC curves were shown in Figure 3-3. For the motion compensation task, all models demonstrated improvement in the image quality of

motion-corrupted images. MT-cGAN achieved the best performance across the quantitative metrics of NRMSE, PSNR, and SSIM, as shown in Figure 3-4A. Representative images from each model's output were shown in Figure 3-4B. The baseline model exhibited more visible blurring and residual artifacts, suggesting that including AWGN and parallel imaging during the simulation process is beneficial for enhancing model robustness and performance. Compared to the baseline and ST-cGAN models, MT-cGAN provided superior artifact reduction and better preservation of fine structures, such as cortical folds and small vessels, as indicated by red arrows.

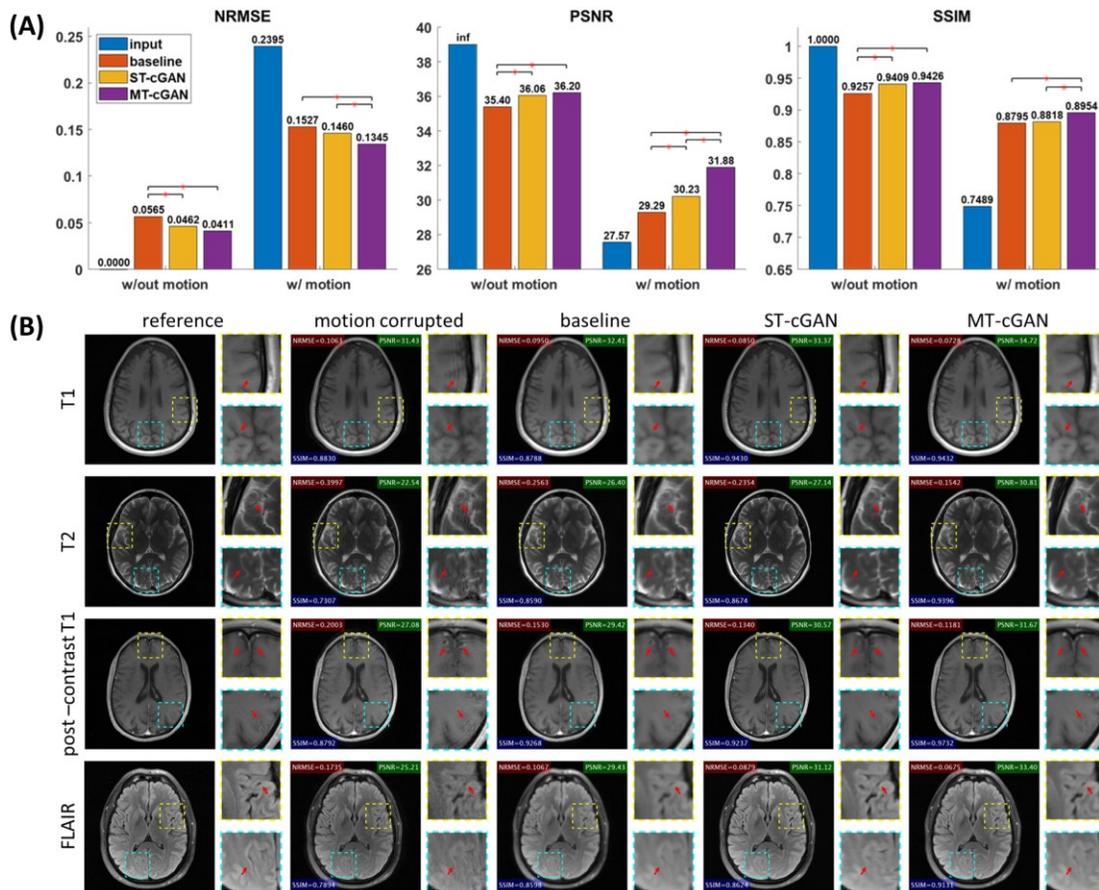


Figure 3-4: Quantitative evaluation results and representative images for the motion compensation task on the simulated test dataset. (A) MT-cGAN achieved the best NRMSE, PSNR, and SSIM on motion-compensated images. Asterisks indicate a significant difference ($p < 0.05$). (B) MT-cGAN showed fewer residual artifacts and better fine structural preservation (marked by red arrows).

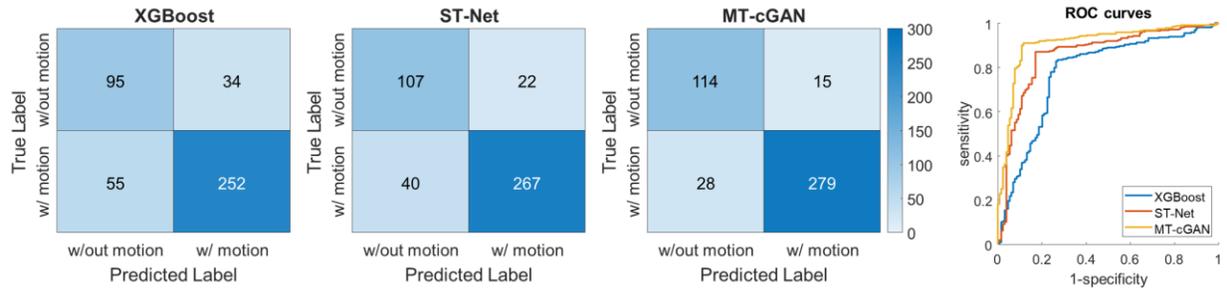


Figure 3-5: Confusion matrices and ROC curves for the motion detection task on the MR-ART dataset. The AUC for XGBoost, ST-Net, and MT-cGAN is 0.7761, 0.8578, and 0.9090, respectively. The MT-cGAN achieved the highest performance.

On the MR-ART dataset, MT-cGAN achieved the highest performance in the motion detection task, with an accuracy of 90.14%, sensitivity of 90.88%, and specificity of 88.37%. This outperformed both XGBoost, with an accuracy of 79.59%, sensitivity of 82.08%, and specificity of 73.64%, and ST-Net, with an accuracy of 85.78%, sensitivity of 86.97%, and specificity of 82.95%. The confusion matrices and ROC curves were shown in Figure 3-5. As indicated by the ROC curves, MT-cGAN achieved the highest AUC at 0.9090, compared to 0.8578 for ST-Net and 0.7761 for XGBoost. For motion compensation, all models showed improved metrics on motion-corrupted images. MT-cGAN achieved the lowest EFC and CJV on motion-corrected images. Figure 3-6A provided the quantitative evaluation results, and Figure 3-6B provided representative images from each model. The output of MT-cGAN showed fewer residual artifacts than the baseline and ST-cGAN models, as indicated by red arrows.

The local in-vivo data consists of 24 imaging volumes (T1-weighted and T2-weighted, both motion-free and motion-corrupted) from six subjects. Due to the small dataset size, the evaluation results serve as an indicative comparison rather than a comprehensive performance assessment. For the motion detection task, both ST-Net and MT-cGAN achieved an accuracy of 83.33%, outperforming the XGBoost classifier, which achieved an accuracy of 70.83%. For the motion compensation task, MT-cGAN demonstrated the best performance across three quantitative

metrics. As shown in Figure 3-7A, MT-cGAN achieved the lowest NRMSE and the highest PSNR and SSIM on the motion-corrected images. Representative images were shown in Figure 3-7B. The MT-cGAN outputs showed reduced visual blurring and better preservation of fine structural details compared to the baseline and ST-cGAN models.

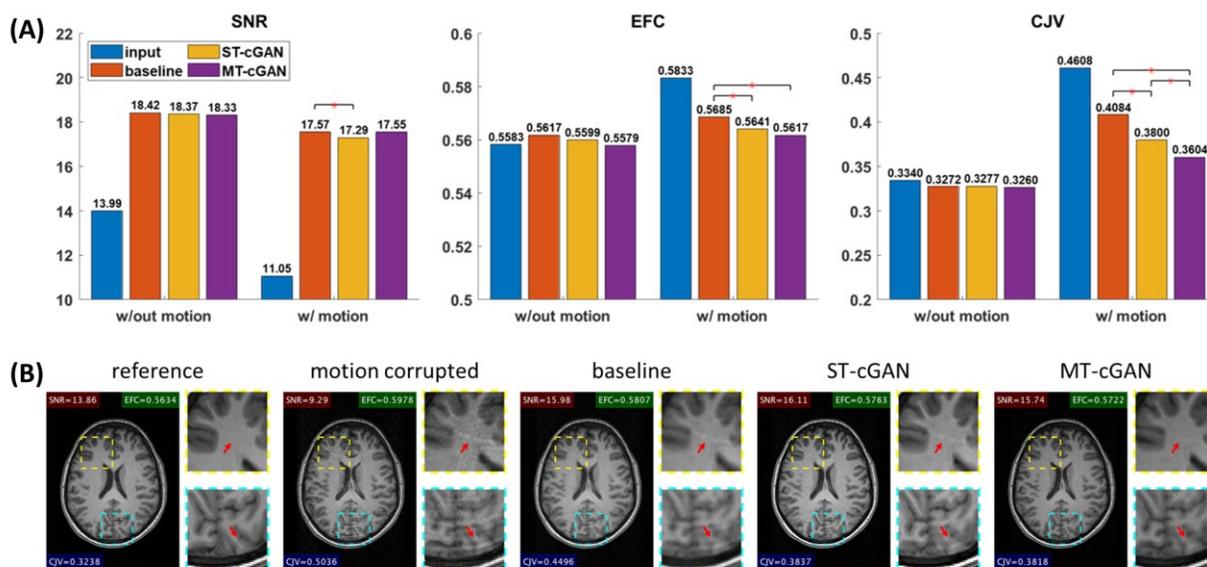


Figure 3-6: Quantitative evaluation results and representative images for the motion compensation task on the MR-ART dataset. (A) For compensation of images with motion artifacts, MT-cGAN achieved the lowest EFC and CJV, while the baseline model attained the highest SNR. Asterisks indicate a significant difference ($p < 0.05$). (B) The output of MT-cGAN exhibit minimal residual artifacts as indicated by red arrows.

The computational efficiency of MT-cGAN was evaluated in terms of training and inference time. Training MT-cGAN for 50 epochs on a system equipped with two NVIDIA Titan Xp GPUs took approximately 63 hours. For deployment, MT-cGAN can process one volume of size $320 \times 320 \times 8$ in approximately 1.2 seconds on the same system.

3.4 Discussion

In this chapter, we developed and evaluated a MT-cGAN framework for simultaneous motion detection and compensation in brain MRI. Our MT-cGAN model demonstrated high efficacy in

reducing motion artifacts and accurately detecting motion-corrupted images. By leveraging MTL and including AWGN and parallel imaging schemes in the training data simulation, MT-cGAN shows promising advantages over STL models and conventional methods in both tasks. This outcome highlights the potential of our approach to provide automatic motion detection and compensation in real clinical settings.

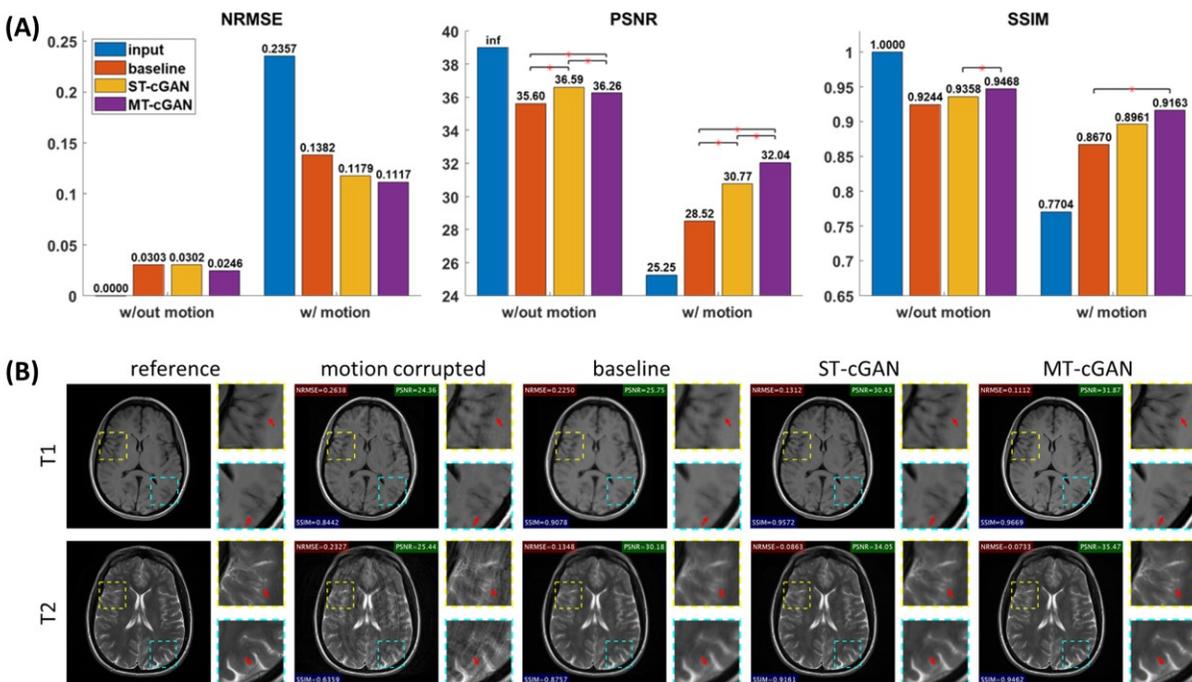


Figure 3-7: Quantitative evaluation results and representative images for the motion compensation task on the local in-vivo data. (A) MT-cGAN achieved the best NRMSE, PSNR, and SSIM on compensating motion-corrupted images. Asterisks indicate a significant difference ($p < 0.05$). (B) Compared to the baseline model and ST-cGAN, the outputs of MT-cGAN showed reduced visual blurring and superior preservation of fine structures (marked by red arrows).

Our training data generation included a comprehensive motion simulation process. We incorporated periodic, linear, and sudden motion, capturing a wide spectrum of potential subject movements in brain MRI. This diversity in motion simulation ensured that the model learned to generalize across different types of motions. Prior studies have often used a simpler motion simulation strategy by combining k -space segments at different motion states, which may limit the model's adaptability to real brain MRI data. Another key feature of our simulation process was the

inclusion of AWGN and parallel imaging schemes. These additions provided the model with realistic training data that more closely resembles clinical conditions. These augmentations were proved beneficial for improving the model’s robustness in clinical applications, where MRI scans are often affected by noise and residual parallel imaging artifacts^{131,132}. However, the current motion simulation process was restricted to rigid motion. Non-rigid motion is also present in clinical MRI, especially for cardiac or abdominal imaging¹³³. Handling non-rigid motion would be a valuable step in making the model more clinically useful and extending its potential to other body parts.

The conditional GAN architecture was chosen due to its ability to generate high-fidelity images, as demonstrated in previous studies^{111,112,115}. Compared to traditional CNNs trained with L2 loss, the combined L1+MS-SSIM and adversarial loss helps the generator to produce sharp and realistic images, addressing the common problem of blurriness often observed in the image restoration^{96,126,134}. Recently, attention-based architectures have shown promise in various imaging tasks^{135–138}. Future work includes exploring alternative architectures like transformers, as they provide the ability to capture long-range dependencies and spatial hierarchies, which may further enhance the model’s performance.

The superior performance of MT-cGAN on both tasks demonstrated the effectiveness of MTL framework. Unlike STL models, which independently handle either motion detection or compensation, MTL captures the complementary nature of two tasks, as they both focus on identifying and addressing motion artifacts. This approach helped the encoder to learn motion-related feature maps in the latent space and reduced the potential for overfitting. One crucial issue in MTL is balancing the loss functions for different tasks. In this study, we optimized static

regularization parameters through grid search, but advanced approaches, such as adaptive weighting¹³⁹ or an uncertainty-based method¹⁴⁰ may yield improved results.

One notable limitation for current study is the relatively small size of the local in-vivo test dataset. A more diverse dataset from multiple institutions with different scanning parameters and different pre-processing procedures should help to validate the generalizability of the model. Additionally, as phase information is crucial in MRI, incorporating phase data into the network may improve artifact characterization. However, few existing motion datasets provide raw phase information, underscoring the need for datasets with both magnitude and phase data for more comprehensive training.

3.5 Conclusion

The MT-cGAN model presented a robust approach for simultaneous motion detection and compensation in brain MRI. By combining realistic data simulation and MTL framework, MT-cGAN showed its potential to improve diagnostic image quality and reduce rescan rates, contributing to more efficient and accurate brain MRI.

Chapter 4: Convolutional Neural Network for Field Inhomogeneity and Concomitant Gradient Field Correction in Spiral Imaging

4.1 Introduction

Compared to Cartesian imaging, spiral imaging offers the advantages in motion robustness and shorter scan time due to its efficient coverage of k -space and natural oversampling of central k -space¹⁴¹⁻¹⁴⁴, making it an excellent alternative scheme for k -space sampling. However, spiral imaging is particularly susceptible to off-resonance effects, leading to blurring and geometric distortions in the acquired images²⁶. Off-resonance effects in MRI arise from several sources, including field inhomogeneities and concomitant fields³³. Field inhomogeneities typically result from imperfections in the main magnetic field B_0 , subject-induced susceptibility variations, and eddy currents. These inhomogeneities introduce spatially varying frequency shifts, especially in regions near tissue-air interfaces, leading to a cumulative phase error during the long spiral readout. Concomitant fields, on the other hand, are induced by the applied gradients according to Maxwell's equations. Although the concomitant fields are generally weaker than the primary gradients, their effects can become noticeable in low-field, high-gradient systems if uncorrected¹⁴⁵.

Over the years, various methods have been developed to address off-resonance artifacts. Field map-based methods involve acquiring a map that characterizes off-resonance frequencies across the image, which is then used to apply corrections through conjugate phase reconstruction (CPR). However, exact CPR is very time consuming and several approximation methods have been proposed. Noll et al²⁷ introduced an efficient correction scheme by segmenting the collected

data based on collection time. Irarrazabal et al¹⁴⁶ proposed to use an estimated linear field map for inhomogeneity correction requiring little extra computation time. Man et al²⁹ utilized multifrequency interpolation (MFI) to reduce computational demands. Schomberg¹⁴⁷ compared several CPR approximation methods in terms of their speed and accuracy. Sutton et al²⁸ described an iterative reconstruction algorithm based on the min-max time interpolation.

Automatic deblurring methods try to reduce the image blurring by optimizing some objective function, which do not require the field map. Noll et al³¹ developed the original autofocus method by minimizing the image imaginary part. Man et al³² and Moriguchi et al¹⁴⁸ introduced multi-stage and block-by-block correction strategies to improve computational efficiency. More recently, deep learning techniques have demonstrated potential in automatic deblurring. Zeng et al¹⁴⁹ implemented a deep residual network for off-resonance correction in 3D cone scans. Lim et al¹⁵⁰ applied convolutional neural networks (CNNs) to reduce blurring in spiral real-time speech MRI.

However, most of these works focused primarily on correcting field inhomogeneity. In this study, we proposed a CNN to simultaneously correct field inhomogeneity and concomitant field effects in spiral MRI. The proposed method does not require the field map acquisition or concomitant field estimation. We synthesized off-resonance artifacts based on simulated field maps and concomitant fields and used the synthesized images for network training. The network performance was evaluated on both synthetic and in-vivo data.

4.2 Methods

4.2.1 Training data generation

We used the fastMRI database⁸⁰ to create a dataset for training our CNN. The fastMRI dataset provides raw k -space data that includes the phase information, which is essential for simulating off-resonance artifacts. We randomly selected 1000 volumes (500 T1-weighted and 500 T2-weighted) as the training dataset. Another 200 volumes were randomly chosen as the testing dataset.

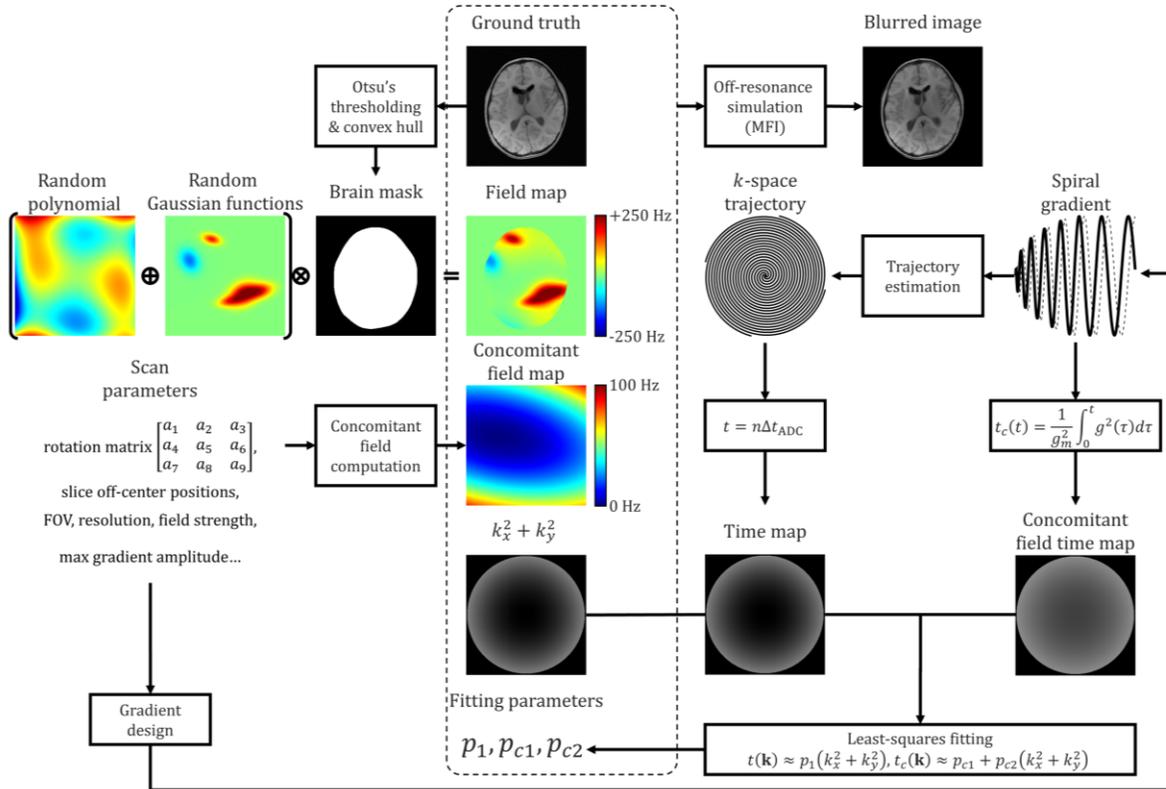


Figure 4-1: Off-resonance artifacts simulation process. The field map was simulated by combining random polynomial and Gaussian functions, and the concomitant field map was calculated based on the scan parameters. The k -space trajectory and corresponding gradient waveforms were generated and used for computing the time maps. Least-squares fitting was adopted to estimate the parameters in mapping the time maps to a quadratic function. The final blurred image with off-resonance artifacts was obtained by MFI.

To simulate off-resonance artifacts in spiral MRI, we first generated field maps by combining random polynomials and several Gaussian functions. For concomitant field simulation, we utilized the simplified equations described in King et al¹⁹, ignoring the third and higher order terms:

$$\Delta\omega_c(\mathbf{x}) \approx \frac{\gamma g_m^2}{4B_0} (F_1x^2 + F_2y^2 + F_3z^2 + F_4yz + F_5xz + F_6xy) \quad (4-1)$$

where $\mathbf{x} = [x, y, z]^T$ are the logical image coordinates, B_0 is the main magnetic field strength, g_m is the maximal readout gradient amplitude. The values of F_n are defined in King et al¹⁵¹ and depend on the elements in the rotation matrix \mathbf{A} between the actual scanner coordinates and the logical image coordinates, where

$$\mathbf{A} = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix} \quad (4-2)$$

The MR signal equation for an object $m(\mathbf{x})$ in the presence of B_0 inhomogeneity and concomitant fields can be written as:

$$s(t) = \int_{\mathbf{x}} m(\mathbf{x}) e^{-i2\pi\mathbf{k}(t)\cdot\mathbf{x}} e^{-i\Delta\omega(\mathbf{x})t} e^{-i\Phi_c(\mathbf{x},t)} d\mathbf{x} \quad (4-3)$$

where $\mathbf{k}(t) = \frac{\gamma}{2\pi} \int_0^t g(\tau) d\tau$ is the k -space trajectory, $g^2(t) = g_x^2(t) + g_y^2(t)$, $g_x(t)$ and $g_y(t)$ are the readout gradients in the logical image coordinates, $\Delta\omega(\mathbf{x})$ is the angular off-resonance frequency of B_0 inhomogeneity, and $\Phi_c(\mathbf{x}, t)$ is the additional phase accrual due to concomitant fields. According to Equation 4-1, $\Phi_c(\mathbf{x}, t)$ can be expressed as:

$$\Phi_c(\mathbf{x}, t) = \Delta\omega_c(\mathbf{x})t_c(t) \quad (4-4)$$

where $t_c(t) = \frac{1}{g_m^2} \int_0^t g^2(\tau) d\tau$. As proposed by Ahunbay et al¹⁵² and Cheng et al¹⁵³, since both t and $t_c(t)$ can be approximated with a quadratic function $k_x^2 + k_y^2$, the total off-resonance phase can be written as:

$$\Phi(\mathbf{x}, t) = p_{c1}\Delta\omega_c(\mathbf{x}) + (p_1\Delta\omega(\mathbf{x}) + p_{c2}\Delta\omega_c(\mathbf{x}))(k_x^2 + k_y^2) \quad (4-5)$$

where p_1 , p_{c1} , and p_{c2} are constants obtained by least-squares fitting.

We generated a spiral trajectory based on the image field of view (FOV) and resolution and calculated the corresponding spiral gradients. Based on these components, we simulated the blurred image by applying MFI on the ground truth, as shown in Figure 4-1. This approach allows for simultaneous simulation of B_0 inhomogeneity and concomitant field effect. Detailed parameters used in the simulation were listed in Table 4-1.

Parameter	Sampling range
Main field strength B_0	1.5 T, 3 T
Field map off resonance frequency	-300 Hz ~ 300 Hz
Slice off-center positions	-10 cm ~ 10 cm
Slice oblique angles	-15° ~ 15°
No. of spiral interleaves	4 ~ 64
Readout length	4 ms ~ 28 ms

Table 4-1: Description of the parameters used in the off-resonance simulation.

4.2.2 Network architecture

Inspired by the autofocus method for off-resonance correction³¹, we implemented a AutofocusNet that operates on multiple demodulated images. The blurred input image was first demodulated at M different frequencies, where M is a hyperparameter set to 21 to provide a sufficient range for

capturing off-resonance effects. This produced a series of 21 complex-valued images, which were then concatenated to form the input for the network.

The architecture of our network consists of three residual blocks, each designed to handle complex-valued data. Within each residual block, there are two 3×3 complex-valued convolutional layers (CConv)⁷⁶, each with a depth of 128 filters. These convolutional layers are followed by complex-valued rectified linear unit (CReLU)⁷⁶ activations to introduce nonlinearity. A skip connection is added between the input and output of each residual block, enabling the network to learn residual mappings, which improves its stability and convergence¹⁵⁴. The final deblurred image is generated by a 1×1 CConv, which combines the outputs from the residual blocks into a single corrected image, as shown in Figure 4-2. This architecture was designed to balance complexity with computational efficiency, leveraging complex-valued operations to exploit the phase information in off-resonance correction.

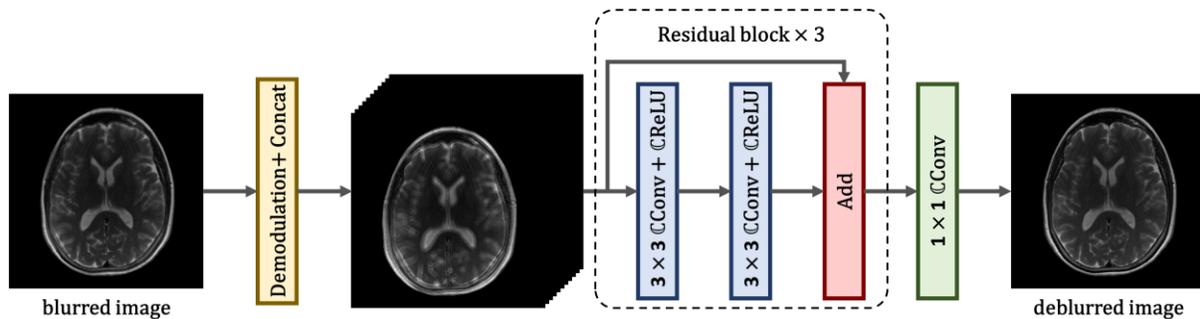


Figure 4-2: Architecture for AutofocusNet. The input blurred image is demodulated at multiple frequencies to produce a series of images, which are concatenated and fed into the network. The network consists of three residual blocks. The final layer is a 1×1 CConv to generate the deblurred output image.

The entire network was implemented in PyTorch⁸⁴ and optimized using an L1 loss function, measuring the difference between the network output and the ground truth. The Adam optimizer⁸⁵ was employed with a learning rate of 0.0001 and momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We trained the network for 200 epochs with roughly 30 hours on a NVIDIA Titan Xp GPU.

4.2.3 Evaluation

The performance of our AutofocusNet model was evaluated on both simulated and real phantom and in vivo data. For the simulated testing data, we used the same process described in the Training data generation Section to produce images off-resonance artifacts. We compared our method to the autofocus method, which does not require field map acquisition or concomitant field calculation. To ensure a fair comparison, the autofocus method also used demodulated images at $M = 21$ different frequencies, matching our preprocessing in AutofocusNet. Quantitative evaluation metrics, including normalized root-mean-squared error (NRMSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM) were computed to assess the performance across methods.

For real data validation, we acquired phantom and in vivo data on a Siemens 1.5 T Avanto scanner with spiral trajectories. We compared our method to the autofocus method and the Chebyshev approximation-based semiautomatic method developed by Chen et al¹⁵⁵. The semiautomatic method requires knowledge of the field map and concomitant fields. For each scan, two single-shot spirals with an echo delay of 1 ms were first acquired, allowing the computation of a low-resolution field map. The concomitant fields can be calculated from theory based on the applied spiral gradients.

4.3 Results

On the simulated testing dataset, AutofocusNet outperformed the conventional autofocus method across all quantitative evaluation metrics. Specifically, AutofocusNet achieved an NRMSE of 0.0343, PSNR of 36.99, and SSIM of 0.9520, whereas the autofocus method achieved an NRMSE of 0.0596, PSNR of 32.62, and SSIM of 0.9284, as shown in the Figure 4-3A. Representative

images were shown in Figure 4-3B. The output of the autofocus method still showed residual blurring and off-resonance artifacts, particularly in regions with high off-resonance frequency. The output of AutofocusNet showed sharper anatomical structures and higher image quality.

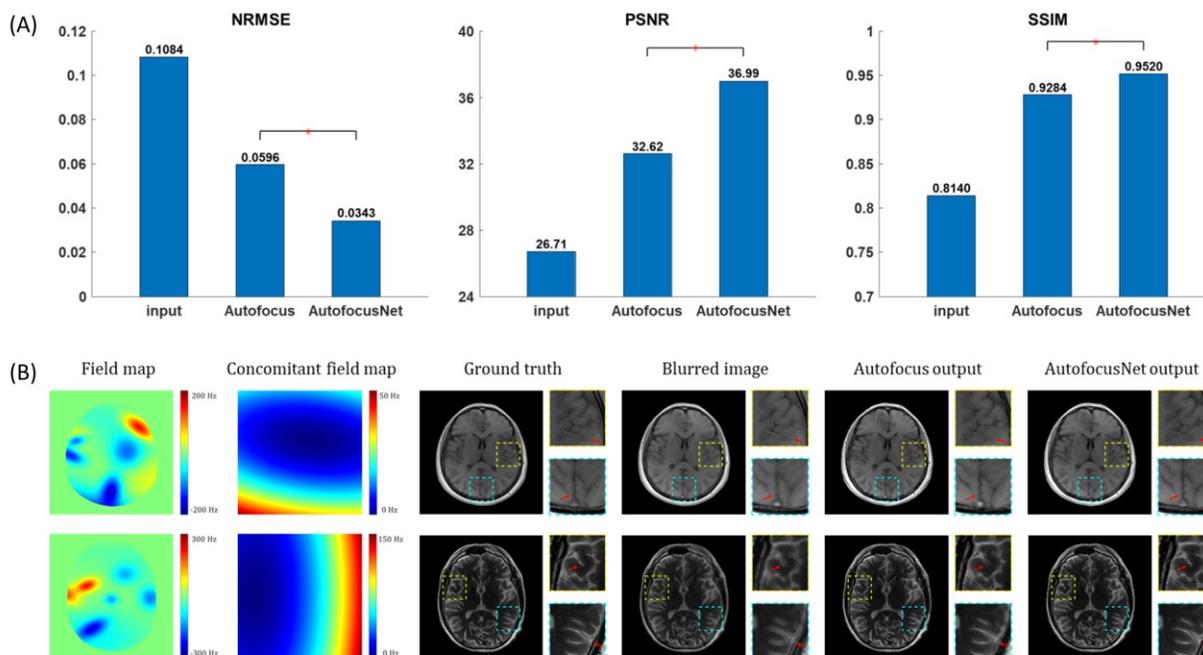


Figure 4-3: Quantitative evaluation results and representative images on the simulated test dataset. (A) AutofocusNet achieved the better NRMSE, PSNR, and SSIM. Asterisks indicate a significant difference ($p < 0.05$). (B) The output of the autofocus method still showed blurring and artifacts in regions with high off-resonance frequencies, while the output of AutofocusNet showed sharper structures with reduced artifacts (marked by red arrows).

The performance of proposed method was further evaluated using phantom and in vivo data, with representative images shown in Figure 4-4. These data were acquired using spiral trajectories with the readout length ranging from 8 ms to 16 ms, and the imaging slice off-center distance ranging from 3 cm to 8 cm along the transverse direction. The autofocus method showed residual blurring and artifacts. In certain regions, the output appeared even more degraded, likely due to the algorithm encountering spurious minima during optimizing the objective function¹⁵⁶. Compared to the autofocus method, AutofocusNet produced images with sharper structures and reduced artifacts. AutofocusNet achieved comparable results to the semiautomatic method without

using the field map or computing the concomitant field, making it a more efficient option for off-resonance correction.

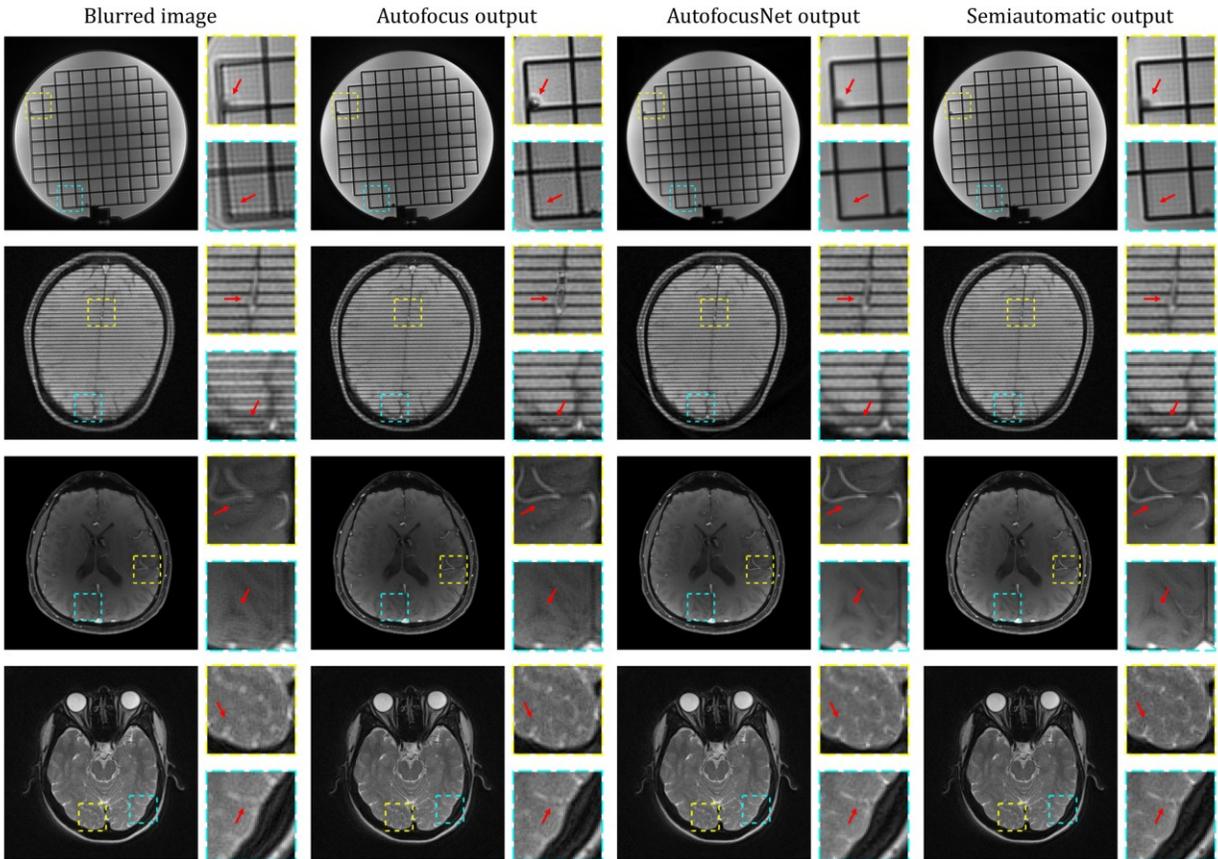


Figure 4-4: Comparison of model performance on phantom and in vivo data. Top-to-bottom rows: phantom image, in vivo head image with tagging lines to aid visualization of blurring, T1-weighted head image, T2-weighted head image. The autofocus method showed residual blurring and artifacts (marked by red arrows). The AutofocusNet achieved comparable results to the semiautomatic method.

4.4 Discussion

Off-resonance artifacts present a significant challenge in spiral MRI. In this study, we introduced AutofocusNet, a deep-learning-based approach for correcting off-resonance artifacts in spiral MRI without requiring a field map or concomitant field computation. No literature to date has discussed

simultaneous field inhomogeneity and concomitant field correction with CNNs. This model provides a practical and efficient alternative to conventional off-resonance correction methods.

On the simulated testing dataset, AutofocusNet outperformed the conventional autofocus method, achieving lower NRMSE, higher PSNR, and higher SSIM. On phantom and in vivo data, AutofocusNet maintained its performance and generated images with less blurring and fewer residual artifacts. The autofocus method relies on optimizing an objective function to minimize blurring. However, this process is sensitive to the choice of the objective function and can easily get trapped in spurious minima. AutofocusNet addressed this issue by training the network to learn the mapping between the demodulated blurred images and the clean image.

While AutofocusNet showed promising results, there are several areas for potential improvement and further exploration. First, the network architecture could be further optimized. Lim et al¹⁵⁷ showed that as incorporating depthwise separable convolutional layers can enhance the model's ability to capture relationships across channels and improve the deblurring performance. Second, integrating AutofocusNet within an iterative reconstruction framework may offer further benefits, as iterative methods could refine the correction process by leveraging AutofocusNet as a prior. This approach could be useful in cases with extreme field map variations, where a single correction might be insufficient. Another area for future work is to test AutofocusNet on low-field data, since Concomitant field effects are more pronounced in low-field MRI. Evaluating the model's performance on low-field data would help in assessing its robustness and effectiveness across different imaging conditions.

4.5 Conclusion

In conclusion, AutofocusNet offers a practical and effective solution for off-resonance correction in spiral MRI without using the field map or computing the concomitant fields during the

reconstruction. By training the model using synthetic data with simulated field inhomogeneity and concomitant field effects, our approach has demonstrated improved off-resonance correction than the conventional autofocus method.

Chapter 5: Complex Valued Cascading Cross-Domain Convolutional Neural Network for Reconstructing Undersampled CMR Images

5.1 Introduction

Cardiac magnetic resonance imaging (CMR) has emerged as an important tool for non-invasively evaluating cardiovascular diseases, offering unique insights into cardiac anatomy, function, tissue composition, and blood flow dynamics. Unlike other imaging modalities, CMR provides high-resolution images of the heart and surrounding vasculature without ionizing radiation, making it safer for repeated examinations¹⁵⁸. This characteristic, along with CMR's exceptional soft-tissue contrast, allows for detailed visualization of myocardial tissue properties and detection of subtle pathological changes. CMR is widely recommended for various clinical applications, including evaluating left ventricular ejection fraction, detecting diffuse myocardial fibrosis, and analyzing myocardial perfusion¹⁵⁹. The European Society of Cardiology recognizes CMR as the gold standard for the left ventricle ejection fraction measurement¹⁶⁰.

CMR encompasses a range of specialized imaging techniques for different clinical needs, allowing comprehensive cardiac assessments. Cine MRI provides real-time visualization of cardiac motion, which is essential for morphologic assessment of heart and valves¹⁶¹. Myocardial perfusion imaging evaluates the blood flow through the heart muscle, helping to detect ischemic areas indicative of coronary artery disease¹⁶². Late Gadolinium Enhancement imaging is widely used to identify myocardial scarring and fibrosis by visualizing pathologic areas where the contrast agents are retained¹⁶³. T1 and T2 mapping techniques quantify the characteristics of myocardial

tissue, allowing for early detection of pathological changes in the myocardium that may not be visible with other imaging methods¹⁶⁴.

Despite its comprehensive capabilities, CMR is inherently constrained by its slow acquisition speed due to the complicated imaging operations (e.g., breath-hold) and reliance on gated acquisition methods, where the data acquisition is synchronized with cardiac and respiratory cycles to minimize motion artifacts¹⁵⁹. The prolonged scan time leads to patient discomfort and compromised image quality, impacting diagnostic accuracy and limiting CMR's applicability in routine clinical settings. Consequently, accelerating CMR acquisition and reconstruction has become a primary research focus, aiming to make CMR more accessible and practical for both patients and clinicians.

Various advanced techniques have been developed to accelerate CMR acquisition by undersampling the k -space data, effectively reducing the scan time while retaining the image quality. Parallel imaging (PI) leverages multi-coil receiver arrays to acquire undersampled k -space data and exploit the data redundancy from the multi-coil acquisition. While PI techniques, like SENSE² and GRAPPA³, allows for substantial acceleration, they are often limited by noise amplification, particularly with higher acceleration ratios (R)¹⁶⁵. Compressed sensing (CS)⁵ is based on the assumption that MRI data is sparse or compressible in certain domains, such as wavelet or total variation domains. By enforcing sparsity constraints during the nonlinear reconstruction, CS can achieve high acceleration rates by recovering the image from incoherent aliasing artifacts. Moreover, the combination of PI and CS, known as PICS, has been developed to harness the strengths of both techniques¹⁶⁶. This hybrid approach achieves faster acquisition speeds by nearly two-fold than PI or CS alone. To date, these conventional methods have been investigated in several applications, for their ability to achieve good image quality with high

acceleration rates. However, the sophisticated reconstruction algorithms require extensive computational resources. Offline image reconstructions are typically performed within several minutes, thus limiting the clinical application of low latency real-time imaging.

In last ten years, various deep-learning-based methods have been developed for MR image reconstruction. Deep-learning-based methods show exceptional potential in undersampled CMR data reconstruction, offering superior image quality and significantly reducing reconstruction times. Different network architectures have been designed and adapted for CMR applications. Convolutional neural networks (CNNs) are widely applied in accelerated CMR reconstruction^{167,168}. By learning the mapping between undersampled images and fully-sampled images through the training data, CNNs have shown improved quality in reducing the aliasing artifacts. Generative adversarial networks (GANs) have also been employed in synthesizing high-fidelity CMR images^{169,170}. By utilizing the adversarial framework, GANs can achieve high-quality reconstructions that closely resemble fully sampled images. Recently, unrolled networks have gained increasing attention. These models mimic traditional iterative reconstruction algorithms by alternating between the networks and data consistency. For example, a combination of ESPIRiT reconstruction and a CNN that handles complex-valued data as separate channels was proposed and evaluated on retrospective 12× undersampled datasets¹⁷¹. However, the inherent interrelationship between the components of the complex k -space values has not been fully exploited. Thus, an improved complex-valued cascading network architecture has been introduced, where the architectural paradigm exhibited better performance compared to CS-based approaches, as demonstrated on the prospective undersampled datasets¹⁷². However, both the features of cascading and 4D require considerable computational demands, which may potentially exceed the resource constraints of commonplace server configurations.

In this work, we proposed a novel complex-valued cascading cross-domain convolutional neural network, termed as C³-Net, for an improved balance between computation demands and image quality on accelerated CMR datasets. The proposed network was first trained with fully and undersampled datasets and then compared to the results from other networks and against the fully sampled images as the reference.

5.2 Methods

5.2.1 Network architecture

Let \mathbf{x} represents a 2D complex-valued cardiac cine slice. Our purpose is to reconstruct \mathbf{x} from the multi-coil undersampled k -space \mathbf{y}_u , such that:

$$\mathbf{y}_u = \mathbf{uFSx} \quad (5-1)$$

where \mathbf{u} is the binary k -space undersampling mask, \mathbf{F} is the Fourier transform (FT), and \mathbf{S} is the sensitivity maps, which can be derived from the fully sampled calibration region through ESPIRiT¹²⁴.

When dealing with data acquired with an undersampling ratio beyond the capability of the coil hardware, Equation 5-1 is underdetermined. To tackle the ill-posed inversion problem, we propose a cascading convolutional neural network (CNN), which resembles a classical iterative algorithm with a fixed number of iterations. The proposed C³-Net alternates between the restoration step and the data consistency (DC) step, as shown in Figure 5-1. The restoration part consists of two major components: k -space subnetwork (K-Net) and image subnetwork (I-Net). At the i -th iteration, the K-Net generates an estimation of the true k -space:

$$\mathbf{y}_{\text{knet}} = f_{\text{knet}}(\mathbf{y}_{\text{in}}|\boldsymbol{\theta}_{\text{knet}}) \quad (5-2)$$

where f_{knet} is the forward mapping of the K-Net parameterized by $\boldsymbol{\theta}_{\text{knet}}$, and \mathbf{y}_{in} is the input k -space, of which \mathbf{y}_{u} is at the first iteration. Then, the output k -space \mathbf{y}_{knet} is transformed into the image domain and combined by the sensitivity maps $\mathbf{x}_{\text{in}} = \mathbf{S}^H \mathbf{F}^H \mathbf{y}_{\text{knet}}$, where H represents the conjugate transpose. The I-Net generates an estimation of the true image:

$$\mathbf{x}_{\text{inet}} = f_{\text{inet}}(\mathbf{x}_{\text{in}} | \boldsymbol{\theta}_{\text{inet}}) \quad (5-3)$$

where f_{inet} is the forward mapping of the I-Net parameterized by $\boldsymbol{\theta}_{\text{inet}}$. To incorporate the data consistency, for k -space entries that are initially missing, we use the predicted values from the I-Net; for the k -space entries that are initially sampled, we simply replace the predicted values with the original values:

$$\mathbf{y}_{\text{dc}} = \begin{cases} \mathbf{F} \mathbf{S} \mathbf{x}_{\text{inet}}, & \mathbf{u}_i = 0 \\ \mathbf{y}_{\text{u}}, & \mathbf{u}_i = 1 \end{cases} \quad (5-4)$$

Both the K-Net and I-Net use a U-Net⁹² as the backbone network structure. The complex U-Net (\mathbb{C} U-Net) consists of a series of encoding and decoding blocks. Each encoding or decoding block consists of two complex-valued convolutional layers (\mathbb{C} Conv)⁷⁶ with kernel size 3×3 . Each convolutional layer is followed by a radial instance normalization (IN)¹⁷³ and a complex-valued rectified linear unit (\mathbb{C} ReLU)⁷⁶. The radial IN scales the magnitude while maintaining the phase information, and the \mathbb{C} ReLU function activates the real and imaginary parts, separately. A 2×2 average pooling is applied at the end of each encoding block, while a 2×2 upsampling is performed at the start of each decoding block. The skip connections between the encoding block and the corresponding decoding block are important to expedite training and avoid vanishing gradients. The final output is generated using a \mathbb{C} Conv with kernel size 1×1 . The number of filters n_f of the first \mathbb{C} Conv is set to 8, and the number of iterations K is set to 2 in this work due to the limit of GPU memory.

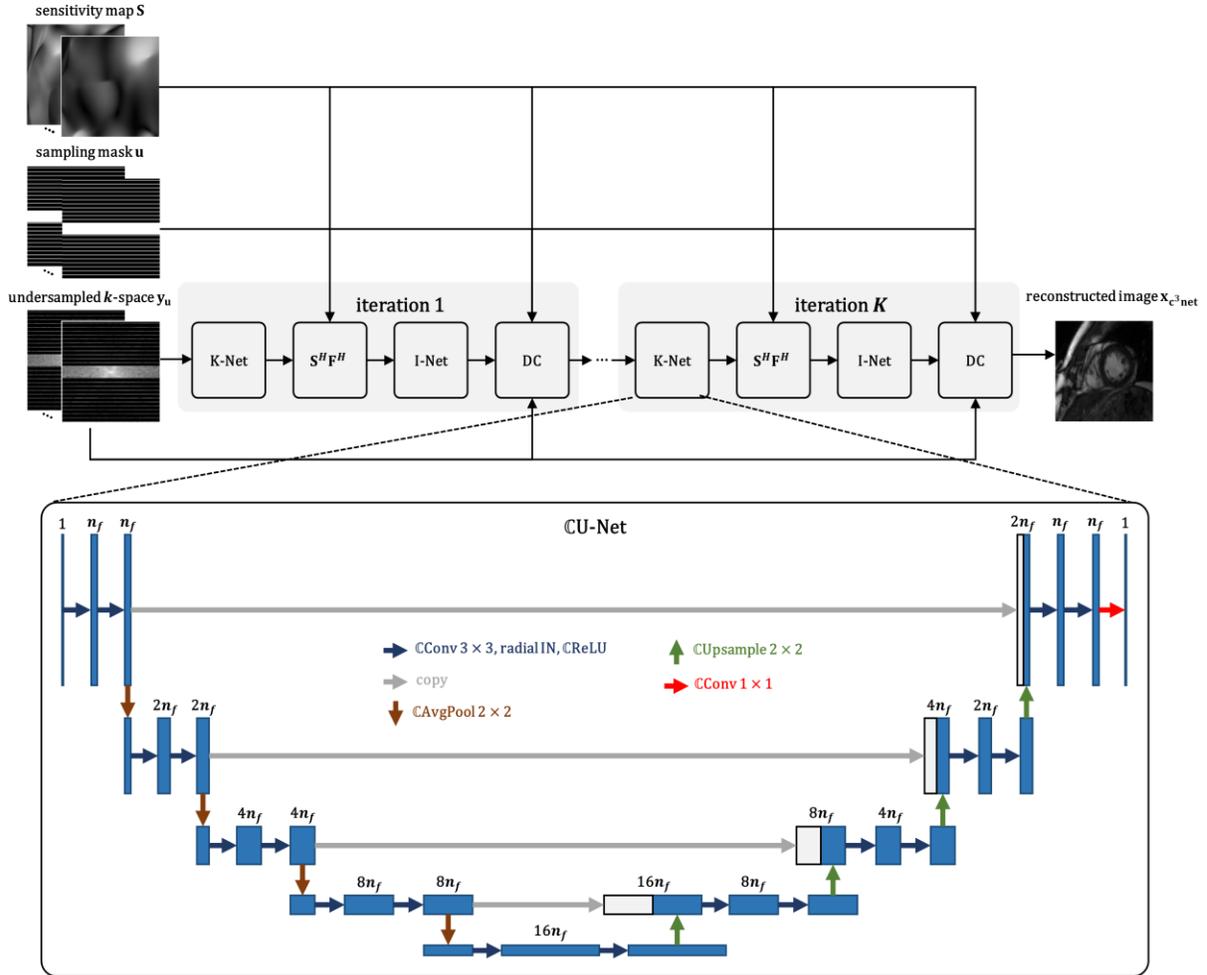


Figure 5-1: Illustration of the proposed framework for undersampled CMR image reconstruction. The C^3 -Net alternates between the restoration step and DC step. The restoration part consists of k -space subnetwork and image subnetwork, which use complex U-Net as the backbone network architecture.

5.2.2 Training

The training, validation, and testing datasets used in this work were built from the CMRxRecon challenge dataset (<https://cmrxrecon.github.io/>). The raw CMRxRecon training dataset includes 120 fully sampled multi-coil cine MRIs obtained on 3T scanners¹⁷⁴. Detailed descriptions can be found on the project website. We randomly divided the dataset into training (90 subjects), validation (10 subjects), and testing (20 subjects) subsets. For training, a total of 15192 2D slices were generated by splitting up each 4D sample (matrix size in readout direction, matrix size in

phase encoding direction, slice number, time frame). Random flipping along readout and phase encoding directions was employed as training augmentation to further expand the size of training dataset. The raw k -space data for each 2D slice was scaled to have its magnitude between 0 to 1. During training, the undersampling ratio R was randomly selected between 4 to 12 and the equispaced undersampling mask was generated on-the-fly for each 2D slice. The central 24 phase encoding lines were always fully sampled as the autocalibration signal (ACS) region. The sensitivity maps were pre-computed from the time-averaged ACS using ESPIRiT.

The C^3 -Net was implemented in the open-source machine learning library PyTorch⁸⁴. All subnetworks in the reconstruction pipeline were jointly trained in an end-to-end manner using a mixed L1 and structural similarity index (SSIM) loss⁹⁶:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N |\mathbf{x}_{c^3_{\text{net}}} - \mathbf{x}_{\text{ref}}| + \alpha(1 - \text{SSIM}(|\mathbf{x}_{c^3_{\text{net}}}|, |\mathbf{x}_{\text{ref}}|)) \quad (5-5)$$

where $\mathbf{x}_{c^3_{\text{net}}}$ is the complex-valued output of C^3 -Net, \mathbf{x}_{ref} is the complex-valued reference image reconstructed from the fully sampled k -space data, N is the total number of training pairs, and α is the weight parameter (empirically set to 1). Training was carried out with an Adam optimizer⁸⁵ for 50 epochs with a learning rate of 0.0001, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$. Due to the limit of GPU memory, the training batch size was set to 1.

5.2.3 Evaluation

To evaluate C^3 -Net performance compared to other deep learning methods, a baseline U-Net and a real-valued cascading cross-domain CNN, termed as C^2 -Net, were trained with the same training setup. The proposed C^3 -Net was also compared against the L1-ESPIRiT reconstruction using the code provided by the CMRxRecon organizers. For quantitative assessment, the peak signal-to-

noise ratio (PSNR), SSIM, and normalized mean squared error (NMSE) of the magnitude images were calculated.

5.3 Results

The comparison results of all methods are reported in Table 5-1. using quantitative metrics (PSNR, SSIM, and NMSE) across different acceleration ratios ($R = 4, 8, \text{ and } 10$). The proposed C^3 -Net almost yields the best image quality over ESPIRiT, U-Net, and C^2 -Net, especially at higher acceleration ratios, as indicated by the bold fonts which show the best reconstruction performance.

The visualization results of one short-axis and three long-axis views of cardiac MR images are depicted in Figure 5-2 and Figure 5-3, respectively. In Figure 5-2, ESPIRiT shows good image performance at $R = 4$ with less artifacts, because the acceleration ratio is far smaller than the number of coils used for data acquisition. As the acceleration ratio increases (e.g., 8 and 10), images reconstructed from ESPIRiT degrades, with noticeable residual aliasing artifacts. By contrast, the proposed method presents the best results in both spatial and temporal dimension, as can be clearly seen from the difference images with much reduced aliasing artifacts and from $x-t$ profiles where the cardiac motion is well preserved along temporal domain. Figure 5-3 provides examples of reconstruction results from 2-, 3-, and 4-chamber views at an acceleration ratio of 10. Our model excels, particularly on 2- and 4-chambers datasets, where the artifacts surrounding the heart wall and adipose regions are notably mitigated. The factors underpinning the suboptimal performance which were observed in 3-chamber images for both deep-learning-based methods will be discussed in the following section.

SAX ($R = 4$)	Zero-filling	L1-ESPIRiT	U-Net	C ² -Net	C ³ -Net
PSNR [dB]	29.77	38.52	36.28	37.22	38.53
SSIM	0.859	0.937	0.957	0.967	0.972
NMSE ($\times 10^2$)	8.411	1.407	1.872	1.876	1.636
SAX ($R = 8$)	Zero-filling	L1-ESPIRiT	U-Net	C ² -Net	C ³ -Net
PSNR [dB]	29.05	32.59	34.88	35.12	36.63
SSIM	0.845	0.849	0.941	0.946	0.956
NMSE ($\times 10^2$)	9.203	3.050	1.996	2.145	1.729
SAX ($R = 10$)	Zero-filling	L1-ESPIRiT	U-Net	C ² -Net	C ³ -Net
PSNR [dB]	28.89	31.75	34.49	34.41	36.19
SSIM	0.843	0.837	0.937	0.938	0.951
NMSE ($\times 10^2$)	9.727	3.941	1.969	2.518	1.661
LAX ($R = 4$)	Zero-filling	L1-ESPIRiT	U-Net	C ² -Net	C ³ -Net
PSNR [dB]	29.76	41.55	34.51	34.66	36.67
SSIM	0.821	0.949	0.930	0.896	0.927
NMSE ($\times 10^2$)	8.393	1.514	4.876	8.205	6.194
LAX ($R = 8$)	Zero-filling	L1-ESPIRiT	U-Net	C ² -Net	C ³ -Net
PSNR [dB]	29.08	32.70	33.09	32.74	33.12
SSIM	0.818	0.838	0.920	0.881	0.924
NMSE ($\times 10^2$)	8.590	3.574	4.058	7.706	4.043
LAX ($R = 10$)	Zero-filling	L1-ESPIRiT	U-Net	C ² -Net	C ³ -Net
PSNR [dB]	28.84	31.70	32.60	32.31	32.39
SSIM	0.815	0.824	0.915	0.878	0.919
NMSE ($\times 10^2$)	8.806	4.667	4.135	7.667	4.273

Table 5-1: Quantitative assessment of reconstruction performance on the CMRxRecon dataset.

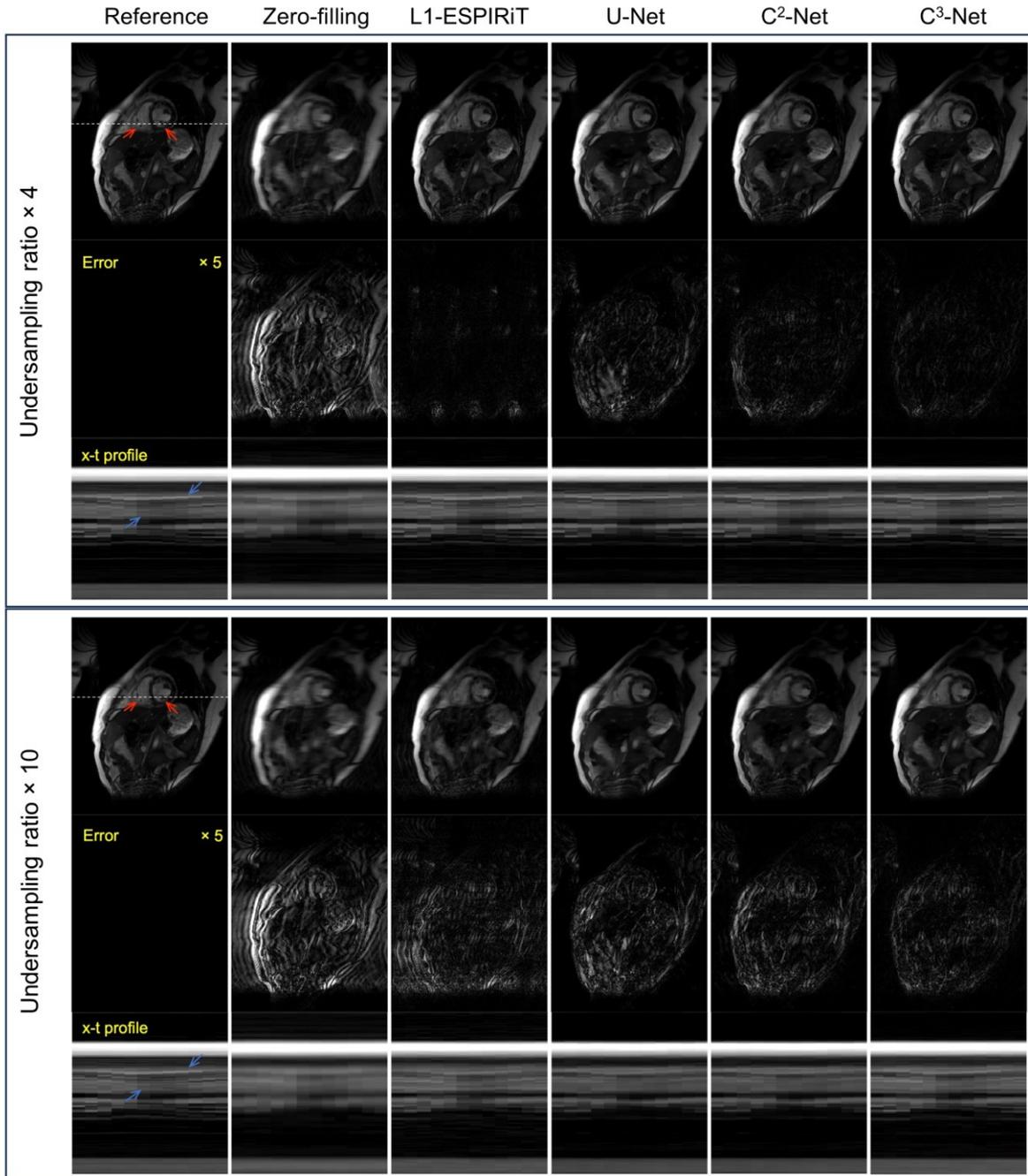


Figure 5-2: Comparison of one short-axis slice (end-diastole frame) generated using five reconstruction methods, from datasets retrospectively undersampled with accelerated ratios of 4 and 10. From left to right are images reconstructed from fully-sampled k -space as the reference, zero-filling, L1-ESPIRiT, U-Net, C²-Net and C³-Net, respectively. From top to bottom are magnitude image (top), absolute difference image relative to the fully sampled reference (middle), and x - t plots below the white lines (bottom). The error maps are windowed by scaling the image intensity by a factor of 5. The red arrows point to structures that show fine papillary muscles in ventricles and the blue arrows indicate preserved temporal fidelity.

5.4 Discussion

In this study, we designed and implemented a cascading network that operates in both the k -space domain and the image domain. The K-Net estimates the missing k -space values utilizing the information from neighboring sampled points. The I-Net reduces the residual aliasing artifacts and further improves the image quality. The DC layer ensures that the already acquired k -space samples remain unchanged and enhances the output image fidelity. Compared to the U-Net acting in the image domain without the DC layer, the cross-domain cascading networks (C²-Net and C³-Net) showed superior performance on the testing dataset and were less prone to generate unrealistic small structures on the output images. Also, all operations used in C³-Net were complex-valued, enabling the network to fully exploit the complex-valued input data⁷⁶. Compared to C²-Net which performs operations on the real and imaginary channels separately, C³-Net achieved better metrics and generated output with less residual artifacts.

For the DC layer, we employed a simple but effective operation in our study: replacing the predicted values with the original values for sampled k -space points. However, if nonnegligible noise exists in the acquisition, this method may fail, as shown in the 3-chamber view in Figure 5-3. Utilizing a linear combination of the predicted values and the original values weighted by a fixed or trainable parameter^{175,176} is expected to improve the network performance on noisy data.

The sampling pattern may play an important role in the performance of image reconstruction. Although in this work the same sampling mask was used for retrospective undersampled datasets, using a variable sampling pattern, such as Cartesian Poisson sampling, could potentially improve the de-aliasing performance when using CS- or DL-based approaches. Also, non-Cartesian sampling patterns, such as radial or spiral sampling, have been demonstrated to provide advantages of shorter scan time and higher motion insensitivity^{141,177} than Cartesian

acquisitions, which are critical for dynamic imaging. An approach to apply the proposed C^3 -Net to non-Cartesian scenarios involves pre-gridding the radial or spiral k -space data onto a Cartesian grid^{178,179} prior to inputting it into the K-Net. Subsequently, the training procedure remains consistent with that of the Cartesian case.

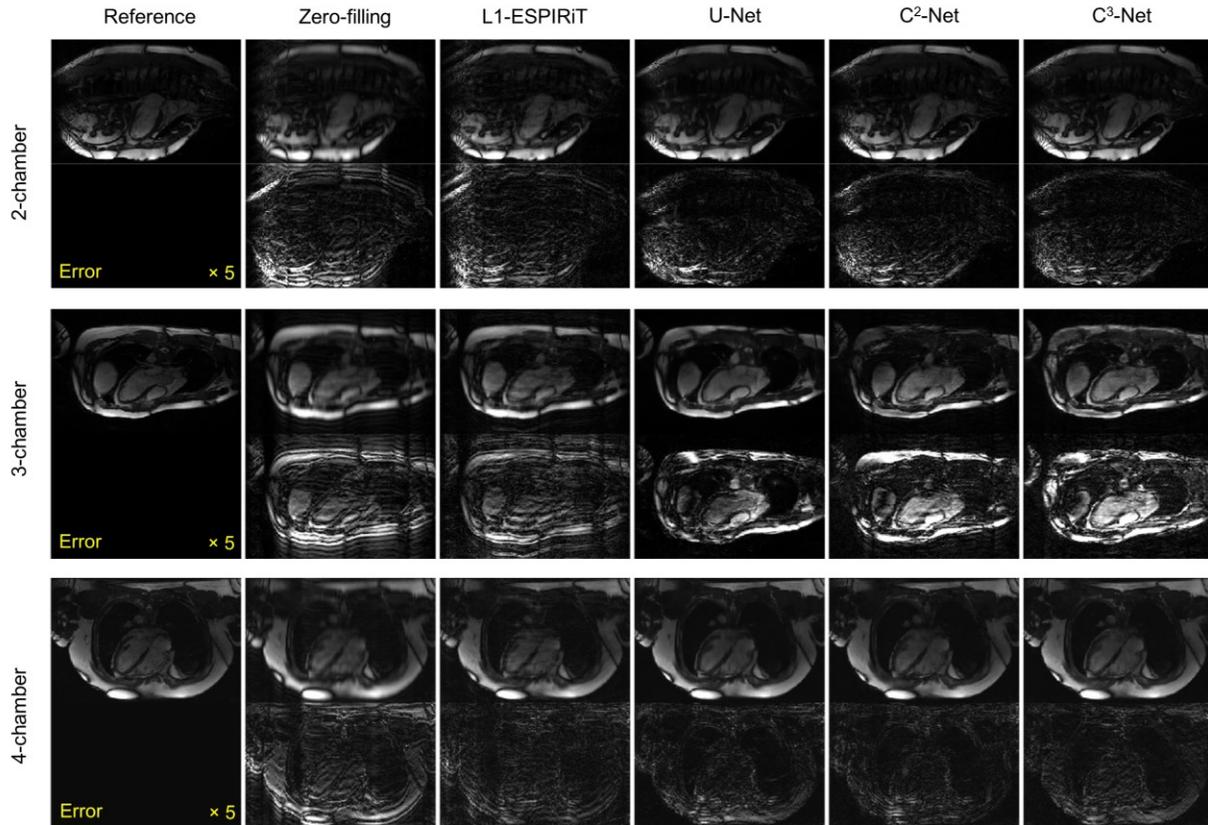


Figure 5-3: Comparison of long-axis cardiac images from 2-chamber (top), 3-chamber (middle), and 4-chamber views (bottom), reconstructed using fully-sampled k -space as the reference, zero-filling, L1-ESPIRiT, U-Net, C^2 -Net, and C^3 -Net, respectively. Absolute difference images relative to the fully sampled reference are shown for performance comparison, as well. The error maps are windowed by scaling the image intensity by a factor of 5.

5.5 Conclusion

The proposed C^3 -Net integrates both the complex-value of MR data and the coupled domain information (k -space domain and image domain) in the CNN model, providing a significant

improvement of image quality at high acceleration rates in comparison with the state-of-the-art methods (L1-ESPIRiT, U-Net, and C²-Net).

Chapter 6: Conclusions and Future Directions

6.1 Conclusions

This dissertation presents a series of deep learning models designed to address significant challenges in MRI, such as SNR issues, motion artifacts, off-resonance artifacts, and undersampled image reconstruction. These challenges often degrade MRI image quality and impact diagnostic accuracy, limiting its clinical potential. The proposed deep learning methods demonstrate improved performance over comparison methods in enhancing image quality and artifact reduction.

Key contributions include the introduction of a complex-valued denoising convolutional neural network (CDnCNN), which utilizes the complex nature of MRI data and integrates noise level information in the model to improve denoising performance in low-field MRI and address the spatially variant noise. A multi-task conditional GAN (MT-cGAN) is presented for simultaneous motion detection and compensation in brain MRI, leveraging multi-task learning framework to improve model robustness and performance on both tasks. For field inhomogeneity and concomitant field, a deep-learning-based method is presented to correct the off-resonance artifacts arising from both sources, without requiring field map acquisition or concomitant field calculation. Finally, the complex-valued cascading cross-domain network (C³-Net) demonstrates effective and efficient reconstruction for undersampled cardiac MRI, which can achieve higher acceleration rates than parallel imaging and compressed sensing while maintain image quality.

Together, these contributions show the potential of deep learning in MRI image processing and artifact reduction. The outcomes of this work, including the development of advanced models, training data simulation strategies, and streamlined deployment workflows, help to overcome

existing MRI shortcomings, offer more accurate and reliable diagnostic information, and broaden the application scope of MRI.

6.2 Future Directions

6.2.1 Inline implementations

Inline implementations of deep learning models can significantly expand its applicability in MRI scans. By integrating the models into existing scan workflows, they can process the data from the scanner in near real-time, bridging the gap between the data acquisition and image generation. This is crucial for dynamic imaging, such as cardiac MRI, where immediate visualization aids in diagnostic decisions. Inline implementations also allow dynamic adjustment based on the network output, such as performing a rescan if motion artifacts are detected on the images. There exist some examples of inline frameworks, like Gadgetron¹⁸⁰, Berkeley Advanced Reconstruction Toolbox – BART (<https://mricon.github.io/bart/>), and Siemens Framework for Image Reconstruction Environments (FIRE), enabling iterative reconstruction algorithms and deep learning models to be integrated into the scan workflow. While inline implementations offer numerous advantages, they also present challenges, such as the computational demands of real-time processing and the integration of advanced algorithms into existing old hardware.

6.2.2 Nonrigid motion compensation

In Chapter 3, a multi-task conditional GAN for motion detection and compensation is introduced, where the main focus is to address the rigid motion in brain imaging. However, nonrigid motion also presents a significant challenge in MRI, especially in dynamic imaging applications, such as cardiac and abdominal imaging. Unlike rigid motion, where the movement involves global

translation or rotation of an object, nonrigid motion refers to local deformation and displacement within the imaging field, resulting in more complex artifacts appearance. For example, abdominal MRI is often affected by respiratory motion, which introduces nonrigid deformation in organs such as liver, kidneys, and pancreas.

Simulating nonrigid motion is critical for developing and testing motion compensation models. Potential methods include modeling the motion by mathematical expressions, using motion fields from real patient data, or generating synthetic deformation fields. Figure 6-1 shows an example of applying synthetic global affine transformation and B-spline-based local deformation to an abdominal image. However, compared to rigid motion simulation, the computational demand for nonrigid motion simulation is much higher, particularly for 3D motion. Generating sufficient realistic data for training the deep-learning-based methods is a major bottleneck.

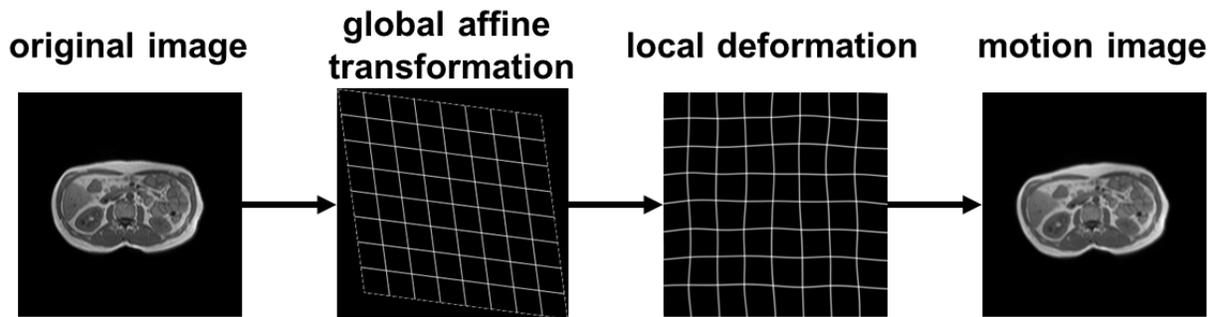


Figure 6-1: Nonrigid motion simulation with global affine transformation and local deformation.

6.2.3 Novel network architectures

Vision Transformers (ViTs)³⁸ have emerged as a powerful alternative to convolutional neural networks (CNNs) in image processing tasks. ViTs adapt the Transformer architecture, which is first developed for natural language processing, to image analysis. Unlike CNNs, which rely on convolutions to capture local spatial relationships, ViTs use self-attention mechanisms to model

both local and global dependencies, allowing them to excel in tasks requiring long-range context. The core of a ViT is the self-attention, which enables the network to focus on different parts of the input image based on their relevance to the task. The self-attention mechanism calculates relationships between every pair of input tokens, capturing both local and long-range dependencies. In ViTs, the input image is divided into non-overlapping patches, which are flattened and embedded into a sequence of tokens. These tokens are then passed through the ViTs' layers, where the self-attention operation is applied.

Compared to CNNs, ViTs provide several benefits. CNNs inherently focus on local features due to the limited receptive field of convolutional filters. While deeper layers capture broader context, this hierarchical approach can lose finer details. On the other hand, ViTs use self-attention to model relationships across the entire image at every layer, making them well-suited for tasks requiring holistic understanding. ViTs also scale effectively with larger datasets and higher computational resources. CNNs typically require architectural modifications for scalability (e.g., skip connections), ViTs maintain a consistent structure and can achieve improved performance simply by increasing the model size and training data.

In the context of MRI, where data often includes complex spatial and frequency relationships, ViTs present an exciting opportunity to overcome some limitations of CNNs. Their ability to model global relationships and process global context makes them a promising tool for tasks like image reconstruction and artifact correction. One major challenge for their application is that ViTs typically require large datasets for training to achieve optimal performance, which can be difficult in the field of MRI.

Appendix – List of Publications

First Author Manuscripts

1. **Dou Q**, Yan K, Chen S, Wang Z, Feng X, Meyer CH. C³-Net: Complex-Valued Cascading Cross-Domain Convolutional Neural Network for Reconstructing Undersampled CMR Images. *International Workshop on Statistical Atlases and Computational Models of the Heart* (pp. 390-399). Cham: Springer Nature Switzerland, 2023.
2. **Dou Q**, Wang Z, Feng X, Campbell-Washburn AE, Mugler JP III, Meyer CH. MRI denoising with a non-blind deep complex-valued convolutional neural network. *NMR in Biomedicine*. 2024;e5291.
3. **Dou Q**, Feng X, Meyer CH. Multi-Task Learning for Simultaneous Motion Detection and Compensation in Brain Imaging. In preparation (Target: *Magn Reson Med*).

Other Manuscripts

1. Feng X, **Dou Q**, Tustison N, Meyer CH. Brain tumor segmentation with uncertainty estimation and overall survival prediction. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Revised Selected Papers, Part I 5* (pp. 304-314). Springer International Publishing, 2010.

Conference Abstracts

1. **Dou Q**, Feng X, Wang Z, Weller DS, Meyer CH. Deep learning motion compensation for Cartesian and spiral trajectories. *ISMRM Scientific Meeting & Exhibition* (Vol. 4447). 2019.

2. **Dou Q**, Feng X, Wang Z, Weller DS, Meyer CH. Deep learning motion compensation for Cartesian and spiral trajectories. *ISMRM Workshop on: Data Sampling & Image Reconstruction*. 2020.
3. **Dou Q**, Wang Z, Feng X, Mugler JP III, Meyer CH. Retrospective motion compensation for spiral brain imaging with a deep convolutional neural network. *ISMRM Scientific Meeting & Exhibition* (Vol. 1359). 2021.
4. **Dou Q**, Feng X, Patel SH, Meyer CH. Prognostic value of MR imaging features derived from automatic segmentation in glioblastoma. *ISMRM Scientific Meeting & Exhibition* (Vol. 3259). 2021.
5. **Dou Q**, Chen Q, Rong Y, Feng X. Patch-Based DCNN Method for CBCT Image Enhancement. *International Journal of Radiation Oncology, Biology, Physics*. 2021;111(3): e90-e91.
6. **Dou Q**, Wang Z, Feng X, Ramasawmy R, Mugler JP, Campbell-Washburn AE, Meyer CH. Low-field MRI denoising with a deep complex-valued convolutional neural network. *ISMRM Workshop on: Low Field MRI*. 2022.
7. Yan K, Wang Z, **Dou Q**, Chen S, Meyer CH. Applying advanced denoisers to enhance highly undersampled MRI reconstruction under plug-and-play ADMM framework. *ISMRM Scientific Meeting & Exhibition* (Vol. 1163). 2022.
8. **Dou Q**, Feng X, Meyer CH. Deep learning-based brain MRI reconstruction with realistic noise. *ISMRM Scientific Meeting & Exhibition* (Vol. 3470). 2022.
9. **Dou Q**, Wang Z, Feng X, Meyer CH. Automatic Off-Resonance Correction for Spiral Imaging with a Convolutional Neural Network. *ISMRM Scientific Meeting & Exhibition* (Vol. 5022). 2022.

10. **Dou Q**, Wang Z, Feng X, Meyer CH. MRI Denoising with a Non-Blind Deep Complex-Valued Convolutional Neural Network. *ISMRM Scientific Meeting & Exhibition* (Vol. 3887). 2023.

Invention Disclosures

1. **Dou Q**, Wang Z, Feng X, Mugler JP III, Meyer CH. Motion compensation for MRI imaging. *U.S. Patent Application No. 17/733,967*.
2. **Dou Q**, Wang Z, Feng X, Meyer CH. Method and system for low-field MRI denoising with a deep complex-valued convolutional neural network. *U.S. Patent Application No. 18/305,296*.

Bibliography

1. Fessler JA, Sutton BP. Nonuniform fast Fourier transforms using min-max interpolation. *IEEE Trans Signal Process.* 2003;51(2):560-574.
2. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: Sensitivity encoding for fast MRI. *Magn Reson Med.* 1999;42(5):952-962.
3. Griswold MA, Jakob PM, Heidemann RM, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn Reson Med.* 2002;47(6):1202-1210.
4. Lustig M, Donoho DL, Santos JM, Pauly JM. Compressed Sensing MRI. *IEEE Signal Process Mag.* 2008;25(2):72-82.
5. Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn Reson Med.* 2007;58(6):1182-1195.
6. National Electrical Manufacturers Association. NEMA Standards Publication MS1–2008. Determination of Signal-to-Noise Ratio (SNR) in Diagnostic Magnetic Resonance Imaging. Published online 2008.
7. Menon V, Lim KO, Anderson JH, Johnson J, Pfefferbaum A. Design and efficacy of a head-coil bite bar for reducing movement-related artifacts during functional MRI scanning. *Behav Res Methods Instrum Comput.* 1997;29(4):589-594.
8. Bettinardi V, Scardaoni R, Gilardi MC, et al. Head holder for PET, CT, and MR studies. *J Comput Assist Tomogr.* 1991;15(5):886-892.
9. Malviya S, Voepel-Lewis T, Eldevik OP, Rockwell DT, Wong JH, Tait AR. Sedation and general anaesthesia in children undergoing MRI and CT: adverse events and outcomes. *Br J Anaesth.* 2000;84(6):743-748.
10. Windram J, Grosse-Wortmann L, Shariat M, Greer ML, Crawford MW, Yoo SJ. Cardiovascular MRI without sedation or general anesthesia using a feed-and-sleep technique in neonates and infants. *Pediatr Radiol.* 2012;42(2):183-187.
11. Lanzer P, Barta C, Botvinick EH, Wiesendanger HU, Modin G, Higgins CB. ECG-synchronized cardiac MR imaging: method and evaluation. *Radiology.* 1985;155(3):681-686.
12. Ehman RL, McNamara MT, Pallack M, Hricak H, Higgins CB. Magnetic resonance imaging with respiratory gating: techniques and advantages. *AJR Am J Roentgenol.* 1984;143(6):1175-1182.

13. Qin L, van Gelderen P, Derbyshire JA, et al. Prospective head-movement correction for high-resolution MRI using an in-bore optical tracking system. *Magn Reson Med.* 2009;62(4):924-934.
14. Zaitsev M, Dold C, Sakas G, Hennig J, Speck O. Magnetic resonance imaging of freely moving objects: prospective real-time motion correction using an external optical motion tracking system. *NeuroImage.* 2006;31(3):1038-1050.
15. Ehman RL, Felmlee JP. Adaptive technique for high-definition MR imaging of moving structures. *Radiology.* 1989;173(1):255-263.
16. Hu X, Kim SG. Reduction of signal fluctuation in functional MRI using navigator echoes. *Magn Reson Med.* 1994;31(5):495-503.
17. Sachs TS, Meyer CH, Hu BS, Kohli J, Nishimura DG, Macovski A. Real-time motion detection in spiral MRI using navigators. *Magn Reson Med.* 1994;32(5):639-645.
18. Fu ZW, Wang Y, Grimm RC, et al. Orbital navigator echoes for motion measurements in magnetic resonance imaging. *Magn Reson Med.* 1995;34(5):746-753.
19. Welch EB, Manduca A, Grimm RC, Ward HA, Jack CR. Spherical navigator echoes for full 3D rigid body motion measurement in MRI. *Magn Reson Med.* 2002;47(1):32-41.
20. van der Kouwe AJW, Benner T, Dale AM. Real-time rigid body motion correction and shimming using cloverleaf navigators. *Magn Reson Med.* 2006;56(5):1019-1032.
21. Maclaren J, Herbst M, Speck O, Zaitsev M. Prospective motion correction in brain imaging: a review. *Magn Reson Med.* 2013;69(3):621-636.
22. Johnson PM, Liu J, Wade T, Tavallaei MA, Drangova M. Retrospective 3D motion correction using spherical navigator echoes. *Magn Reson Imaging.* 2016;34(9):1274-1282.
23. Robson PM, Grant AK, Madhuranthakam AJ, Lattanzi R, Sodickson DK, McKenzie CA. Comprehensive Quantification of Signal-to-Noise Ratio and g-Factor for Image-Based and k-Space-Based Parallel Imaging Reconstructions. *Magn Reson Med Off J Soc Magn Reson Med Soc Magn Reson Med.* 2008;60(4):895.
24. Aja-Fernández S, Vegas-Sánchez-Ferrero G, Tristán-Vega A. Noise estimation in parallel MRI: GRAPPA and SENSE. *Magn Reson Imaging.* 2014;32(3):281-290.
25. Dietrich O, Raya JG, Reeder SB, Reiser MF, Schoenberg SO. Measurement of signal-to-noise ratios in MR images: Influence of multichannel coils, parallel imaging, and reconstruction filters. *J Magn Reson Imaging.* 2007;26(2):375-385.
26. Haskell MW, Nielsen JF, Noll DC. Off-resonance artifact correction for MRI: A review. *NMR Biomed.* 2023;36(5):e4867.

27. Noll DC, Meyer CH, Pauly JM, Nishimura DG, Macovski A. A homogeneity correction method for magnetic resonance imaging with time-varying gradients. *IEEE Trans Med Imaging*. 1991;10(4):629-637.
28. Sutton BP, Noll DC, Fessler JA. Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities. *IEEE Trans Med Imaging*. 2003;22(2):178-188.
29. Man LC, Pauly JM, Macovski A. Multifrequency interpolation for fast off-resonance correction. *Magn Reson Med*. 1997;37(5):785-792.
30. Wang D, Zwart NR, Pipe JG. Joint water-fat separation and deblurring for spiral imaging. *Magn Reson Med*. 2018;79(6):3218-3228.
31. Noll DC, Pauly JM, Meyer CH, Nishimura DG, Macovskj A. Deblurring for non-2D fourier transform magnetic resonance imaging. *Magn Reson Med*. 1992;25(2):319-333.
32. Man LC, Pauly JM, Macovski A. Improved automatic off-resonance correction without a field map in spiral imaging. *Magn Reson Med*. 1997;37(6):906-913.
33. Bernstein MA, Zhou XJ, Polzin JA, et al. Concomitant gradient terms in phase contrast MR: Analysis and correction. *Magn Reson Med*. 1998;39(2):300-308.
34. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks. Published online June 10, 2014.
35. Mirza M, Osindero S. Conditional Generative Adversarial Nets. Published online November 6, 2014.
36. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-Image Translation with Conditional Adversarial Networks. Published online November 26, 2018.
37. Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need. Published online August 2, 2023.
38. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. Published online June 3, 2021.
39. Hammernik K, Klatzer T, Kobler E, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med*. 2018;79(6):3055-3071.
40. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. *Nature*. 2018;555(7697):487-492.
41. Heckel R, Jacob M, Chaudhari A, Perlman O, Shimron E. Deep learning for accelerated and robust MRI reconstruction. *Magn Reson Mater Phys Biol Med*. 2024;37(3):335-368.
42. Chaudhari AS, Fang Z, Kogan F, et al. Super-resolution musculoskeletal MRI using deep learning. *Magn Reson Med*. 2018;80(5):2139-2154.

43. Lyu Q, Shan H, Steber C, et al. Multi-Contrast Super-Resolution MRI Through a Progressive Network. *IEEE Trans Med Imaging*. 2020;39(9):2738-2749.
44. Ji Z, Zou B, Kui X, et al. Deep learning-based magnetic resonance image super-resolution: a survey. *Neural Comput Appl*. 2024;36(21):12725-12752.
45. Dar SUH, Yurt M, Karacan L, Erdem A, Erdem E, Çukur T. Image Synthesis in Multi-Contrast MRI With Conditional Generative Adversarial Networks. *IEEE Trans Med Imaging*. 2019;38(10):2375-2388.
46. Wang K, Doneva M, Meineke J, et al. High-fidelity direct contrast synthesis from magnetic resonance fingerprinting. *Magn Reson Med*. 2023;90(5):2116-2129.
47. Dai X, Lei Y, Fu Y, et al. Multimodal MRI synthesis using unified generative adversarial networks. *Med Phys*. 2020;47(12):6343-6354.
48. Ayde R, Vornehm M, Zhao Y, Knoll F, Wu EX, Sarracanie M. MRI at low field: A review of software solutions for improving SNR. *NMR Biomed*. Published online October 7, 2024:e5268.
49. Spieker V, Eichhorn H, Hammernik K, et al. Deep Learning for Retrospective Motion Correction in MRI: A Comprehensive Review. *IEEE Trans Med Imaging*. 2024;43(2):846-859.
50. Zhang Q, Ruan G, Yang W, et al. MRI Gibbs-ringing artifact reduction by means of machine learning using convolutional neural networks. *Magn Reson Med*. 2019;82(6):2133-2145.
51. Marques JP, Simonis FFJ, Webb AG. Low-field MRI: An MR physics perspective. *J Magn Reson Imaging JMRI*. 2019;49(6):1528-1542.
52. Campbell-Washburn AE, Ramasawmy R, Restivo MC, et al. Opportunities in Interventional and Diagnostic Imaging by Using High-Performance Low-Field-Strength MRI. *Radiology*. 2019;293(2):384-393.
53. Wald LL, McDaniel PC, Witzel T, Stockmann JP, Cooley CZ. Low-cost and portable MRI. *J Magn Reson Imaging JMRI*. 2020;52(3):686-696.
54. Sheth KN, Mazurek MH, Yuen MM, et al. Assessment of Brain Injury Using Portable, Low-Field Magnetic Resonance Imaging at the Bedside of Critically Ill Patients. *JAMA Neurol*. 2020;78(1):41-47.
55. Edelstein WA, Bottomley PA, Pfeifer LM. A signal-to-noise calibration procedure for NMR imaging systems. *Med Phys*. 1984;11(2):180-185.
56. Gudbjartsson H, Patz S. The Rician Distribution of Noisy MRI Data. *Magn Reson Med*. 1995;34(6):910.

57. Buades A, Coll B, Morel JM. A non-local algorithm for image denoising. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. Vol 2. ; 2005:60-65 vol. 2.
58. Darbon J, Cunha A, Chan TF, Osher S, Jensen GJ. Fast nonlocal filtering applied to electron cryomicroscopy. In: *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. ; 2008:1331-1334.
59. Buades A, Coll B, Morel JM. Non-Local Means Denoising. *Image Process Line*. 2011;1:208-212.
60. Wiest-Daesslé N, Prima S, Coupé P, Morrissey SP, Barillot C. Rician noise removal by non-Local Means filtering for low signal-to-noise ratio MRI: applications to DT-MRI. *Med Image Comput Comput-Assist Interv MICCAI Int Conf Med Image Comput Comput-Assist Interv*. 2008;11(Pt 2):171-179.
61. Coupe P, Yger P, Prima S, Hellier P, Kervrann C, Barillot C. An Optimized Blockwise Nonlocal Means Denoising Filter for 3-D Magnetic Resonance Images. *IEEE Trans Med Imaging*. 2008;27(4):425-441.
62. Dabov K, Foi A, Katkovnik V, Egiazarian K. Image denoising with block-matching and 3D filtering. In: *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*. Vol 6064. SPIE; 2006:354-365.
63. Dabov K, Foi A, Katkovnik V, Egiazarian K. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. *IEEE Trans Image Process*. 2007;16(8):2080-2095.
64. Eksioğlu EM. Decoupled Algorithm for MRI Reconstruction Using Nonlocal Block Matching Model: BM3D-MRI. *J Math Imaging Vis*. 2016;56(3):430-440.
65. Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Trans Image Process*. 2017;26(7):3142-3155.
66. Zhang K, Zuo W, Zhang L. FFDNet: Toward a Fast and Flexible Solution for CNN based Image Denoising. Published online May 22, 2018.
67. Quan Y, Chen Y, Shao Y, Teng H, Xu Y, Ji H. Image denoising using complex-valued deep CNN. *Pattern Recognit*. 2021;111:107639.
68. Manjón JV, Coupé P. MRI denoising using deep learning. In: *International Workshop on Patch-Based Techniques in Medical Imaging*. Springer; 2018:12-19.
69. Tripathi PC, Bag S. CNN-DMRI: A Convolutional Neural Network for Denoising of Magnetic Resonance Images. *Pattern Recognit Lett*. 2020;135:57-63.
70. Li S, Zhou J, Liang D, Liu Q. MRI denoising using progressively distribution-based neural network. *Magn Reson Imaging*. 2020;71:55-68.

71. Tian M, Song K. Boosting Magnetic Resonance Image Denoising With Generative Adversarial Networks. *IEEE Access*. 2021;9:62266-62275.
72. Koonjoo N, Zhu B, Bagnall GC, Bhutto D, Rosen MS. Boosting the signal-to-noise of low-field MRI with deep learning image reconstruction. *Sci Rep*. 2021;11(1):8248.
73. Nitta T. On the critical points of the complex-valued neural network. In: *Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP '02. Vol 3.* ; 2002:1099-1103 vol.3.
74. Arjovsky M, Shah A, Bengio Y. Unitary evolution recurrent neural networks. In: *International Conference on Machine Learning*. PMLR; 2016:1120-1128.
75. Danihelka I, Wayne G, Uria B, Kalchbrenner N, Graves A. Associative long short-term memory. In: *International Conference on Machine Learning*. PMLR; 2016:1986-1994.
76. Trabelsi C, Bilaniuk O, Zhang Y, et al. Deep Complex Networks. *ArXiv170509792 Cs*. Published online February 25, 2018.
77. Wang S, Cheng H, Ying L, et al. DeepcomplexMRI: Exploiting deep residual network for fast parallel MR imaging with complex convolution. *Magn Reson Imaging*. 2020;68:136-147.
78. El-Rewaidy H, Neisius U, Mancio J, et al. Deep complex convolutional network for fast reconstruction of 3D late gadolinium enhancement cardiac MRI. *NMR Biomed*. 2020;33(7):e4312.
79. Cole E, Cheng J, Pauly J, Vasanawala S. Analysis of deep complex-valued convolutional neural networks for MRI reconstruction and phase-focused applications. *Magn Reson Med*. 2021;86(2):1093-1109.
80. Zbontar J, Knoll F, Sriram A, et al. fastMRI: An Open Dataset and Benchmarks for Accelerated MRI. Published online December 11, 2019.
81. Webb A, O'Reilly T. Tackling SNR at low-field: a review of hardware approaches for point-of-care systems. *Magn Reson Mater Phys Biol Med*. 2023;36(3):375-393.
82. Walsh DO, Gmitro AF, Marcellin MW. Adaptive reconstruction of phased array MR imagery. *Magn Reson Med*. 2000;43(5):682-690.
83. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *International Conference on Machine Learning*. PMLR; 2015:448-456.
84. Paszke A, Gross S, Massa F, et al. Pytorch: An imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst*. 2019;32.

85. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *ArXiv14126980 Cs*. Published online January 29, 2017.
86. Donoho DL, Johnstone IM. Ideal Spatial Adaptation by Wavelet Shrinkage. *Biometrika*. 1994;81(3):425-455.
87. Walt S van der, Schönberger JL, Nunez-Iglesias J, et al. scikit-image: image processing in Python. *PeerJ*. 2014;2:e453.
88. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13(4):600-612.
89. Lyu M, Mei L, Huang S, et al. M4Raw: A multi-contrast, multi-repetition, multi-channel MRI k-space dataset for low-field MRI research. *Sci Data*. 2023;10(1):264.
90. Wang Z, Allen SP, Feng X, Mugler JP, Meyer CH. SPRING-RIO TSE: 2D T2 -Weighted Turbo Spin-Echo brain imaging using SPiral RINGs with retraced in/out trajectories. *Magn Reson Med*. 2022;88(2):601-616.
91. Otsu N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans Syst Man Cybern*. 1979;9(1):62-66.
92. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2015:234-241.
93. Feng X, Tustison NJ, Patel SH, Meyer CH. Brain Tumor Segmentation Using an Ensemble of 3D U-Nets and Overall Survival Prediction Using Radiomic Features. *Front Comput Neurosci*. 2020;14.
94. Hyun CM, Kim HP, Lee SM, Lee S, Seo JK. Deep learning for undersampled MRI reconstruction. *Phys Med Biol*. 2018;63(13):135007.
95. Souza R, Bento M, Nogovitsyn N, et al. Dual-domain cascade of U-nets for multi-channel magnetic resonance image reconstruction. *Magn Reson Imaging*. 2020;71:140-153.
96. Zhao H, Gallo O, Frosio I, Kautz J. Loss Functions for Image Restoration With Neural Networks. *IEEE Trans Comput Imaging*. 2017;3(1):47-57.
97. Ledig C, Theis L, Huszár F, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. ; 2017:105-114.
98. Yang Q, Yan P, Zhang Y, et al. Low-Dose CT Image Denoising Using a Generative Adversarial Network With Wasserstein Distance and Perceptual Loss. *IEEE Trans Med Imaging*. 2018;37(6):1348-1357.

99. Wang Y, Song X, Chen K. Channel and Space Attention Neural Network for Image Denoising. *IEEE Signal Process Lett.* 2021;28:424-428.
100. Delakis I, Hammad O, Kitney RI. Wavelet-based de-noising algorithm for images acquired with parallel magnetic resonance imaging (MRI). *Phys Med Biol.* 2007;52(13):3741-3751.
101. Guo W, Huang F. Adaptive total variation based filtering for MRI images with spatially inhomogeneous noise and artifacts. In: *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro.* ; 2009:101-104.
102. Coupé P, Manjón JV, Gedamu E, Arnold D, Robles M, Collins DL. Robust Rician noise estimation for MR images. *Med Image Anal.* 2010;14(4):483-493.
103. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. *Commun ACM.* 2020;63(11):139-144.
104. Mardani M, Gong E, Cheng JY, et al. Deep Generative Adversarial Neural Networks for Compressive Sensing MRI. *IEEE Trans Med Imaging.* 2019;38(1):167-179.
105. Koch KM, Sherafati M, Arpinar VE, et al. Analysis and Evaluation of a Deep Learning Reconstruction Approach with Denoising for Orthopedic MRI. *Radiol Artif Intell.* 2021;3(6):e200278.
106. Karkalousos D, Noteboom S, Hulst HE, Vos FM, Caan MWA. Assessment of data consistency through cascades of independently recurrent inference machines for fast and robust accelerated MRI reconstruction. *Phys Med Biol.* 2022;67(12).
107. Dosenbach NUF, Koller JM, Earl EA, et al. Real-time motion analytics during brain MRI improve data quality and reduce costs. *NeuroImage.* 2017;161:80-93.
108. Andre JB, Bresnahan BW, Mossa-Basha M, et al. Toward Quantifying the Prevalence, Severity, and Cost Associated With Patient Motion During Clinical MR Examinations. *J Am Coll Radiol JACR.* 2015;12(7):689-695.
109. Atkinson D, Hill DLG, Stoyle PNR, Summers PE, Keevil SF. Automatic correction of motion artifacts in magnetic resonance images using an entropy focus criterion. *IEEE Trans Med Imaging.* 1997;16(6):903-910.
110. Loktyushin A, Nickisch H, Pohmann R, Schölkopf B. Blind retrospective motion correction of MR images. *Magn Reson Med.* 2013;70(6):1608-1618.
111. Johnson PM, Drangova M. Conditional generative adversarial network for 3D rigid-body motion correction in MRI. *Magn Reson Med.* 2019;82(3):901-910.
112. Küstner T, Armanious K, Yang J, Yang B, Schick F, Gatidis S. Retrospective correction of motion-affected MR images using deep learning frameworks. *Magn Reson Med.* 2019;82(4):1527-1540.

113. Lee J, Kim B, Park H. MC2-Net: motion correction network for multi-contrast brain MRI. *Magn Reson Med.* 2021;86(2):1077-1092.
114. Pawar K, Chen Z, Shah NJ, Egan GF. Suppressing motion artefacts in MRI using an Inception-ResNet network with motion simulation augmentation. *NMR Biomed.* 2022;35(4):e4225.
115. Hewlett M, Petrov I, Johnson PM, Drangova M. Deep-learning-based motion correction using multichannel MRI data: a study using simulated artifacts in the fastMRI dataset. *NMR Biomed.* 2024;37(10):e5179.
116. Vakli P, Weiss B, Szalma J, et al. Automatic brain MRI motion artifact detection based on end-to-end deep learning is similarly effective as traditional machine learning trained on image quality metrics. *Med Image Anal.* 2023;88:102850.
117. Fantini I, Yasuda C, Bento M, Rittner L, Cendes F, Lotufo R. Automatic MR image quality evaluation using a Deep CNN: A reference-free method to rate motion artifacts in neuroimaging. *Comput Med Imaging Graph.* 2021;90:101897.
118. Wang G, Gong E, Banerjee S, et al. Synthesize High-Quality Multi-Contrast Magnetic Resonance Imaging From Multi-Echo Acquisition Using Multi-Task Deep Generative Model. *IEEE Trans Med Imaging.* 2020;39(10):3089-3099.
119. Zhou Y, Chen H, Li Y, et al. Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images. *Med Image Anal.* 2021;70:101918.
120. Zhang Y, Yang Q. An overview of multi-task learning. *Natl Sci Rev.* 2018;5(1):30-43.
121. Zhang Y, Yang Q. A Survey on Multi-Task Learning. *IEEE Trans Knowl Data Eng.* 2022;34(12):5586-5609.
122. Godenschweger F, Kägebein U, Stucht D, et al. Motion correction in MRI of the brain. *Phys Med Biol.* 2016;61(5):R32-56.
123. Batchelor PG, Atkinson D, Irarrazaval P, Hill DLG, Hajnal J, Larkman D. Matrix description of general motion correction applied to multishot images. *Magn Reson Med.* 2005;54(5):1273-1280.
124. Uecker M, Lai P, Murphy MJ, et al. ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. *Magn Reson Med.* 2014;71(3):990-1001.
125. Xu B, Wang N, Chen T, Li M. Empirical Evaluation of Rectified Activations in Convolutional Network. Published online November 27, 2015.
126. Pezzotti N, Yousefi S, Elmahdy MS, et al. An Adaptive Intelligence Algorithm for Undersampled Knee MRI Reconstruction. Published online October 27, 2020.

127. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Published online June 10, 2016.
128. Esteban O, Birman D, Schaer M, Koyejo OO, Poldrack RA, Gorgolewski KJ. MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLOS ONE*. 2017;12(9):e0184661.
129. Nárai Á, Hermann P, Auer T, et al. Movement-related artefacts (MR-ART) dataset of matched motion-corrupted and clean structural MRI brain scans. *Sci Data*. 2022;9(1):630.
130. Ganzetti M, Wenderoth N, Mantini D. Intensity Inhomogeneity Correction of Structural MR Images: A Data-Driven Approach to Define Input Algorithm Parameters. *Front Neuroinformatics*. 2016;10.
131. Yanasak NE, Kelly MJ. MR Imaging Artifacts and Parallel Imaging Techniques with Calibration Scanning: A New Twist on Old Problems. *RadioGraphics*. 2014;34(2):532-548.
132. Sartoretti T, Reischauer C, Sartoretti E, Binkert C, Najafi A, Sartoretti-Schefer S. Common artefacts encountered on images acquired with combined compressed sensing and SENSE. *Insights Imaging*. 2018;9(6):1107.
133. Zaitsev M, Maclaren J, Herbst M. Motion artifacts in MRI: A complex problem with many partial solutions. *J Magn Reson Imaging*. 2015;42(4):887-901.
134. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-Image Translation with Conditional Adversarial Networks. *ArXiv161107004 Cs*. Published online November 26, 2018.
135. Huang J, Fang Y, Wu Y, et al. Swin transformer for fast MRI. *Neurocomputing*. 2022;493:281-304.
136. Li Y, Yang J, Yu T, Chi J, Liu F. Global attention-enabled texture enhancement network for MR image reconstruction. *Magn Reson Med*. 2023;90(5):1919-1931.
137. Guo P, Mei Y, Zhou J, Jiang S, Patel VM. ReconFormer: Accelerated MRI Reconstruction Using Recurrent Transformer. *IEEE Trans Med Imaging*. 2024;43(1):582-593.
138. Shou Q, Zhao C, Shao X, et al. Transformer-based deep learning denoising of single and multi-delay 3D arterial spin labeling. *Magn Reson Med*. 2024;91(2):803-818.
139. Crawshaw M. Multi-Task Learning with Deep Neural Networks: A Survey. Published online September 10, 2020.
140. Kendall A, Gal Y, Cipolla R. Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. Published online April 24, 2018.
141. Meyer CH, Hu BS, Nishimura DG, Macovski A. Fast Spiral Coronary Artery Imaging. *Magn Reson Med*. 1992;28(2):202-213.

142. Nishimura DG, Irarrazabal P, Meyer CH. A Velocity k-Space Analysis of Flow Effects in Echo-Planar and Spiral Imaging. *Magn Reson Med.* 1995;33(4):549-556.
143. Glover GH. Spiral imaging in fMRI. *NeuroImage.* 2012;62(2):706-712.
144. Lee Y, Wilm BJ, Brunner DO, et al. On the signal-to-noise ratio benefit of spiral acquisition in diffusion MRI. *Magn Reson Med.* 2021;85(4):1924-1937.
145. Volegov PL, Mosher JC, Espy MA, Kraus RH. On concomitant gradients in low-field MRI. *J Magn Reson San Diego Calif 1997.* 2005;175(1):103-113.
146. Irarrazabal P, Meyer CH, Nishimura DG, Macovski A. Inhomogeneity correction using an estimated linear field map. *Magn Reson Med.* 1996;35(2):278-282.
147. Schomberg H. Off-resonance correction of MR images. *IEEE Trans Med Imaging.* 1999;18(6):481-495.
148. Moriguchi H, Dale BM, Lewin JS, Duerk JL. Block regional off-resonance correction (BRORC): A fast and effective deblurring method for spiral imaging. *Magn Reson Med.* 2003;50(3):643-648.
149. Zeng DY, Shaikh J, Holmes S, et al. Deep residual network for off-resonance artifact correction with application to pediatric body MRA with 3D cones. *Magn Reson Med.* 2019;82(4):1398-1411.
150. Lim Y, Bliesener Y, Narayanan S, Nayak KS. Deblurring for spiral real-time MRI using convolutional neural networks. *Magn Reson Med.* 2020;84(6):3438-3452.
151. King KF, Ganin A, Zhou XJ, Bernstein MA. Concomitant gradient field effects in spiral scans. *Magn Reson Med.* 1999;41(1):103-112.
152. Ahunbay E, Pipe JG. Rapid method for deblurring spiral MR images. *Magn Reson Med.* 2000;44(3):491-494.
153. Cheng JY, Santos JM, Pauly JM. Fast concomitant gradient field and field inhomogeneity correction for spiral cardiac imaging. *Magn Reson Med.* 2011;66(2):390-401.
154. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. Published online December 10, 2015.
155. Chen W, Sica CT, Meyer CH. Fast conjugate phase image reconstruction based on a Chebyshev approximation to correct for B0 field inhomogeneity and concomitant gradients. *Magn Reson Med.* 2008;60(5):1104-1111.
156. Allen SP, Feng X, Fielden SW, Meyer CH. Correcting image blur in spiral, retraced in/out (RIO) acquisitions using a maximized energy objective. *Magn Reson Med.* 2019;81(3):1806-1817.

157. Lim Y, Narayanan SS, Nayak KS. Attention-gated convolutional neural networks for off-resonance correction of spiral real-time MRI. Published online February 14, 2021.
158. Ismail TF, Strugnell W, Coletti C, et al. Cardiac MR: From Theory to Practice. *Front Cardiovasc Med.* 2022;9.
159. Wang C, Li Y, Lv J, et al. Recommendation for Cardiac Magnetic Resonance Imaging-Based Phenotypic Study: Imaging Part. *Phenomixs Cham Switz.* 2021;1(4):151-170.
160. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J.* 2021;42(36):3599-3726.
161. Curtis AD, Cheng HLM. Primer and Historical Review on Rapid Cardiac CINE MRI. *J Magn Reson Imaging JMRI.* 2022;55(2):373-388.
162. Patel AR, Salerno M, Kwong RY, Singh A, Heydari B, Kramer CM. Stress Cardiac Magnetic Resonance Myocardial Perfusion Imaging: JACC Review Topic of the Week. *J Am Coll Cardiol.* 2021;78(16):1655.
163. Kuruvilla S, Adenaw N, Katwal AB, Lipinski MJ, Kramer CM, Salerno M. Late gadolinium enhancement on cardiac magnetic resonance predicts adverse cardiovascular outcomes in nonischemic cardiomyopathy: a systematic review and meta-analysis. *Circ Cardiovasc Imaging.* 2014;7(2):250-258.
164. Kim PK, Hong YJ, Im DJ, et al. Myocardial T1 and T2 Mapping: Techniques and Clinical Applications. *Korean J Radiol.* 2017;18(1):113.
165. Deshmane A, Gulani V, Griswold MA, Seiberlich N. Parallel MR Imaging. *J Magn Reson Imaging JMRI.* 2012;36(1):55.
166. Otazo R, Kim D, Axel L, Sodickson DK. Combination of Compressed Sensing and Parallel Imaging for Highly Accelerated First-Pass Cardiac Perfusion MRI. *Magn Reson Med Off J Soc Magn Reson Med Soc Magn Reson Med.* 2010;64(3):767.
167. Qin C, Schlemper J, Caballero J, Price AN, Hajnal JV, Rueckert D. Convolutional Recurrent Neural Networks for Dynamic MR Image Reconstruction. *IEEE Trans Med Imaging.* 2019;38(1):280-290.
168. El-Rewaidy H, Fahmy AS, Pashakhanloo F, et al. Multi-domain convolutional neural network (MD-CNN) for radial reconstruction of dynamic cardiac MRI. *Magn Reson Med.* 2021;85(3):1195-1208.
169. Zhao M, Wei Y, Wong KKL. A Generative Adversarial Network technique for high-quality super-resolution reconstruction of cardiac magnetic resonance images. *Magn Reson Imaging.* 2022;85:153-160.

170. Yoon S, Nakamori S, Amyar A, et al. Accelerated Cardiac MRI Cine with Use of Resolution Enhancement Generative Adversarial Inline Neural Network. *Radiology*. Published online May 30, 2023.
171. Sandino CM, Lai P, Vasanaawala SS, Cheng JY. Accelerating cardiac cine MRI using a deep learning-based ESPIRiT reconstruction. *Magn Reson Med*. 2021;85(1):152-167.
172. Küstner T, Fuin N, Hammernik K, et al. CINENet: deep learning-based 3D cardiac CINE MRI reconstruction with multi-coil complex-valued 4D spatio-temporal convolutions. *Sci Rep*. 2020;10(1):13710.
173. Ulyanov D, Vedaldi A, Lempitsky V. Instance Normalization: The Missing Ingredient for Fast Stylization. Published online November 6, 2017.
174. Wang C, Lyu J, Wang S, et al. CMRxRecon: An open cardiac MRI dataset for the competition of accelerated image reconstruction. Published online September 19, 2023.
175. Schlemper J, Caballero J, Hajnal JV, Price AN, Rueckert D. A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction. *IEEE Trans Med Imaging*. 2018;37(2):491-503.
176. Eo T, Jun Y, Kim T, Jang J, Lee HJ, Hwang D. KIKI-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. *Magn Reson Med*. 2018;80(5):2188-2201.
177. Feng L, Axel L, Chandarana H, Block KT, Sodickson DK, Otazo R. XD-GRASP: Golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing. *Magn Reson Med*. 2016;75(2):775-788.
178. Seiberlich N, Breuer FA, Blaimer M, Barkauskas K, Jakob PM, Griswold MA. Non-Cartesian data reconstruction using GRAPPA operator gridding (GROG). *Magn Reson Med*. 2007;58(6):1257-1265.
179. Jackson JI, Meyer CH, Nishimura DG, Macovski A. Selection of a convolution function for Fourier inversion using gridding (computerised tomography application). *IEEE Trans Med Imaging*. 1991;10(3):473-478.
180. Hansen MS, Sørensen TS. Gadgetron: An open source framework for medical image reconstruction. *Magn Reson Med*. 2013;69(6):1768-1776.