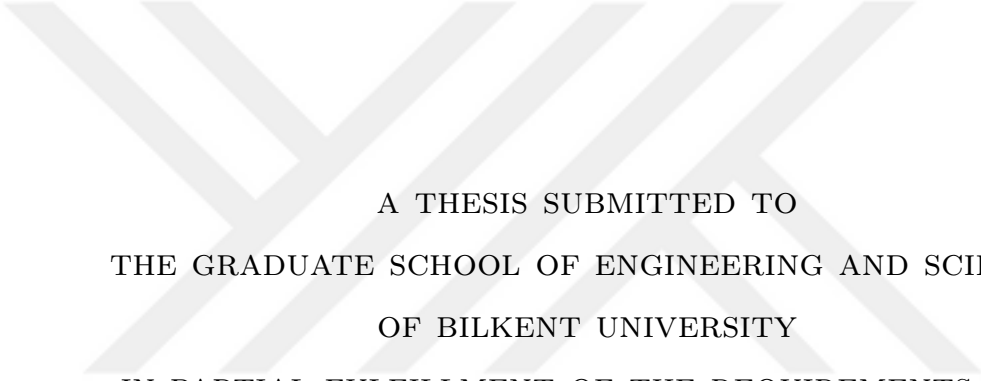


PHYSICS-DRIVEN DEEP LEARNING FOR MEDICAL IMAGE RECONSTRUCTION



A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

By
Bilal Kabas
August 2025

PHYSICS-DRIVEN DEEP LEARNING FOR MEDICAL IMAGE
RECONSTRUCTION

By Bilal Kabas

August 2025

We certify that we have read this thesis and that in our opinion it is fully adequate,
in scope and in quality, as a thesis for the degree of Master of Science.

Tolga Çukur(Advisor)

Emine Ülkü Sarıtaş Çukur

Emre Akbaş

Approved for the Graduate School of Engineering and Science:

Orhan Arıkan
Director of the Graduate School

ABSTRACT

PHYSICS-DRIVEN DEEP LEARNING FOR MEDICAL IMAGE RECONSTRUCTION

Bilal Kabas

M.S. in Electrical and Electronics Engineering

Advisor: Tolga Çukur

August 2025

Medical image reconstruction from undersampled acquisitions is an ill-posed problem involving inversion of the imaging operator linking measurement and image domains. Physics-driven (PD) models have gained prominence in reconstruction tasks due to their desirable performance and generalization. These models jointly promote data fidelity and artifact suppression, typically by combining data-consistency mechanisms with learned network modules. Artifact suppression depends on the network’s ability to disentangle artifacts from true tissue signals, both of which can exhibit contextual structure across diverse spatial scales. Convolutional neural networks (CNNs) are strong in capturing local correlations, albeit relatively insensitive to non-local context. While transformers promise to alleviate this limitation, practical implementations frequently involve design compromises to reduce computational cost by balancing local and non-local sensitivity, occasionally resulting in performance comparable to or trailing that of CNNs. To enhance contextual sensitivity without incurring high complexity, we introduce a novel physics-driven autoregressive state-space model (MambaRoll) for medical image reconstruction. In each cascade of its unrolled architecture, MambaRoll employs a physics-driven state-space module (PD-SSM) to aggregate contextual features efficiently at a given spatial scale, and autoregressively predicts finer-scale feature maps conditioned on coarser-scale features to capture multi-scale context. Learning across scales is further enhanced via a deep multi-scale decoding (DMSD) loss tailored to the autoregressive prediction task. Demonstrations on accelerated MRI and sparse-view CT reconstructions show that MambaRoll consistently outperforms state-of-the-art data-driven and physics-driven methods based on CNN, transformer, and SSM backbones.

Keywords: Medical image reconstruction, physics-driven, state space, autoregressive.

ÖZET

TIBBİ GÖRÜNTÜ GERİÇATIMI İÇİN FİZİK TABANLI DERİN ÖĞRENME

Bilal Kabas

Elektrik ve Elektronik Mühendisliği, Yüksek Lisans

Tez Danışmanı: Tolga Çukur

Ağustos 2025

Az örneklenmiş ölçümlerden elde edilen tıbbi görüntülerin geriçatımı, ölçüm ve görüntü uzayları arasında bağlantı kuran görüntüleme operatörünün terslenmesini içeren kötü konulmuş bir problemdir. Fizik tabanlı (PD) modeller, başarılı performansları ve genelleme kabiliyetlerinden dolayı görüntü geriçatımı görevlerinde öne çıkmaktadır. Bu modeller genellikle veri tutarlılığı mekanizmalarını öğrenilmiş sinir ağı modülleriyle birleştirerek, veri uyumluluğunu ve artefakt bastırmayı birlikte sağlar. Artefakt bastırma, yapay sinir ağının gerçek doku sinyallerini artefaktlardan ayırma yeteneğine bağlıdır ve her ikisi de farklı uzamsal ölçeklerde bağlama dayalı yapılar sergileyebilir. Konvolüsyonel sinir ağları (CNN), lokal korelasyonları yakalamada güçlüdür; ancak uzamsal olarak uzak bağlamlara karşı nispeten duyarsızdır. Dönüştürücüler (transformer) bu sınırlamayı aşma potansiyeli taşısa da, pratikteki uygulamalar çoğunlukla lokal ve lokal olmayan duyarlılığı dengeleyerek hesaplama maliyetini azaltma yönünde tasarım ödünleri içerir; bu da bazen CNN'lerin dengi ya da gerisinde bir performans ile sonuçlanır. Bu çalışmada tıbbi görüntü geriçatımı için, yüksek hesaplama karmaşıklığına yol açmadan bağlamsal duyarlılığı artırmak amacıyla, yeni bir fizik tabanlı otoregresif durum-uzay modeli olan MambaRoll önerilmektedir. MambaRoll'un açılmış mimarisinin her kademesinde, belirli bir uzamsal ölçekte bağlama dayalı özellikleri verimli şekilde toplamak ve daha detaylı ölçekteki özellik haritalarını daha kaba ölçekli özelliklere koşullu olarak otoregresif biçimde tahmin ederek çok ölçekli bağlamı yakalamak amacıyla fizik tabanlı bir durum-uzay modülü (PD-SSM) kullanılır. Ölçekler arası öğrenme, otoregresif tahmin görevine özel olarak tasarlanmış derin çok ölçekli çözümleme (DMSD) kayıp fonksiyonu ile pekiştirilmiştir. Hızlandırılmış MRI ve seyrek projeksiyonlu BT geriçatımında yapılan deneyler, MambaRoll'un CNN, dönüştürücü

ve SSM kullanan, veri tabanlı ve fizik tabanlı güncel yöntemlerden tutarlı olarak daha üstün performanslı olduğunu göstermektedir.



Anahtar sözcükler: Tıbbi görüntü geriçatımı, fizik tabanlı, durum uzayı, otoregresif.

Acknowledgement

I would first like to express my deepest gratitude to my advisor, Prof. Tolga Çukur, for his invaluable guidance, encouragement, and support throughout this journey. His patience, trust, and belief in me have been a constant source of strength. His insights and mentorship have not only shaped my research but also played a crucial role in my academic and personal growth. I feel truly fortunate to have had him as my advisor.

I am also deeply thankful to my jury members, Prof. Emine Ülkü Sarıtaş and Prof. Emre Akbaş, for their thoughtful comments and contributions to this thesis. Their support and guidance have been greatly appreciated.

My heartfelt thanks go to my closest friends, Abdulsamet, Berat, Mustafa, and Alman Batuhan, who have stood by me not only as friends but as brothers. Their companionship and support have made this journey lighter and more meaningful. Berat, I will remember you every time I import pandas. And Abdulsamet, whenever I feel overwhelmed, I will recall your encouragement; your unwavering support has made this journey possible.

I would also like to acknowledge my lab mates, Fuat, Ömer, Valiyeh, and Şaban, for their collaboration, kindness, and support throughout this journey.

I am grateful as well to Prof. Bülent Yılmaz and Prof. Samet Güler, whose guidance during my undergraduate years helped prepare me for this path. My dearest friend Samet was also a constant source of inspiration during those years, and I truly miss his companionship.

Finally, I owe my deepest gratitude to my mum and dad, whose endless love, sacrifices, and support have carried me through every stage of life. To Elif, thank you for your constant encouragement, patience, and love, which have been a source of strength and motivation.

To all of you, I extend my most sincere thanks and appreciation.

Contents

1	Introduction	1
2	Preliminaries	5
2.1	Fundamentals of MRI	5
2.1.1	Physical Principles	6
2.1.2	Image Formation	9
2.1.3	Image Reconstruction	11
2.2	Accelerated MRI	12
2.3	Fundamentals of CT Imaging	14
2.3.1	Physical Principles	14
2.3.2	Image Formation	16
2.3.3	Image Reconstruction	17
2.4	Sparse-view CT	18
3	Related Work	20

3.1	Physics-Driven Models	20
3.2	State Space Models	21
3.3	Autoregressive Models	22
4	Theory	23
4.1	Medical Image Reconstruction	23
4.2	MambaRoll	24
4.2.1	Network Architecture	26
4.2.2	Deep Multi-Scale Decoding Loss	29
5	Methods	31
5.1	Datasets	31
5.2	Architectural Details	32
5.3	Competing Methods	33
5.3.1	Accelerated MRI Reconstruction	33
5.3.2	Sparse-view CT Reconstruction	34
5.4	Modeling Procedures	35
6	Results	37
6.1	Accelerated MRI Reconstruction	37
6.2	Sparse-view CT Reconstruction	39

6.3	Generalization Under Acquisition	
	Variability	41
6.4	Contextual Awareness	44
6.5	Computational Complexity	45
6.6	Ablation Studies	47
6.6.1	Network Architecture	47
6.6.2	Loss Function	48
6.6.3	Design Parameters	49
7	Discussion	52
8	Conclusion	54

List of Figures

- 2.1 Comparison of T_1 -weighted and T_2 -weighted MRI images. The T_1 -weighted image (left) emphasizes anatomical detail and depicts fat as bright and fluids as dark. The T_2 -weighted image (right) highlights fluid-containing structures, which appear bright. 9
- 2.2 K-space and image space in MRI. The left image shows the magnitude of the acquired k-space data $|S|$, represented in spatial frequency coordinates (k_x, k_y) . The right image displays the corresponding magnitude image $|I|$ in the spatial domain (x, y) , obtained by applying an inverse Fourier transform to the k-space data. 12
- 2.3 A sparse signal reconstructed using undersampled k-space. Equi-spaced undersampling exhibits coherent aliasing, whereas random sampling allows information recovery due to incoherent aliasing. . 13
- 2.4 Deep compressed sensing employs an unrolled architecture, interleaving learnable network modules G_θ with data consistency (DC) layers for reconstruction from undersampled acquisitions (IFT: inverse Fourier transform). 14

2.5	Projection space and image space in computed tomography (CT). The left image shows the sinogram (projection space), where each line represents a one-dimensional projection $p(\theta), t$ of the sample at angle θ . The right image depicts the reconstructed attenuation map $\hat{\mu}(x, y)$ in spatial coordinates, obtained by applying filtered backprojection.	17
2.6	Comparison between full-view and sparse-view CT reconstruction using filtered backprojection (FBP). The left panel shows the sinogram (projection data) and corresponding reconstruction from a full-view acquisition. The right panel depicts the projection data with reduced angular sampling (sparse-view), and its FBP-based reconstruction, which suffers from visible streaking artifacts and reduced image quality due to the ill-posedness of the inverse problem under insufficient data.	19

- 4.1 MambaRoll is a novel physics-driven autoregressive state space model, which utilizes an unrolled architecture with K cascades to map a linear reconstruction of undersampled data $\hat{x}_0$ to the learning-based reconstruction $\hat{x}_K$. In each cascade, MambaRoll recovers contextualized feature maps at high spatial resolution g_S progressively across S spatial scales by leveraging autoregressive prediction of finer-scale feature maps from coarser scales (see Fig. 4.2 for details). At a given scale s , recovery is performed by a novel physics-driven state space module (PD-SSM) that efficiently aggregates contextual features while maintaining consistency to acquired data. PD-SSM comprises a convolutional encoder to map high-resolution input feature maps $f_{<s}$ onto the desired scale, a compressed SSM block to capture context across feature maps efficiently, a convolutional decoder to recollect high-resolution feature maps, and a residual data-consistency (DC) block to enforce fidelity to acquired data. Following autoregressive predictions, a refinement module based on convolutional layers is used to enhance restoration of detailed texture. 25
- 4.2 Illustration of autoregressive prediction and deep supervision of multi-scale representations in MambaRoll. To explicitly link multi-scale representations via an autoregressive prior, the PD-SSM module at a given scale s predicts the finer-scale feature map (f_s) conditioned on the preceding set of feature maps obtained at coarser scales ($f_{<s} := [f_0 \ f_1 \ \dots \ f_{s-1}]$). For model supervision, a deep multi-scale decoding loss (DMSD) loss is used in combination with a conventional reconstruction loss, where $\hat{x}_K$ is derived via the refinement module (RM). DMSD supervises intermediate reconstructions ($\tilde{u}_s^K$) at each scale to further enhance multi-scale learning. 27

6.1	Accelerated MRI reconstructions of a representative cross-section from a T_2 -weighted acquisition at $R = 12$. Reconstructed images from competing methods are shown along with a linear reconstruction of undersampled data (ZF), and the reference image derived from fully-sampled data. Zoom-in windows and error maps are included to highlight differences in tissue depiction. PSNR and SSIM values are listed below each reconstruction.	39
6.2	Sparse-view CT reconstructions of a representative cross-section at $R = 8$. Reconstructed images from competing methods are shown along with a linear reconstruction of undersampled data (FBP), and the reference image derived from fully-sampled data.	41
6.3	Accelerated MRI reconstructions of a representative cross-section from a FLAIR-weighted acquisition at $R = 8$ with models trained at $R = 12$. Reconstructed images from competing methods are shown along with a linear reconstruction of undersampled data (ZF), and the reference image derived from fully-sampled data. . .	43
6.4	Sparse-view CT reconstructions of a representative cross-section at $R = 8$ with models trained at $R = 6$. Reconstructed images from competing methods are shown along with a linear reconstruction of undersampled data (FBP), and the reference image derived from fully-sampled data.	43
6.5	Gradient-based effective receptive field (ERF) visualizations for MRI (top row) and CT (bottom row) reconstruction across competing methods. ERF maps highlight the input image regions that most influence the reconstruction at the fixed target location at image center. Greener areas indicate higher contribution, whereas yellower areas indicate lower contribution (see colorbar).	45

6.6	Performance versus efficiency in MRI (left panel) and CT (right panel) reconstructions. Each method is represented by a circular marker whose center denotes the performance-latency trade-off (where latency is taken as inference time), and whose diameter indicates the memory load (see scale bars).	46
6.7	Performance of MambaRoll variants on representative reconstruction tasks (FLAIR, $R = 4$ for MRI and $R = 4$ for CT). Results shown for the proposed Deep MSD loss (deep multi-scale decoding loss marked with squares), along with Shallow SS, Shallow MS, and Deep MSL variants.	48
6.8	Performance of MambaRoll variants that prescribed varying numbers of spatial scales. Results displayed for the representative reconstruction tasks of FLAIR, $R = 4$ for MRI and $R = 4$ for CT. Squares mark the parameter values that were originally selected to implement MambaRoll.	49
6.9	Performance of MambaRoll variants that prescribed varying compression factors (a) on FLAIR, $R = 4$ for MRI; (b) on $R = 4$ for CT. Performance metrics (PSNR, SSIM) are displayed in the right panel, whereas efficiency metrics (memory load and latency) are displayed in the left panel.	50

List of Tables

5.1	Categorization of the competing methods on accelerated MRI reconstruction. (DD: data-driven, PD: physics-driven, CNN: convolutional, ViT: transformer, SSM: state-space)	34
5.2	Categorization of the competing methods on sparse-view CT reconstruction. (DD: data-driven, PD: physics-driven, CNN: convolutional, ViT: transformer, SSM: state-space)	35
6.1	Accelerated MRI reconstruction performance at $R = 4-12$. PSNR (dB) and SSIM (%) listed as mean $\pm$ std across the test set. Boldface marks the top performing method.	37
6.2	Average accelerated MRI reconstruction performance across acceleration rates. PSNR (dB) and SSIM (%) listed as mean $\pm$ std across the test set. Boldface marks the top performing method.	38
6.3	Sparse-view CT reconstruction performance at $R = 4-8$. Boldface marks the top performing method.	40
6.4	Average Sparse-view CT reconstruction performance across acceleration rates. Boldface marks the top performing method.	40
6.5	Generalization performance in MRI reconstruction under variability in R values between training and test sets ($R_{\text{training}} \rightarrow R_{\text{test}}$).	42

6.6	Generalization performance in CT reconstruction under variability in R values between training and test sets ($R_{\text{training}} \rightarrow R_{\text{test}}$). . . .	42
6.7	Average run time (msec) and memory load (MB) per cross-section for MRI reconstruction during training and inference phases. . . .	46
6.8	Average run time (msec) and memory load (MB) per cross-section for CT reconstruction during training and inference phases. . . .	46
6.9	Performance of MambaRoll variants on representative reconstruction tasks. A ‘w/o PD-SSM’ variant ablating PD-SSM modules, a ‘w/o AR’ variant ablating autoregressive processing, a ‘w/o comp. SSM’ variant that ablated compressed SSM blocks, and a ‘w/o DC’ variant ablating data-consistency modules were considered. . . .	48

Chapter 1

Introduction

In many tomographic modalities, tissue structure is spatially encoded prior to acquisition, placing raw data in a distinct measurement domain (e.g., k-space in MRI, projection domain in CT). Higher spatial resolution can be achieved by collecting more unique encodings, but this increases scan time, patient discomfort, and/or radiation dose. A common strategy to improve efficiency is undersampling—acquiring only a subset of k-space points or projection views [1]. However, reconstructing images from undersampled data is inherently ill-posed, requiring inversion of the imaging operator that links measurement and image domains [2]. Traditional methods often yield residual aliasing artifacts and noise during such inversion [3], highlighting the need for more robust reconstruction techniques.

Learning-based methods have made notable progress in reconstruction over the recent years [4, 5]. A group of data-driven (DD) methods trains networks to directly map input images linearly recovered from undersampled data to ground-truth images derived from fully-sampled acquisitions, hence building conditional image priors without explicitly modeling the imaging operator [6, 7, 8, 9, 10, 11, 12]. While DD methods benefit from task-specific feature extraction innate to deep learning, ignoring the imaging operator used to acquire data can impair generalization [13]. Alternatively, generative models typically learn unconditional priors to synthesize high-quality images [14, 15, 16, 17, 18, 19].

These generative methods can reconstruct images by optimizing joint objectives for prior regularization and data consistency [20, 21], occasionally causing sub-optimal performance and extended inference times [22].

Among learning-based reconstruction frameworks, physics-driven (PD) approaches have emerged as particularly effective, offering strong performance and generalization [23]. PD methods employ a joint objective based on learned network modules for artifact suppression and data-consistency steps guided by the imaging operator [24, 25, 26, 27, 28, 29]. Naturally, their performance hinges on the ability of network modules to distinguish artifacts from anatomical features—demanding sensitivity to contextual relationships spanning diverse spatial scales in medical images [30]. CNNs capture local patterns well but struggle with non-local context [31, 24, 8, 32]. While transformers promise improved modeling of long-range relationships, their high computational costs often lead to simplified designs that constrain either spatial precision or contextual sensitivity [33, 34, 35, 36, 37, 38, 39, 40, 41]. This underscores the need for architectures that more effectively balance contextual sensitivity and computational efficiency.

State space models (SSMs) are emerging as a promising alternative in this regard [42]. SSMs use rectilinear scans to rasterize images into one-dimensional (1D) sequences, and model inter-pixel dependencies as a linear dynamical system with learnable parameters. This framework promises stronger sensitivity to long-range context than CNNs and higher efficiency than transformers. Few recent studies have proposed SSM-based reconstruction models prior to our study. A DD architecture incorporating a single-scale SSM module has been adopted for medical image reconstruction [43], though the neglect of imaging physics in DD methods may limit generalization performance across different imaging operators. More recently, DD [44] and PD [45] SSM architectures have been developed specifically for MRI reconstruction. However, these studies have also adopted single-spatial-scale SSM modules that can exhibit suboptimal sensitivity to interactions between distant pixels, as such pixels may be separated by many intermediate elements when the image is converted to a 1D sequence [46]. This can compromise the ability of existing SSM methods to capture a comprehensive set of contextual features across spatial scales.

Given that many existing reconstruction models inherently process multi-scale feature representations (e.g., UNet, Swin Transformer), extending SSMs to operate on multi-scale feature maps could, in principle, alleviate the limitations of single-scale processing. However, without explicit mechanisms for linking information across scales and supervising intermediate representations, multi-scale feature maps alone may be insufficient for high-fidelity reconstruction. On one hand, naive linking strategies—such as simple additive and residual connections—can lead to inconsistencies and artifacts that degrade image quality [47]. On the other hand, conventional supervision focused solely on the final reconstructed image might provide suboptimal guidance for learning accurate representations at intermediate scales [48].

To address these challenges, we introduce MambaRoll, a novel physics-driven autoregressive state space model for performant and efficient medical image reconstruction. Rather than independently processing different spatial scales, MambaRoll conducts autoregressive prediction of multi-scale feature maps, where each finer scale is conditioned on and refines the prediction from the preceding coarser scales. For this purpose, an unrolled architecture is equipped with novel Physics-driven State Space Modules (PD-SSMs) that reconstruct conditioned multi-scale representations and preserve data fidelity, while maintaining high efficiency via compressed SSM blocks with space-to-depth transformation. MambaRoll’s autoregressive design links feature maps to ensure consistent information flow from global anatomical structure to local tissue details. To further enhance multi-scale learning, we introduce a Deep Multi-Scale Decoding (DMSD) loss that supervises intermediate reconstructions at each scale, providing direct feedback for learning multi-scale feature hierarchies rather than relying solely on supervision based on the final output. Comprehensive experiments on accelerated MRI and sparse-view CT demonstrate that MambaRoll surpasses state-of-the-art CNN-, transformer-, and SSM-based methods. *To our knowledge, MambaRoll is the first framework to unify physics-driven multi-scale SSM design with an autoregressive prior, enabling generalization across medical image reconstruction tasks.*

The content of this thesis reflects the studies described in the following publication [49]:

B. Kabas, F. Arslan, V. A. Nezhad, S. Ozturk, E. U. Saritas, and T. Çukur, “Physics-driven autoregressive state space models for medical image reconstruction,” arXiv:2412.09331, 2024.

MambaRoll involves a novel PD-SSM architecture designed to enhance contextual sensitivity and improve fidelity in medical image reconstruction. MambaRoll’s unrolled architecture autoregressively predicts feature maps at finer spatial scales, conditioned on coarser-scale representations to facilitate hierarchical context aggregation. The novel PD-SSM module in MambaRoll facilitates autoregressive predictions by mitigating artifacts at each scale while enforcing data consistency through residual correction at the full scale. To further guide hierarchical learning, we introduce a DMSD loss that provides deep supervision on decoded images at multiple spatial scales.

Outline

The rest of this thesis is organized as follows: Chapter 2 is dedicated to the necessary background and preliminaries, covering the fundamentals of MRI and CT imaging, as well as accelerated MRI and sparse-view CT. Chapter 3 reviews the relevant literature, including physics-driven models, state space models, and autoregressive methods. Chapter 4 presents the theoretical foundations of medical image reconstruction and introduces the proposed MambaRoll architecture in detail. Chapter 5 outlines the experimental methodology, datasets, architectural choices, and baseline models. Chapter 6 reports the experimental results on accelerated MRI and sparse-view CT. Chapter 7 discusses the implications and limitations of the findings along with future directions to improve the proposed method. Finally, Chapter 8 concludes the thesis.

Chapter 2

Preliminaries

2.1 Fundamentals of MRI

Magnetic Resonance Imaging (MRI) is a non-invasive and non-ionizing medical imaging technique that utilizes the principles of nuclear magnetic resonance (NMR) to capture detailed images of internal structures within the human body. MRI uses strong magnetic fields and radiofrequency (RF) pulses, which do not involve ionizing radiation and hence pose no health risks, unlike X-ray-based imaging modalities. Unlike modalities such as magnetic particle imaging (MPI) [50, 51, 52, 53, 54, 55], MRI does not strictly require the use of contrast agents, as many diagnostic applications can be performed with native tissue contrast alone. MRI is a powerful diagnostic tool widely used in clinical settings for the accurate assessment of various medical conditions. Common clinical and research applications of MRI include brain imaging for tumors and stroke [56, 57], functional MRI (fMRI) [58, 59, 60, 61, 62] for mapping brain activity, and magnetic resonance angiography [63, 64, 65, 66, 67] for visualizing blood vessels.

2.1.1 Physical Principles

2.1.1.1 Nuclear Magnetic Resonance

MRI is based on the principles of NMR, which exploits the magnetic properties of atomic nuclei that arise from their quantum mechanical characteristics. Specifically, spin characteristics of nuclei enable MRI. Several atoms can be used for MRI such as sodium (^{23}Na), carbon (^{13}C), phosphorus (^{31}P), and hydrogen (^1H) which have non-zero nuclear spin, consequently, can interact with a magnetic field and generate a detectable NMR signal. In clinical settings, MRI uses hydrogen atoms due to their high abundance stemming from the fact that water accounts for 50-70% of the human body and the primary ingredient of tissues and fat. The most common form of the hydrogen atom consists of a single proton and an electron. Since its nucleus lacks a neutron, hydrogen is often referred to as a *proton*.

Classical physics adequately describes proton behavior in a nucleus to explain the basics of MRI. Protons are positively charged and spinning particles. These properties enable protons to create a magnetic field and have a magnetic moment like a tiny bar magnet. In the presence of an external magnetic field, protons exhibit *precession* in addition to their intrinsic spin. In precession, the magnetic moments of protons wobble around the direction of the external magnetic field, analogous to the wobbling motion of a spinning top affected by gravity. The frequency at which the precession occurs is called the *Larmor frequency*:

$$\omega = \gamma B \quad (2.1)$$

where B is the magnitude of the external magnetic field and γ is the gyromagnetic ratio, which is a constant specific to each nucleus.

2.1.1.2 External Magnetic Field

The vector sum of the magnetic moments of individual protons is the net magnetization, denoted as $\mathbf{M}_0$. In the absence of an external magnetic field, the

magnetic moments of protons are oriented randomly, resulting in zero net magnetization. When placed in an external magnetic field, often referred to as the $\mathbf{B}_0$ field, the protons tend to align their magnetic moments with the direction of the field and precess around it. Therefore, the net magnetization vector is aligned with the $\mathbf{B}_0$ field in the $+z$ direction, which is referred to as the *equilibrium* state. The $\mathbf{B}_0$ field is a very strong static magnetic field generated by superconducting coils, typically ranging from 1.5 to 3 Tesla in clinical MRI scanners.

2.1.1.3 Excitation and Relaxation Mechanisms

The core principle of MRI and its contrast mechanism is that, after being excited by a specific radiofrequency (RF) pulse, protons emit signals as they return to the equilibrium by changing the orientation of their magnetic moments, which is influenced by tissue characteristics. MRI uses 90° and 180° RF pulses, each with a specific purpose. An RF pulse can excite protons only when transmitted at Larmor frequency, resonating with their precessional motion in the magnetic field. Specifically, the RF pulse must rotate within the transmit coil at Larmor frequency to enable proton excitation. This electromagnetic wave is accompanied by a magnetic field, denoted by $\mathbf{B}_1$. A 90° RF pulse at Larmor frequency flips the magnetic moment vectors of protons into the transverse (xy) plane, hence tipping the net magnetization away from the longitudinal (z) axis by 90° . Since all protons precess in phase in the presence of the $\mathbf{B}_1$ field, the magnitude of the net magnetization does not change, which equals M_0 . For simplicity, a rotating frame of reference, where the $\mathbf{B}_1$ field appears stationary in the transverse plane, is used instead of the laboratory (stationary) frame to further analyze the behavior of the magnetic moments and the net magnetization vector.

When the RF pulse is turned off, protons start returning to their equilibrium state where their magnetic moments are aligned with the $\mathbf{B}_0$ field. During this process, the energy absorbed by the protons is emitted back, a phenomenon known as *relaxation*. In the T_1 (longitudinal) relaxation, the net magnetization

in the longitudinal axis M_z is restored, which is described by

$$M_z(t) = M_0(1 - e^{-t/T_1}) \quad (2.2)$$

where T_1 is the time constant. In the T_2 (transverse) relaxation, the net magnetization in the transverse plane M_{xy} decays to zero, which is described by

$$M_{xy}(t) = M_0 e^{-t/T_2} \quad (2.3)$$

where T_2 is the time constant. T_1 relaxation involves energy exchange with the surrounding lattice, whereas T_2 decay is caused by dephasing among protons due to intrinsic spin-spin interactions. Both time constants vary depending on the tissue type. For example, longitudinal magnetization recovers much faster in fat than in cerebrospinal fluid (CSF), resulting in a shorter T_1 time for fat.

In practical applications, T_2 is much shorter than T_1 as protons dephase quickly after RF excitation due to local magnetic field *inhomogeneities* in addition to spin-spin interactions, thus referred to as T_2^* . Inhomogeneities in the $\mathbf{B}_0$ field arise from imperfections in the external magnetic field and variations in the magnetic properties of different tissue structures. For example, a 1 μT deviation causes a 180° phase shift in the magnetic moments of protons in approximately 12 ms. Note that for a 1 T MRI scanner, T_1 approximately ranges from 200 to 3000 ms depending on tissues.

After RF excitation, protons precessing in the transverse plane generate a time-varying magnetic flux in the receive coil, which induces a voltage. This detectable signal is called the *free induction decay* (FID), which can be expressed as

$$S_{FID}(t) = S_0 e^{-t/T_2} e^{-\gamma \Delta B_0 t} = S_0 e^{-t/T_2^*} \quad (2.4)$$

where ΔB_0 denotes inhomogeneities in the $\mathbf{B}_0$ field and S_0 is the initial signal amplitude which is proportional to M_0 .

The contrast mechanism of MRI is governed by the *repetition time* (TR) and the *echo time* (TE). Sample images for T_1 and T_2 contrasts are shown in Fig. 2.1. The repetition time is the time delay between consecutive RF pulses, whereas the echo time is the time interval between the center of the RF pulse and the peak

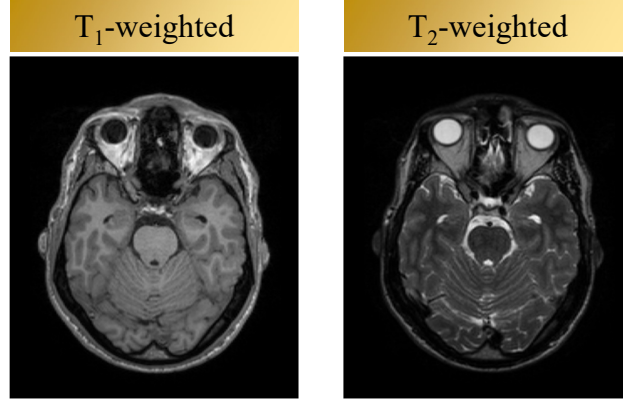


Figure 2.1: Comparison of T_1 -weighted and T_2 -weighted MRI images. The T_1 -weighted image (left) emphasizes anatomical detail and depicts fat as bright and fluids as dark. The T_2 -weighted image (right) highlights fluid-containing structures, which appear bright.

of the measured echo signal. FID occurs rapidly due to T_2^* dephasing, posing challenges for accurate detection. To counteract this, a 180° RF pulse is applied at $t = TE/2$ following the initial 90° excitation pulse. This refocusing pulse reverses the phase shift caused by magnetic field inhomogeneities (ΔB), resulting in the formation of a spin echo at time $t = TE$. Consequently, the T_2 relaxation is measured accurately, minimizing the effects of T_2^* dephasing. Therefore, the overall signal equation can be written as

$$S(TR, TE) = S_0(1 - e^{-TR/T_1})e^{-TE/T_2}. \quad (2.5)$$

2.1.2 Image Formation

MRI image formation results from the application of specific radiofrequency pulse sequences such as spin echo, gradient echo, and steady-state free precession [68, 69, 70, 71, 72, 73, 74, 75, 76], that manipulate nuclear spins and magnetic field gradients, allowing spatial encoding and contrast generation based on tissue relaxation properties.

MRI spatially encodes voxels within a sample using gradient coils that generate linearly varying magnetic fields along the x , y , and z directions. This is achieved

in three main steps: slice selection, frequency encoding, and phase encoding.

Slice selection enables imaging of a specific cross-section within a sample through a gradient magnetic field G_z in the longitudinal axis, which is applied during the RF excitation pulse. Therefore, the Larmor frequency becomes position-dependent:

$$w(z) = \gamma(B_0 + G_z z) \quad (2.6)$$

$$\Delta w = \gamma G_z \Delta z \quad (2.7)$$

where Δw is the frequency bandwidth of the RF excitation, and Δz is the slice thickness, which involves excited protons.

Following slice selection, phase encoding is used to spatially resolve the emitted signal from protons along the y -axis within the excited cross-section. Thus, the Larmor frequency is expressed in terms of the magnetic field gradient G_y , which is applied before the signal readout, and position y :

$$w(y) = \gamma(B_0 + G_y y) \quad (2.8)$$

The exposure to the gradient field G_y causes spins to accumulate a phase shift dependent on y position, which is computed as follows

$$\phi(y) = \int_0^{\Delta t} \omega(y) dt = \Delta \omega(y) \cdot \Delta t = \gamma G_y y \Delta t \quad (2.9)$$

where Δt is the duration of the applied gradient field.

Finally, frequency encoding resolves spatial information along the x -axis by applying a gradient magnetic field G_x during signal readout. This gradient causes the Larmor frequency to vary with position along x :

$$w(x) = \gamma(B_0 + G_x x) \quad (2.10)$$

Combining phase and frequency encoding gradients, the local magnetic field applied to each voxel within the selected cross-section is

$$B(x, y, t) = B_0 + G_x(t)x + G_y(t)y. \quad (2.11)$$

Therefore, the Larmor frequency at position (x, y) is

$$\omega(x, y, t) = \gamma B(x, y, t) = \gamma (B_0 + G_x(t)x + G_y(t)y), \quad (2.12)$$

$$\Delta\omega(x, y, t) = \gamma (G_x(t)x + G_y(t)y). \quad (2.13)$$

Subsequently, the phase shift at position (x, y) is

$$\phi(x, y, t) = \int_0^t \Delta\omega(x, y, \tau) d\tau = \gamma \left(x \int_0^t G_x(\tau) d\tau + y \int_0^t G_y(\tau) d\tau \right). \quad (2.14)$$

Assume k_x and k_y are accumulated gradients over time:

$$k_x(t) = \frac{1}{2\pi} \gamma \int_0^t G_x(\tau) d\tau, \quad k_y(t) = \frac{1}{2\pi} \gamma \int_0^t G_y(\tau) d\tau. \quad (2.15)$$

So the phase is rewritten as

$$\phi(x, y, t) = 2\pi (k_x(t)x + k_y(t)y). \quad (2.16)$$

The detected signal in the receiver coil at position (x, y) is the magnetization phasor at that location:

$$S(x, y, t) = M(x, y) e^{i\phi(x, y, t)} \quad (2.17)$$

where $M(x, y)$ is the transverse magnetization. So the detected signal at time t is the integral of all spin magnetic moments across x and y axes:

$$S(t) = \iint M(x, y) e^{i\phi(x, y, t)} dx dy. \quad (2.18)$$

Subsequently, inserting 2.16 into above formula yields:

$$S(t) = S(k_x, k_y) = \iint M(x, y) e^{-i2\pi(k_x x + k_y y)} dx dy \quad (2.19)$$

which means the measured signal in MRI at a given spatial frequency coordinate (k_x, k_y) is the 2D Fourier transform of the transverse magnetization, implying that the MRI signal is collected in the frequency domain (k-space).

2.1.3 Image Reconstruction

The MR image I is the inverse Fourier transform of the collected k-space data:

$$I(x, y) = \mathcal{F}^{-1}\{S(k_x, k_y)\} = \iint S(k_x, k_y) \cdot e^{i2\pi(k_x x + k_y y)} dk_x dk_y. \quad (2.20)$$

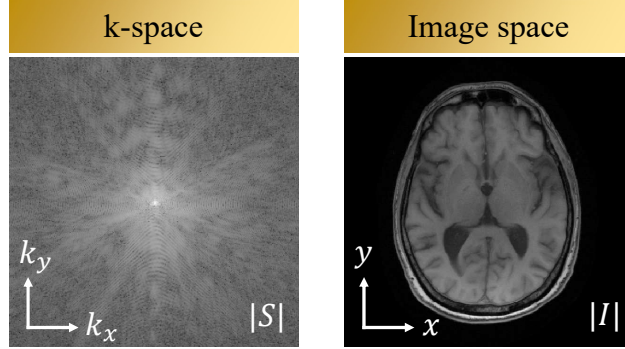


Figure 2.2: K-space and image space in MRI. The left image shows the magnitude of the acquired k-space data $|S|$, represented in spatial frequency coordinates (k_x, k_y) . The right image displays the corresponding magnitude image $|I|$ in the spatial domain (x, y) , obtained by applying an inverse Fourier transform to the k-space data.

Therefore, the fast Fourier transform (FFT) can be employed for image reconstruction in conventional Cartesian MRI (Fig. 2.2). In more advanced scenarios, additional reconstruction approaches include model-based iterative techniques such as total variation (TV) minimization [77] and POCS (projection onto convex set)[78]; parallel imaging algorithms such as SENSE [79] and GRAPPA [80]; and methods designed for undersampled acquisitions, notably compressed sensing [81].

2.2 Accelerated MRI

Fully-sampling k-space points leads to prolonged scan times, which can cause patient discomfort and motion artifacts. Instead, acquiring fewer k-space data points accelerates imaging but introduces aliasing and artifacts due to undersampling. Specifically, grid-based (equispaced) undersampling in k-space results in coherent aliasing artifacts, which leads to information loss. While parallel imaging [79, 80, 82] techniques can be employed to accelerate MR imaging, compressed sensing MRI [81, 83, 84, 85, 86] addresses this problem via a random k-space undersampling scheme in which incoherent aliasing artifacts allow signal recovery (Fig. 2.3). The compressed sensing theory is established upon the three main

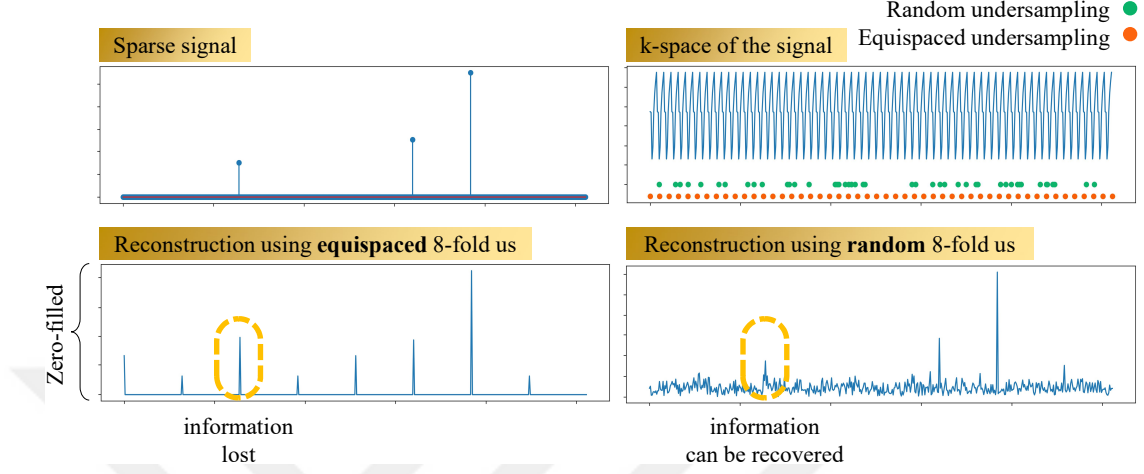


Figure 2.3: A sparse signal reconstructed using undersampled k-space. Equispaced undersampling exhibits coherent aliasing, whereas random sampling allows information recovery due to incoherent aliasing.

requirements:

1. **Transform sparsity:** the data is sparse and hence compressible in the transform domain, i.e., k-space.
2. **Incoherent sampling:** pseudo-random or variable density undersampling leads to incoherent artifacts which can then be recovered.
3. **Nonlinear reconstruction:** this involves solving a nonlinear optimization problem to find a maximally sparse solution in the transform domain, which is also consistent with the acquired data points.

The compressed sensing MRI formulation is

$$\hat{x} = \arg \min_x \frac{1}{2} \|\mathcal{F}_u x - y\|_2^2 + \lambda \|\Psi(x)\|_1 \quad (2.21)$$

where y is the acquired k-space points, $\mathcal{F}_u$ is undersampled Fourier transform, Ψ is sparsifying transform, λ is regularization weight, and $\hat{x}$ is the reconstructed image. In deep compressed sensing, the regularization term $\Psi(x)$ is learned via a neural network often employing an unrolled architecture (Fig. 2.4).

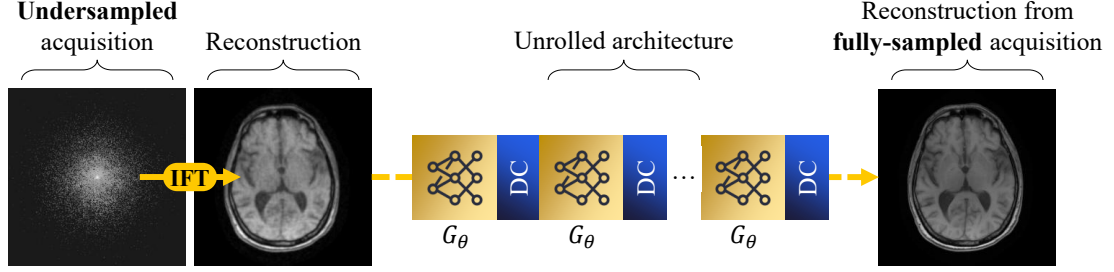


Figure 2.4: Deep compressed sensing employs an unrolled architecture, interleaving learnable network modules G_θ with data consistency (DC) layers for reconstruction from undersampled acquisitions (IFT: inverse Fourier transform).

2.3 Fundamentals of CT Imaging

Computed tomography (CT) is a non-invasive medical imaging modality that utilizes X-rays to obtain three-dimensional (3D) images of internal body structures, which enables volumetric inspection. In traditional X-ray imaging, 3D body structures are projected into a 2D plane, which involves overlaps between tissues. CT employs X-ray beams rotating around the subject, which are then processed by computers to reconstruct 3D images.

X-ray is an ionizing radiation that can remove electrons from atoms. Free electrons and ions can cause alterations in chemical structures. Therefore, ionization poses biological risks to living organisms. However, powerful diagnostic capabilities of CT outweigh the associated risks. CT provides critical diagnostics such as imaging tumors, bone fractures, and eroded bones.

2.3.1 Physical Principles

X-rays are produced inside a vacuum tube, which consists of a cathode filament (negative electrode) and an anode (positive electrode), which are often made from tungsten. When a high voltage is applied to the electrodes, electrons are released by the cathode filament into the tube via a process called thermionic emission. Subsequently, free electrons accelerate towards the anode and create X-rays as they collide. In a vacuum tube, X-rays are generated in two distinct ways. The

characteristic X-rays are produced when free electrons strike atoms and eject one of the inner shell electrons. As electrons in higher energy states fill these vacancies, photons are emitted. X-rays can also be produced by bremsstrahlung in which electrons are decelerated by nuclei and residual energy is emitted as photon.

As X-rays traverse tissues, their intensity is attenuated. The patient dose is determined by the absorbed photon energy. X-ray beams interact with atoms in tissues through various mechanisms. X-ray beams may strike inner shell electrons and lead to vacancies that are subsequently filled by electrons in higher energy levels. This transition releases secondary X-rays, a process known as photoelectric absorption. The photon energy and the atomic number of the tissue determine the photoelectric absorption, which is crucial in image contrast. In CT images, Regions with higher energy absorption appear brighter, whereas areas with lower absorption appear darker. In another interaction mechanism, X-rays target outer shell electrons and eject them, causing ionization. Incident beams are then dispersed in different directions, which is known as Compton scattering. This mechanism contributes to image noise and patient dose. Additionally, in coherent scattering, low-energy X-rays are deflected by electrons without causing ejection, altering the beam's direction.

X-ray beams are exponentially attenuated as they traverse a material, following the Beer–Lambert law:

$$I = I_0 e^{-\mu l} \quad (2.22)$$

where I_0 is the initial beam intensity, μ is the attenuation coefficient per unit thickness in cm^{-1} , and l is the distance in cm that beams travels. In a medium with varying attenuation coefficients, the relationship is expressed by a line integral as follows

$$I = I_0 \exp \left(- \int_L \mu(x, y, z) dl \right) \quad (2.23)$$

where $\mu(x, y, z)$ is the attenuation coefficient at (x, y, z) position, L is the path, and dl is the differential path length that beams traverse. Subsequently, dividing

both sides by I_0 and taking the natural logarithm in 2.23 yields

$$-\ln\left(\frac{I}{I_0}\right) = \int_L \mu(x, y, z) dl. \quad (2.24)$$

This line integral corresponds to the Radon transform of the 3D attenuation coefficient $\mu(x, y, z)$. In CT imaging, where cross-sectional slices are sequentially scanned, this reduces to the Radon transform of the 2D attenuation map $\mu(x, y)$. The attenuation map directly represents internal body structures as attenuation coefficients vary depending on tissue types.

The Beer-Lambert law assumes monochromatic X-rays, whereas beams produced in a vacuum tube are polychromatic, i.e., have a wide energy spectrum due to bremsstrahlung. Dense tissues preferentially absorb low-energy X-ray photons, increasing the average energy of the transmitted beam, a phenomenon known as beam hardening. This results in image artifacts in CT, such as cupping or streaking, due to non-linear attenuation.

2.3.2 Image Formation

The Radon transform of the cross-sectional attenuation map $\mu(x, y)$ is

$$p(\theta, t) = \int_{L(\theta, t)} \mu(x, y) dl = -\ln\left(\frac{I(\theta, t)}{I_0}\right), \quad (2.25)$$

$$L(\theta, t) = \{(x, y) \in \mathbb{R} \times \mathbb{R} : x \cos \theta + y \sin \theta = t\}$$

where $p(\theta, t)$ is the projection data (sinogram) for the beam transmitted at angle θ and offset t . The detector in a CT scanner measures $I(\theta, t)$, which is the attenuated intensity of a beam along the line at (θ, t) . By computing $-\ln(I(\theta, t)/I_0)$, we obtain the Radon transform of the attenuation map from raw intensity measurements. CT image reconstruction algorithms estimate $\mu(x, y)$ via the inverse Radon transform.

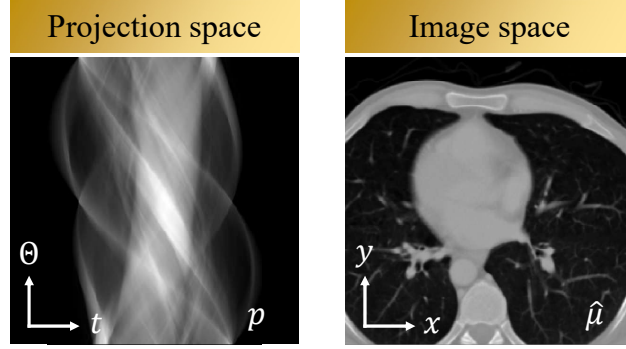


Figure 2.5: Projection space and image space in computed tomography (CT). The left image shows the sinogram (projection space), where each line represents a one-dimensional projection $p(\theta), t$ of the sample at angle θ . The right image depicts the reconstructed attenuation map $\hat{\mu}(x, y)$ in spatial coordinates, obtained by applying filtered backprojection.

2.3.3 Image Reconstruction

The CT image reconstruction aims to generate the underlying attenuation map $\mu(x, y)$ from raw intensity measurements I (Fig. 2.5). Traditional CT reconstruction techniques can be grouped into two main classes: analytical methods and iterative methods.

Analytical methods solve the inverse Radon transform in closed form. The Fourier slice theorem establishes the theoretical foundation for analytical solutions by describing the relationship between the projection domain and the image domain through the Fourier transform. According to the Fourier slice theorem, the one-dimensional Fourier transform of a projection acquired at angle θ corresponds to a line (or slice) through the two-dimensional Fourier transform of the object's attenuation map, taken along the same angle θ and passing through the origin in the frequency domain. This is expressed by

$$P(\theta, \omega) = M(\omega \cos \theta, \omega \sin \theta), \quad (2.26)$$

$$P(\theta, \omega) = \mathcal{F}\{p(\theta, t)\}, \quad M(u, v) = \mathcal{F}\{\mu(x, y)\}.$$

Collecting one-dimensional Fourier transforms of projections from $\theta = 0$ to $\theta = \pi$ fills frequencies on the two-dimensional Fourier space. Then, the inverse Fourier transform can be performed to obtain the CT image. However, direct Fourier

transform yields blurry images. Instead, in the filtered backprojection (FBP), projections are filtered by a high-pass ramp filter of $H(w) = |w|$:

$$\tilde{P}(\theta, \omega) = |\omega| \cdot P(\theta, \omega) \quad (2.27)$$

Subsequently, taking the inverse Fourier transform gives the filtered projections:

$$\tilde{p}(\theta, t) = \mathcal{F}^{-1}[\tilde{P}(\theta, \omega)]. \quad (2.28)$$

Finally, backprojection is performed to reconstruct the image from the filtered projections. The reconstructed pixel intensity at position (x, y) is obtained by integrating the contributions from all projection angles:

$$\hat{\mu}(x, y) = \int_0^\pi \tilde{p}(\theta, x \cos \theta + y \sin \theta) d\theta. \quad (2.29)$$

On the other hand, iterative methods employ numerical optimization schemes in which the initial attenuation map estimation is refined step-by-step. Algebraic reconstruction technique (ART) [87], simultaneous iterative reconstruction technique (SIRT) [88], and total variation minimization [89] are among iterative CT reconstruction techniques.

2.4 Sparse-view CT

In sparse-view, or few-view CT, the patient radiation dose and scan time are effectively reduced by acquiring a smaller number of projections compared to a full-scan protocol.

The CT image reconstruction problem can be expressed in terms of a linear system:

$$\mathcal{A}\mu + \varepsilon = p, \quad (2.30)$$

where μ is the attenuation map, p denotes measurements (projections), ε is measurement noise, and $\mathcal{A}$ is the imaging operator that links image and measurement domains. When fewer projections are acquired, this system of linear equations becomes undetermined.

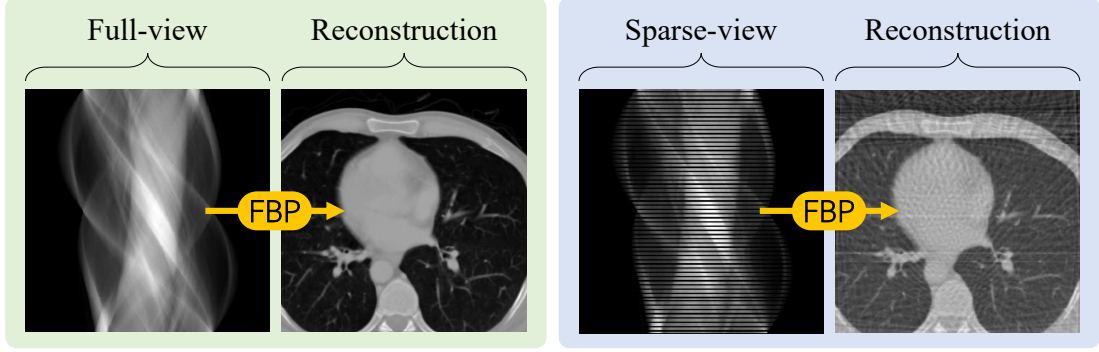


Figure 2.6: Comparison between full-view and sparse-view CT reconstruction using filtered backprojection (FBP). The left panel shows the sinogram (projection data) and corresponding reconstruction from a full-view acquisition. The right panel depicts the projection data with reduced angular sampling (sparse-view), and its FBP-based reconstruction, which suffers from visible streaking artifacts and reduced image quality due to the ill-posedness of the inverse problem under insufficient data.

Image reconstruction via FBP with a limited number of projections results in heavy streaking artifacts, which significantly reduce image quality and hinder reliable diagnostic interpretation (Fig. 2.6). Similar to the accelerated MRI, the sparse-view CT reconstruction can be performed via compressed sensing:

$$\hat{\mu} = \arg \min_{\mu} \frac{1}{2} \|\mathcal{A}x - p_{sub}\|_2^2 + \lambda \|\Psi(x)\|_1 \quad (2.31)$$

where $p_{sub} \subset p$ is the undersampled projections, and Ψ is a sparsifying transform. Recent studies utilize deep learning techniques to solve the sparse-view CT reconstruction problem.

Chapter 3

Related Work

Deep learning has driven significant advances in medical imaging tasks such as classification, synthesis, and reconstruction [90, 91, 92, 93, 94, 95, 96, 97, 98, 99]. Recent works on medical image synthesis employ generative models such as generative adversarial networks (GANs) [100, 101, 102, 103, 104], and diffusion models [105, 106, 107, 108]. While transformer-based architectures [30, 109, 110, 111, 112] have shown superior performance over earlier CNN-based methods, SSMs [113, 114, 115, 116] have recently been demonstrated to outperform both in a range of downstream medical imaging tasks.

3.1 Physics-Driven Models

Among learning-based approaches for medical image reconstruction, PD models strike a good balance between performance and generalization [117, 118]. However, their effectiveness in reducing undersampling artifacts depends largely on the network modules responsible for regularization. Tissue signals exhibit correlations spanning short to long spatial distances. As such, artifacts—arising from the interaction between the spatial distribution of tissue signals and imaging operators—also tend to show spatial correlations [11]. Capturing these contextual

features can be crucial for reliably separating artifacts from underlying anatomy.

Early PD models employed CNNs with compact filters to process feature maps with high efficiency and local accuracy [8, 24, 119]. However, such models can struggle with artifacts that exhibit distributed patterns—for example, those arising from heavy undersampling in MRI or non-Cartesian sampling in CT. Later PD models introduced transformers based on self-attention filters [34, 37]. While self-attention promises to effectively capture long-range context, it incurs substantial computational overhead. This often constrains transformers to operate on relatively coarse image patches—compromising fine structural details [38]—or to use a limited field of view, which reduces sensitivity to broader anatomical context [35]. Despite efforts to mitigate this trade-off, achieving a balanced sensitivity to both fine local details and extended spatial context remains a significant challenge for transformer-based PD models [40].

3.2 State Space Models

Recently, SSMs have emerged as an efficient alternative to transformers with a refined balance between local and non-local contextual sensitivity [43]. *Building on these strengths, we introduce MambaRoll as the first physics-driven autoregressive SSM model for medical image reconstruction.* Several earlier studies have explored SSMs commonly for MRI reconstruction [43, 44, 45], with one study extending to CT and PET [43]. These works provide valuable demonstrations of SSM’s potential in image reconstruction. However, they typically employ either DD [43, 44] or PD [45] designs based on single-spatial-scale SSM blocks, which can cause suboptimal capture of multi-scale contextual features evident in medical images. In contrast, MambaRoll introduces PD-SSM modules that operate progressively over spatial scales, enabling explicit linking of multi-scale contextual features.

While a few recent studies have begun to explore the use of SSMs for MRI reconstruction, MambaRoll carries principal novelties in terms of architecture and

application scope. One study proposes a DD SSM method with hierarchically tokenized scan trajectories to enhance contextual modeling [120]. Unlike [120], MambaRoll employs a PD approach and explicitly integrates an autoregressive prior across spatial scales to explicitly link multi-scale feature maps and thereby improve reconstruction fidelity. Another study proposes a PD method with spatial and spectral SSM modules for improved modeling of global context [121]. Yet, both SSM modules are confined to a single spatial scale, potentially limiting their contextual reach. MambaRoll instead introduces multi-scale PD-SSM modules and a DMSD training loss to enhance sensitivity to multi-scale contextual features.

Finally, while most prior SSM-based methods are limited to MRI, with the exception of [43], MambaRoll is designed to generalize beyond MRI to other medical image reconstruction tasks under a unified framework.

3.3 Autoregressive Models

Autoregressive modeling decomposes the joint probability distribution of a sequence of elements into a product of conditional distributions, such that each element is predicted conditioned on the preceding subset of elements within the sequence. Several previous reconstruction studies based on CNN architectures have applied autoregressive modeling to capture dependencies between elements in raw pixel sequences [122, 123] or between temporal steps in diffusion generative models [124]. In contrast, MambaRoll introduces a PD framework that integrates autoregressive modeling to capture dependencies between feature maps at different spatial scales, and leverages a novel SSM backbone for this purpose. *This represents, to the best of our knowledge, the first integration of spatial-scale autoregression in a medical image reconstruction model.* With these technical advances, we provide demonstrations of MambaRoll for performant MRI and CT reconstructions.

Chapter 4

Theory

4.1 Medical Image Reconstruction

In many modalities, acquired data and the underlying image that reflects the spatial distribution of tissue structure are associated through a linear system [125]:

$$\mathcal{A}x + \varepsilon = y, \quad (4.1)$$

where $x \in \mathbb{C}^N$ denotes the underlying complex image in vector form (N : number of pixels), $y \in \mathbb{C}^M$ are acquired complex data in vector form (M : number of measurements), ε is measurement noise, and $\mathcal{A} \in \mathbb{C}^{M \times N}$ is the imaging operator that describes the relationship between image and measurement domains. For MRI scans, $\mathcal{A} = \mathcal{P}\mathcal{F}\mathcal{C}$ where $\mathcal{P}$ denotes the k-space sampling pattern, $\mathcal{F}$ denotes Fourier transformation, and $\mathcal{C}$ denotes coil sensitivities [23]. Accordingly, $\mathcal{A}$ can be computed based on $\mathcal{P}$ and $\mathcal{F}$ that depend on known scan parameters, and $\mathcal{C}$ that can be estimated from data in a central calibration region. For CT scans, $\mathcal{A}$ can be taken as a Radon transformation $\mathcal{R}$ [1]. Image reconstruction from acquired measurements then involves solution of the linear system in Eq. 4.1 by inverting the imaging operator. Such inversion becomes increasingly more ill-conditioned as the number of measurements decline with respect to the number of pixels that are to be resolved [126]. Thus, in challenging tasks such as accelerated

MRI and sparse-view CT reconstruction, image priors are employed to improve problem conditioning by regularizing reconstructions:

$$\hat{x} = \arg \min_x \|\mathcal{A}x - y\|_2^2 + R(x), \quad (4.2)$$

where $R(x)$ is an image prior that regularizes reconstructions based on the distribution of high-quality medical images.

In PD models, the solution of Eq. 4.2 is typically operationalized as iterations through cascades of an unrolled architecture G_{θ_k} , such that the output of cascade $k \in [1 \ K]$ depends on the output of the previous cascade [31]:

$$\hat{x}_k = G_{\theta_k}(\hat{x}_{k-1}; \mathcal{A}, y), \quad (4.3)$$

where θ_k denotes the parameters of the k th cascade, and $\hat{x}_0 = \mathcal{A}^\dagger y$ is a linear reconstruction of undersampled data with $\dagger$ denoting the Hermitian adjoint operator. For a given cascade, $G_{\theta_k}(\hat{x}_{k-1}; \mathcal{A}, y) := \Psi_{dc}(\Psi_{nm}(\hat{x}_{k-1}); \mathcal{A}, y)$ interleaves projection through a network module for regularization, with projection through a data-consistency module for fidelity to acquired data [127]:

$$v_{out} = \Psi_{nm}(v_{in}) = \arg \min_{v_{out}} R_{\theta_k}(v_{out}|v_{in}), \quad (4.4)$$

$$v_{out} = \Psi_{dc}(v_{in}; \mathcal{A}, y) = v_{in} + \mathcal{A}^\dagger(y - \mathcal{A}v_{in}), \quad (4.5)$$

where the projection in Eq. 4.4 is a forward pass through the network module, v_{in} and v_{out} denote the projection input and output, respectively. Performance in PD models depends critically on the ability of Ψ_{nm} in separating undersampling artifacts from underlying tissue structure.

4.2 MambaRoll

MambaRoll embodies a set of PD-SSM modules to recover high-resolution medical images gradually across multiple spatial scales, followed by a refinement module that enhances recovery of fine-grained tissue structure. Inspired by recent advances in autoregressive modeling for computer vision tasks [128], we design our

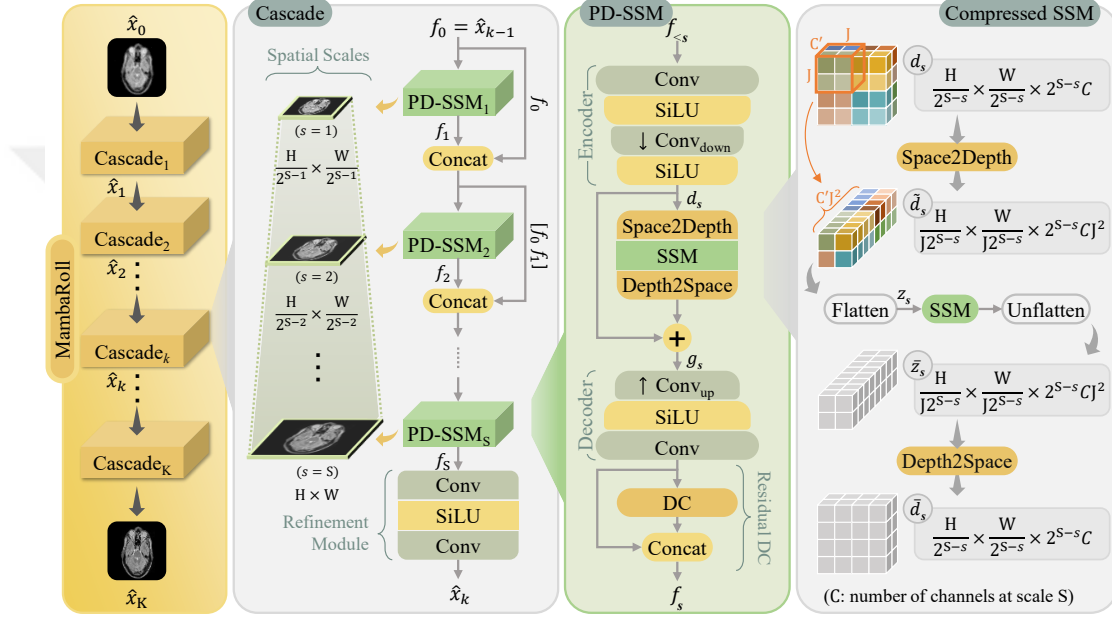


Figure 4.1: MambaRoll is a novel physics-driven autoregressive state space model, which utilizes an unrolled architecture with K cascades to map a linear reconstruction of undersampled data $\hat{x}_0$ to the learning-based reconstruction $\hat{x}_K$. In each cascade, MambaRoll recovers contextualized feature maps at high spatial resolution g_s progressively across S spatial scales by leveraging autoregressive prediction of finer-scale feature maps from coarser scales (see Fig. 4.2 for details). At a given scale s , recovery is performed by a novel physics-driven state space module (PD-SSM) that efficiently aggregates contextual features while maintaining consistency to acquired data. PD-SSM comprises a convolutional encoder to map high-resolution input feature maps $f_{<s}$ onto the desired scale, a compressed SSM block to capture context across feature maps efficiently, a convolutional decoder to recollect high-resolution feature maps, and a residual data-consistency (DC) block to enforce fidelity to acquired data. Following autoregressive predictions, a refinement module based on convolutional layers is used to enhance restoration of detailed texture.

PD-SSM modules to autoregressively predict feature maps at finer spatial scales conditioned on coarser-scale representations. This design serves to enhance contextual representation by progressively aggregating multi-scale features captured by SSM, while remaining consistent with the acquired data.

4.2.1 Network Architecture

Receiving a linear reconstruction of undersampled data as a two-channel input $\hat{x}_0 = \mathcal{A}^\dagger y \in \mathbb{R}^{H \times W \times 2}$ with channels storing real and imaginary components (H : image height, W : image width), MambaRoll projects its input through K cascades comprising *novel PD-SSM modules* and refinement modules to compute the reconstruction $\hat{x}_K \in \mathbb{R}^{H \times W \times 2}$. At cascade k , the input image $f_0 = \hat{x}_{k-1} \in \mathbb{R}^{H \times W \times 2}$ is first projected through S consecutive PD-SSM modules that autoregressively process feature maps across multiple spatial scales (Fig. 4.1):

$$f_s = \text{PD-SSM}_s(f_{<s}), \quad \text{for } s \in \{1, 2, \dots, S\}, \quad (4.6)$$

where s denotes the scale index, $f_s \in \mathbb{R}^{H \times W \times 4}$ for $s \geq 1$ (4 channels due to residual data-consistency blocks expressed in Eq. 4.15), and $f_{<s} := [f_0 \ f_1 \ \dots \ f_{s-1}] \in \mathbb{R}^{H \times W \times (4s-2)}$ is formed by concatenation of feature maps derived by PD-SSM modules operating at earlier scales. To ensure that PD-SSM modules are able to process feature maps while maintaining fidelity to acquired data, f_s for all s carry the same spatial dimensionality as the input image $\hat{x}_0$. Thus, processing of feature maps at different spatial scales is achieved by adjusting resolution internally within PD-SSM modules (Fig. 4.2), as described below. Afterwards, a refinement module is employed to enhance recovery of structural details based on convolutional layers:

$$\hat{x}_k = \text{RM}(f_S) := \text{Conv}(\sigma(\text{Conv}(f_S))), \quad (4.7)$$

where $\hat{x}_k \in \mathbb{R}^{H \times W \times 2}$ is the output of the k th cascade, and σ is a SiLU activation function.

PD-SSM modules. Diverging from conventional SSM modules, PD-SSM modules are based on a novel design that enables progressive recovery of images

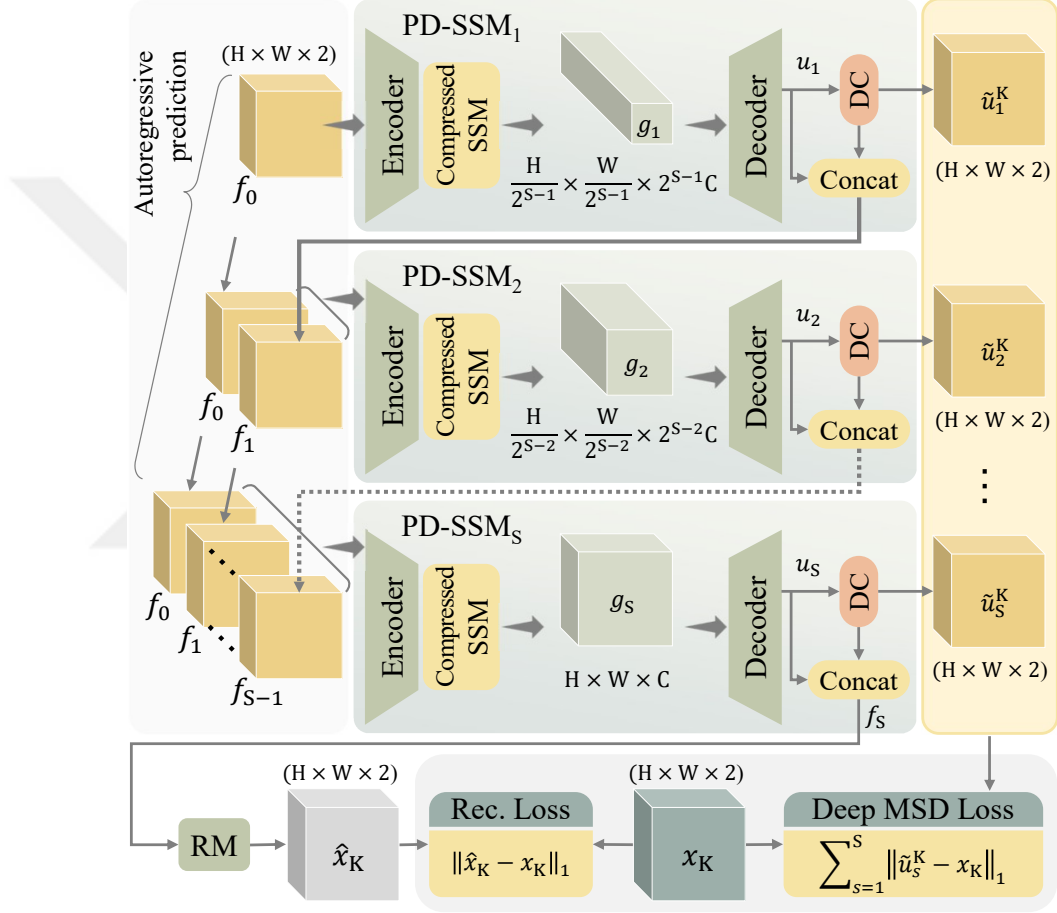


Figure 4.2: Illustration of autoregressive prediction and deep supervision of multi-scale representations in MambaRoll. To explicitly link multi-scale representations via an autoregressive prior, the PD-SSM module at a given scale s predicts the finer-scale feature map (f_s) conditioned on the preceding set of feature maps obtained at coarser scales ($f_{<s} := [f_0 \ f_1 \ \dots \ f_{s-1}]$). For model supervision, a deep multi-scale decoding loss (DMSD) loss is used in combination with a conventional reconstruction loss, where $\hat{x}_K$ is derived via the refinement module (RM). DMSD supervises intermediate reconstructions ($\tilde{u}_s^K$) at each scale to further enhance multi-scale learning.

across multiple spatial scales while maintaining fidelity to acquired data at each scale. In particular, a given PD-SSM_s module at scale s first projects its input $f_{<s} \in \mathbb{R}^{H \times W \times 4s-2}$ through a convolutional encoder to map high-resolution feature maps onto the desired scale:

$$d_s = \text{Enc}_s(f_{<s}) := \sigma(\text{Conv}_{\text{down}}(\sigma(\text{Conv}(f_{<s})))), \quad (4.8)$$

where $d_s \in \mathbb{R}^{H_s \times W_s \times C_s}$ with $H_s = H/2^{(S-s)}$, $W_s = W/2^{(S-s)}$, $C_s = 2^{(S-s)}C$ (C : number of channels at scale S), and $\text{Conv}_{\text{down}}$ performs spatial downsampling while Conv expands channel dimensionality to C_s .

Afterwards, a compressed SSM block with residual connection is employed to contextual features at the s th scale as $g_s = \text{SSM}(d_s)$, while improving efficiency in state-space modeling. Computational complexity of conventional SSMs scale with the sequence length (i.e., the number of image pixels). To improve efficiency, our compressed SSM blocks applied a space-to-depth transformation on input feature map d_s to shorten the sequence length by a factor of J^2 , by tiling $J \times J$ pixel patches along the channel dimension [129]:

$$\tilde{d}_s[w_1, w_2, \tilde{c}] = d_s[J(w_1 - 1) + v_1, J(w_2 - 1) + v_2, c], \quad (4.9)$$

$$c = \left\lceil \frac{\tilde{c}}{J^2} \right\rceil, \quad v_1 = \left\lceil \frac{((\tilde{c} - 1) \bmod J^2) + 1}{J} \right\rceil,$$

$$v_2 = ((\tilde{c} - 1) \bmod J) + 1.$$

where $w_1 \in \{1, \dots, H_s/J\}$, $w_2 \in \{1, \dots, W_s/J\}$, and $\tilde{c} \in \{1, \dots, C_s J^2\}$. The resultant feature map $\tilde{d}_s \in \mathbb{R}^{H_s/J \times W_s/J \times C_s J^2}$ is transformed onto a sequence $z_s \in \mathbb{R}^{H_s W_s / J^2 \times C_s J^2}$ via a two-dimensional sweep scan, and independently processed across channels with a discretized state-space model:

$$h_s[n, c] = \mathbf{A}h_s[n-1, c] + \mathbf{B}z_s[n, c], \quad (4.10)$$

$$\bar{z}_s[n, c] = \mathbf{C}h_s[n, c], \quad (4.11)$$

where $n \in [1, H_s W_s / J^2]$ denotes sequence index, $h_s[n, c] \in \mathbb{R}^{D \times 1}$ denotes the scale-specific hidden state at step n , and $\mathbf{A} \in \mathbb{R}^{D \times D}$, $\mathbf{B} \in \mathbb{R}^{D \times 1}$, $\mathbf{C} \in \mathbb{R}^{1 \times D}$ are learnable parameters (D : the state dimensionality). In Eqs. 4.10-4.11, $\mathbf{B}$ and $\mathbf{C}$ are defined as functions of z_s as in selective SSM variants to enable input-dependent filtering [46]. After the SSM layer, the output sequence $\bar{z}_s$ is expanded

and depth-to-space transformed:

$$\begin{aligned}\bar{d}_s[J(w_1 - 1) + v_1, J(w_2 - 1) + v_2, c] &= \text{expand}(\bar{z}_s)[w_1, w_2, \tilde{c}], \\ \tilde{c} &= (c - 1) \cdot J^2 + (v_1 - 1) \cdot J + v_2, \quad v_1, v_2 \in \{1, \dots, J\}.\end{aligned}\tag{4.12}$$

The restored feature map $\bar{d}_s \in \mathbb{R}^{H_s \times W_s \times C_s}$ is residually added onto the input:

$$g_s = \bar{d}_s + d_s.\tag{4.13}$$

To enable PD-SSM modules to enforce fidelity to acquired data, the contextualized feature map g_s is projected through a convolutional decoder to recollect a high-resolution image:

$$u_s = \text{Dec}_s(g_s) := \text{Conv}(\sigma(\text{Conv}_{\text{up}}(g_s))),\tag{4.14}$$

where $u_s \in \mathbb{R}^{H \times W \times 2}$, Conv_{up} performs spatial upsampling while Conv reduces channel dimensionality to 2 to store real and imaginary components. Finally, the recollected image can be projected through a residual data-consistency block that concatenates u_s with a data-consistent version of itself:

$$\tilde{u}_s = \text{DC}(u_s) := u_s + \mathcal{A}^\dagger(y - \mathcal{A}u_s),\tag{4.15}$$

$$f_s = \text{resDC}(u_s) := [u_s \quad \tilde{u}_s],\tag{4.16}$$

where the second term ensures that reconstructed data are consistent with acquired data y for the subset of collected spatial encodings in measurement domain. This design allows the model to have fine-grained control over the emphasis on data consistency to optimize task performance.

4.2.2 Deep Multi-Scale Decoding Loss

A core innovation in MambaRoll lies in its autoregressive processing across spatial scales that promotes contextual aggregation. Yet, the effectiveness of each scale’s prediction is crucially dependent on its alignment with the underlying anatomical structure. To strengthen learning at intermediate scales, we introduce a *Deep Multi-Scale Decoding (DMSD)* loss. Specifically, the feature map g_s^K at scale

s within the final cascade (i.e., K th) of the unrolled architecture is decoded to a full-resolution image $u_s^K = \text{Dec}_s(g_s^K)$ through a learned decoder (Fig. 4.2). These decoded images are projected through the data-consistency block and then supervised against the ground-truth image x_K to provide direct learning signals at multiple scales:

$$\mathcal{L}_{\text{DMSD}} = \sum_{s=1}^S \|\tilde{u}_s^K - x_K\|_1 = \sum_{s=1}^S \|\text{DC}(\text{Dec}_s(g_s^K)) - x_K\|_1. \quad (4.17)$$

By supervising a strictly data-consistency enforced version of decoded images (i.e., $\tilde{u}_s^K$) rather than the residually-corrected images (i.e., f_s^K), the DMSD loss directly promotes the formation of scale-specific features that are more closely aligned with the ground-truth image.

The total loss for MambaRoll then combines the reconstruction loss on the predicted image with Eq. 4.17:

$$\mathcal{L}_{\text{total}} = \|\hat{x}_K - x_K\|_1 + \mathcal{L}_{\text{DMSD}}, \quad (4.18)$$

where $\hat{x}_K$ denoted the model’s predicted image. Note that the DMSD loss departs from conventional supervision by treating scale-wise predictions not merely as intermediate latent steps, but as independently decodable image estimates. By grounding images decoded at separate spatial scales individually to the ground-truth image, this approach facilitates mitigation of residual artifacts within the feature maps at coarser spatial scales.

Chapter 5

Methods

5.1 Datasets

We conducted evaluations on two publicly available datasets comprising a variety of imaging protocols. For each dataset, subjects were selected sequentially from the original repositories to avoid potential bias from manual filtering. Data were randomly split into independent training, validation and test sets at the subject level, ensuring that no subject appeared in more than one set.

Accelerated MRI. Experiments were conducted on brain MRI data from the fastMRI [130] dataset. For each subject, three volumes with T_1 , T_2 and FLAIR contrasts were analyzed as multi-coil complex k-space data, where each volume comprised 10 cross-sections of size 320×320 . Multi-coil data were coil compressed onto 5 virtual coils while preserving 90% of data variance [131], and coil sensitivities were estimated via ESPIRiT [132]. Data were split into training, validation, test sets of (100, 10, 40) subjects, corresponding to (3000, 300, 1200) cross-sectional samples. Accelerated MRI scans were obtained via two-dimensional variable-density undersampling at rates $R = 4 - 12$ [81]. Linear reconstructions of undersampled data were obtained by zero-filling missing k-space samples, followed by inverse Fourier transformation (ZF).

Sparse-view CT. Experiments were conducted on lung CT data from the LoDoPaB-CT dataset [133]. For each subject, a single volume was analyzed comprising 90 cross-sections of size 352×352 . Data were split into non-overlapping training, validation, test sets composed of (30, 6, 12) subjects, corresponding to (2700, 540, 1080) cross-sectional samples. To obtain sparse-view CT scans, ground-truth reconstructions in LoDoPaB-CT were Radon transformed based on a two-dimensional parallel-beam geometry, and sinograms were uniformly under-sampled at rates $R = 4 - 8$ across the angular dimension. Linear reconstructions of undersampled data were obtained via filtered back projection (FBP).

5.2 Architectural Details

MambaRoll used PD-SSM modules for autoregressive prediction across $S = 3$ spatial scales (at 0.25, 0.5, 1 scale of the original image resolution), and a refinement module to provide the cascade output. Each PD-SSM module received feature maps from the previous cascade as input, along with features maps at earlier scales in the current cascade. The PD-SSM module comprised a convolutional encoder to downsample feature maps to the target scale, a compressed SSM block to model spatial context, a convolutional decoder to upsample feature maps, and a residual data-consistency block. The encoder used expansion factors across the channel dimension of (4, 2, 1), whereas the decoder used expansion factors of (0.25, 0.5, 1) across scales. The SSM block used a compression factor of $J = 4$ and a sweep-scan trajectory to project feature maps onto a sequence, and a state-dimensionality of $D = 16$. The refinement module comprised a convolutional block. All convolution layers used a 3×3 kernel size and SiLU activation functions.

5.3 Competing Methods

MambaRoll was demonstrated against state-of-the-art baselines in MRI and sparse-view CT reconstruction, including both data-driven and physics-driven methods. For each method, key hyperparameters including the number of cascades in the unrolled architecture (K) and learning rate were selected to maximize performance on the validation set. Models were trained via the Adam optimizer for 50 epochs, $(\beta_1, \beta_2) = (0.9, 0.999)$, and cosine annealing of learning rates in the range $[\eta_{min}, \eta_{max}]$.

5.3.1 Accelerated MRI Reconstruction

Six MRI reconstruction baselines were considered including convolutional (MoDL, PD-UNet), transformer (SwinMR, DSFormer), and SSM (MambaMIR, MambaRecon) methods which are categorized in Table 5.1.

MambaRoll: The proposed model was implemented with $K=5$, $\eta_{min}=10^{-5}$, $\eta_{max}=2 \times 10^{-4}$.

MambaRecon: A physics-driven SSM model based on visual state space blocks was considered [45]. Cross-validated parameters were $K=5$, $\eta_{min}=10^{-6}$, and $\eta_{max}=10^{-3}$.

MambaMIR: A data-driven SSM model based on arbitrary-masked state space blocks was considered [43]. Cross-validated parameters were $\eta=2 \times 10^{-5}$.

DSFormer: A physics-driven transformer model with shifted windowing multi-head self-attention blocks was considered [39]. Cross-validated parameters were $K=5$, and $\eta=2 \times 10^{-4}$.

SwinMR: A data-driven transformer model with shifted windowing multi-head self-attention blocks was considered [35]. Cross-validated parameters were $\eta_{min}=6 \times 10^{-6}$, and $\eta_{max}=2 \times 10^{-4}$.

Table 5.1: Categorization of the competing methods on accelerated MRI reconstruction. (DD: data-driven, PD: physics-driven, CNN: convolutional, ViT: transformer, SSM: state-space)

	Framework		Architecture		
	DD	PD	CNN	ViT	SSM
UNet [135]	✓		✓		
MoDL [31]		✓	✓		
SwinMR [35]	✓			✓	
DSFormer [39]		✓		✓	
MambaMIR [43]	✓				✓
MambaRecon [45]		✓			✓
MambaRoll		✓			✓

UNet: A data-driven convolutional model based on a UNet backbone was considered [134, 135]. Cross-validated parameters were $\eta_{min}=10^{-6}$, and $\eta_{max}=10^{-3}$.

MoDL: A physics-driven convolutional model based on convolutional layers processing feature maps at the original image resolution was considered [31]. Cross-validated parameters were $K=5$, $\eta_{min}=10^{-5}$, and $\eta_{max}=2\times 10^{-4}$.

5.3.2 Sparse-view CT Reconstruction

Six CT reconstruction baselines were considered including convolutional (FBP-ConvNet, MetaInv-Net), transformer (USwin, RegFormer), and SSM (MambaMIR, PD-Mamba) methods which are categorized in Table 5.2.

MambaRoll: The proposed model was implemented with $K=3$, $\eta_{min}=10^{-5}$, $\eta_{max}=2\times 10^{-4}$.

PD-Mamba: A physics-driven model was considered that interleaved convolutional [31] and visual state space blocks [46] with data-consistency blocks. Cross-validated parameters were $K=3$, $\eta_{min}=10^{-5}$, and $\eta_{max}=2\times 10^{-4}$.

MambaMIR: A data-driven SSM model based on arbitrary-masked state space blocks was considered [43]. Cross-validated parameters were $\eta=2\times 10^{-5}$.

Table 5.2: Categorization of the competing methods on sparse-view CT reconstruction. (DD: data-driven, PD: physics-driven, CNN: convolutional, ViT: transformer, SSM: state-space)

	Framework		Architecture		
	DD	PD	CNN	ViT	SSM
FBPConvNet [20]	✓		✓		
MetaInv-Net [27]		✓	✓		
USwin [136]	✓			✓	
RegFormer [34]		✓		✓	
MambaMIR [43]	✓				✓
PD-Mamba [46]		✓			✓
MambaRoll		✓			✓

RegFormer: A physics-driven transformer model with shifted windowing multi-head self-attention blocks was considered [34]. Cross-validated parameters were $K=3$, and $\eta=10^{-4}$.

USwin: A data-driven transformer model with shifted windowing multi-head self-attention blocks was considered [136]. Cross-validated parameters were $\eta=10^{-4}$.

MetaInv-Net: A physics-driven convolutional model based on convolutional layers processing feature maps at the original image resolution was considered [27]. Cross-validated parameters were $K=3$, $\eta_{min}=10^{-6}$, and $\eta_{max}=5 \times 10^{-5}$.

FBPConvNet: A data-driven convolutional model based on a UNet backbone was considered [20]. Cross-validated parameters were $\eta_{min}=10^{-6}$, and $\eta_{max}=10^{-3}$.

5.4 Modeling Procedures

Network models used two separate input-output channels to represent real and imaginary components of MR images, and a single input-output channel for CT images. Modeling was performed via the PyTorch framework on an Nvidia RTX 4000 Ada Generation GPU. Models were trained and tested on two-dimensional

cross sections. Experiments confirmed that model performance consistently converged with the training set sizes used in this thesis. Reconstruction performance was assessed by quantifying peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) between recovered and reference images. Reference images were derived via Fourier reconstruction of fully-sampled acquisitions for MRI scans, and via filtered back-projection of fully-sampled sinograms in CT scans. Significance of performance differences among competing methods were assessed via Wilcoxon signed-rank tests.

Chapter 6

Results

6.1 Accelerated MRI Reconstruction

Demonstrations were first performed for accelerated MRI reconstruction. MambaRoll was compared against state-of-the-art models based on convolutional (MoDL, UNet), transformer (SwinMR, DSFormer), and SSM (MambaMIR, MambaRecon) backbones. Performance metrics for reconstructions of undersampled data at rates $R = 4 - 12$ are listed in Table 6.1 and 6.2. We find that MambaRoll significantly outperforms all competing methods across reconstruction tasks ($p < 0.05$). On average across tasks, MambaRoll offers (PSNR, SSIM)

Table 6.1: Accelerated MRI reconstruction performance at $R = 4 - 12$. PSNR (dB) and SSIM (%) listed as mean $\pm$ std across the test set. Boldface marks the top performing method.

	$R = 4$		$R = 8$		$R = 12$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
UNet	36.25 $\pm$ 2.42	94.27 $\pm$ 3.53	33.93 $\pm$ 2.38	91.63 $\pm$ 4.98	30.29 $\pm$ 5.43	89.63 $\pm$ 5.69
MoDL	41.76 $\pm$ 2.55	97.64 $\pm$ 2.08	37.56 $\pm$ 2.60	95.03 $\pm$ 3.69	35.87 $\pm$ 2.41	93.19 $\pm$ 4.48
SwinMR	37.31 $\pm$ 2.38	94.63 $\pm$ 2.61	34.10 $\pm$ 2.18	91.58 $\pm$ 3.90	32.60 $\pm$ 2.06	89.65 $\pm$ 4.47
DSFormer	43.13 $\pm$ 3.12	97.91 $\pm$ 2.12	38.69 $\pm$ 2.84	95.04 $\pm$ 3.74	36.90 $\pm$ 2.71	93.23 $\pm$ 4.62
MambaMIR	38.58 $\pm$ 2.60	95.12 $\pm$ 3.23	36.05 $\pm$ 2.56	92.66 $\pm$ 4.88	34.70 $\pm$ 2.47	91.38 $\pm$ 5.61
MambaRecon	43.45 $\pm$ 3.31	98.16 $\pm$ 2.31	38.74 $\pm$ 2.85	95.57 $\pm$ 3.94	37.03 $\pm$ 2.72	93.99 $\pm$ 4.69
MambaRoll	45.55$\pm$3.71	98.57$\pm$3.57	39.51$\pm$3.04	96.12$\pm$3.71	37.68$\pm$2.77	94.61$\pm$4.37

Table 6.2: Average accelerated MRI reconstruction performance across acceleration rates. PSNR (dB) and SSIM (%) listed as mean $\pm$ std across the test set. Boldface marks the top performing method.

	PSNR	SSIM
UNet	33.49	91.84
MoDL	38.39	95.28
SwinMR	34.67	91.95
DSFormer	39.57	95.39
MambaMIR	36.44	93.05
MambaRecon	39.74	95.91
MambaRoll	40.91	96.43

improvements of 4.97dB, 2.87% over convolutional; 3.79dB, 2.76% over transformer; and 2.82dB, 1.95% over SSM baselines. It is also important to note that physics-driven methods (MambaRoll, MambaRecon, DSFormer, MoDL) generally yield superior performance against data-driven methods (MambaMIR, SwinMR, UNet). Taken together, these results demonstrate that incorporating physics-driven modules enhances reliability in MRI reconstruction, and that the autoregressive state-space architecture of MambaRoll outperforms convolutional, transformer, and conventional SSM backbones in recovering high-quality MR images.

Representative reconstructions from competing methods are shown in Fig. 6.1. Data-driven baselines (MambaMIR, SwinMR) exhibit pronounced noise, residual artifacts, and visible blur. While physics-driven baselines are less affected by noise and blur, they display distinct artifacts: DSFormer presents ringing artifacts, MambaRecon shows pixellation primarily near high-intensity tissue regions, MoDL and to a higher degree UNet exhibit edge blurring. In contrast, MambaRoll achieves enhanced tissue delineation with reduced noise and artifact levels. These gains can be attributed to the autoregressive state-space approach in MambaRoll that better captures contextual features to attain superior reconstructions.

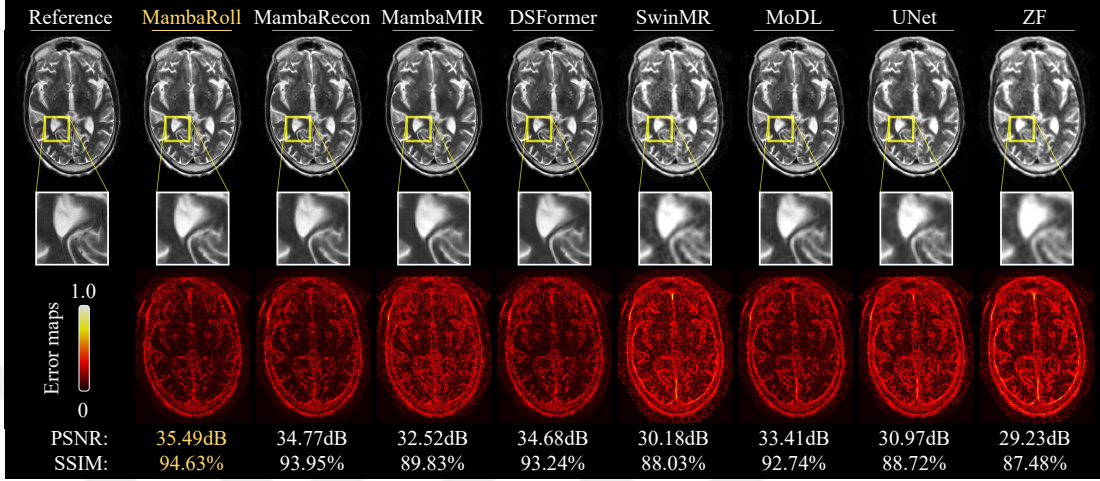


Figure 6.1: Accelerated MRI reconstructions of a representative cross-section from a T₂-weighted acquisition at $R = 12$. Reconstructed images from competing methods are shown along with a linear reconstruction of undersampled data (ZF), and the reference image derived from fully-sampled data. Zoom-in windows and error maps are included to highlight differences in tissue depiction. PSNR and SSIM values are listed below each reconstruction.

6.2 Sparse-view CT Reconstruction

Next, demonstrations were performed for sparse-view CT reconstruction. MambaRoll was compared against state-of-the-art models based on convolutional (FBPConvNet, MetaInv-Net), transformer (USwin, RegFormer), and SSM (MambaMIR, PD-Mamba) backbones. Performance metrics for reconstructions of undersampled data at rates $R = 4 - 8$ are listed in Table 6.3 and 6.4. We find that MambaRoll significantly outperforms all competing methods, consistently across reconstruction tasks ($p < 0.05$). On average across tasks, MambaRoll offers (PSNR, SSIM) improvements of 4.84dB, 6.75% over convolutional; 2.80dB, 2.64% over transformer; and 2.44dB, 3.70% over SSM baselines. Although physics-driven methods still generally tend to offer performance benefits over their data-driven counterparts (e.g., PD-Mamba vs MambaMIR, RegFormer vs USwin), performance differences are relatively moderate compared to those in MRI reconstruction. Moreover, while MambaRoll offered comparable performance benefits over baselines in MRI, its benefits over convolutional baselines are substantially stronger in CT. This results can be attributed to the streaking-like

Table 6.3: Sparse-view CT reconstruction performance at $R = 4 - 8$. Boldface marks the top performing method.

	$R = 4$		$R = 8$		$R = 12$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
FBPConvNet	34.41±1.65	85.68±4.92	33.85±1.73	84.86±5.75	32.49±1.57	81.76±5.94
MetaInv-Net	36.31±2.23	90.03±5.72	33.10±1.68	84.82±6.61	31.17±1.38	80.90±6.82
USwin	36.41±2.58	91.96±5.33	33.48±1.94	87.41±6.24	32.11±1.68	84.21±6.48
RegFormer	39.88±3.23	93.33±4.89	36.85±2.58	89.49±5.99	34.79±2.17	86.29±6.39
MambaMIR	37.42±2.40	89.57±4.77	35.42±2.20	86.73±5.89	33.92±1.99	83.74±6.19
PD-Mamba	38.99±2.97	92.64±5.08	35.90±2.33	88.35±6.18	34.06±1.99	85.27±6.65
MambaRoll	40.72±3.56	94.05±4.86	38.00±2.93	91.15±5.97	36.44±2.62	89.06±6.49

Table 6.4: Average Sparse-view CT reconstruction performance across acceleration rates. Boldface marks the top performing method.

	PSNR	SSIM
FBPConvNet	33.58	84.10
MetaInv-Net	33.53	85.25
USwin	34.00	87.86
RegFormer	37.17	89.70
MambaMIR	35.59	86.68
PD-Mamba	36.32	88.75
MambaRoll	38.39	91.42

aliasing artifacts spanning over long distances in sparse-view CT images, compromising the performance of convolutional methods that have locality biases. Overall, our results indicate that, in the context of sparse-view CT reconstruction, the autoregressive state-space design of MambaRoll proves more effective than convolutional, transformer-based, and traditional SSM architectures in reconstructing high-fidelity images from limited projections.

Representative images from sparse-view CT reconstructions are presented in Fig. 6.2. Data-driven baselines (MambaMIR, USwin, FBPConvNet) show notable reconstruction errors due to spatial blurring and residual streaking artifacts. Although physics-driven baselines (PD-Mamba, RegFormer, MetaInvNet) tend to yield improved performance, they still suffer from moderate streaking artifacts that occasionally manifest as contrast inaccuracies near regions of bright tissue signals. In comparison, MambaRoll more effectively mitigates streaking artifacts in CT images, yielding reconstructions that closely resemble the ground truth in tissue depiction and contrast. Corroborating findings on MRI reconstruction,

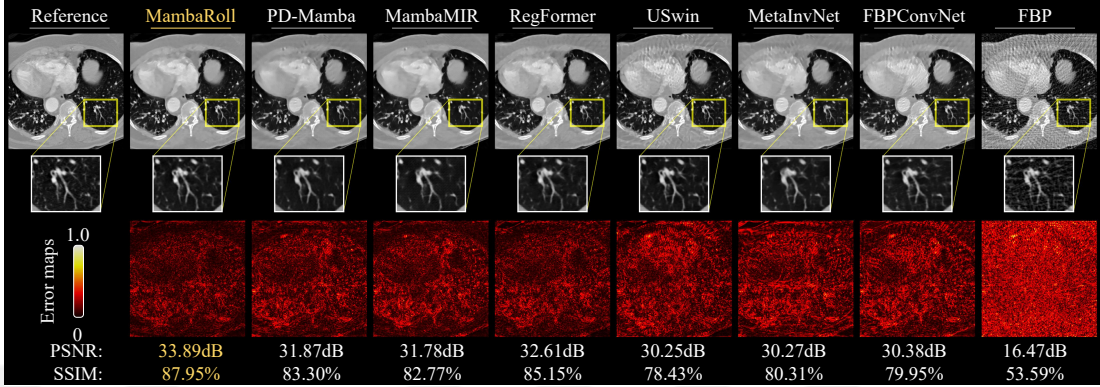


Figure 6.2: Sparse-view CT reconstructions of a representative cross-section at $R = 8$. Reconstructed images from competing methods are shown along with a linear reconstruction of undersampled data (FBP), and the reference image derived from fully-sampled data.

these results indicate that MambaRoll’s autoregressive state-space architecture help improve capture of complex spatial dependencies to enable superior artifact suppression and structural fidelity.

6.3 Generalization Under Acquisition Variability

To investigate the robustness of reconstruction models under distributional shifts, we evaluated generalization performance across varying undersampling rates—an important aspect of domain variability in accelerated imaging. As representative scenarios, we considered MRI reconstruction with models trained at $R = 12$ and tested at $R = 8$, and vice versa; similarly, we considered CT reconstruction with models trained at $R = 8$ and tested at $R = 6$, and vice versa. These cases were selected to capture both upward and downward shifts in undersampling factors. Metrics for generalization performance are listed in Table 6.5 for MRI reconstruction tasks, and in Table 6.6 for CT reconstruction tasks. We find that MambaRoll significantly outperforms respective baselines in both MRI and CT tasks ($p < 0.05$). On average across MRI tasks, MambaRoll offers (PSNR, SSIM)

Table 6.5: Generalization performance in MRI reconstruction under variability in R values between training and test sets ($R_{\text{training}} \rightarrow R_{\text{test}}$).

	$R : 12 \rightarrow 8$		$R : 8 \rightarrow 12$	
	PSNR	SSIM	PSNR	SSIM
UNet	32.30 $\pm$ 2.14	87.79 $\pm$ 5.68	31.31 $\pm$ 1.98	86.84 $\pm$ 6.07
MoDL	36.41 $\pm$ 2.54	92.95 $\pm$ 4.00	34.09 $\pm$ 2.18	90.85 $\pm$ 4.72
SwinMR	33.70 $\pm$ 2.02	90.45 $\pm$ 4.33	32.01 $\pm$ 1.90	87.97 $\pm$ 5.16
DSFormer	35.23 $\pm$ 2.37	92.58 $\pm$ 4.06	35.00 $\pm$ 2.44	91.02 $\pm$ 4.76
MambaMIR	32.95 $\pm$ 2.21	88.98 $\pm$ 5.34	33.01 $\pm$ 2.20	87.89 $\pm$ 5.70
MambaRecon	36.52 $\pm$ 2.72	93.36 $\pm$ 4.30	35.45 $\pm$ 2.49	91.99 $\pm$ 4.60
MambaRoll	37.36$\pm$2.94	94.30$\pm$3.80	35.93$\pm$2.55	92.68$\pm$4.38

Table 6.6: Generalization performance in CT reconstruction under variability in R values between training and test sets ($R_{\text{training}} \rightarrow R_{\text{test}}$).

	$R : 8 \rightarrow 6$		$R : 6 \rightarrow 8$	
	PSNR	SSIM	PSNR	SSIM
FBPConvNet	33.40 $\pm$ 1.64	83.96 $\pm$ 5.89	31.29 $\pm$ 1.48	79.44 $\pm$ 6.21
MetaInvNet	32.40 $\pm$ 1.34	83.50 $\pm$ 6.62	30.72 $\pm$ 1.46	79.96 $\pm$ 6.68
USwin	34.80 $\pm$ 1.88	87.67 $\pm$ 6.22	28.91 $\pm$ 1.94	78.04 $\pm$ 6.55
RegFormer	35.75 $\pm$ 2.18	88.39 $\pm$ 6.01	33.90 $\pm$ 1.93	85.03 $\pm$ 6.29
MambaMIR	34.60 $\pm$ 1.98	84.95 $\pm$ 6.11	32.80 $\pm$ 1.67	82.29 $\pm$ 5.87
PD-Mamba	35.07 $\pm$ 2.00	87.22 $\pm$ 6.35	33.38 $\pm$ 1.84	84.05 $\pm$ 6.42
MambaRoll	37.53$\pm$2.76	90.75$\pm$6.11	34.68$\pm$2.37	85.81$\pm$6.34

improvements of 3.12dB, 3.88% over convolutional; 2.66dB, 2.99% over transformer; and 2.16dB, 2.94% over SSM baselines. On average across CT tasks, MambaRoll offers (PSNR, SSIM) improvements of 4.15dB, 6.56% over convolutional; 2.77dB, 3.50% over transformer; and 2.14dB, 3.65% over SSM baselines.

Representative images are depicted in Fig. 6.3 for MRI reconstruction, and in Fig. 6.4 for CT reconstruction. In both sets of results, data-driven methods show elevated noise and residual artifacts, while physics-driven baselines better suppress noise but exhibit a moderate degree of modality-specific artifacts. Consistent with performance gains confirmed by quantitative evaluations, these results indicate that MambaRoll achieves strong generalization across undersampling factors, outperforming baseline models under mismatched training and testing configurations.

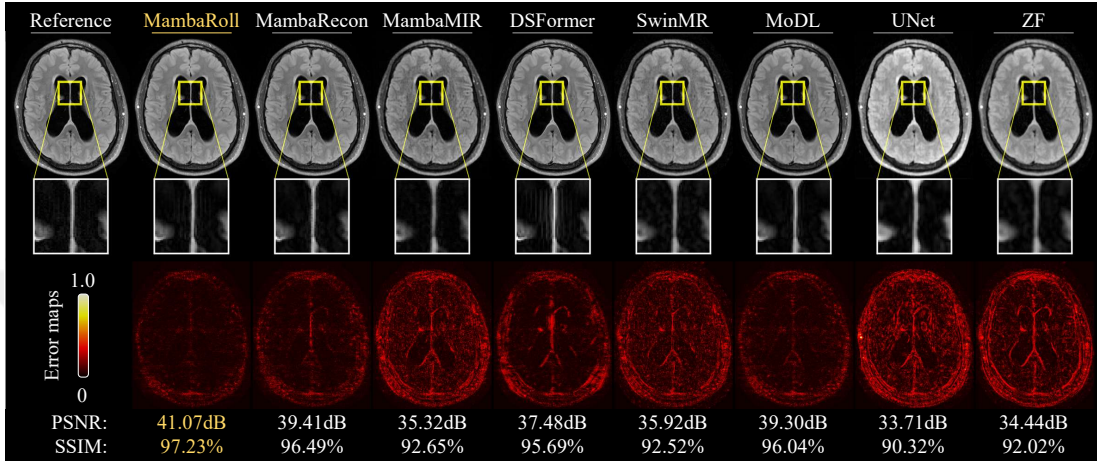


Figure 6.3: Accelerated MRI reconstructions of a representative cross-section from a FLAIR-weighted acquisition at $R = 8$ with models trained at $R = 12$. Reconstructed images from competing methods are shown along with a linear reconstruction of undersampled data (ZF), and the reference image derived from fully-sampled data.

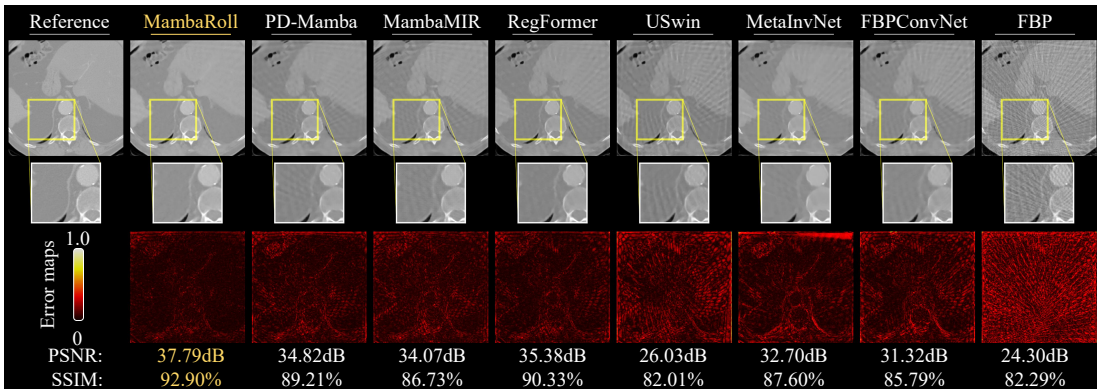


Figure 6.4: Sparse-view CT reconstructions of a representative cross-section at $R = 8$ with models trained at $R = 6$. Reconstructed images from competing methods are shown along with a linear reconstruction of undersampled data (FBP), and the reference image derived from fully-sampled data.

6.4 Contextual Awareness

To better understand how different models utilize spatial context during reconstruction, we examined contextual awareness — defined here as the model’s ability to integrate information from spatially distant but contextually dependent regions of the input image. In medical image reconstruction, such contextual integration is critical for capturing extended anatomical structures and reducing local ambiguities, especially under undersampled or corrupted inputs. For this purpose, we visualized effective receptive fields (ERFs) by backpropagating gradients from the reconstructed output at a fixed spatial location to the input image. This analysis reveals which input regions most influence the model’s prediction, serving as a proxy for its spatial sensitivity. Fig. 6.5 displays the ERFs of competing methods for MRI and CT reconstruction. Overall, MambaRoll exhibits the broadest and most coherent ERF coverage across both imaging modalities, indicating consistent multi-scale contextual integration. This is in contrast to conventional SSMs, whose ERFs are moderately expansive but often focalized across horizontal and vertical lines reflecting biases from the raster-scans used to sequentialize image pixels. Meanwhile, transformer-based models show relatively compact or fragmented ERFs — especially those using local window attention (e.g., SwinMR, USwin), corroborating findings in the literature. This suggests that despite their theoretical capacity for global attention, practical transformer implementations may underutilize broad spatial cues without specific architectural enhancements. CNN-based models displayed variable ERFs: deeper models with UNet-like structures (e.g., UNet, MetaInvNet) achieved relatively large fields due to multi-resolution processing and skip connections, whereas more basic designs using consistent feature map resolution across network stages (e.g., MoDL, FBPCnvNet) remained highly localized. Collectively, these findings highlight that by combining autoregressive modeling with multi-scale state space processing, MambaRoll more coherently captures long-range dependencies. This broad spatial sensitivity is particularly valuable in medical image reconstruction, where anatomical coherence and global structure play crucial roles in accurate recovery.

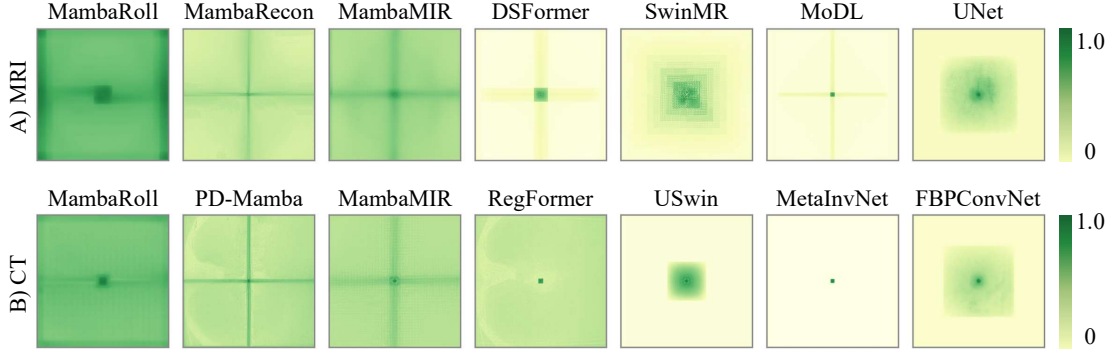


Figure 6.5: Gradient-based effective receptive field (ERF) visualizations for MRI (top row) and CT (bottom row) reconstruction across competing methods. ERF maps highlight the input image regions that most influence the reconstruction at the fixed target location at image center. Greener areas indicate higher contribution, whereas yellower areas indicate lower contribution (see colorbar).

6.5 Computational Complexity

Computational complexity is an important consideration for practical adoption of learning-based models in medical image reconstruction. Run times and memory load of competing methods in training and inference phases are listed in Table 6.7 for MRI reconstruction and in Table 6.8 for CT reconstruction. CNN-based methods generally demonstrate fast run times and low memory demand, although with some exceptions. In MRI reconstruction, UNet employs a relatively complex backbone that restricts efficiency, while in CT reconstruction, MetaInvNet shows higher compute load as a PD architecture due to the intrinsic complexity of Radon transformation within data-consistency modules. Transformer-based methods typically incur elevated compute burden, with notably higher run times and memory demand despite implementations devised to maintain efficiency (e.g., windowed attention mechanisms). While RegFormer has competitive run times to CNNs in CT reconstruction, it has high memory demand as a PD architecture. Meanwhile, SSM-based methods generally demonstrate efficiency levels that fall between CNNs and transformers. Among SSMs, MambaRoll demonstrates the highest computational efficiency, with run times and memory demand closest to CNN methods. These results highlight the inherent efficiency advantages of SSMs over transformers.

Table 6.7: Average run time (msec) and memory load (MB) per cross-section for MRI reconstruction during training and inference phases.

	Run time (msec)		Memory (MB)	
	Training	Inference	Training	Inference
UNet	42.37	11.56	1516	550
MoDL	50.60	14.99	1398	304
SwinMR	149.60	49.85	3864	870
DSFormer	750.21	248.89	16090	1338
MambaMIR	121.34	36.84	2382	474
MambaRecon	139.89	36.55	2304	462
MambaRoll	96.09	29.67	2568	358

Table 6.8: Average run time (msec) and memory load (MB) per cross-section for CT reconstruction during training and inference phases.

	Run time (msec)		Memory (MB)	
	Training	Inference	Training	Inference
FBPConvNet	84.93	13.23	1962	974
MetaInvNet	77.57	29.16	5144	3288
USwin	246.06	80.59	5926	1082
RegFormer	108.64	48.20	7890	5884
MambaMIR	143.05	41.29	2830	532
PD-Mamba	303.37	90.93	9264	4244
MambaRoll	134.23	49.79	6144	3374

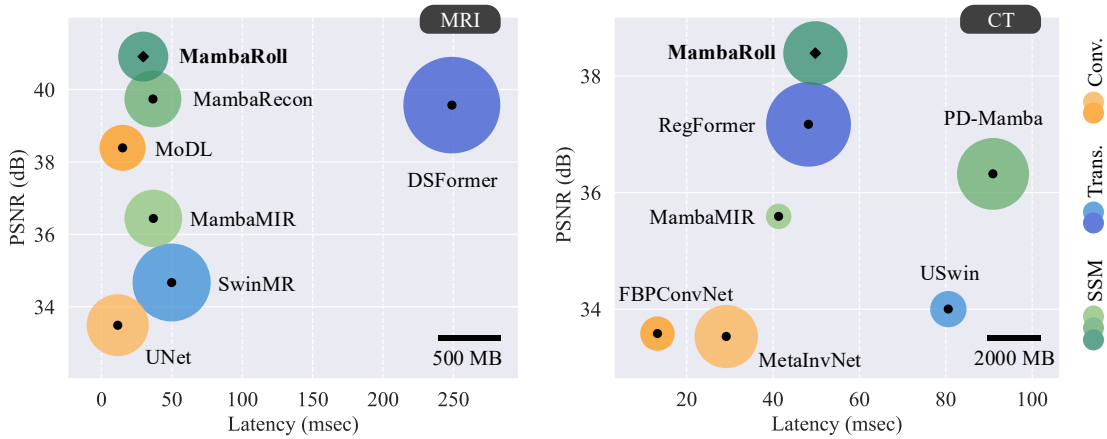


Figure 6.6: Performance versus efficiency in MRI (left panel) and CT (right panel) reconstructions. Each method is represented by a circular marker whose center denotes the performance-latency trade-off (where latency is taken as inference time), and whose diameter indicates the memory load (see scale bars).

Comparative visualizations of performance and efficiency across competing methods are presented in Fig. 6.6. While some CNN-based methods achieve higher computational efficiency (lower latency and/or memory load), this efficiency comes at the expense of reconstruction quality (lower PSNR and SSIM). Meanwhile, MambaRoll offers an attractive balance between reconstruction performance (as already demonstrated in Sections 6.1-6.2) and computational efficiency. For CT reconstructions specifically, physics-driven models (PD-Mamba, RegFormer) incur additional computational costs due to forward and inverse Radon transformations in their data-consistency modules, a challenge effectively addressed by MambaRoll’s efficient design. Taken together, these results suggest particular promise for MambaRoll for high fidelity and efficiency in medical image reconstruction.

6.6 Ablation Studies

6.6.1 Network Architecture

We first conducted ablation studies to evaluate the influence of major architectural elements of MambaRoll to reconstruction performance. To assess the importance of novel PD-SSM modules, we trained a ‘w/o PD-SSM’ variant that removed the PD-SSM modules albeit retained the refinement modules so as to create an unrolled architecture based on a convolutional network module. To assess the importance of autoregressive modeling across spatial scales, we trained a ‘w/o AR’ variant that used PD-SSM modules operating at a single spatial scale where feature maps retained the resolution of original input images. To assess the importance of SSM layers in capturing contextual representations, we trained a ‘w/o comp. SSM’ variant that ablated the compressed SSM blocks from the PD-SSM modules. To assess the importance of data-consistency blocks, we trained a ‘w/o DC’ variant that ablated DC blocks albeit retaining their residual connections to create a purely data-driven SSM model. Table 6.9 lists performance

Table 6.9: Performance of MambaRoll variants on representative reconstruction tasks. A ‘w/o PD-SSM’ variant ablating PD-SSM modules, a ‘w/o AR’ variant ablating autoregressive processing, a ‘w/o comp. SSM’ variant that ablated compressed SSM blocks, and a ‘w/o DC’ variant ablating data-consistency modules were considered.

	fastMRI - FLAIR, $R = 4$		CT, $R = 4$	
	PSNR	SSIM	PSNR	SSIM
MambaRoll	43.86$\pm$3.32	98.43$\pm$1.43	40.72$\pm$3.56	94.05$\pm$4.86
w/o PD-SSM	40.49 $\pm$ 2.70	97.08 $\pm$ 1.98	31.00 $\pm$ 1.27	83.50 $\pm$ 5.00
w/o AR	42.62 $\pm$ 3.11	98.01 $\pm$ 1.66	37.11 $\pm$ 2.56	90.53 $\pm$ 5.33
w/o comp. SSM	41.43 $\pm$ 3.11	97.44 $\pm$ 2.13	35.65 $\pm$ 2.41	88.51 $\pm$ 6.17
w/o DC	36.88 $\pm$ 2.50	93.44 $\pm$ 3.38	38.66 $\pm$ 2.99	92.14 $\pm$ 5.15

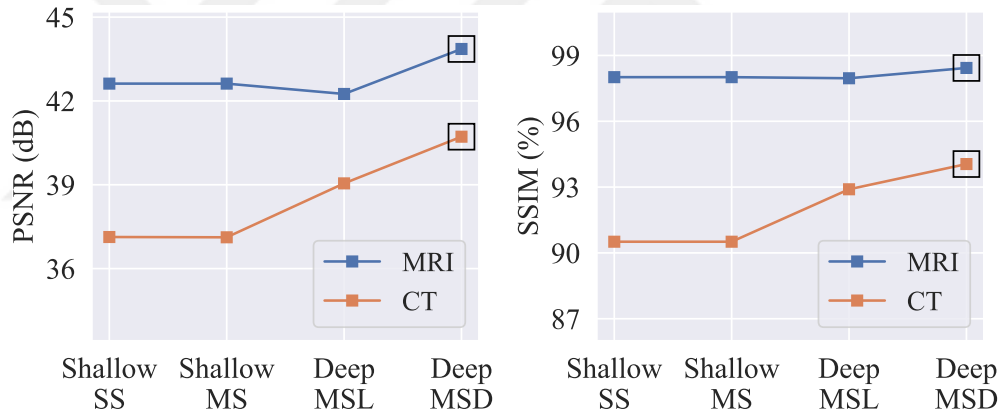


Figure 6.7: Performance of MambaRoll variants on representative reconstruction tasks (FLAIR, $R = 4$ for MRI and $R = 4$ for CT). Results shown for the proposed Deep MSD loss (deep multi-scale decoding loss marked with squares), along with Shallow SS, Shallow MS, and Deep MSL variants.

metrics for all MambaRoll variants. We find that MambaRoll yields superior performance against all variants, corroborating the importance of all architectural design elements in high-fidelity reconstruction.

6.6.2 Loss Function

Next, we conducted ablation studies to evaluate the importance of the DMSD loss on reconstruction performance. For this purpose, we trained several variants

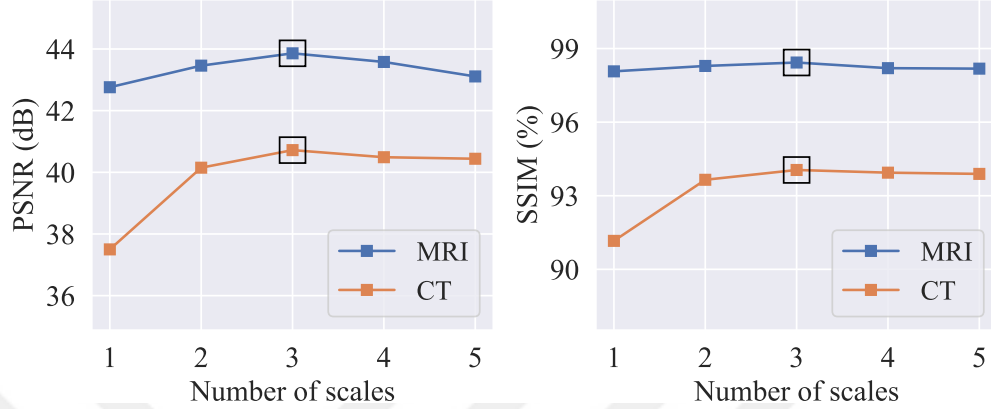


Figure 6.8: Performance of MambaRoll variants that prescribed varying numbers of spatial scales. Results displayed for the representative reconstruction tasks of FLAIR, $R = 4$ for MRI and $R = 4$ for CT. Squares mark the parameter values that were originally selected to implement MambaRoll.

of MambaRoll with alternative types of supervision. A ‘Shallow SS’ variant was built that only used a single-scale loss term on the network output. A ‘Shallow MS’ variant was built that used the single-scale loss term on the network output alongside multi-scale loss terms on the spatially downsampled versions of the output (with a total of S scales). A ‘Deep MSL’ variant was built that used a deep multi-scale loss on the latent feature maps extracted by PD-SSM modules (i.e., g_s), instead of the decoded images as in MambaRoll. Figure 6.7 depicts performance metrics for all variants. We find that the proposed DMSD loss yields superior performance against all variants, corroborating the importance of the proposed training objective in facilitating learning of the autoregressive prediction task.

6.6.3 Design Parameters

We also conducted ablation studies to assess the influence of the number of spatial scales for autoregressive modeling, and the compression factor employed in the PD-SSM modules. MambaRoll variants were built by varying the design parameters across a range. As shown in Fig. 6.8, performance increases with growing number of scales till 3, without apparent performance or efficiency benefits at

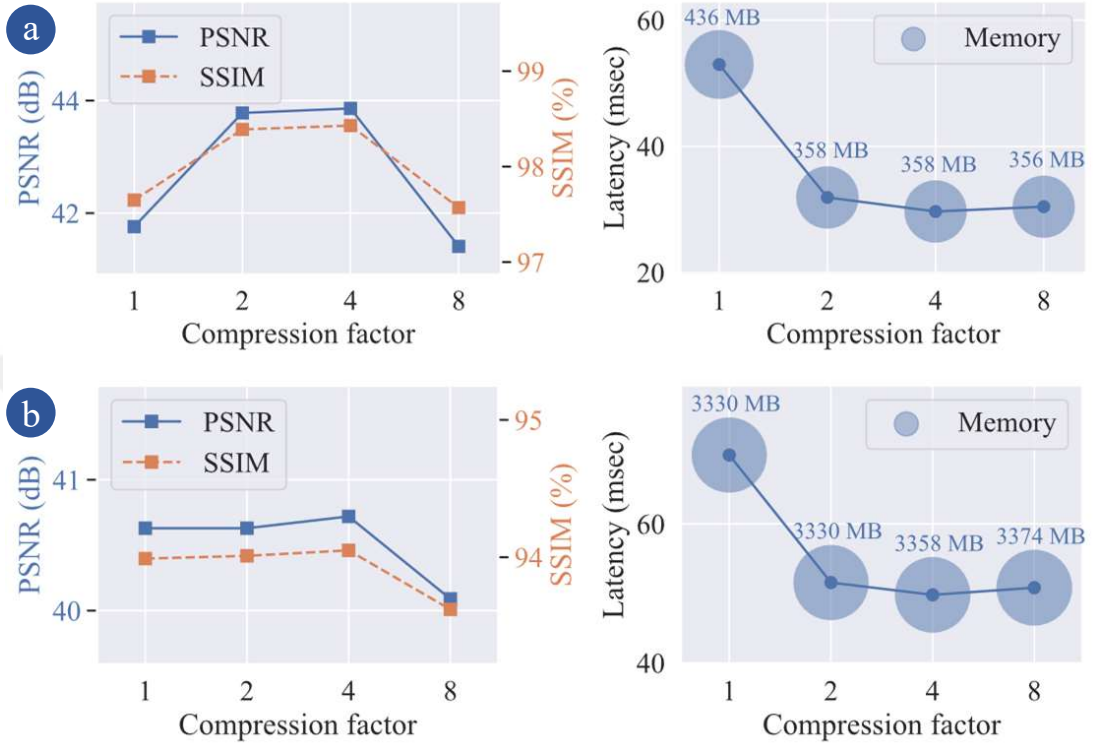


Figure 6.9: Performance of MambaRoll variants that prescribed varying compression factors **(a)** on FLAIR, $R = 4$ for MRI; **(b)** on $R = 4$ for CT. Performance metrics (PSNR, SSIM) are displayed in the right panel, whereas efficiency metrics (memory load and latency) are displayed in the left panel.

higher numbers of scales beyond this saturation point. These results indicate that the selected configuration for spatial scales offers an optimal trade-off between performance and efficiency.

Meanwhile, Fig. 6.9 comparatively visualizes the reconstruction performance and computational efficiency of MambaRoll variants across different compression factors in compressed SSM blocks. While larger compression factors theoretically enhance contextual information exchange across distant image regions by increasing spatial sampling step sizes, they also expand feature maps along the channel dimension, potentially elevating computational overhead. Our analysis reveals that a compression factor of 4 provides optimal performance improvements while

maintaining high computational efficiency. Factors beyond 4 degrade reconstruction performance without offering additional efficiency benefits, likely due to excessive information dispersion that disrupts the local spatial coherence necessary for effective medical image reconstruction.



Chapter 7

Discussion

The performance and utility of the proposed method could benefit from several technical improvements. A first group of improvements concern the learning setup for medical image reconstruction. Here, demonstrations were performed in a supervised setup using fully-sampled acquisitions as ground truth. To enable training on datasets with only undersampled data, self-supervised learning approaches could be adopted [137, 103, 138]. For mixed datasets pooling fully-sampled and undersampled across separate subject pools, a hybrid strategy combining self-supervised training on undersampled data followed by supervised fine-tuning might be used [8]. Alternatively, cycle-consistent learning could enhance training on such unpaired datasets [105]. When curating any training data is difficult altogether, test-time training could be used for scan-specific reconstruction, using data-consistency losses, though at increased inference cost [139, 140, 19].

A second group of improvements concern the loss functions used to express the learning objective. Here we observed high-quality MRI and CT reconstructions with models trained via autoregressive image prediction guided via a deep multi-scale supervision loss. Corroborating recent work, our results suggest that autoregressive modeling might be a promising alternative to generative methods based on adversarial and diffusive losses [128]. Yet, combining autoregressive loss

terms with other generative losses might help further improve realism and sensitivity to fine details [107]. While our primary focus in the current study was on architectural design of reconstruction models, MambaRoll can be combined with generative learning objectives to examine their potential benefits in future work [141, 4].

A third group of improvements concern the reconstruction tasks examined. Here we built reconstruction models that operate on single-modality acquisitions (a single MRI contrast or a single CT contrast). Recent work suggests SSMs can also aid in joint reconstruction of multiple modalities [44]. A basic approach to extend MambaRoll for multi-modal reconstruction would be to assign separate channels per modality in the input-output layers and across data-consistency layers in PD-SSM modules [16, 142]. Separate model branches could also be employed to process each modality [5], while cross-modality interactions are mediated by attentional or state-space modules for fine-grained control of feature representations shared between modalities [44]. Here we focused on cross-sectional reconstructions from one- or two-dimensionally undersampled acquisitions, as commonly exercised in the literature for a strong balance between accuracy and efficiency [43]. Future work may explore volumetric reconstruction to systematically assess potential benefits of SSMs in explicit volumetric modeling.

A final group of improvements concerns the architectural design of MambaRoll. Each PD-SSM module integrates a linear SSM layer between scale-specific convolutional layers that downsample or upsample feature maps, enabling autoregressive state-space modeling across spatial scales. Our experiments suggest that this design outperforms existing SSM architectures based on conventional Mamba modules (e.g., MambaMIR, MambaRecon), which use a popular gating mechanism to modulate SSM outputs [46]. Note that our findings contribute to a growing body of evidence that hybrid designs—such as combining SSM and transformer modules—may offer performance gains beyond using them in isolation. Another avenue involves dual-domain approaches that combine spatial and spectral processing to enhance task performance [143, 120]. Further work is needed to systematically explore the benefits of such hybrid architectures in medical image reconstruction.

Chapter 8

Conclusion

In this thesis, we introduced a novel deep learning method for medical image reconstruction, MambaRoll, based on physics-driven autoregressive state-space modeling. MambaRoll leverages autoregressive prediction across multiple spatial scales along with physics-driven SSM modules to effectively aggregate contextual features in medical images. To enhance learning efficacy, we also incorporated a deep multi-scale decoding loss tailored to the autoregressive prediction task. Demonstrations on accelerated MRI and sparse-view CT reconstructions indicate that the proposed model consistently outperforms state-of-the-art methods in terms of image quality, without compromising computational efficiency. Therefore, MambaRoll holds great potential for improving the utility of learning-based medical image reconstruction.

Bibliography

- [1] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, “Image reconstruction is a new frontier of machine learning,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, p. 1289–1296, 2018.
- [2] Y. Du, R. Dharmakumar, and S. A. Tsaftaris, “The MRI Scanner as a Diagnostic: Image-Less Active Sampling,” in *International Conference on Medical Image Computing and Computer-assisted Intervention*, pp. 467–476, 2024.
- [3] S. K. Zhou, H. Greenspan, C. Davatzikos, J. S. Duncan, B. Van Ginneken, A. Madabhushi, J. L. Prince, D. Rueckert, and R. M. Summers, “A Review of Deep Learning in Medical Imaging: Imaging Traits, Technology Trends, Case Studies With Progress Highlights, and Future Promises,” *Proc. IEEE*, vol. 109, no. 5, pp. 820–838, 2021.
- [4] M. U. Mirza, O. Dalmaz, H. A. Bedel, G. Elmas, Y. Korkmaz, A. Gungor, S. U. Dar, and T. Çukur, “Learning fourier-constrained diffusion bridges for mri reconstruction,” *arXiv preprint arXiv:2308.01096*, 2023.
- [5] M. Yurt, M. Ozbey, S. U. Dar, B. Tinaz, K. K. Oguz, and T. Cukur, “Progressively volumetrized deep generative models for data-efficient contextual learning of MR image recovery,” *Medical Image Analysis*, vol. 78, p. 102429, 2022.
- [6] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, and D. Liang, “Accelerating magnetic resonance imaging via deep learning,” in *IEEE 13th International Symposium on Biomedical Imaging*, pp. 514–517, 2016.

- [7] D. Lee, J. Yoo, S. Tak, and J. C. Ye, “Deep residual learning for accelerated mri using magnitude and phase networks,” *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1985–1995, 2018.
- [8] S. U. H. Dar, M. Özbey, A. B. Çath, and T. Çukur, “A transfer-learning approach for accelerated mri using deep neural networks,” *Magnetic Resonance in Medicine*, vol. 84, no. 2, pp. 663–685, 2020.
- [9] Y. Han and J. C. Ye, “Framing u-net via deep convolutional framelets: Application to sparse-view ct,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1418–1429, 2018.
- [10] W. Wu, D. Hu, C. Niu, H. Yu, V. Vardhanabhuti, and G. Wang, “DRONE: Dual-Domain Residual-based Optimization NETwork for Sparse-View CT Reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 11, pp. 3002–3014, 2021.
- [11] Y. Korkmaz, S. U. Dar, M. Yurt, M. Özbey, and T. Cukur, “Unsupervised mri reconstruction via zero-shot learned adversarial transformers,” *IEEE Transactions on Medical Imaging*, vol. 41, no. 7, pp. 1747–1763, 2022.
- [12] M. Lee, H. Kim, and H.-J. Kim, “Sparse-view ct reconstruction based on multi-level wavelet convolution neural network,” *Physica Medica*, vol. 80, pp. 352–362, 2020.
- [13] Z. Zhang, X. Liang, X. Dong, Y. Xie, and G. Cao, “A sparse-view ct reconstruction method based on combination of densenet and deconvolution,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1407–1417, 2018.
- [14] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, and D. Firmin, “DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1310–1321, 2018.

- [15] M. Mardani, E. Gong, J. Y. Cheng, S. Vasanawala, G. Zaharchuk, L. Xing, and J. M. Pauly, “Deep generative adversarial neural networks for compressive sensing MRI,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 1, pp. 167–179, 2019.
- [16] S. U. Dar, M. Yurt, M. Shahdloo, M. E. Ildız, B. Tınaz, and T. Çukur, “Prior-guided image reconstruction for accelerated multi-contrast mri via generative adversarial networks,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1072–1087, 2020.
- [17] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. Tamir, “Robust compressed sensing mri with deep generative priors,” in *Advances in Neural Information Processing Systems*, vol. 34, pp. 14938–14954, 2021.
- [18] W. Xia, W. Cong, and G. Wang, “Patch-based denoising diffusion probabilistic model for sparse-view ct reconstruction. arxiv,” *arXiv preprint arXiv:2211.10388*, 2022.
- [19] A. Güngör, S. U. Dar, Ş. Öztürk, Y. Korkmaz, H. A. Bedel, G. Elmas, M. Ozbey, and T. Çukur, “Adaptive diffusion priors for accelerated mri reconstruction,” *Medical Image Analysis*, vol. 88, p. 102872, 2023.
- [20] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, “Deep convolutional neural network for inverse problems in imaging,” *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.
- [21] U. S. Kamilov, C. A. Bouman, G. T. Buzzard, and B. Wohlberg, “Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications,” *IEEE Signal Processing Magazine*, vol. 40, no. 1, pp. 85–97, 2023.
- [22] G. Elmas, S. U. Dar, Y. Korkmaz, E. Ceyani, B. Susam, M. Ozbey, S. Avestimehr, and T. Çukur, “Federated learning of generative image priors for mri reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 42, no. 7, pp. 1996–2009, 2023.

- [23] K. Hammernik, T. Küstner, B. Yaman, Z. Huang, D. Rueckert, F. Knoll, and M. Akçakaya, “Physics-driven deep learning for computational magnetic resonance imaging: Combining physics and machine learning for improved medical imaging,” *IEEE Signal Processing Magazine*, vol. 40, no. 1, pp. 98–114, 2023.
- [24] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, “A deep cascade of convolutional neural networks for dynamic mr image reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 491–503, 2017.
- [25] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, “Learning a variational network for reconstruction of accelerated MRI data,” *Magnetic Resonance in Medicine*, vol. 79, no. 6, pp. 3055–3071, 2017.
- [26] Y. Zhang, D. Hu, S. Hao, J. Liu, G. Quan, Y. Zhang, X. Ji, and Y. Chen, “Dream-net: Deep residual error iterative minimization network for sparse-view ct reconstruction,” *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 1, pp. 480–491, 2023.
- [27] H. Zhang, B. Liu, H. Yu, and B. Dong, “MetaInv-Net: Meta Inversion Network for Sparse View CT Image Reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 2, pp. 621–634, 2021.
- [28] B. Zhou and S. K. Zhou, “Dudornet: learning a dual-domain recurrent network for fast mri reconstruction with deep t1 prior,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4273–4282, 2020.
- [29] S. A. H. Hosseini, B. Yaman, S. Moeller, M. Hong, and M. Akçakaya, “Dense recurrent neural networks for accelerated mri: History-cognizant unrolling of optimization algorithms,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1280–1291, 2020.

- [30] O. Dalmaz, M. Yurt, and T. Çukur, “Resvit: Residual vision transformers for multimodal medical image synthesis,” *IEEE Transactions on Medical Imaging*, vol. 41, no. 10, pp. 2598–2614, 2022.
- [31] H. K. Aggarwal, M. P. Mani, and M. Jacob, “MoDL: Model-Based deep learning architecture for inverse problems,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 394–405, 2019.
- [32] J. Xiang, Y. Dong, and Y. Yang, “FISTA-Net: Learning a Fast Iterative Shrinkage Thresholding Network for Inverse Problems in Imaging,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1329–1339, 2021.
- [33] C. Wang, K. Shang, H. Zhang, Q. Li, and S. K. Zhou, “Dudotrans: Dual-domain transformer for sparse-view ct reconstruction,” in *Machine Learning for Medical Image Reconstruction*, pp. 84–94, 2022.
- [34] W. Xia, Z. Yang, Z. Lu, Z. Wang, and Y. Zhang, “Regformer: A local–nonlocal regularization-based model for sparse-view ct reconstruction,” *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 8, no. 2, pp. 184–194, 2024.
- [35] J. Huang, Y. Fang, Y. Wu, H. Wu, Z. Gao, Y. Li, J. Del Ser, J. Xia, and G. Yang, “Swin transformer for fast MRI,” *Neurocomputing*, vol. 493, pp. 281–304, 2022.
- [36] Y. Korkmaz, M. Yurt, S. U. H. Dar, M. Özbey, and T. Cukur, “Deep mri reconstruction with generative vision transformers,” in *International Workshop on Machine Learning for Medical Image Reconstruction*, pp. 54–64, Springer, 2021.
- [37] P. Guo, Y. Mei, J. Zhou, S. Jiang, and V. M. Patel, “Reconformer: Accelerated mri reconstruction using recurrent transformer,” *IEEE Transactions on Medical Imaging*, 2024.
- [38] K. Lin and R. Heckel, “Vision transformers enable fast and robust accelerated mri,” in *International Conference on medical imaging with deep learning*, pp. 774–795, PMLR, 2022.

- [39] B. Zhou, N. Dey, J. Schlemper, S. S. M. Salehi, C. Liu, J. S. Duncan, and M. Sofka, “Dsformer: A dual-domain self-supervised transformer for accelerated multi-contrast mri reconstruction,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 4966–4975, 2023.
- [40] Y. Korkmaz, T. Cukur, and V. M. Patel, “Self-supervised mri reconstruction with unrolled diffusion models,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 491–501, Springer, 2023.
- [41] C.-M. Feng, Y. Yan, G. Chen, Y. Xu, Y. Hu, L. Shao, and H. Fu, “Multimodal transformer for accelerated mr imaging,” *IEEE Transactions on Medical Imaging*, vol. 42, no. 10, pp. 2804–2816, 2023.
- [42] L. Zhu, B. Liao, Q. Zhang, X. Wang, W. Liu, and X. Wang, “Vision mamba: efficient visual representation learning with bidirectional state space model,” in *International Conference on Machine Learning*, 2024.
- [43] J. Huang, L. Yang, F. Wang, Y. Wu, Y. Nan, W. Wu, C. Wang, K. Shi, A. I. Aviles-Rivero, C.-B. Schönlieb, *et al.*, “Enhancing global sensitivity and uncertainty quantification in medical image reconstruction with monte carlo arbitrary-masked mamba,” *Medical Image Analysis*, vol. 99, p. 103334, 2025.
- [44] J. Zou, L. Liu, Q. Chen, S. Wang, Z. Hu, X. Xing, and J. Qin, “Mmr-mamba: Multi-modal mri reconstruction with mamba and spatial-frequency information fusion,” *Medical Image Analysis*, vol. 102, p. 103549, 2025.
- [45] Y. Korkmaz and V. M. Patel, “Mambarecon: Mri reconstruction with structured state space models,” in *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 4142–4152, 2025.
- [46] Y. Liu, Y. Tian, Y. Zhao, H. Yu, L. Xie, Y. Wang, Q. Ye, J. Jiao, and Y. Liu, “Vmamba: Visual state space model,” *Advances in Neural Information Processing Systems*, vol. 37, pp. 103031–103063, 2024.

- [47] M. Yurt, S. U. Dar, A. Erdem, E. Erdem, K. K. Oguz, and T. Çukur, “mustGAN: multi-stream Generative Adversarial Networks for MR Image Synthesis,” *Medical Image Analysis*, vol. 70, p. 101944, 2021.
- [48] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [49] B. Kabas, F. Arslan, V. A. Nezhad, S. Ozturk, E. U. Saritas, and T. Çukur, “Physics-driven autoregressive state space models for medical image reconstruction,” *arXiv:2412.09331*, 2024.
- [50] A. Güngör, B. Askin, D. A. Soydan, E. U. Saritas, C. B. Top, and T. Çukur, “TranSMS: Transformers for Super-Resolution Calibration in Magnetic Particle Imaging,” *IEEE Transactions on Medical Imaging*, vol. 41, no. 12, pp. 3562–3574, 2022.
- [51] B. Askin, A. Güngör, D. Alptekin Soydan, E. U. Saritas, C. B. Top, and T. Cukur, “Pp-mpi: A deep plug-and-play prior for magnetic particle imaging reconstruction,” in *International Workshop on Machine Learning for Medical Image Reconstruction*, pp. 105–114, Springer, 2022.
- [52] A. Gungor, B. Aşkın, D. A. Soydan, E. U. Saritas, C. B. Top, and T. Çukur, “A denoiser scaling technique for plug-and-play mpi reconstruction,” *International Journal on Magnetic Particle Imaging IJMPI*, vol. 9, no. 1 Suppl 1, 2023.
- [53] S. Ilbey, C. B. Top, A. Güngör, T. Çukur, E. U. Saritas, and H. E. Güven, “Fast system calibration with coded calibration scenes for magnetic particle imaging,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 9, pp. 2070–2080, 2019.
- [54] O. B. Demirel, T. Kilic, T. Çukur, and E. U. Saritas, “Anatomical measurements correlate with individual magnetostimulation thresholds for khz-range homogeneous magnetic fields,” *Medical Physics*, vol. 47, no. 4, pp. 1836–1844, 2020.

- [55] A. Ozaslan, A. Alacaoglu, O. Demirel, T. Çukur, and E. Saritas, “Fully automated gridding reconstruction for non-cartesian x-space magnetic particle imaging,” *Physics in Medicine & Biology*, vol. 64, no. 16, p. 165018, 2019.
- [56] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, *et al.*, “The multi-modal brain tumor image segmentation benchmark (brats),” *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.
- [57] I. Despotović, B. Goossens, and W. Philips, “Mri segmentation of the human brain: challenges, methods, and applications,” *Computational and Mathematical Methods in Medicine*, vol. 2015, no. 1, p. 450341, 2015.
- [58] T. Çukur, S. Nishimoto, A. G. Huth, and J. L. Gallant, “Attention during natural vision warps semantic representation across the human brain,” *Nature Neuroscience*, vol. 16, no. 6, pp. 763–770, 2013.
- [59] T. Çukur, A. G. Huth, S. Nishimoto, and J. L. Gallant, “Functional subdomains within human ffa,” *Journal of Neuroscience*, vol. 33, no. 42, pp. 16748–16766, 2013.
- [60] E. Çelik, S. U. H. Dar, Ö. Yılmaz, Ü. Keleş, and T. Çukur, “Spatially informed voxelwise modeling for naturalistic fmri experiments,” *NeuroImage*, vol. 186, pp. 741–757, 2019.
- [61] M. Shahdloo, E. Çelik, and T. Çukur, “Biased competition in semantic representation during natural visual search,” *NeuroImage*, vol. 216, p. 116383, 2020.
- [62] Ö. Yılmaz, E. Çelik, and T. Çukur, “Informed feature regularization in voxelwise modeling for naturalistic fmri experiments,” *European Journal of Neuroscience*, vol. 52, no. 5, pp. 3394–3410, 2020.
- [63] T. Çukur, M. Lustig, and D. G. Nishimura, “Improving non-contrast-enhanced steady-state free precession angiography with compressed sensing,” *Magnetic Resonance in Medicine*, vol. 61, no. 5, pp. 1122–1131, 2009.

- [64] T. Çukur, J. H. Lee, N. K. Bangerter, B. A. Hargreaves, and D. G. Nishimura, “Non-contrast-enhanced flow-independent peripheral mr angiography with balanced ssfp,” *Magnetic Resonance in Medicine*, vol. 61, no. 6, pp. 1533–1539, 2009.
- [65] T. Cukur, A. Shimakawa, H. Yu, B. A. Hargreaves, B. S. Hu, D. G. Nishimura, and J. H. Brittain, “Magnetization-prepared ideal bssfp: a flow-independent technique for noncontrast-enhanced peripheral angiography,” *Journal of Magnetic Resonance Imaging*, vol. 33, no. 4, pp. 931–939, 2011.
- [66] N. K. Bangerter, T. Cukur, B. A. Hargreaves, B. S. Hu, J. H. Brittain, D. Park, G. E. Gold, and D. G. Nishimura, “Three-dimensional fluid-suppressed t2-prep flow-independent peripheral angiography using balanced ssfp,” *Magnetic Resonance Imaging*, vol. 29, no. 8, pp. 1119–1124, 2011.
- [67] E. Ilicak, S. Cetin, E. Bulut, K. K. Oguz, E. U. Saritas, G. Unal, and T. Çukur, “Targeted vessel reconstruction in non-contrast-enhanced steady-state free precession angiography,” *NMR in Biomedicine*, vol. 29, no. 5, pp. 532–544, 2016.
- [68] T. Çukur, N. K. Bangerter, and D. G. Nishimura, “Enhanced spectral shaping in steady-state free precession imaging,” *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1216–1223, 2007.
- [69] T. Çukur and D. G. Nishimura, “Fat-water separation with alternating repetition time balanced ssfp,” *Magnetic Resonance in Medicine*, vol. 60, no. 2, pp. 479–484, 2008.
- [70] T. Çukur, M. Lustig, and D. G. Nishimura, “Multiple-profile homogeneous image combination: application to phase-cycled ssfp and multicoil imaging,” *Magnetic Resonance in Medicine*, vol. 60, no. 3, pp. 732–738, 2008.
- [71] T. Çukur and D. G. Nishimura, “Multiple repetition time balanced steady-state free precession imaging,” *Magnetic Resonance in Medicine*, vol. 62, no. 1, pp. 193–204, 2009.

- [72] T. Çukur, M. Yamada, W. R. Overall, P. Yang, and D. G. Nishimura, “Positive contrast with alternating repetition time ssfp (parts): A fast imaging technique for spio-labeled cells,” *Magnetic Resonance in Medicine*, vol. 63, no. 2, pp. 427–437, 2010.
- [73] T. Çukur, “Spectrally selective imaging with wideband balanced steady-state free precession mri,” *Magnetic Resonance in Medicine*, vol. 75, no. 3, pp. 1132–1141, 2016.
- [74] E. Ilicak, L. K. Senel, E. Biyik, and T. Çukur, “Profile-encoding reconstruction for multiple-acquisition balanced steady-state free precession imaging,” *Magnetic Resonance in Medicine*, vol. 78, no. 4, pp. 1316–1329, 2017.
- [75] E. Biyik, E. Ilicak, and T. Cukur, “Reconstruction by calibration over tensors for multi-coil multi-acquisition balanced ssfp imaging,” *Magnetic Resonance in Medicine*, vol. 79, no. 5, pp. 2542–2554, 2018.
- [76] K. Keskin, U. Yilmaz, and T. Çukur, “Constrained ellipse fitting for efficient parameter mapping with phase-cycled bssfp mri,” *IEEE Transactions on Medical Imaging*, vol. 41, no. 1, pp. 14–26, 2021.
- [77] K. T. Block, M. Uecker, and J. Frahm, “Undersampled radial mri with multiple coils. iterative image reconstruction using a total variation constraint,” *Magnetic Resonance in Medicine*, vol. 57, no. 6, pp. 1086–1098, 2007.
- [78] E. M. Haacke, E. Linskogj, and W. Lin, “A fast, iterative, partial-fourier technique capable of local phase recovery,” *Journal of Magnetic Resonance (1969)*, vol. 92, no. 1, pp. 126–145, 1991.
- [79] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, “Sense: sensitivity encoding for fast mri,” *Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 952–962, 1999.
- [80] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, “Generalized autocalibrating partially parallel acquisitions (grappa),” *Magnetic Resonance in Medicine*, vol. 47, no. 6, pp. 1202–1210, 2002.

- [81] M. Lustig, D. Donoho, and J. M. Pauly, “Sparse mri: The application of compressed sensing for rapid mr imaging,” *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [82] T. Cukur, J. M. Santos, J. M. Pauly, and D. G. Nishimura, “Variable-density parallel imaging with partially localized coil sensitivities,” *IEEE Transactions on Medical Imaging*, vol. 29, no. 5, pp. 1173–1181, 2010.
- [83] T. Çukur, “Accelerated phase-cycled ssfp imaging with compressed sensing,” *IEEE Transactions on Medical Imaging*, vol. 34, no. 1, pp. 107–115, 2014.
- [84] M. Shahdloo, E. Ilicak, M. Tofghi, E. U. Saritas, A. E. Çetin, and T. Çukur, “Projection onto epigraph sets for rapid self-tuning compressed sensing mri,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 7, pp. 1677–1689, 2018.
- [85] T. Cukur, M. Lustig, E. U. Saritas, and D. G. Nishimura, “Signal compensation and compressed sensing for magnetization-prepared mr angiography,” *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1017–1027, 2011.
- [86] L. K. Senel, T. Kilic, A. Gungor, E. Kopanoglu, H. E. Guven, E. U. Saritas, A. Koc, and T. Cukur, “Statistically segregated k-space sampling for accelerating multiple-acquisition mri,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 7, pp. 1701–1714, 2019.
- [87] R. Gordon, R. Bender, and G. T. Herman, “Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and x-ray photography,” *Journal of Theoretical Biology*, vol. 29, no. 3, pp. 471–481, 1970.
- [88] P. Gilbert, “Iterative methods for the three-dimensional reconstruction of an object from projections,” *Journal of Theoretical Biology*, vol. 36, no. 1, pp. 105–117, 1972.
- [89] E. Y. Sidky and X. Pan, “Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization,” *Physics in Medicine & Biology*, vol. 53, no. 17, p. 4777, 2008.

- [90] Ş. Öztürk and T. Cukur, “Deep clustering via center-oriented margin free-triplet loss for skin lesion detection in highly imbalanced datasets,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 9, pp. 4679–4690, 2022.
- [91] V. A. Nezhad, G. Elmas, F. Arslan, B. Kabaş, and T. Çukur, “Generalizable deep mri reconstruction with cross-site data synthesis,” in *2024 32nd Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, IEEE, 2024.
- [92] V. A. Nezhad, G. Elmas, B. Kabas, F. Arslan, and T. Çukur, “Generative autoregressive transformers for model-agnostic federated mri reconstruction,” *arXiv preprint arXiv:2502.04521*, 2025.
- [93] D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, “Medical image synthesis with deep convolutional adversarial networks,” *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 12, pp. 2720–2730, 2018.
- [94] I. Sivgin, H. A. Bedel, S. Ozturk, and T. Çukur, “A plug-in graph neural network to boost temporal sensitivity in fmri analysis,” *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 9, pp. 5323–5334, 2024.
- [95] H. A. Bedel and T. Çukur, “Dreamr: Diffusion-driven counterfactual explanation for functional mri,” *IEEE Transactions on Medical Imaging*, 2024.
- [96] A. Zaid Alkilani, T. Çukur, and E. U. Saritas, “Fd-net: An unsupervised deep forward-distortion model for susceptibility artifact correction in epi,” *Magnetic Resonance in Medicine*, vol. 91, no. 1, pp. 280–296, 2024.
- [97] C. M. Hyun, H. P. Kim, S. M. Lee, S. Lee, and J. K. Seo, “Deep learning for undersampled mri reconstruction,” *Physics in Medicine & Biology*, vol. 63, no. 13, p. 135007, 2018.
- [98] O. Dalmaz, A. D. Desai, R. Heckel, T. Cukur, A. S. Chaudhari, and B. A. Hargreaves, “Efficient noise calculation in deep learning-based mri reconstructions,” *arXiv preprint arXiv:2505.02007*, 2025.

- [99] S. Ozturk, M. B. Yilmaz, M. Kara, M. T. Yavuz, A. Koç, and T. Çukur, “Meta-entity driven triplet mining for aligning medical vision-language models,” *arXiv preprint arXiv:2504.15929*, 2025.
- [100] S. U. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, and T. Cukur, “Image synthesis in multi-contrast mri with conditional generative adversarial networks,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2375–2388, 2019.
- [101] M. Yurt, S. U. Dar, A. Erdem, E. Erdem, K. K. Oguz, and T. Çukur, “must-gan: multi-stream generative adversarial networks for mr image synthesis,” *Medical Image Analysis*, vol. 70, p. 101944, 2021.
- [102] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, “Medical image synthesis with context-aware generative adversarial networks,” in *International Conference on Medical Image Computing and Computer-assisted Intervention*, pp. 417–425, Springer, 2017.
- [103] M. Yurt, O. Dalmaz, S. Dar, M. Ozbey, B. Tımaz, K. Oguz, and T. Çukur, “Semi-Supervised Learning of MRI Synthesis Without Fully-Sampled Ground Truths,” *IEEE Transactions on Medical Imaging*, vol. 41, no. 12, pp. 3895–3906, 2022.
- [104] O. Dalmaz, M. U. Mirza, G. Elmas, M. Ozbey, S. U. Dar, E. Ceyani, K. K. Oguz, S. Avestimehr, and T. Çukur, “One model to unite them all: Personalized federated learning of multi-contrast mri synthesis,” *Medical Image Analysis*, vol. 94, p. 103121, 2024.
- [105] M. Ozbey, O. Dalmaz, S. U. H. Dar, H. A. Bedel, S. Ozturk, A. Gungor, and T. Çukur, “Unsupervised medical image translation with adversarial diffusion models,” *IEEE Transactions on Medical Imaging*, vol. 42, no. 12, pp. 3524–3539, 2023.
- [106] S. Pan, T. Wang, R. L. Qiu, M. Axente, C.-W. Chang, J. Peng, A. B. Patel, J. Shelton, S. A. Patel, J. Roper, *et al.*, “2d medical image synthesis using transformer-based denoising diffusion probabilistic model,” *Physics in Medicine & Biology*, vol. 68, no. 10, p. 105004, 2023.

- [107] F. Arslan, B. Kabas, O. Dalmaz, M. Ozbey, and T. Çukur, “Self-consistent recursive diffusion bridge for medical image translation,” *Medical Image Analysis*, vol. 106, p. 103747, 2025.
- [108] B. Kabas, F. Arslan, V. A. Nezhad, and T. Çukur, “Multi-contrast mr image synthesis with a brownian diffusion model,” in *2024 32nd Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, IEEE, 2024.
- [109] Ş. Öztürk, M. Y. Turalı, and T. Çukur, “Hydravit: Adaptive multi-branch transformer for multi-label disease classification from chest x-ray images,” *Biomedical Signal Processing and Control*, vol. 100, p. 106959, 2025.
- [110] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, “Unetr: Transformers for 3d medical image segmentation,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 574–584, 2022.
- [111] I. Aytekin, O. Dalmaz, K. Gonc, H. Ankishan, E. U. Saritas, U. Bagci, H. Celik, and T. Çukur, “Covid-19 detection from respiratory sounds with hierarchical spectrogram transformers,” *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 3, pp. 1273–1284, 2023.
- [112] H. A. Bedel, I. Sivgin, O. Dalmaz, S. U. Dar, and T. Çukur, “Bolt: Fused window transformers for fmri time series analysis,” *Medical Image Analysis*, vol. 88, p. 102841, 2023.
- [113] J. Liu, H. Yang, H.-Y. Zhou, L. Yu, Y. Liang, Y. Yu, S. Zhang, H. Zheng, and S. Wang, “Swin-umamba†: Adapting mamba-based vision foundation models for medical image segmentation,” *IEEE Transactions on Medical Imaging*, 2024.
- [114] Z. Xing, T. Ye, Y. Yang, G. Liu, and L. Zhu, “Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation,” in *International Conference on Medical Image Computing and Computer-assisted Intervention*, pp. 578–588, Springer, 2024.

- [115] Ş. Öztürk, O. C. Duran, and T. Cukur, “Denomamba: A fused state-space model for low-dose ct denoising,” *arXiv preprint arXiv:2409.13094*, 2024.
- [116] O. F. Atli, B. Kabas, F. Arslan, A. C. Demirtas, M. Yurt, O. Dalmaz, and T. Cukur, “I2I-Mamba: multi-modal medical image synthesis via selective state space modeling,” *arXiv preprint arXiv:2405.14022*, 2024.
- [117] G. Wang, T. Luo, J.-F. Nielsen, D. C. Noll, and J. A. Fessler, “B-spline parameterized joint optimization of reconstruction and k-space trajectories (bjork) for accelerated 2d mri,” *IEEE Transactions on Medical Imaging*, vol. 41, no. 9, pp. 2318–2330, 2022.
- [118] Z. Ramzi, C. G R, J.-L. Starck, and P. Ciuciu, “NC-PDNet: A Density-Compensated Unrolled Network for 2D and 3D Non-Cartesian MRI Reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 41, no. 7, pp. 1625–1638, 2022.
- [119] Y. Xue, Y. Du, G. Carloni, E. Pachetti, C. Jordan, and S. A. Tsaftaris, “Cine cardiac mri reconstruction using a convolutional recurrent network with refinement,” in *STACOM*, pp. 421–432, 2024.
- [120] Y. Meng, Z. Yang, Z. Song, and Y. Shi, “DH-Mamba: Exploring Dual-domain Hierarchical State Space Models for MRI Reconstruction,” *arXiv:2501.08163*, 2025.
- [121] T. Hou, H. Zhu, B. Huang, K. Chen, and Z. Zheng, “Causal inertia proximal mamba network for magnetic resonance image reconstruction,” *Medical Image Analysis*, vol. 105, p. 103649, 2025.
- [122] G. Luo, N. Zhao, W. Jiang, E. S. Hui, and P. Cao, “Mri reconstruction using deep bayesian estimation,” *Magnetic Resonance in Medicine*, vol. 84, no. 4, pp. 2246–2261, 2020.
- [123] T. H. Kim, P. Garg, and J. P. Haldar, “LORAKI: Autocalibrated recurrent neural networks for autoregressive MRI reconstruction in k-space,” *arXiv preprint arXiv:1904.09390*, 2019.

- [124] G. Luo, S. Huang, and M. Uecker, “Autoregressive image diffusion: Generation of image sequence and application in MRI,” in *Advances in Neural Information Processing Systems*, 2024.
- [125] J. C. Ye, Y. C. Eldar, and M. Unser, *Deep Learning for Biomedical Image Reconstruction*. Cambridge: Cambridge University Press, 2023.
- [126] G. Wang, J. C. Ye, and B. D. Man, “Deep learning for tomographic image reconstruction,” *Nature Machine Intelligence*, vol. 2, no. 12, p. 737–748, 2020.
- [127] T. Çukur, S. U. Dar, V. A. Nezhad, Y. Jun, T. H. Kim, S. Fujita, and B. Bilgic, “A tutorial on mri reconstruction: From modern methods to clinical implications,” *arXiv preprint arXiv:2507.16715*, 2025.
- [128] K. Tian, Y. Jiang, Z. Yuan, B. Peng, and L. Wang, “Visual autoregressive modeling: Scalable image generation via next-scale prediction,” *Advances in Neural Information Processing Systems*, vol. 37, pp. 84839–84865, 2024.
- [129] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1874–1883, 2016.
- [130] F. Knoll, J. Zbontar, A. Sriram, M. J. Muckley, M. Bruno, A. Defazio, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana, Z. Zhang, M. Drozdal, A. Romero, M. Rabbat, P. Vincent, J. Pinkerton, D. Wang, N. Yakubova, E. Owens, C. L. Zitnick, M. P. Recht, D. K. Sodickson, and Y. W. Lui, “fastMRI: A publicly available raw k-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning,” *Radiology: Artificial Intelligence*, vol. 2, no. 1, p. e190007, 2020.
- [131] T. Zhang, J. M. Pauly, S. S. Vasanawala, and M. Lustig, “Coil compression for accelerated imaging with cartesian sampling,” *Magnetic Resonance in Medicine*, vol. 69, no. 2, pp. 571–582, 2013.

- [132] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig, “Espirit—an eigenvalue approach to autocalibrating parallel mri: where sense meets grappa,” *Magnetic Resonance in Medicine*, vol. 71, no. 3, pp. 990–1001, 2014.
- [133] J. Leuschner, M. Schmidt, D. O. Baguer, and P. Maass, “Lodopab-ct, a benchmark dataset for low-dose computed tomography reconstruction,” *Scientific Data*, vol. 8, no. 1, p. 109, 2021.
- [134] A. Aghabiglou and E. M. Eksioğlu, “Deep unfolding architecture for mri reconstruction enhanced by adaptive noise maps,” *Biomedical Signal Processing and Control*, vol. 78, p. 104016, 2022.
- [135] P. Guo, P. Wang, J. Zhou, S. Jiang, and V. M. Patel, “Multi-institutional collaborations for improving deep learning-based magnetic resonance image reconstruction using federated learning,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2423–2432, 2021.
- [136] Y. Xu, S. Han, D. Wang, G. Wang, J. S. Maltz, and H. Yu, “Hybrid u-net and swin-transformer network for limited-angle cardiac computed tomography,” *Physics in Medicine & Biology*, vol. 69, no. 10, p. 105012, 2024.
- [137] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya, “Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data,” *Magnetic Resonance in Medicine*, vol. 84, no. 6, pp. 3172–3191, 2020.
- [138] J. Liu, Y. Sun, C. Eldeniz, W. Gan, H. An, and U. S. Kamilov, “Rare: Image reconstruction using deep priors learned without groundtruth,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1088–1099, 2020.
- [139] D. Narnhofer, K. Hammernik, F. Knoll, and T. Pock, “Inverse gans for accelerated mri reconstruction,” in *Wavelets and Sparsity XVIII*, vol. 11138, pp. 381–392, SPIE, 2019.

- [140] M. Z. Darestani and R. Heckel, “Accelerated mri with un-trained neural networks,” *IEEE Transactions on Computational Imaging*, vol. 7, pp. 724–733, 2021.
- [141] C. Cao, Z.-X. Cui, Y. Wang, S. Liu, T. Chen, H. Zheng, D. Liang, and Y. Zhu, “High-frequency space diffusion model for accelerated mri,” *IEEE Transactions on Medical Imaging*, vol. 43, no. 5, pp. 1853–1865, 2024.
- [142] L. Xiang, Y. Chen, W. Chang, Y. Zhan, W. Lin, Q. Wang, and D. Shen, “Deep-learning-based multi-modal fusion for fast MR reconstruction,” *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 7, pp. 2105–2114, 2019.
- [143] T. Eo, Y. Jun, T. Kim, J. Jang, H.-J. Lee, and D. Hwang, “KIKI-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images,” *Magnetic Resonance in Medicine*, vol. 80, no. 5, pp. 2188–2201, 2018.