
Stability and Generalizability in Learning-Based Medical Image Reconstruction

Author Weitong Zhang^{1,2,3,4}
Submitted September 30th 2025

Supervisors Prof. Bernhard Kainz^{1,5}
Prof. Dimitrios C. Karampinos^{4,6}

¹Department of Computing, Imperial College London, UK.

²UKRI Centre for Doctoral Training in AI for Healthcare, Imperial College London, UK.

³Imperial -TUM Joint Academy of Doctoral Studies (JADS), Imperial College London, UK.

⁴Klinikum rechts der Isar, Technical University of Munich, GER.

⁵Friedrich-Alexander University Erlangen-Nürnberg, GER.

⁶École Polytechnique Fédérale de Lausanne (EPFL), Switzerland.

A thesis submitted in fulfilment of the requirements for the degree of *Philosophiae Doctor*
in Computing Research.

Biomedical Image Analysis Group, Department of Computing,
Imperial College London

Biomedical Image Analysis Group,
Department of Computing,
Address 1,
Imperial College London,
180 Queens Gate,
South Kensington,
SW7 2AZ

Weitong Zhang © 2025

"A boundary is not that at which something stops, but
that from which something begins its presencing."

Martin Heidegger

To my parents and partner.

Declaration

I, Weitong Zhang, hereby declare that the work described in this thesis is my own, except where specifically acknowledged.

The template of this thesis is designed by Dr. Robert Robinson ¹⁷¹ <https://github.com/mlnotebook>.

Copyright

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution 4.0 International License (CC BY). Under this license, you may copy and redistribute the material in any medium or format for both commercial and non-commercial purposes. You may also create and distribute modified versions of the work. This on the condition that you credit the author. When reusing or sharing this work, ensure you make the license terms clear to others by naming the license and linking to the license text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this license or permitted under UK Copyright Law.

Acknowledgements

My deepest gratitude goes to my supervisor, Prof. Bernhard Kainz. No words can truly express how grateful I am for your guidance, patience, and unwavering support throughout my PhD journey. From the beginning, you gave me the freedom to explore, the space to grow, and the confidence to believe in my own ideas. You have not only been a brilliant mentor but also an inspiration. Your passion for research, openness, and ability to see the bigger picture have shaped the way I think and work. Thank you for your encouragement, for challenging me when I needed it most, and for always leading by example.

I am also deeply grateful to supervisors at TUM, DSI and CDTAI4Healthcare for their invaluable support and mentorship. Your insights and encouragement have been instrumental in my academic growth. The discussions we have shared and the knowledge you have imparted will stay with me beyond this PhD.

To my colleagues and collaborators, thank you for making this journey more than just research. Science is never a solitary endeavor, and I have been lucky to work with brilliant and kind people. The late-night debugging sessions, the conference trips, the countless coffee breaks filled with both frustration and laughter, all of these moments have made this journey unforgettable.

Beyond academia, I would like to thank go to my friends for their unwavering support, for keeping me sane when things felt overwhelming, and for reminding me that there is life beyond research. To my family, there are no words to express my gratitude. Your love, patience and belief in me have been my greatest source of strength. No matter how far I've been, I have always carried your support with me, and this achievement is as much yours as it is mine.

To my partner, who has been there through every moment of this PhD journey, no one knows better than you the true depth of this experience. You've witnessed all my struggles, breakthroughs, doubts, and victories firsthand. During those countless nights when I worked until dawn, you were silently by my side; when I faced setbacks, you gave me the courage to keep going; and when I celebrated small wins, you were always my most genuine cheerleader. Your love and trust have been my strongest foundation, helping me find direction when I was lost and regaining strength when I was exhausted. The journey we've shared together is far more precious than this dissertation could ever be. Thank you for showing me that no matter what challenges I face, I never have to face them alone.

A PhD is not just about academic research, it is also a journey of perseverance, discovery, and growth. It has tested my patience, my resilience, and my ability to embrace uncertainty. There

were moments of doubt, of exhaustion, of wondering if I would ever reach the finish line. But through it all, I have learned that progress is not always linear, that failure is often the best teacher, and that the people we meet along the way matter more than the destination itself.

And finally, to myself, I want to take a moment to acknowledge the effort, the long nights, the countless revisions, and the persistence that brought me here. Through every challenge, I kept going. Through every setback, I found the strength to continue. This journey was not easy, but I have made it. I am proud of the person I have become, and I am grateful for the determination that has carried me through.

Abstract

Deep learning has improved medical image reconstruction, but clinical use is limited by poor reliability under domain shift (new scanners, protocols, motion, noise) and scarce expert labels. This thesis builds reconstruction methods that generalize beyond training data, need less supervision, and can indicate when outputs are unreliable.

It contributes: (1) unsupervised/self-supervised MRI reconstruction that models motion and acquisition degradation without paired ground truth; (2) test-time adaptation without source data or labels, using physics constraints and self-consistency to recalibrate to new domains; (3) a stable diffusion-based reconstruction framework based on measure-consistent stochastic dynamics, improving robustness and reducing sampling error while clarifying failure modes. Uncertainty and mismatch detection are integrated throughout to support safer deployment.

Contents

1	Introduction	19
1.1	Motivation	19
1.2	Related Work	23
1.2.1	Learning-Based Medical Image Reconstruction	23
1.2.2	Unsupervised and Self-Supervised Reconstruction Methods	23
1.2.3	Domain Shift and Test-Time Adaptation in Medical Imaging	24
1.2.4	Generative Models for Inverse Problems	24
1.2.5	Stability, Uncertainty, and Reliability of Diffusion Models	25
1.2.6	Positioning of The Thesis	25
1.3	Contributions	27
1.4	Thesis overview	29
1.5	Publications	31
2	Background	34
2.1	Deep Learning for Computer Vision	35
2.1.1	Neural Network Fundamentals	35
2.1.2	Inverse Problems in Imaging	35
2.1.3	Learning Objectives and Parameter Estimation.	36
2.1.4	Machine Learning Fundamentals	37
2.1.4.1	Learning Paradigms	40
2.1.5	Convolutional Neural Networks	41
2.1.5.1	Convolutional Layers	41
2.1.5.2	Pooling Layers	42
2.1.5.3	Normalization Layers	42
2.1.5.4	CNN Architectures	43
2.2	Deep Learning for Medical Image Computing	45
2.2.1	Overview of Deep Learning Applications in Medical Imaging	45
2.2.1.1	Discriminative Learning Paradigms	45
2.2.1.2	Object Detection and Instance Segmentation	47
2.2.1.3	Inverse Problems in Medical Imaging	47
2.2.1.4	Image Registration	49
2.2.1.5	Generative Modeling	50

2.2.2	Fundamental Challenges in Medical Deep Learning	51
3	Motion-Corrected Quantitative Body MRI Super-resolution	53
3.1	Background	55
3.2	Respiratory Motion Correction and Super-Resolution for 3D Abdominal MRI	62
3.2.1	Overview	62
3.2.2	Introduction	62
3.2.3	Degradation Models and Assumptions	63
3.2.4	Method	65
3.2.5	Experiments	67
3.2.6	Summary	71
4	Test-time-Adapted Quantitative Body MRI Super-resolution	72
4.1	Physics-informed Unsupervised Test-time Adaptation for MRI Super-Resolution	73
4.1.1	Overview	73
4.1.2	Related Work	74
4.1.3	Introduction	75
4.1.4	Method	76
4.1.5	Experiments	79
4.1.6	Summary	83
5	Generalizable and Stable Diffusion Models for Inverse Problems	84
5.1	Background:	85
5.1.1	Related Work	85
5.2	Stability and Generalizability in SDE Diffusion Models with Measure-Preserving Dynamics	87
5.2.1	Overview	87
5.2.2	Introduction	87
5.2.3	Preliminaries	89
5.2.4	Intuition of Measure-Preserving Dynamics in SDE:	91
5.2.5	Mathematical Preliminaries	91
5.2.6	Instability Analysis: Transitional SGMs with Corrupted SDE Diffusion	93
5.2.6.1	Analysis of Proposition 1	94
5.2.6.2	Proof for reverse-time D^3GM process:	96
5.2.6.3	Proof for Proposition 2	97
5.2.7	Towards Stability: Measure-preserving Dynamics in SDE Diffusion	102
5.2.7.1	Implementation details	104
5.2.7.2	Additional Insights	107
5.2.8	Evaluation and results	109

5.2.8.1	Stability: Illustrations of the Measure-preserving Dynamics within Diffusion Models	110
5.2.8.2	Generalizability: D ³ GM for real-world data	112
5.2.9	Discussion and limitations	113
5.2.10	Summary	115
6	Conclusions	123
6.1	Thesis Summary	124
6.1.1	Motion-Corrected Quantitative Body MRI Super-resolution	124
6.1.2	Test-time-Adapted Quantitative Body MRI Super-resolution	124
6.1.3	Generalizable and Stable Diffusion Models for Inverse Problems	125
6.2	Limitations	126
6.2.1	Motion Correction and Super-Resolution	126
6.2.2	Test-time Adaptation Approach	126
6.2.3	Dynamics-aware Diffusion Models	126
6.3	Further Work	127
6.3.1	Vision-Language Models for Medical Image Understanding	127
6.3.2	Safety and Robustness Verification	128
6.3.3	Privacy-Preserving Reconstruction Systems	128
6.3.4	Autonomous Agents for Adaptive Protocols	128
	References	133

List of Figures

2.1	Overview of an ANN that takes an image as input and predicts its label. The network processes input images in a hierarchical way: from low-level textures and edges to high-level semantic concepts. Image adapted from [99][150]	38
2.2	Left: Mechanism of convolution layers: Each convolution kernel only aggregates and processes information of a small patch of the input. A map of responses of this kernel is obtained by sliding it along every spatial location of the input. Outputs from different kernels are stacked together, along the new depth (channel) dimension. Right: Visualization of shallow-layer kernels learned on ImageNet. Image source: [113]	41
3.1	Examples of 2D bowel MRI scans of a Crohn's disease subject in IBD dataset. Bowel MRI sequences are acquired with Coronal T2 weighted images and Axial T2 weighted images. The Coronal/Axial contains an HR stack of 2D images while the other two planes are LR images. The intersection of the green(X), blue(Y) and red(Z) lines indicates the corrupted area that has been affected by the motion. . .	56
3.2	During training, the input comprises corrupted MRI with motion artifacts and blurring effects. During inference, only the first pair of LR encoder E_{CLR} and SR decoder D_{SR} will be utilized to generate high-quality, motion-free, and super-resolved images.	65
3.3	GLRB is used to construct an effective feature extraction backbone based on encoders and decoders with different inputs. Additionally, the SR decoder and the QHR encoder include additional Pixel-shuffle and downsampling layers.	66
3.4	(a) The proposed PMRM sequence of respiratory simulation for motion corruption for Crohn's diseases. (b) The staged reconstruction process involving application of PMRM, MC result, and ground truth (GT) on TCGA-LIHC dataset. The MC matrix is calculated and applied on HR for visual comparison with simulated motion perturbations directly. (c) The MoCoSR results on clinical IBD data without GT. . .	69
3.5	Qualitative evaluation for SR on IBD dataset using representative methods of interpolation, model-based, learning-based, ablation, and MoCoSR.	69

3.6	Distribution relationships between reconstructed 3D bowel data and original HR data in a downstream Crohn's disease diagnosis task. (a) MoCoSR preserves diagnostic features and reconstructs a close representation of original data distribution. (b) The comparison method results in the loss and concealment of discriminative features. (c) Incorrectly reconstructed data misleads shifts in the distribution of SR data, which affects downstream task results.	70
4.1	UTAR combines a physical imaging chain degradation model (PIC) with a conditional consistency learning strategy (CCL), alleviating severe domain shifts during the testing phase using a source-free domain adaptation approach.	77
4.2	Illustration of the proposed physics-informed imaging chain degradation model (PIC). For the frequency domain, it includes a <i>sinc</i> filter simulating a low-resolution acquisition, and a sequence-specific degradation model tailored to the target domain, here: relaxation effects along the phase-encoding direction. For the spatial domain, it includes domain shuffled degradation for improved realism.	78
4.3	Visual comparison of different methods on the three datasets across target and source domain. (a) is the LR target, (b) and (f) are SR methods trained within the same domain, (c)-(e) are representative TTA methods.	81
4.4	An investigation of the power spectrum density (PSD) for MRI SR (averaged over 50 brain samples). (a) Generic encoder-decoder SR model under-performs in the target domain due to image acquisition domain gaps. (b) TTA can improve the initial capabilities of the model. However, they are still insufficient to reconstruct higher frequencies. (c) UTAR provides an overall improvement and narrowed the domain gap, enabling generic models to be adapted to domain shifts.	82

5.1	Dynamics-aware SDE Diffusion Generative Model (D ³ GM). When extending transitional SDEs to random dynamical systems (RDS), their measure-preserving property should be kept to maintain stability. This corresponds to driving the SDE towards the drift term μ (LQ). There is a <i>Temporal Distribution Discrepancy</i> which results from the gap between the forward estimation x_T and the low quality image in the SDE. With the distribution aligned between x_T and μ , the SDE can be made more robust to inverse problems. Reconstruction results for low quality (LQ) images after application of our D ³ GM method, on different tasks, compared to the ground truth (GT) on two domains - The frequency domain: MRI Reconstruction (undersampling factor 8x, 16x, frequency masks are colored red); MRI Super-resolution (up-scaling factor of X4, cross-domain evaluation). Due to the frequency-domain acquisition and the ill-posed nature of k-space undersampling, reconstruction artifacts in MRI differ fundamentally from those in natural image restoration, making this task particularly challenging for generic diffusion-based models.	103
5.2	Reverse Initialization with Basin of attraction.	108
5.3	Sampling trajectories of score-based SDE diffusion models. We visualize the evolution of sample trajectories under (from left to right) a standard score-based SGM, transitional SGMs including a Coefficient-Decoupled SDE and an Ornstein-Uhlenbeck (OU) SDE, and the proposed D ³ GM. While standard and transitional SGMs exhibit unstable trajectories or deviation from the target distribution due to accumulated temporal distribution discrepancy, D ³ GM demonstrates stable convergence towards the desired attractor. This behavior illustrates the effect of measure-preserving dynamics in maintaining stability during reverse diffusion processes at the inference stage.	111
5.4	O-HAZE and Dense-Haze.	112
5.5	Rain100H and Rain100L.	113
5.6	Qualitative results for (a) deraining and (b) dehazing.	113
5.7	Deraining results with light rain images of our method.	116
5.8	Deraining results with heavy rain images of our method.	117
5.9	Dehazing results with real hazy images of our method.	118
5.10	MRI reconstruction results with undersampling rate x8 and x16, on Frequency-encoding and Phase-encoding directions.	119
5.11	MRI super-resolution results with in-domain images of our method.	120
5.12	MRI super-resolution results with cross-domain (different imaging devices and degradation methods) images of our method.	121
5.13	MRI reconstruction results with undersampling rate x8 and x16 of our method, on Phase-encoding and Frequency-encoding directions.	122

6.1	MESHAgent's framework for cardiac imaging phenotype analysis through multi-agent reasoning. Operating and validating in three stages: (i) specialist agents analyze cardiac structures from imaging measurements; (ii) agents discover confounders via distributed analysis; (iii) collaborative reasoning generates phenome-wide associations, validated through clinical metrics. [241]	131
6.2	Illustration of strategic decoding through a Bayesian Decoding Game (BDG) for the Switzerland capital query (Q.) with candidates (A.), with initial rankings (left) and convergence dynamics (right) between Generator (solid) and Verifier (dashed). The collusive Nash equilibrium validates 'Zurich' (incorrect) while rejecting 'Bern' (correct) through game-theoretic interactions based on choice patterns rather than factual correctness.	131
6.3	Overview of the BDG under a signaling game structure. (a) Communication Topologies: comparing Single-agent Inconsistency with internal contradictions, Multi-agent Collusion that reinforces errors, and BDG Strategic Consensus achieving correct convergence through σ -separation. (b) BDG Signalling Tree: Environment (Env) sends private signals to generator ($G_{C/I}$), who generates candidates ($y_1, \dots, y_n$) for information transmission. The verifier makes judgments (C/I) without observing original signals, enabling strategic decoding through Markovian Strategy Update (dashed) until equilibrium.	132

List of Tables

3.1	IBD data set: the upper shows MRI data acquisition parameters. The lower shows the number of patients in each severity level of inflammation.	68
3.2	Quantitative evaluation of extensive methods on TCGA-LIHC and IBD data sets in terms of SSIM and PSNR for 3D abdominal MRI SR	70
3.3	Crohn’s disease classification performance on different 3D data. MoCoSR has a negligible effect on the downstream classification task as shown by high p-values in contrast to LR, MRESR, and CMRSR which produce significantly lower performance.	71
4.1	Quantitative results on the three unseen datasets Guys and IOP of IXI, PDFS of fastMRI under different enlargement scales. → indicates adaptation from source domain to target domain. Note that CST [38] is supervised and has extra access into the source data, UTAR is unsupervised and source-free.	81
4.2	PIC, CCL and UTAR for various domain shifts with a 4× SR model.	82
5.1	Differences between state-of-the-art SDE diffusion-based approaches.	90
5.2	Datasets parameters and split setting. IOP details of the scan parameters are not available.	104
5.3	Evaluation of the Lipschitz continuity.	104
5.4	Exemplary deraining results of (A) SGM, (B) transitional SGMs: Coefficient Decoupled SDE, OU SDE, and D ³ GM.	105
5.5	$\mu_t(x_t, t)$, $v_t(x_t, t)$ solutions with various θ_t , σ_t .	108
5.6	Quantitative results for Rain100H and Rain100L.(best in bold and second best underlined)	110
5.7	Quantitative results for O-HAZE and Dense-Haze.	110
5.8	Quantitative results for fastMRI dataset with acceleration rates x8 and x16.	113
5.9	Quantitative results for IXI MRI SR on unseen datasets.	113
5.10	Comparison to tailor-made approaches.	115
5.11	D ³ GM vs. deraining works.	115

Chapter 1

Introduction

1.1 Motivation

In recent years, deep learning has become a key method in medical image analysis, achieving high accuracy in tasks such as segmentation, classification, and reconstruction. For example, the U-Net architecture [173] has been applied to organ and lesion segmentation, and deep convolutional networks have achieved expert level accuracy in breast cancer [5] and skin cancer diagnosis. Although these results have been validated on curated datasets, the integration of such models into clinical workflows remains limited. This limitation arises from insufficient clinical validation, challenges in model interpretability, variations in imaging protocols, and the absence of regulatory frameworks that ensure safety and reliability.

Many algorithms have been developed and evaluated on public benchmarks, but few have undergone prospective clinical validation or been applied in routine practice. This gap highlights the disconnect between algorithm development and practical deployment. A major reason for this limitation is the poor generalization of deep learning models. Most models assume that training and testing data share similar distributions, while in reality, medical imaging data vary across hospitals, scanner vendors, imaging protocols, and patient populations. These domain shifts alter the input distributions and are intrinsic to clinical data acquisition. As a result, models that perform well during development often degrade when used in new settings, raising concerns about reliability and limiting their adoption in clinical decision-making.

The development of supervised deep learning models is constrained by the scarcity of high quality annotated data. Producing high quality annotations requires clinical expertise and careful review, which is particularly demanding for high resolution volumetric or motion affected imaging sequences. For instance, Peng et al. note that "the scarcity of high quality annotated training data is often unavailable for clinical and health care tasks" [156]. Furthermore, many target domains, such as rare pathologies or low resource regions, lack annotated datasets entirely, making supervised methods difficult to apply. These limitations restrict the scalability of supervised learning and hinder its deployment in new or evolving clinical environments.

A second major challenge is model generalization under domain shift. Medical imaging data vary across hospitals, scanner vendors, acquisition protocols, and patient populations. A recent systematic study found that data from different scanners almost always produce performance drops

compared with same scanner data, with the effect most severe in MRI [69]. Because many models assume that training and testing distributions are identical, this variability reduces reliability in unseen settings.

Another limitation is the absence of mechanisms for uncertainty quantification. Many existing models are optimized for accuracy on in distribution test sets but provide no guarantee when applied to out of distribution inputs. As a result, they may produce confidently incorrect predictions without any indicator of reliability. In high stakes settings such as radiology, this behavior undermines safety and trust. As discussed by [115], trustworthy clinical AI must provide uncertainty estimates to support decision making.

These observations motivate the central research question of this thesis:

Can we design deep learning systems for medical image reconstruction that remain reliable under domain shift, require minimal supervision, and can quantify their uncertainty when operating beyond the training distribution?

To address this question, the thesis investigates three tightly connected research directions:

- Domain adaptation and domain generalization strategies that enable models to adapt to new acquisition settings without full retraining.
- Representation learning methods that produce features stable across acquisition shifts and imaging domains.
- Reconstruction oriented generative models that are both visually accurate and aware of their own uncertainty.

This thesis addresses these challenges through three interrelated research directions. Each direction targets a specific bottleneck that limits the reliability and generalization of deep learning in medical image reconstruction. Together, they aim to improve the robustness, adaptability, and clinical relevance of learning based reconstruction systems under realistic constraints, including sparse supervision, distributional shift, and limited retraining capacity.

1. **Unsupervised motion-robust super-resolution**

Motion artifacts are among the most common sources of quality degradation in MRI, particularly in abdominal and fetal imaging, where involuntary motion and limited patient compliance are unavoidable [194]. These artifacts not only degrade visual quality but also compromise quantitative accuracy and downstream clinical interpretation. Super-resolution techniques offer a potential solution by enhancing spatial resolution from low-quality acquisitions; however, most existing approaches rely on supervised learning with paired high- and low-resolution data, which are difficult or impossible to obtain in the presence of motion.

Unsupervised and self-supervised learning paradigms provide a promising alternative by re-

laxing the requirement for fully paired training data and enabling learning directly from routinely acquired clinical images. In the context of motion-corrupted MRI, such approaches can exploit internal data consistency, physical acquisition constraints, or cross-slice redundancy to recover high-resolution representations without explicit ground-truth supervision. This research direction investigates unsupervised motion-robust super-resolution frameworks that explicitly account for motion-induced degradation, aiming to improve reconstruction quality and robustness in realistic clinical settings where annotated data are scarce or unavailable.

2. **Test-time domain adaptation** The challenge of domain shift extends beyond performance degradation; it also raises questions of reliability and adaptability in real-world deployment. Test time domain adaptation aims to address this issue by enabling models to recalibrate themselves during inference without access to source data or manual labels. Unlike traditional retraining or fine-tuning strategies, test time adaptation treats deployment as a continuous learning process in which the model dynamically aligns its internal representations to the target distribution. In medical imaging, this paradigm is particularly relevant because clinical data distributions evolve over time due to variations in scanner calibration, acquisition protocols, and patient demographics. Methods based on entropy minimization, normalization statistics alignment, or self-supervised consistency have shown that even limited adaptation at inference can substantially improve robustness [211]. However, the clinical application of these methods remains limited, as most approaches assume stable acquisition conditions and ignore physical constraints inherent to imaging systems. This research explores how test time adaptation can be extended to medical image reconstruction by integrating domain-informed priors and physics-based consistency into the adaptation loop. The objective is to develop models that can autonomously adjust to domain variations while preserving diagnostic fidelity and interpretability, thereby enhancing the reliability of AI-driven reconstruction in practical clinical environments.
3. **Stable and generalizable diffusion models** Generative modeling has become a powerful framework for solving ill posed inverse problems, particularly with the advent of diffusion based models [81]. These models have demonstrated strong visual fidelity and flexibility across a wide range of imaging tasks [81]. While recent advances highlight their generative capacity, their stability and reliability in clinical contexts remain open questions [104]. Empirical evidence suggests that diffusion models can be sensitive to perturbations in the forward process and to mismatches between training and deployment distributions. This sensitivity compromises robustness in scenarios with high noise levels, undersampled inputs, or previously unseen anatomies. To make these models viable for medical applications, it is essential to obtain a principled understanding of their failure modes and to design formulations that emphasize stability and generalization under distributional uncertainty. This motivates the development of measure consistent stochastic formulations that retain the expressiveness of diffusion models while improving their numerical and semantic robustness [218].

These directions are unified by a common methodological and clinical motivation: to move beyond conventional benchmarks and build learning systems that can function reliably in imperfect, variable, and under-annotated real-world environments. By targeting reconstruction under motion, adaptation without supervision, and stability in generative inference, this thesis aims to contribute foundational insights toward the development of AI tools that are not only technically advanced but also operationally viable in clinical deployment.

1.2 Related Work

This thesis builds upon several intersecting research areas, including learning-based medical image reconstruction, robustness under domain shift, and probabilistic generative modeling. While significant progress has been made in each of these directions, their integration under realistic clinical constraints remains incomplete. This section reviews relevant prior work and clarifies the gaps that motivate the contributions of this thesis.

1.2.1 Learning-Based Medical Image Reconstruction

Medical image reconstruction can be formulated as an ill-posed inverse problem, where the goal is to recover a latent image from incomplete or corrupted measurements. Classical reconstruction methods rely on explicit regularization, such as total variation or sparsity in predefined transform domains. While these approaches offer theoretical guarantees, they often struggle to capture the complex anatomical variability present in clinical data.

Deep learning has substantially reshaped this landscape. Convolutional neural networks, particularly encoder-decoder architectures such as U-Net [173], have been widely adopted for reconstruction and super-resolution tasks in MRI and CT. These models learn implicit image priors from data and have demonstrated strong performance under controlled experimental settings. Subsequent work has extended these ideas through cascaded architectures, unrolled optimization networks, and hybrid model-based learning frameworks that incorporate data consistency constraints.

Despite these advances, most learning-based reconstruction methods rely on fully supervised training with paired low- and high-quality data. In many clinical scenarios, such ground truth is unavailable or ill-defined. Motion-corrupted acquisitions, accelerated scans, and retrospective undersampling frequently lack a reliable reference image. As a result, supervised reconstruction models often exhibit limited generalization and may introduce hallucinated structures when applied outside their training distribution.

These limitations have motivated interest in alternative learning paradigms that reduce reliance on paired supervision and improve robustness under realistic acquisition conditions.

1.2.2 Unsupervised and Self-Supervised Reconstruction Methods

To address the scarcity of high-quality annotations, unsupervised and self-supervised learning approaches have been explored for medical image reconstruction. These methods exploit internal data consistency, physical constraints of the acquisition process, or redundancy across slices and contrasts. Examples include cycle-consistency-based frameworks, k-space self-supervision, and methods that leverage undersampling masks or noise realizations as implicit supervision signals.

In MRI, self-supervised approaches have demonstrated the ability to learn reconstruction mappings directly from undersampled data without access to fully sampled references. While these methods significantly reduce annotation requirements, they often rely on assumptions such as stationary noise, consistent acquisition protocols, or mild degradation patterns. In the presence of severe motion, domain shift across scanners, or complex corruption processes, their performance can degrade substantially.

Moreover, many unsupervised reconstruction methods optimize pixel-wise or frequency-domain losses without explicitly modeling uncertainty or stability. As a result, they provide limited insight into the reliability of their predictions, which is a critical requirement for clinical deployment.

1.2.3 Domain Shift and Test-Time Adaptation in Medical Imaging

Domain shift is a well-recognized challenge in medical imaging. Variations in scanner vendors, field strengths, acquisition protocols, and patient populations can lead to substantial changes in data distributions. Empirical studies have shown that models trained on data from a single site often suffer performance drops when evaluated on unseen scanners, with MRI being particularly sensitive to such shifts [69].

Domain adaptation techniques aim to mitigate this issue by aligning feature representations across domains, using strategies such as adversarial learning, feature normalization, or latent space regularization. While these approaches can improve robustness, many require access to source data or labeled target samples, which may be impractical in clinical settings due to privacy or logistical constraints.

Test-time adaptation (TTA) offers an alternative paradigm, allowing models to adapt during inference using only unlabeled target data. Methods based on entropy minimization, batch normalization statistics alignment, or self-supervised consistency have shown promising results in classification and segmentation tasks [211]. However, their application to image reconstruction remains relatively underexplored.

Reconstruction differs fundamentally from discriminative tasks: the output space is high-dimensional, the inverse problem is ill-posed, and physical acquisition constraints play a central role. Most existing TTA methods do not explicitly incorporate these factors, limiting their effectiveness and interpretability when applied to medical image reconstruction. These challenges highlight the need for adaptation strategies that are informed by both data distributions and acquisition physics.

1.2.4 Generative Models for Inverse Problems

Generative modeling provides a principled framework for inverse problems by learning data-driven priors over images. Early approaches based on variational autoencoders and generative adversarial networks have been applied to MRI reconstruction and super-resolution, often producing visually

appealing results. However, adversarial training can be unstable, and GAN-based methods are prone to hallucinations, raising concerns about their clinical reliability.

Diffusion models have emerged as a powerful alternative, offering improved stability and likelihood-based training [81]. By modeling a gradual denoising process governed by stochastic differential equations, diffusion models can generate high-fidelity samples and have been successfully extended to conditional and inverse problem settings. Recent work has explored diffusion-based reconstruction by incorporating measurement consistency, posterior sampling, or score-based optimization.

Despite their promise, diffusion models introduce new challenges in medical imaging. Their performance can be sensitive to noise schedules, forward process design, and mismatches between training and deployment distributions. In MRI, where degradation occurs in the frequency domain rather than the image domain, diffusion models trained on natural image corruptions may fail to capture the true acquisition process. Furthermore, the iterative nature of diffusion sampling can amplify small modeling errors, leading to instability or degraded reconstructions under domain shift [104].

1.2.5 Stability, Uncertainty, and Reliability of Diffusion Models

Recent studies have begun to examine the reliability and failure modes of diffusion models, particularly in safety-critical applications. While diffusion models often produce visually plausible outputs, they do not inherently guarantee stability under perturbations or provide calibrated uncertainty estimates. In medical imaging, this limitation is particularly problematic, as confidently incorrect reconstructions can mislead downstream clinical decisions.

Several approaches have been proposed to improve robustness, including modified stochastic dynamics, constrained sampling, and uncertainty-aware formulations. However, many existing methods remain empirical, lacking a principled connection between the stochastic process design and the preservation of meaningful data distributions. Theoretical analyses of diffusion dynamics under domain shift and complex degradation remain limited.

These observations motivate the exploration of diffusion formulations that emphasize stability, measure consistency, and robustness to distributional mismatch, particularly in the context of inverse problems with ill-defined ground truth.

1.2.6 Positioning of The Thesis

The body of existing work reviewed above has substantially advanced learning-based reconstruction, unsupervised training strategies, domain adaptation, and diffusion-based generative modeling. However, these developments have largely progressed along parallel directions, with limited integration under the constraints encountered in real clinical imaging. In particular, current meth-

ods often address supervision scarcity, domain shift, or generative modeling stability in isolation, while practical deployment typically involves all of these challenges simultaneously.

This thesis is positioned around the premise that robustness in medical image reconstruction cannot be achieved through architectural design or training strategy alone. Instead, it requires a coherent treatment of data scarcity, distributional shift, and stochastic inference dynamics as interdependent components of the reconstruction problem. From this perspective, the contributions of the thesis are not defined by a single model class, but by a unifying methodological viewpoint that emphasizes stability, adaptability, and reliability under imperfect conditions.

Concretely, the thesis investigates three complementary research directions. First, it explores unsupervised and self-supervised reconstruction methods that explicitly account for motion-induced degradation, where high-quality reference images are unavailable or ill-defined. Second, it examines test-time domain adaptation for inverse problems, framing deployment as a dynamic process in which models continuously adjust to evolving acquisition distributions without access to source data or manual labels. Third, it develops measure-consistent diffusion formulations that prioritize stability and distributional fidelity in generative inference, particularly in settings where the forward degradation process deviates from assumptions commonly made in natural image modeling.

Together, these directions form a coherent framework for learning-based medical image reconstruction under realistic clinical constraints. Rather than optimizing performance on narrowly controlled benchmarks, the thesis focuses on identifying principles that improve robustness, generalization, and interpretability when models are exposed to motion, domain shift, and uncertainty. In doing so, it aims to bridge the gap between methodological advances in machine learning and the reliability requirements of clinical imaging practice.

1.3 Contributions

Building upon the motivation and challenges outlined in the introduction, this thesis makes several key contributions to the field of generalizable and data-efficient deep learning for medical image reconstruction, with a focus on motion-corrected quantitative body MRI. These contributions aim to bridge the gap between research algorithms and clinical deployment by improving robustness, generalization, and accountability of reconstruction frameworks under domain shifts and limited supervision.

1. Motion-corrected super-resolution without paired supervision

We propose a novel unsupervised joint motion-correction and super-resolution (MoCoSR) framework for 3D abdominal MRI, which simultaneously mitigates respiratory and bowel-motion artifacts while enhancing spatial resolution. Unlike conventional supervised reconstruction pipelines that rely on expensive, motion-free high-resolution ground truth data, MoCoSR introduces a cycle-consistent dual-domain learning scheme that leverages latent representations from low-resolution corrupted inputs. The method disentangles motion and resolution degradation through multi-scale feature consistency in the latent space. Extensive experiments on both TCGA-LIHC and IBD datasets demonstrate that MoCoSR achieves superior reconstruction quality (SSIM $\uparrow$, PSNR $\uparrow$) while preserving diagnostic fidelity for Crohn’s disease assessment, showing no statistically significant bias in downstream disease classification.

2. Physics-informed test-time adaptation for MRI super-resolution

To address the poor generalization of SR models across scanners, acquisition protocols, and patient cohorts, we introduce UTAR (Unsupervised Test-time Adaptation with physics-informed degradation modeling). UTAR integrates a dual-domain physical imaging chain model (PIC) that explicitly simulates the k-space acquisition process and a Conditional Consistency Learning (CCL) mechanism that enforces spectral alignment during inference. This approach enables on-the-fly adaptation of pre-trained SR models to unseen target domains without any access to source-domain data, preserving data privacy and regulatory compliance. Quantitative and qualitative evaluations on IXI and fastMRI datasets show that UTAR consistently reduces domain gap and improves high-frequency fidelity under severe acquisition shifts, outperforming both supervised and source-dependent TTA methods. The method represents a practical step toward source-free and deployable medical image enhancement under real-world variability.

3. Dynamics-aware diffusion model for stable inverse problem solving

We further develop D3GM (Dynamics-aware SDE Diffusion Generative Model), a novel generative formulation grounded in the theory of measure-preserving random dynamical systems (RDS). Unlike conventional diffusion models, which often exhibit instability and distributional drift under heterogeneous degradations, D3GM enforces measure-consistent stochastic dynamics, ensuring temporal stability and probabilistic invariance throughout the diffusion process. We formally characterize the Temporal Distribution Discrepancy inherent in transitional SGMs and provide theoretical guarantees for stability and convergence. Empirically, D3GM achieves state-of-the-art reconstruction and restoration performance across multiple inverse problems including MRI reconstruction, cross-domain MRI super-resolution, dehazing, and deraining, demonstrating robustness and generalization beyond domain-specific tuning. This contribution provides a theoretically grounded pathway toward stable and generalizable diffusion-based solvers for ill-posed medical imaging problems.

The thesis integrates data efficiency, physical modeling, and theoretical stability into a unified research direction that moves beyond task-specific improvements toward a cohesive understanding of generalization in deep medical imaging. By combining methodological rigor with clinical relevance, the presented work lays a foundation for the next generation of adaptive and trustworthy reconstruction systems capable of operating across heterogeneous acquisition settings and patient populations.

1.4 Thesis overview

Following the presentation of the main contributions in the previous section, this chapter provides an overview of the thesis structure and the logical progression of the research. The remainder of this thesis is designed to progressively develop the conceptual framework, methodologies, and empirical validations that support the proposed advances in generalizable medical image reconstruction. Rather than presenting isolated techniques, the chapters follow a coherent narrative from foundational principles to specific algorithmic contributions and theoretical insights. The overall organization of the thesis is outlined below.

Chapter 2 establishes the theoretical and technical background. It reviews essential concepts in deep learning and neural network architectures, followed by a discussion of inverse problems in medical image reconstruction. This chapter also surveys recent advances in deep learning-based MRI reconstruction, highlighting the challenges of generalization, data scarcity, and domain variability that motivate the developments in subsequent chapters.

Chapters 3 to 5 form the core of this thesis and present the proposed methodologies for improving the generalization and robustness of deep learning models for quantitative body MRI. As summarized in Fig. 1.2, the research evolves across three complementary directions: (a) learning from unpaired and motion-corrupted data (Chapter 3) (b) learning through test-time adaptation with physics-informed priors (Chapter 4) (c) learning stable generative models under heterogeneous degradations (Chapter 5). Together, these directions progressively expand the generalization capability of learning-based reconstruction systems from data-level adaptation to theoretically principled stability.

Chapter 3 introduces MoCoSR, an unsupervised framework that jointly performs motion correction and super-resolution for 3D abdominal MRI. By leveraging a cycle-consistent dual-domain learning scheme, MoCoSR disentangles motion and resolution degradation without requiring paired ground-truth data. The method significantly improves image fidelity and diagnostic reliability on clinical datasets, offering a practical and data-efficient solution for motion-robust MRI reconstruction.

Chapter 4 presents UTAR, a physics-informed test-time adaptation framework that enhances cross-domain generalization of super-resolution models. It incorporates a physical imaging chain model and a conditional consistency learning mechanism to adapt pre-trained models to unseen scanners and protocols without source-domain access. This enables source-free adaptation under realistic domain shifts, validated by extensive experiments on multiple public MRI datasets.

Chapter 5 proposes D3GM (Dynamics-aware SDE Diffusion Generative Model), a theoretically grounded approach to stable and generalizable inverse problem solving. Built upon the theory of measure-preserving random dynamical systems, D3GM ensures probabilistic consistency and temporal sta-

bility in the diffusion process. Theoretical analysis and empirical evaluations demonstrate that D3GM achieves state-of-the-art performance across MRI reconstruction and various image restoration tasks under diverse degradations.

Finally, Chapter 6 concludes the thesis by summarizing the overall findings, discussing the theoretical and practical implications of the presented work, and outlining potential future directions. These include integrating multimodal vision–language representations, developing privacy-preserving reconstruction systems, and ensuring the reliability and interpretability of AI-based medical imaging solutions in clinical applications.

Overall, the structure of the thesis reflects an iterative process of developing, validating, and generalizing learning-based reconstruction models under increasingly realistic and unconstrained conditions. This progression from methodological innovation to clinical applicability embodies the central theme of this research: building generalizable, trustworthy, and physically consistent AI systems for medical imaging.

1.5 Publications

Parts of this thesis have been published previously as peer-reviewed scientific publications. This thesis is mainly based on the following works:

1. **Zhang, Weitong**, Berke Basaran, Qingjie Meng, Matthew Baugh, Jonathan Stelter, Phillip Lung, Uday Patel, Wenjia Bai, Dimitrios Karampinos, and Bernhard Kainz. "Mocosr: Respiratory motion correction and super-resolution for 3D abdominal MRI." In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 121-131. Cham: Springer Nature Switzerland, 2023. (Poster) [\[240\]](#)
2. **Zhang, Weitong**, Jonathan Stelter, Liu Li, Avinash Kori, Cheng Ouyang, Chen Chen, Florian Knoll, Dimitrios Karampinos, and Bernhard Kainz. "Physics-informed Unsupervised Test-time Adaptation for MRI Super-Resolution." In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) Workshop on Advancing Data Solutions in Medical Imaging AI, 2024. (Oral)
3. **Zhang, Weitong**, Chengqi Zang, Liu Li, Sarah Cechnicka, Cheng Ouyang, and Bernhard Kainz. "Stability and Generalizability in SDE Diffusion Models with Measure-Preserving Dynamics." Advances in Neural Information Processing Systems 37 (NeurIPS) 2024: 81606-81644. (Poster) [\[243\]](#)

The following works were also performed during the course of the PhD. They are related to the topic of the thesis but are not presented in detail:

1. **Zhang, Weitong**, Zang C, Kainz B. "From Self-Check to Consensus: Bayesian Strategic Decoding in Large Language Models." Advances in Neural Information Processing Systems (NeurIPS), 2025.
2. **Zhang, Weitong**, Qiao M., Zang C, Niederer S, Matthews P, Bai W., Kainz B. "Multi-Agent Reasoning for Cardiovascular Imaging Phenotype Analysis." In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2025 (Poster, Highlight) [\[241\]](#).
3. Qiao M., Zheng J., **Zhang, Weitong**, Ma Q., Li L, Bai W. "Mesh4D: A Motion-Aware Multi-View Variational Autoencoder for 3D+t Mesh Reconstruction." In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2025 .
4. Dombrowski, Mischa, **Weitong Zhang**, Sarah Cechnicka, Hadrien Reynaud, and Bernhard Kainz. "Image Generation Diversity Issues and How to Tame Them." Conference on Computer Vision and Pattern Recognition (CVPR), 2025. (Poster) [\[46\]](#)
5. **Zhang, Weitong**, Zang C, Schmidt M, Blythman R. "BiD: Behavioral Agents in Dynamic Auctions." International Conference on Learning Representations, International Conference on Learning Representations (ICLR) 2025 Workshop World Models, 2025. (Poster)

6. **Zhang, Weitong**, Zang C, Schmidt M, Blythman R. "Strategic LLM Decoding through Bayesian Games." International Conference on Learning Representations (ICLR), LLM reasoning and planning, 2025. (Poster)
7. Cechnicka S, **Zhang, Weitong**, Dombrowski M., Baugh M., Li Z., Roufosse C., Kainz B. "Graph Conditioned Diffusion for Controllable Histopathology Image Generation." In International Conference on Medical Image Computing and Computer-Assisted Intervention, 2025.
8. Li Z, **Zhang W**, Cechnicka S, Kainz B. Data-Efficient Generation for Dataset Distillation. European Conference on Computer Vision (ECCV), Global Dataset Distillation Challenge, 2024. (Rank 1, Oral) [\[122\]](#)
9. Stelter, Jonathan, Kilian Weiss, Lisa Steinhelfer, Veronika Spieker, Elizabeth Huaroc Moquilaza, Weitong Zhang, Marcus R. Makowski et al. "Simultaneous whole-liver water T 1 and T 2 mapping with isotropic resolution during free-breathing." NMR in Biomedicine 37, no. 12 (2024): e5216. [\[196\]](#)
10. Stelter, Jonathan, Kilian Weiss, Jakob Meineke, Veronika Spieker, Weitong Zhang, Julia A. Schnabel, Rickmer F. Braren, Bernhard Kainz, and Dimitrios C. Karampinos. "Subspace modeling enables time-efficient B1-corrected volumetric water-specific relaxometry with a long echo train CASPR trajectory." [\[195\]](#)
11. Li, Liu, Hanchun Wang, Matthew Baugh, Qiang Ma, Weitong Zhang, Cheng Ouyang, Daniel Rueckert, and Bernhard Kainz. "Universal Topology Refinement for Medical Image Segmentation with Polynomial Feature Synthesis." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 670-680. Cham: Springer Nature Switzerland, 2024.(Poster) [\[120\]](#)
12. Li, Liu, Qiang Ma, Cheng Ouyang, Zeju Li, Qingjie Meng, Weitong Zhang, Mengyun Qiao et al. "Robust segmentation via topology violation detection and feature synthesis." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 67-77. Cham: Springer Nature Switzerland, 2023. (Poster) [\[119\]](#)
13. Li, Liu, Qiang Ma, Zeju Li, Cheng Ouyang, Weitong Zhang, Anthony Price, Vanessa Kyriakopoulou et al. "Fetal cortex segmentation with topology and thickness loss constraints." In Workshop on the Ethical and Philosophical Issues in Medical Imaging, pp. 123-133. Cham: Springer Nature Switzerland, 2022. (Poster) [\[118\]](#)
14. Liu, Che, Sibao Cheng, Chen Chen, Mengyun Qiao, Weitong Zhang, Anand Shah, Wenjia Bai, and Rossella Arcucci. "M-flag: Medical vision-language pre-training with frozen language models and latent space geometry optimization." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 637-647. Cham: Springer Nature Switzerland, 2023. (Poster) [\[127\]](#)

15. Basaran, Berke Doga, Weitong Zhang, Mengyun Qiao, Bernhard Kainz, Paul M. Matthews, and Wenjia Bai. "LesionMix: A lesion-level data augmentation method for medical image segmentation." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 73-83. Cham: Springer Nature Switzerland, 2023. (Oral) [\[14\]](#)
16. Guo, Jingjie, Weitong Zhang, Matthew Sinclair, Daniel Rueckert, and Chen Chen. "Pay attention to the atlas: Atlas-guided test-time adaptation method for robust 3d medical image segmentation." In International Conference on Medical Image Computing and Computer-Assisted Intervention Workshop on Big task small data, 2023. (Oral) [\[71\]](#)

Chapter 2

Background

This chapter provides the theoretical, algorithmic, and practical background required to understand the developments presented in Chapters 3 to 5. It begins with an overview of magnetic resonance imaging (MRI) physics and signal acquisition, formulates image reconstruction as an inverse problem, and reviews classical and learning-based reconstruction methods. It then discusses motion artifacts and super-resolution in quantitative body MRI, introduces the concepts of domain shifts and model generalization, and presents modern generative priors with an emphasis on diffusion-based approaches. The chapter concludes with evaluation metrics, dataset governance, clinical translation, and reproducibility considerations that underpin the methodological rigor of this thesis.

2.1 Deep Learning for Computer Vision

2.1.1 Neural Network Fundamentals

Machine learning encompasses a research domain focused on developing computational approaches that enable quantitative prediction through statistical modeling [21]. Unlike traditional programming paradigms, machine learning algorithms autonomously derive their predictive rules from collections of observational data, commonly known as the *training dataset*.

Following established conventions in the field, we denote the input to a machine learning model as $\mathbf{x} \in \mathcal{X}$, where $\mathcal{X}$ represents the input space containing all possible input instances. Similarly, we represent the output or prediction as $\mathbf{y} \in \mathcal{Y}$, with $\mathcal{Y}$ denoting the output space comprising all potential prediction values.

The specific structure of $\mathbf{x}$ and $\mathbf{y}$ varies substantially across application domains. In the context of image classification, for instance, the input $\mathbf{x}$ typically consists of an RGB image represented as a three-dimensional tensor $\mathbf{x} \in \mathbb{R}^{H \times W \times 3}$, where H and W correspond to the vertical and horizontal pixel dimensions respectively, and the value 3 indicates the three color channels (red, green, and blue). For such classification tasks, the corresponding output $\mathbf{y} \in \mathcal{N} = \{1, 2, 3, \dots, N_C\}$ represents a discrete categorical variable indicating the assigned class from among N_C possible categories—for example, $\mathbf{y} = 1$ might designate the “feline” class while $\mathbf{y} = 2$ could represent the “canine” class.

At its core, a machine learning model constitutes a mapping function $f : \mathcal{X} \rightarrow \mathcal{Y}$ that transforms elements from the input space into elements of the output space. This function f is frequently described as a *hypothesis*, with the collection of all possible functions that could be learned denoted as the *hypothesis space* $\mathcal{F}$. Typically, f is parameterized by a vector of adjustable values $\boldsymbol{\theta}$ (alternatively termed *model weights* or *parameters*). We employ the notation $f_{\boldsymbol{\theta}}$ to explicitly indicate this parameterization. The structural complexity of $f_{\boldsymbol{\theta}}$ exhibits considerable variation across application contexts, ranging from elementary linear regression formulations to sophisticated deep neural architectures incorporating billions of tunable parameters.

2.1.2 Inverse Problems in Imaging

Inverse problems are central to many tasks in computational imaging, where the goal is to recover a latent signal from incomplete, noisy, or corrupted observations. Formally, these problems can be expressed as solving for an unknown image x given observations y , where $y = Ax + n$, and A is a (possibly ill-conditioned or non-invertible) forward operator, and n represents noise.

Such problems arise in a wide range of applications, including magnetic resonance imaging (MRI), computed tomography (CT), and image deblurring, where the physical measurement process leads to partial or degraded data. In many of these cases, the inversion is ill-posed: multiple solutions

may fit the observations, or small perturbations in the input may lead to large variations in the output. To resolve this, prior knowledge about the image distribution must be incorporated to guide the reconstruction.

Traditional approaches typically rely on handcrafted priors—such as smoothness, sparsity, or total variation—to regularize the solution space. While these methods have had success in some settings, they are often too rigid or general to capture the rich structure of real-world images, especially in clinical applications where anatomy and pathology vary significantly.

Magnetic Resonance Imaging as an Inverse Problem.

Magnetic resonance imaging (MRI) differs fundamentally from natural image formation in computer vision because it involves sampling in the frequency domain rather than direct spatial acquisition. The measured signal represents a partial Fourier transform of the underlying spin density, and reconstruction therefore constitutes an inverse problem. Given undersampled k-space measurements y , the goal is to recover the latent image x satisfying $y = Ax + n$, where A encodes coil sensitivities and Fourier encoding, and n models measurement noise. When the sampling rate is below the Nyquist criterion, the problem becomes ill-posed and small perturbations in y can cause large errors in x . Classical approaches mitigate this through regularization terms that impose smoothness or sparsity priors, but such handcrafted assumptions often fail to capture the complex statistics of anatomical structures.

This has motivated the development of data-driven approaches, particularly using deep learning, to model image priors implicitly and learn mappings from measurements to reconstructions directly from data. These methods form the basis of most modern inverse problem solvers in computer vision and medical imaging.

2.1.3 Learning Objectives and Parameter Estimation.

Parameter optimization in machine learning systems typically occurs through minimization of targeted criteria functions computed on the training data $\mathcal{D}_{tr}$. These criteria, formalized as *loss functions*, quantify discrepancies between model predictions and desired outputs. Each loss function establishes a numerical penalty that decreases as predictions approach their intended targets, thereby guiding the parameter search toward optimal configurations.

In supervised learning contexts, the loss function $l(\cdot, \cdot)$ directly evaluates disparities between the model's predictions $f_{\theta}(\mathbf{x})$ and corresponding ground truth values $\mathbf{y}$. The theoretically optimal parameter set θ^* minimizes the expected loss across the entire data-generating distribution:

$$\theta^* = \arg \min_{\theta} \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim p_{tr}} [l(f_{\theta}(\mathbf{x}), \mathbf{y})] \quad (2.1)$$

Here, p_{tr} denotes the joint probability distribution from which training examples are drawn. Since

this distribution remains analytically inaccessible, practical implementations approximate the expectation using finite samples. Assuming statistical independence among training observations, this approximation takes the form:

$$\boldsymbol{\theta}^* \approx \arg \min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^N l(f_{\boldsymbol{\theta}}(\mathbf{x}_i), \mathbf{y}_i) \quad (2.2)$$

This methodology, known as *empirical risk minimization* (ERM) [207], substitutes sample averaging for theoretical expectation. For most contemporary models, particularly neural networks, Equation 2.2 lacks analytical solutions and necessitates iterative optimization procedures. Such approaches typically employ gradient-based methods that require loss functions with continuous derivatives with respect to model parameters.

The ERM framework admits probabilistic interpretations under specific conditions. When the loss function solely captures prediction errors without additional constraints, the optimization becomes equivalent to *maximum likelihood estimation* (MLE). However, practical implementations frequently incorporate supplementary terms that encode prior beliefs about plausible parameter distributions. This extension transforms the objective into:

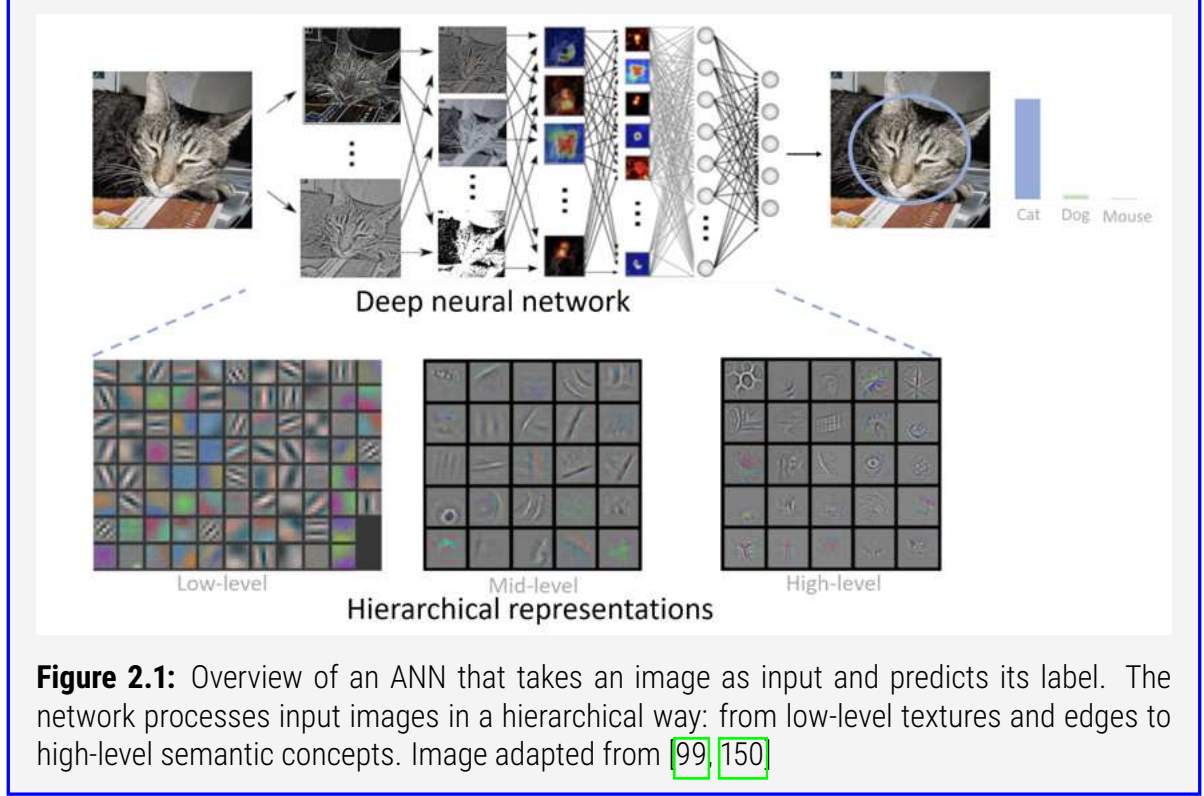
$$\boldsymbol{\theta}^* = \arg \min_{\boldsymbol{\theta}} (\mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim p_{tr}} [l(f_{\boldsymbol{\theta}}(\mathbf{x}), \mathbf{y})] + \lambda \mathcal{R}(\boldsymbol{\theta})) \quad (2.3)$$

The additional term $\mathcal{R}(\boldsymbol{\theta})$ constitutes a *regularization function* that encodes structural preferences for parameter configurations, while λ controls the regularization strength. From a Bayesian perspective, this formulation corresponds to *maximum a posteriori* (MAP) estimation, with the regularization term representing the negative logarithm of a prior probability distribution over parameters. For instance, ℓ_2 regularization ($\mathcal{R}(\boldsymbol{\theta}) = \|\boldsymbol{\theta}\|_2^2$) corresponds to assuming a zero-centered Gaussian prior. Regularization serves a crucial practical purpose beyond its theoretical interpretation: it mitigates overfitting by constraining model complexity, thereby enhancing generalization to unseen data.

2.1.4 Machine Learning Fundamentals

Neural Networks. Artificial neural networks (ANNs) form the computational foundation of modern deep learning systems [19, 80]. Though initially inspired by biological neural structures [175], contemporary ANNs have evolved into specialized mathematical frameworks optimized for pattern recognition and function approximation. These architectures organize computational units into hierarchical processing stages that progressively transform raw inputs into increasingly abstract representations. The fundamental processing unit in neural networks is the artificial neuron, which implements a composition of linear transformation and non-linear activation. This computational motif enables networks to approximate complex non-linear functions through the aggregation of

simpler operations. The hierarchical organization of these units facilitates dimensional reduction across processing stages, creating abstraction hierarchies particularly suited for high-dimensional data processing [19]. As demonstrated in Fig. 2.1 initial network layers extract fundamental features (edges, textures), while deeper layers encode increasingly complex semantic constructs by combining simpler patterns.



Multi-layer Perceptron Structure. The multi-layer perceptron (MLP) [175] represents the canonical feedforward neural architecture. Information flows strictly from input to output through a sequence of fully-connected layers, with each layer performing parallel computations on outputs from the preceding layer. An MLP neuron indexed as unit i in layer l implements the following transformation:

$$z_i^l = \sigma \left(\sum_{j=1}^{n_{l-1}} w_{l,i,j} z_{l-1,j} + b_{l,i} \right) = \sigma(\mathbf{w}_{l,i}^T \mathbf{z}_{l-1} + b_{l,i}) \quad (2.4)$$

Where $\mathbf{w}_{l,i} \in \mathbb{R}^{n_{l-1}}$ represents the weight vector for neuron i in layer l , $b_{l,i} \in \mathbb{R}$ is the bias term, $\mathbf{z}_{l-1} \in \mathbb{R}^{n_{l-1}}$ denotes the output vector from layer $l - 1$, $\sigma(\cdot)$ is a non-linear activation function, and n_{l-1} indicates the number of neurons in layer $l - 1$.

Several activation functions have become standard in neural architectures. The Rectified Linear Unit (ReLU), defined as $\sigma(x) = \max(0, x)$, addresses vanishing gradient problems while maintaining computational efficiency. Sigmoid activations, $\sigma(x) = \frac{1}{1+e^{-x}}$, map inputs to the range $(0, 1)$,

while hyperbolic tangent, $\sigma(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$, provides outputs in the range $(-1, 1)$ with stronger gradients near zero. More recent functions such as GELU, $\sigma(x) = x \cdot \Phi(x)$ where Φ is the standard normal CDF, have shown improved performance in transformer architectures.

The outputs from all neurons in layer l are collected into a vector $\mathbf{z}_l = [z_{l,1}, z_{l,2}, \dots, z_{l,n_l}]^T \in \mathbb{R}^{n_l}$, which serves as input to the subsequent layer. For classification tasks, the network typically concludes with a softmax transformation that converts raw activations into class probabilities:

$$p(y = j|\mathbf{x}) = \frac{e^{z_{L,j}}}{\sum_{k=1}^{n_L} e^{z_{L,k}}} \quad (2.5)$$

Where L designates the final network layer and n_L equals the number of target classes. This normalization ensures outputs sum to unity and can be interpreted as categorical probability distributions.

Optimization Framework. Neural network training involves identifying optimal parameter configurations through iterative optimization. Unlike many statistical methods, neural networks rarely admit closed-form solutions due to their non-linear activation functions and high-dimensional parameter spaces. Instead, gradient-based methods progressively refine parameters using error back-propagation [177], which computes parameter gradients through recursive application of the chain rule. The standard optimization approach employs stochastic gradient descent (SGD), which approximates full-batch gradients using randomly sampled mini-batches:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}_t; \mathcal{B}_t) \quad (2.6)$$

Where $\boldsymbol{\theta}_t$ represents parameters at iteration t , η denotes the learning rate, $\mathcal{B}_t \subset \mathcal{D}_{tr}$ is a randomly sampled mini-batch, and $\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}_t; \mathcal{B}_t)$ is the gradient computed on the mini-batch.

This stochastic approximation provides several advantages: it reduces computational demands per iteration, introduces beneficial noise that can help escape poor local minima, and enables processing of datasets too large to fit in memory. Modern implementations extend basic SGD with momentum terms and adaptive learning rates. Notable algorithms include Momentum SGD, which incorporates velocity terms that accumulate gradients across iterations; Adam [107], which maintains per-parameter learning rates based on first and second moments of gradients; and AdamW, which modifies Adam with decoupled weight decay for improved regularization. Despite the non-convex nature of neural network loss landscapes, these methods reliably identify parameter configurations with strong predictive performance when properly configured and regularized.

2.1.4.1 Learning Paradigms

Machine learning approaches can be systematically categorized based on supervision signal availability, with each paradigm addressing distinct learning scenarios:

Supervised Learning represents the canonical learning framework where models are trained on labeled example pairs $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$. The explicit supervision signal enables direct error computation and gradient-based optimization. This approach dominates practical applications when sufficient labeled data exists, supporting diverse tasks including classification (mapping inputs to discrete category labels), regression (predicting continuous values), and structured prediction (generating complex outputs like bounding boxes or segmentation masks).

Semi-supervised Learning leverages both labeled and unlabeled data, formally represented as $\mathcal{D}_{tr} = \mathcal{D}_{tr}^l \cup \mathcal{D}_{tr}^u$, where $\mathcal{D}_{tr}^l = \{(\mathbf{x}_i^l, \mathbf{y}_i^l)\}_{i=1}^{N_l}$ contains labeled examples and $\mathcal{D}_{tr}^u = \{\mathbf{x}_j^u\}_{j=1}^{N_u}$ comprises unlabeled observations. This approach typically relies on assumptions about data distribution structure, such as continuity (similar inputs should have similar outputs), cluster assumption (points in the same cluster likely share labels), and manifold assumption (data lies on a lower-dimensional manifold). Common techniques include self-training, consistency regularization, and graph-based methods [109, 180]. Semi-supervised learning proves particularly valuable when labeled data acquisition requires expensive expert annotation, as in medical imaging.

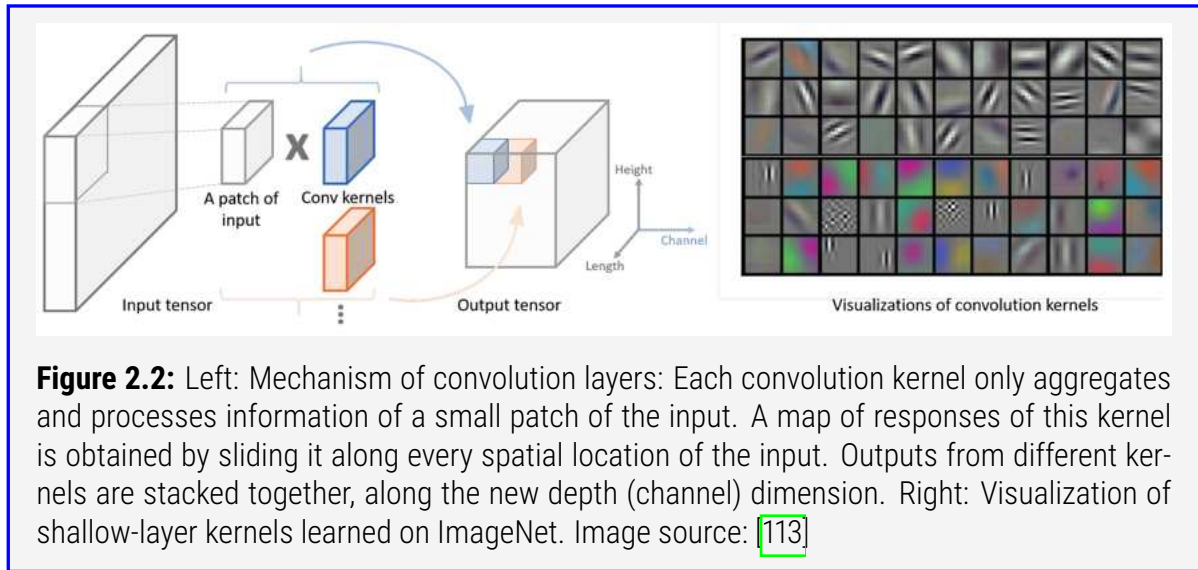
Unsupervised Learning operates exclusively on unlabeled data $\mathcal{D}_{tr}^u = \{\mathbf{x}_i\}_{i=1}^N$, focusing on discovering inherent data properties. Key approaches include clustering (grouping similar instances without predefined categories) [111], dimensionality reduction (creating compact data representations), generative modeling (learning to approximate and sample from data distributions) [66], and self-supervised learning (deriving supervision signals from the data itself) [96]. Recent advances in self-supervised representation learning have proven particularly impactful, enabling models to extract transferable features from unlabeled data that can subsequently enhance performance on downstream supervised tasks.

Reinforcement Learning addresses sequential decision-making problems involving an agent interacting with an environment. The framework comprises state space $\mathcal{S}$ (possible environmental conditions), action space $\mathcal{A}$ (available agent behaviors), reward function $r : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{R}$ (immediate feedback signal), transition dynamics $P(s_{t+1} | s_t, a_t)$ (environment behavior model), and policy $\pi : \mathcal{S} \rightarrow \mathcal{A}$ (agent's decision strategy). The objective is learning an optimal policy π^* that maximizes expected cumulative rewards. Unlike supervised learning, reinforcement learning must balance exploration of unknown behaviors against exploitation of known rewarding actions, often without access to optimal behavior demonstrations. Applications include game playing [55], and autonomous systems [110].

2.1.5 Convolutional Neural Networks

Convolutional neural networks (CNN) [117] are almost dominating computer vision applications, especially for those whose data are defined on Cartesian grids: 2-D and 3-D images and videos. In a CNN, as shown in the left part of Fig. 2.2, each neuron only perceives and processes a small patch of the input at each time (the spatial extent of this small patch is also called *receptive field*). A map of responses of this neuron is obtained by sliding the *same* neuron along every spatial location of the input (this is how convolutional neural network gets its name). This sliding mechanism bestows translational equivariance to the network [117]. The maps of responses by each neuron are stacked together to form the new *channel* (also called *depth*) dimension. These maps are passed to the following layer, and are processed by the neurons of the next layer in a similar manner. The stacked maps of responses can also be interpreted as *feature maps* extracted by shallower layers.

The right part of Fig. 2.2 shows examples of weights (convolutional kernels) learned by neurons in the input layer of a CNN. Without the need for manually defining feature extractors (e.g., SIFT [133]), the neurons discover image features such as edges, corners and colors by themselves, guided by the supervision signal provided by the loss function. Features of deeper layers are more difficult to interpret, and they are commonly believed to encode high-level semantic information, such as object parts and object entities [64].



2.1.5.1 Convolutional Layers

The core operation in CNNs is the convolutional layer, which applies a set of learnable filters to input feature maps. Formally, for an input feature map $\mathbf{X} \in \mathbb{R}^{H \times W \times C_{in}}$, a convolutional layer with C_{out} filters produces an output feature map $\mathbf{Y} \in \mathbb{R}^{H' \times W' \times C_{out}}$ through:

$$Y_{h,w,j} = \sum_{i=0}^{C_{in}-1} \sum_{p=0}^{k-1} \sum_{q=0}^{k-1} X_{h+p,w+q,i} \cdot W_{p,q,i,j} + b_j \quad (2.7)$$

where $W_{p,q,i,j}$ represents the weight at position (p, q) of the i -th input channel in the j -th filter, and b_j is the bias term for the j -th filter. The dimensions H' and W' depend on the stride and padding configurations of the convolutional operation. This formulation highlights several key properties of convolutional layers: **Local Connectivity**: Each output value depends only on a local neighborhood of the input. **Parameter Sharing**: The same filter parameters are applied across the entire spatial dimensions of the input. **Translation Equivariance**: A translation in the input produces a corresponding translation in the output. These properties significantly reduce the number of parameters compared to fully-connected layers while preserving spatial relationships in the data.

2.1.5.2 Pooling Layers

Convolutional neural networks usually also contain *pooling layers* [113] and *normalization layers* [91, 205, 222]. Pooling layers perform downsampling operations on the input image and intermediate feature maps for spatial dimension reduction. For example, a pooling layer with a pooling window of 2×2 and a stride of 2 downsamples its input feature map by a factor of 2 along both vertical and horizontal directions. The most commonly used pooling operation is maximum pooling, which only retains the strongest response of neurons within the pooling window while discarding weaker responses. This operation effectively filters out information less relevant to the task. Mathematically, max-pooling operates as:

$$Y_{h,w,c} = \max_{0 \leq p < s, 0 \leq q < s} X_{h \cdot s + p, w \cdot s + q, c} \quad (2.8)$$

where s is the stride parameter, typically set equal to the pooling window size for non-overlapping pooling.

2.1.5.3 Normalization Layers

Normalization layers stabilize the training process for deep networks by addressing gradient-vanishing or gradient-explosion issues [91]. Gradient vanishing and explosion sometimes occur during the back-propagation stage, where gradients become increasingly small or large as they propagate backward through layers due to multiplications in the chain rule. Normalization layers solve these issues by re-normalizing statistics of feature maps along certain dimensions of the input. Commonly used normalization layers include batch normalization [91], instance normalization [205], group normalization [222], and others. Batch normalization, one of the most widely used normalization techniques, operates as:

$$\hat{X}_{h,w,c} = \gamma_c \frac{X_{h,w,c} - \mu_c}{\sqrt{\sigma_c^2 + \epsilon}} + \beta_c \quad (2.9)$$

where μ_c and σ_c^2 are the mean and variance computed across the batch dimension for each channel

c , while γ_c and β_c are learnable parameters that allow the network to recover the original representation if needed. The small constant ϵ prevents division by zero.

2.1.5.4 CNN Architectures

Compared with MLPs, CNNs feature several advantages that enable them to dominate most computer vision applications: CNNs are more scalable to image data with high spatial dimensions (usually larger than 64×64) compared to MLPs. In MLPs, each neuron weight at the input layer must have the same dimension as the input. In CNNs, the weights of neurons do not need to be as large due to the convolution (sliding) mechanism, making CNNs more parameter efficient. Attributed to the sliding mechanism, CNNs are intrinsically *translationally equivariant*: their response to an object remains roughly consistent regardless of the object's location in the image, therefore making CNNs more robust than MLPs during testing. Also due to spatial equivariance, the spatial relations between pixels/patches in the intermediate feature maps of CNNs are naturally preserved, a consequence of the convolution operation. This property is particularly suitable for tasks requiring strict spatial correspondence between input and output, such as image segmentation, reconstruction, and denoising.

Over the past decade, CNN architectures have evolved significantly, with several key innovations:

Residual Connections: Introduced in ResNet [76], these skip connections allow gradients to flow more effectively through deep networks:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{\mathbf{W}_i\}) + \mathbf{x} \quad (2.10)$$

where $\mathcal{F}$ represents a residual mapping function. This simple modification enabled the training of networks with hundreds of layers.

Inception Modules: Proposed in GoogLeNet [197], these modules process inputs at multiple scales simultaneously using parallel convolutional paths with different kernel sizes.

Depthwise Separable Convolutions: Used in architectures like MobileNet [85], these factorize standard convolutions into depthwise and pointwise operations, significantly reducing computational cost:

$$Y_{h,w,j} = \sum_{i=0}^{C_{in}-1} \left(\sum_{p=0}^{k-1} \sum_{q=0}^{k-1} X_{h+p,w+q,i} \cdot W_{p,q,i} \right) \cdot V_{i,j} \quad (2.11)$$

Attention Mechanisms: Incorporated in networks like SENet [86], these allow the model to focus on relevant features while suppressing irrelevant ones, operating on both channel and spatial dimensions. Modern CNN architectures like EfficientNet [198] combine these innovations with neural

architecture search techniques to achieve optimal trade-offs between accuracy and computational efficiency.

In summary, this chapter has outlined the essential theoretical background spanning inverse problems, learning-based reconstruction, and generalization across domains. The integration of physical modeling and data-driven learning emerges as a recurring theme, enabling robust MRI reconstruction under diverse acquisition conditions. Building upon these foundations, Chapter 3 introduces an unsupervised motion-corrected super-resolution framework, Chapter 4 extends it to physics-informed test-time adaptation across domains, and Chapter 5 generalizes the concept further through a diffusion-based generative model that guarantees stability and measure preservation.

2.2 Deep Learning for Medical Image Computing

The intersection of deep learning methodologies with medical imaging has revolutionized diagnostic and therapeutic approaches in modern healthcare. This section examines the transformative role of deep learning algorithms in medical image analysis, with particular emphasis on segmentation techniques and reconstruction methodologies for magnetic resonance imaging (MRI). The first portion synthesizes current applications across various medical imaging domains, while the latter portion addresses fundamental challenges inherent to deep learning implementations in clinical settings. Particular attention is devoted to the inverse problem as a central focus of this thesis.

2.2.1 Overview of Deep Learning Applications in Medical Imaging

When categorized according to their underlying computational formulations, deep learning applications in medical imaging span several distinct paradigms. Each paradigm addresses specific clinical needs while presenting unique methodological challenges. The following sections elucidate these approaches with emphasis on their mathematical foundations and clinical utility.

2.2.1.1 Discriminative Learning Paradigms

Discriminative approaches in medical image analysis focus on mapping images to clinically relevant outputs through supervised learning methodologies. These approaches can be further categorized based on their output characteristics:

Image-level classification represents the foundational application that catalyzed the contemporary deep learning revolution. Following the remarkable performance of convolutional neural networks on the ImageNet challenge in 2012 [113], researchers rapidly adapted these architectures for medical applications. In clinical contexts, image-level classification typically involves determining the presence or absence of pathological conditions from volumetric or planar imaging data. The diagnostic value of medical imaging frequently extends beyond binary classification to encompass spatial localization of pathologies. Semantic segmentation addresses this need by assigning class labels to individual pixels or voxels within medical images. Unlike natural image segmentation, medical image segmentation demands exceptional precision due to the subtle boundaries between anatomical structures and the potentially severe consequences of misclassification.

Contemporary segmentation architectures evolved from the fundamental insight that effective segmentation requires both contextual understanding and precise spatial localization. Segmentation networks typically operate by sliding classification networks across spatial locations and aggregating predictions to generate pixel-wise classification maps. This methodology enables the simultaneous recognition of anatomical structures and delineation of their boundaries.

The U-Net architecture [173] represents a significant milestone in medical image segmentation, introducing an encoder-decoder structure with symmetric skip connections. This architectural innovation addresses a critical limitation of conventional convolutional networks: the progressive loss of spatial resolution through downsampling operations. The downsampling process—implemented through strided convolutions or pooling operations—enables networks to capture increasingly abstract features while reducing computational complexity. However, this dimensional reduction inherently sacrifices spatial precision. To restore the original spatial dimensions while preserving detailed structural information, U-Net employs several complementary strategies:

The first key component consists of skip connections, which establish direct pathways for transferring feature maps from encoder layers to their corresponding decoder counterparts. These connections provide high-resolution spatial information that guides the upsampling process, preserving fine structural details that would otherwise be lost during downsampling. Complementing these connections, various upsampling mechanisms facilitate dimensional restoration, including transposed convolution [59], interpolation followed by standard convolution, and pixel-shuffling operations [185]. Additionally, the architecture enables multi-scale feature integration by concatenating feature representations from shallow layers (which preserve local details) with those from deeper layers (which capture semantic context), thereby leveraging complementary information sources throughout the network hierarchy.

The inherent architectural logic of U-Net reflects a fundamental insight regarding feature representation in convolutional networks: shallow layers excel at capturing localized structural details but lack semantic context, while deeper layers develop rich semantic understanding at the expense of spatial precision. By establishing direct pathways between corresponding layers, skip connections enable the network to preserve the advantages of both shallow and deep representations. This architectural paradigm has proven so effective that it has inspired numerous variants, including V-Net [143] for volumetric segmentation, Attention U-Net [149] which incorporates attention mechanisms to emphasize relevant features, and nnU-Net [92], which adapts network configurations dynamically based on dataset characteristics.

Recent innovations in segmentation methodology have explored alternative representations beyond pixel-wise classification. Polygon-based approaches [123] model segmentation targets as vertices connected by edges, directly regressing coordinate positions rather than classifying individual pixels. While this representation offers computational efficiency and compact output encoding, it presents challenges when applied to structures with complex topological configurations or multiple disconnected components. The fundamental trade-off between representational flexibility and computational efficiency continues to drive research in this domain.

2.2.1.2 Object Detection and Instance Segmentation

Object detection extends beyond classification by localizing and identifying multiple instances of relevant structures within medical images. This capability is particularly valuable for quantifying discrete pathological findings such as tumors, lesions, and nodules. Unlike semantic segmentation, which assigns class labels without distinguishing between instances of the same class, object detection identifies individual occurrences and typically delineates their spatial extents using bounding boxes. Contemporary object detection frameworks generally follow one of two methodological paradigms:

Two-stage detectors [249] decompose the detection process into sequential steps: first generating region proposals that potentially contain objects of interest, then classifying and refining these proposals. This approach prioritizes detection accuracy at the expense of computational efficiency. The region proposal networks (RPNs) integral to many two-stage detectors analyze feature maps to identify candidate regions based on their "objectness" scores—a measure indicating the likelihood that a region contains a relevant object regardless of its specific class. These candidate regions undergo subsequent refinement through classification and boundary adjustment processes.

Single-stage detectors [165] integrate proposal generation and classification into a unified framework, enabling end-to-end optimization and significantly faster inference. These methods typically predetermine a set of anchor boxes with various aspect ratios and scales, directly regressing both classification probabilities and box coordinate adjustments relative to these anchors. While initially less accurate than their two-stage counterparts, architectural refinements and advanced training methodologies have substantially narrowed this performance gap.

In medical contexts, self-configuring detection pipelines [92] have emerged to address domain-specific challenges such as variable acquisition protocols and institutional heterogeneity. These frameworks automatically optimize preprocessing parameters, anchor configurations, and network hyperparameters based on dataset characteristics, reducing the need for expert intervention in the model development process. Instance segmentation represents an extension of object detection that produces pixel-level masks for each identified instance. A common architectural approach involves integrating a segmentation subnetwork within the object detection framework, applying pixel-wise classification specifically within predicted bounding boxes [87].

2.2.1.3 Inverse Problems in Medical Imaging

Inverse problems constitute a significant category of computational challenges in medical imaging, fundamentally concerned with recovering underlying signals from imperfect measurements. These approaches address the inherent limitations of medical imaging systems, which often capture data that represents an indirect transformation of the anatomical structures of interest. Reconstruction, denoising, super-resolution, and de-aliasing all fall within this problem domain.

MRI reconstruction exemplifies the inverse problem formulation, where the measurement process captures data in the Fourier transform domain (k-space) rather than directly in the spatial domain. The scanner acquires frequency-encoded signals that must be converted to interpretable spatial representations. Due to physical and temporal constraints, k-space measurements are typically undersampled, leading to reconstruction challenges when conventional methods like zero-filled inverse Fourier transform are applied.

The mathematical formulation of these inverse problems generally follows the maximum a posteriori estimation framework:

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} (\|\mathbf{y} - \mathcal{M}(\mathbf{x})\| + \mathcal{R}(\mathbf{x})) \quad (2.12)$$

In this formulation, $\mathbf{x}$ represents the target “clean” image, $\mathbf{y}$ denotes the measured signal, and $\mathcal{M}(\cdot)$ models the measurement (acquisition) process. The first term, commonly referred to as the data fidelity term, ensures consistency between the reconstructed image and the observed measurements. The operator $\|\cdot\|$ typically implements an ℓ_2 metric, though other distance functions may be employed depending on noise characteristics and application requirements.

In practical scenarios, the equation $\mathbf{y} - \mathcal{M}(\mathbf{x}) = 0$ is generally underdetermined, meaning multiple solutions could satisfy the measurement constraints. Directly applying the inverse transformation—such as zero-filled inverse Fourier transform in MRI reconstruction—often produces aliased or degraded images. The regularization term $\mathcal{R}(\cdot)$ addresses this indeterminacy by incorporating prior knowledge about the expected properties of valid solutions.

Conventional regularization approaches have employed analytically defined priors such as sparsity in transformation domains [137] or low-rank structural constraints [72]. Deep learning methodologies have revolutionized this aspect of inverse problem solving by learning data-driven regularization functions directly from training examples [10, 74]. These learned priors implicitly capture complex statistical properties of medical images that prove difficult to formulate explicitly. Several architectural paradigms have emerged for deep learning-based inverse problem solving:

The end-to-end mapping approach represents perhaps the most straightforward implementation, wherein networks directly learn transformations from corrupted or undersampled measurements to fully sampled ground truth images, essentially implementing the function $\mathbf{x} = f_{\theta}(\mathbf{y})$. While conceptually simple, these models often struggle with physical consistency constraints when measurement operators are complex. Addressing this limitation, unrolled optimization architectures explicitly incorporate the physics of the acquisition process by reformulating iterative optimization algorithms as differentiable neural network layers with learnable parameters [2]. This approach maintains consistency with the underlying physical processes while leveraging data-driven optimization of algorithm components. Plug-and-play methods offer yet another alternative, combining conventional optimization frameworks with neural network denoisers that implicitly define reg-

ularization terms [208]. This modular approach enables flexible integration of advanced denoising models within established reconstruction pipelines without requiring end-to-end retraining.

2.2.1.4 Image Registration

Image registration addresses the spatial alignment of corresponding anatomical structures across different images. This process enables the integration of complementary information from multiple imaging modalities, facilitates the monitoring of disease progression over time, and supports precise guidance during interventional procedures. Registration algorithms determine spatial transformations that establish correspondences between anatomical landmarks in source and target images. Deep learning approaches to registration [13, 40] have demonstrated several advantages over conventional optimization-based methods, including significantly faster inference times and improved robustness to initialization conditions. These models typically learn to regress deformation fields or transformation parameters directly from image pairs, optimizing similarity metrics between the transformed source image and the target image. The clinical utility of registration spans numerous applications:

Multimodal fusion represents a primary application domain, wherein structural information from anatomical modalities such as CT or MRI is combined with functional data from metabolic imaging like PET or functional MRI. This integration provides comprehensive insights that enhance both diagnostic accuracy and treatment planning precision [209]. Longitudinal analysis constitutes another critical application, where images acquired at different time points are precisely aligned to enable quantitative assessment of disease progression and treatment efficacy [11]. The temporal dimension introduces additional challenges due to potential anatomical changes between acquisitions, requiring deformable registration approaches that can accommodate structural modifications while maintaining anatomical plausibility. In surgical contexts, intraoperative guidance systems leverage real-time registration between preoperative plans and intraoperative imaging to facilitate precise navigation during minimally invasive procedures [35]. The time constraints of the operating environment necessitate extremely efficient registration algorithms that can adapt to tissue deformation and surgical modifications. Additionally, atlas-based segmentation approaches utilize registration to transfer anatomical annotations from labeled atlases to patient-specific images, enabling automated delineation of structures without requiring manual annotation for each case [23].

The inherent ill-posedness of registration—multiple transformations could potentially align the same anatomical structures—necessitates regularization constraints to ensure anatomically plausible deformations. Deep learning models implicitly learn these constraints from training data, potentially offering more anatomically coherent results than explicit mathematical regularizers.

2.2.1.5 Generative Modeling

Generative modeling represents a paradigm shift from discriminative approaches by focusing on learning the underlying distribution of training data rather than mapping between input-output pairs. Formally, these models approximate the probability distribution p_{tr} of training examples $\mathbf{x}_{tr}$ using parameterized functions, typically implemented as neural networks. The fundamental distinction of generative models lies in their ability to synthesize novel examples by sampling from the learned distribution. Several architectural frameworks have emerged for generative modeling in medical imaging:

Generative Adversarial Networks (GANs) [65] represent one prominent architectural paradigm, employing a competitive training process between generator and discriminator networks. The generator produces synthetic images from random noise vectors, while the discriminator attempts to distinguish between real and generated samples. This adversarial dynamic creates a self-improving system where the generator continuously refines its outputs to better deceive the discriminator, ultimately producing increasingly realistic images that capture the statistical properties of the training data. Variational Autoencoders (VAEs) [108] offer an alternative approach by learning probabilistic encodings of input data as distributions in latent space, from which new samples can be generated through random sampling and subsequent decoding. This methodology provides explicit control over the generation process through manipulation of latent variables, enabling principled interpolation between examples and structured exploration of the learned manifold. Normalizing Flows [169] implement a different strategy through invertible transformations between simple base distributions and complex data distributions, enabling both precise density estimation and efficient sampling. The bijective nature of these transformations preserves information content while transforming between distribution spaces. More recently, Diffusion Models [81] have emerged as a powerful generative framework that progressively transforms data distributions through iterative noise addition and learned denoising processes, demonstrating exceptional generation quality and stability compared to earlier approaches. The applications of generative models in medical imaging extend beyond mere synthesis:

Data augmentation represents a fundamental application area, wherein generative models synthesize additional training examples to improve model robustness and address class imbalance issues particularly prevalent in medical datasets [57]. This approach proves especially valuable for rare pathologies where limited examples are available for supervised learning. Domain adaptation constitutes another critical application domain, where generative models translate images between acquisition protocols or imaging modalities to facilitate cross-domain learning and enable knowledge transfer across institutional boundaries [100]. Anomaly detection leverages the discriminative capability of generative models to identify pathological images by measuring deviations from the learned distribution of normal examples [181]. This approach enables unsupervised identification of abnormalities without requiring explicit annotation of every possible pathological

variant. Additionally, missing data imputation employs generative approaches to reconstruct absent information based on partial observations and learned data distributions [229], enabling more complete utilization of datasets with acquisition artifacts or protocol variations. The integration of conditioning mechanisms into generative frameworks enables controlled synthesis based on specified attributes or auxiliary information, further expanding their utility in medical applications.

2.2.2 Fundamental Challenges in Medical Deep Learning

Despite remarkable achievements, deep learning approaches to medical image analysis face significant challenges that limit their translation into clinical practice. This section examines these challenges, with particular emphasis on the generalization problem that forms the central focus of this thesis.

Limited Supervision and Data Efficiency. Medical imaging datasets typically contain orders of magnitude fewer annotated examples than datasets in natural image domains. This scarcity stems from several factors: the expertise required for accurate annotation, privacy concerns limiting data sharing, the time-intensive nature of manual delineation, and the inherent rarity of certain pathological conditions. Several methodological approaches have emerged to address this fundamental limitation:

The Generalization Challenge. The generalization problem—the focus of this thesis—refers to the degradation in model performance when applied to data distributions that differ from those encountered during training. In medical imaging, this challenge manifests across multiple dimensions:

1. *Scanner variability*: Different imaging equipment produces characteristic artifacts and intensity distributions, even when implementing nominally identical protocols [62].
2. *Protocol heterogeneity*: Variations in acquisition parameters significantly impact image appearance and quality [25].
3. *Institutional differences*: Clinical practices, patient demographics, and disease prevalence vary across medical centers, introducing systematic biases in datasets [236].
4. *Pathological diversity*: The visual presentation of diseases exhibits substantial variability, with rare variants often underrepresented in training data [147].

Conventional machine learning wisdom suggests that diverse training data should improve generalization. However, in medical contexts, where data collection across multiple centers presents logistical and regulatory challenges, alternative approaches must be considered. Domain adaptation methodologies explicitly address distribution shifts by aligning feature representations or applying style transfer techniques [82].

The development of generalizable models fundamentally requires rethinking the inductive biases

encoded in network architectures. Attention mechanisms that focus on relevant anatomical structures while disregarding acquisition-specific artifacts have demonstrated improved cross-domain performance [183]. Similarly, architectural modifications that enforce anatomical consistency constraints can enhance robustness to appearance variations [148].

Interpretability and Clinical Integration. The clinical adoption of deep learning systems demands not only technical performance but also interpretability—the ability to explain model decisions in terms comprehensible to healthcare providers. This requirement stems from both regulatory considerations and the ethical imperative to maintain clinical accountability.

Visualization techniques that highlight influential regions in input images provide intuitive explanations of model behavior. Gradient-based attribution methods [6] and attention maps [213] offer insights into the features driving predictions. However, these post-hoc explanations sometimes fail to capture the complex decision boundaries learned by deep networks. Inherently interpretable architectures represent an alternative approach, incorporating domain knowledge and anatomical constraints directly into model design. Part-based models that explicitly represent anatomical structures and their relationships [247] offer explanations aligned with clinical reasoning patterns. Similarly, graphical models that encode spatial relationships between structures can generate interpretable probabilistic explanations [153].

The integration of deep learning systems into clinical workflows requires careful consideration of human-computer interaction dynamics. Rather than pursuing full automation, hybrid approaches that leverage complementary strengths of algorithms and clinicians often prove most effective [203]. Interactive segmentation tools that combine deep learning predictions with manual refinement capabilities enable efficient workflow integration while maintaining expert oversight.

Research Directions. Deep learning methodologies have transformed medical image analysis across diverse application domains. Continued advancement requires addressing fundamental challenges in data efficiency, generalization, and interpretability. This thesis focuses specifically on the generalization problem for inverse-problem based tasks, developing novel architectural approaches and training methodologies to enhance performance robustness across varying acquisition conditions.

Subsequent chapters will present innovative contributions addressing specific aspects of generalizable systems in medical imaging inverse problem-based tasks, including physics-informed degradation models, unsupervised domain adaptation strategies, and measure-preserving dynamics theory can improve the stability and generalization of diffusion models in inverse problems. These studies provide a unified framework that directly addresses the limitations of current methods in stability and generalization, while offering practical solutions of inverse problems in natural and medical imaging fields for complex real-world applications.

Chapter 3

Motion-Corrected Quantitative Body MRI Super-resolution

Image reconstruction models across domains

This chapter is based on the following publications:

- **Zhang, Weitong**, Berke Basaran, Qingjie Meng, Matthew Baugh, Jonathan Stelter, Phillip Lung, Uday Patel, Wenjia Bai, Dimitrios Karampinos, and Bernhard Kainz. "Mocosr: Respiratory motion correction and super-resolution for 3D abdominal MRI." In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 121-131. Cham: Springer Nature Switzerland, 2023.

The prevalence of gastrointestinal diseases is increasing, especially inflammatory bowel disease. There are no effective means for early diagnosis and no complete cure. In fact, during the diagnosis and treatment of the intestine, Magnetic Resonance Imaging (MRI) occupies a major position in the clinical decision-making process. MRI of the gastrointestinal tract provides high contrast images across soft tissues, however, the involuntary movement of the bowels and respiration result in severe imaging artifacts and limited image resolution. At present, quantitative MRI techniques have achieved great success in tumor staging and therapy outcome assessment. Therefore, the combination of Artificial Intelligence (AI) image processing technology and MRI-based diagnosis and treatment has great potential and benefits in terms of the image quality improvement and quantitative analysis.

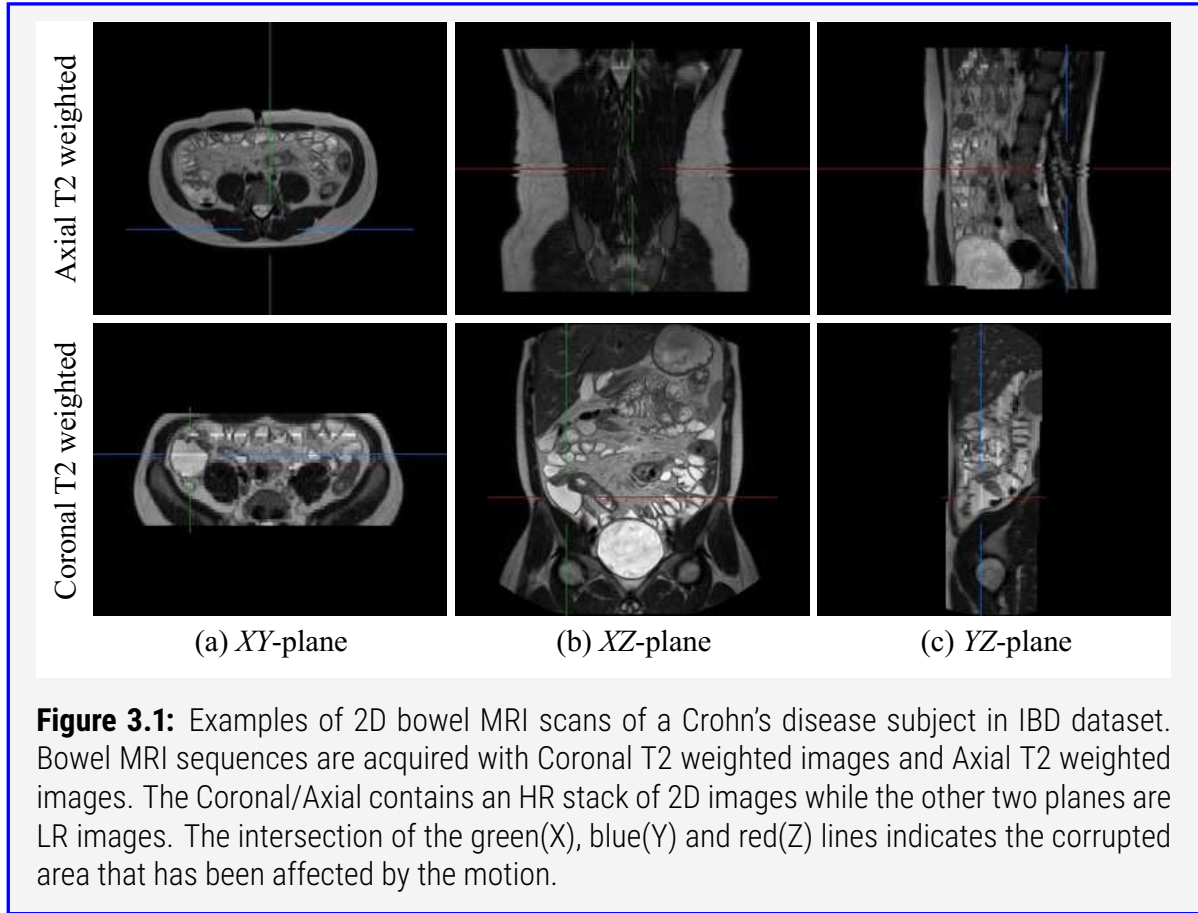
This chapter will take into account the complex physiological movements of the abdomen and the random motion of the patient. Based on a generative algorithm, it aims to achieve a high-quality image reconstruction process that can effectively alleviate the commonly encountered issue of long scanning times in quantitative imaging, and produce high-confidence images corrected by machine learning, specifically of the bowels. Motion-Corrected Quantitative Body MRI is expected to become an automated method to identify inflammatory bowel disease without additional annotations by medical experts. It will also provide biomarkers consistent across patient populations. Once adapted to current clinical imaging practice, this research will aim to assist radiologists in diagnosing and monitoring diseases. We extend the image reconstruction process to blind super-resolution and physiologically informed motion correction to achieve significant MR image quality improvements.

3.1 Background

Abdominal MRI. IBD is a relatively common, but also easily overlooked disease due to the insidious clinical presentation [60] and the long time interval between the initial causative event and diagnosis [179]. One of its manifestations, Crohn's Disease (CD), usually presents symptoms such as abdominal pain, diarrhoea, fatigue, and cramping pain, which may be accompanied by severe complications [157] [166]. The number of patients worldwide is increasing year by year, and according to the latest report in 2021 [124], the cost of treatment far exceeds other common chronic diseases (diabetes, cardiovascular and cerebrovascular diseases), which has become a significant factor of medical care expenditure [124]. Direct health care costs are approximately €3500 per patient per year. Currently, an estimated 1.3 million people suffer from IBD, which equals 0.2% of the whole European population, with most CD patients being under 35 years of age [245]. Although CD cannot be completely cured, early diagnosis can significantly reduce treatment costs and permanent physical damage [201]. Magnetic Resonance Imaging (MRI) plays a crucial role in diagnosing and monitoring CD. In clinical applications and research, spatial information with high resolution MRI is usually preferred because it can provide more significant structure and texture details with a smaller voxel-level volume, thus promoting accurate subsequent analysis and early diagnosis.

IBD usually has symptoms such as abdominal pain, diarrhoea, and cramping pain, and is accompanied by severe complications [166]. Symptoms include inflammation of tissues anywhere along the gastrointestinal tract, and it is often difficult to screen early through the expert experience without the aid of equipment [157]. The first method of standardizing clinical IBD diagnosis is endoscopy, such as the Crohn's Disease Endoscopic Severity Index (CDEIS) [201]. However, the patient's acceptance is relatively poor with the endoscopic-based examination method, and there is the risk of intestinal perforation due to surgery-related discomfort [170]. The radiologist visually inspects the intestine layers for pathological wall thickening, increased T2 intestinal wall signal, and abnormal peristalsis. Rimola Jordi et al. [170] developed a MaRIA score system, which first extracts features through manual annotations by radiologists, and then fits them into the regression model. The MaRIA score strongly correlates with CDEIS, which provides an evaluation basis for AI-based quantitative imaging. [140] used learning-based random forests to segment diseased intestines, which is the first time that image analysis support has been applied to bowel MRI. However, this technique requires radiologists to label and evaluate diseased bowel segments. To address this limitation, [83] developed a residual network focused on the terminal ileum to detect CD. To the best of our knowledge, no other studies have applied deep learning methods to directly diagnose IBD from MRI data.

Motion Artifacts. A fundamental consideration in any MRI imaging procedure is how to balance image resolution, signal-to-noise ratio (SNR) and acquisition time [158]. These three imaging fac-



tors are the result of a complex interaction between image structure, motion offset, MR pulse sequences and a k-space acquisition strategy [220]. The acquisition of MR data is carried out in the Fourier domain (k-space), where each data point corresponds to the frequency content of the image, and where the change in one point in k-space has a substantial effect on the entire image. MRI sequences must therefore take a significant amount of time to collect enough data for reconstructing an image of high quality, which may cause discomfort for the patient. As previously outlined, further deviations in K-space occur when a subject moves during MRI acquisition, which can lead to corruption, blurring and ghosting of the image domain [63]. MRI is therefore always particularly sensitive to subject motion compared to other imaging modalities such as ultrasound or computed tomography [206]. As a result, MRI artifacts due to prolonged scanning, apart from a part originating from physical imaging and hardware limitations [189], another major part is the physiological body motion. According to various patterns of body motion (rigid/non-rigid), associated imaging regions (voluntary and active/involuntary and passive) and the processing scheme (predictability/unpredictability), MRI motion artifacts can be divided into the following patterns: Pseudo-periodic motion, rigid body motion, random non-rigid motion, flow motion.

1. Pseudo-periodic motion produces ghost or replication artifacts at abnormal but predictable locations, with an approximately constant frequency of occurrence. It is common to observe pseudo-periodic movements in cardiac pulsations, thoracic organs, as well as respiratory movements.

Pseudo-periodic motion can be observed in cardiac pulsations, thoracic organs, and respiratory movements. As the heart and chest oscillate periodically, certain object structures recur throughout a long scan of an image. As a result, overlapping artifacts (ghosting), blurring, or localised signal changes may appear in the reconstructed images, depending on the pattern and sequence of motion. Anomalies caused by overlapping artifacts tend to remain at predictable spatial locations, and their characteristic of repeating at a relatively constant frequency is similar to a Fourier transform result. Periodic signals contain additional peaks accumulated during the Fourier transform in addition to the DC peaks that represent the integration of the signal. A sequence of sequential periodic movements at different rates in K-space will result in overlapping shadows of varying saliency with interleaved K-space ordering.

2. Rigid body motion refers to an unconstrained translation and rotation within a volume, resulting in motion artifacts that can be described by considering up to six parameters. Rigid body movements can be recognized in sudden involuntary movements, unconstrained conscious movements, and movements of the fetus.

Rigid body motion consists of one-dimensional translation/rotation (one parameter for mathematical representation), multidimensional translation/rotation (multiple parameters), and completely unconstrained rigid body motion (six parameters). A motion of this kind will result in completely unconstrained cross-directional motion artifacts in the plane, or volume range. As a result of the prolonged scanning, on the one hand, there are involuntary movements of the subject's muscles and organs, most notably in the chest and heart muscles. On the other hand, sudden involuntary movements, unconstrained conscious movements, and movements at the level of the limb trunk such as fetal movements are also capable of causing unpredictable corrupted artifacts during the imaging phase. Tremor can occur in patients affected by neurodegenerative diseases, such as Parkinson's disease. There are more movements in subjects who are disobedient, such as children. Although there is a partial overlap between the imaging causes of rigid and periodic movements, there are significant differences in processing methods and artifacts observed. For example, diaphragmatic movements are sometimes approximated by one-dimensional translations, but involuntary head movements typically require six degrees of freedom, including three translations and three rotations. Such movements changes in the phase or amplitude of the raw data and in the K-space position which can affect the quality of the entire image. In such cases, motion correction based on motion patterns, contrast encoding and sampling schemes can correct rigid body movements. As a result, it is important to examine the relationship between the change in pose in the spatial domain and the change in K-space data acquired in this position.

According to the properties of the Fourier transform, each pixel in an MR image is a weighted sum of all points in k-space. The coordinates of the original data points in k-space and the order in which they were acquired consecutively are defined as sampling schemes or trajectories.

Sampling is commonly conducted in Cartesian coordinates, and other types of trajectory are also can be utilised, including spirals and radials [79]. According to the motion pattern, contrast encoding, and sampling scheme, variations in the phase or amplitude of the raw data and k-space position can negatively affect the entire image. Therefore, it is essential to establish a correlation between the variation of position in the spatial domain and the resultant variation in k-space data acquired in this posture. According to the Fourier shift theorem, object displacement results in a linear phase slope in the direction of motion for the k-space data. However, object rotation can lead to the same rotation of the k-space data [20, 78, 79, 152].

3. Random non-rigid motion is characterized by deformation in all directions with up to 12 degrees of freedom [77], resulting in a variety of damaging motion artifacts in the inter-scan and inter-image. Random non-rigid motion occurs most frequently during peristalsis and in the movement of abdominal organs.

Artifacts caused by random non-rigid motion, in addition to the damage to spatial characteristics caused by rigid body motion across directions, include stretching, compression and shearing along all three axes, with different locations producing inconsistent displacements and deformations. A physiologically induced motion occurs mainly during involuntary bowel movements as well as in different organs throughout the abdomen, causing motion in all directions with up to 12 degrees of freedom. Patient and physiological motion can cause ghosting and/or blurring of image content along the phase-encoding direction. The representation in the image depends primarily on the imaging sequence and the sampling trajectory. However, such motion is inhomogeneous and locally distorted. Motion-induced phase changes in k-space can lead to destructive interference and signal loss because k-space typically contains less structured signals for reference. A further concern is that ghosting and blurring artifacts in the reconstructed image may interfere with the registration and super-resolution in the case of highly undersampled data during the mapping process. To construct reliable non-rigid motion fields, it is critical to resolve motion corrupted data under undersampling.

4. Flow motion causes abnormal signal changes in vessels and imaging areas, resulting in motion artifacts characterized by velocity vector fields with less than three degrees of freedom [232]. Artifacts caused by flow motion are frequently observed in cardiac aortic as well as the cervical and head medullas.

Flow motion flow causes unsaturated magnetisation of the signal as it penetrates the plane and the flow will cause the proton population to be out of phase to varying degrees, resulting in flow artifacts due to signal attenuation. This imaging process also causes abnormal increases and decreases in signal in the vessel and imaging area, resulting in pulsation artifacts. Alternatively, when the scan is perpendicular to the flow direction, varying degrees of signal flow null/inflow enhancement effects can be seen, manifesting as localised low signal/high signal artifacts. For a one-dimensional flow definition, such as laminar flow in a small vessel or cerebral spinal fluid

flow in the neck, velocity is the only parameter necessary to fully describe the flow. More complex flows can be observed in the cerebrospinal fluid, heart and aorta. This requires a two- or three-dimensional velocity vector field for a complete description. Despite containing a trajectory-time function for each particle within the fluid, the displacement field is often not accurate enough for motion measurement purposes.

At present, there is no indication that hardware improvements will be able to solve the subject motion problem during MRI scanning [232]. It is therefore crucial for the formulation and design of post-processing correction solutions to analyze motion patterns and artifact origins. The next step is to investigate the causes and imaging results of various motion artifacts. We will then summarize in the future the key motion artefact model-based and data driven-based correction algorithms that are currently in use or under development. Discussion of strengths, weaknesses, and how they relate to one another can provide useful insight into the design of new methods at the functional, application, and limitation.

Super Resolution for Abdominal MRI. There is consensus [154] that super-resolution in MRI is not achievable in-plane, nor in true 3D acquisitions. MRI have limited resolution due to physical limitations [22, 158] and physiology movement [112, 232], e.g., hardware device, imaging time, desired signal-to-noise ratio (SNR) and body motion. Increasing spatial resolution of magnetic resonance (MR) images typically reduces image SNR and increases imaging time [246], which limits accurate diagnoses. Therefore, Super-Resolution (SR), as a process of generating High-Resolution (HR) images from single or multiple Low-Resolution (LR) images, is capable of reconstructing high-quality MR images, especially for eliminating artifacts from physiological motion and complex intestinal structure. In addition to large slice distances and motion artifacts in MRI, bowel images are characterized by the inherent complexity of anatomical structures and significant individual differences. Anisotropic intestinal data pose a challenge for SR, and the current state-of-the-art methods cannot generate high-quality results for 3D bowel data.

Resolution enhancement often relies on traditional statistical methods, such as interpolation [188], regularization iteration [160], and compressed sensing algorithms [52]. However, such methods are limited by the type of medical data, and their performance is unstable on images with rich structural details [121]. With the development of deep learning, [216] discusses the feasibility of deep learning-based SR methods, where CNNs can model various complex structures while maintaining a high quality of details. Dong et al. introduced SRCNN, which inspired many subsequent methods with applications in medical image analysis [47]. Anand et al. proposed a discrete cosine transform-based image fusion as an SR framework for distinguishing brain tumors [44]. Compared with the motion artifacts of fetal brain [4], cardiac [199], placenta [3] and other domains [105] in medical images, the SR process of intestinal images is greatly affected by complex structural details and significant individual differences.

With generative adversarial networks (GAN) [65], particularly CycleGAN [250], has demonstrated

outstanding performance on medical image analysis [103, 138]. CycleGAN encourages more accurate distribution with the unbalanced optimization [230], and being more efficient and flexible than GAN models, which is vital when dealing with ill-posed/ corrupted data [186]. However, cycle-based frameworks have not been extensively applied to MRI 3D reconstruction challenges due to unstable training processes [132]. Our method utilized with a deep cycle consistency framework, and addresses the problem by designing two pair encoders and decoders with residual connection for optimization and stability. MoCoSR outperforms statistical-based and learning-based SR methods, and achieves stable performance on intestinal images with rich structural details.

Motion Correction for Abdominal MRI. Abdominal MRI supports 3D image reconstruction by acquiring stacks of parallel 2D images [50, 176]. Abdominal imaging sequences typically consist of axial, coronal, and sagittal, with each frame sampled from the upper abdomen to the lower abdomen [83]. These acquired images generally only have a high temporal resolution, high signal-to-noise ratio, and high contrast between soft tissue organs in a single stack of two-dimensional slices, while having lower resolution and signal-to-noise ratio in the other two planes orthogonal to the main plane as shown in Fig. 3.1. Furthermore, each image sequence requires a long scan duration, which contains corrupted regions with involuntary non-rigid organ motion [37] and random rigid body motion [139]. 2D abdominal motion correction only considers in-plane and in-image motion and does not provide inter-slice corruption on the 3D motion for the bowel. This may lead to inaccurate assessment of abdominal function, and biased diagnosis [1, 202]. Therefore, high-quality abdominal 3D MRI reconstruction in the X, Y and Z directions with high resolution, and motion correction for the structure and texture are important.

Different anatomical scan views usually contain complementary information and the combined multiple planes can be more descriptive than a single view. [93] designed a hybrid global-local optimization method for brain images, and investigated correlations between motion correction and different interpolation methods. [212] proposed a latent optimisation framework that jointly performs motion correction and SR for cardiac image segmentation. [184] designed a 3D LSTM network to utilize both local and non-local temporal information in the 4D dynamic cardiac image data for motion detection. [114] developed a flexible HR 4D MRI reconstruction method for periodic motion. Most learning-based motion correction methods are used for brain and cardiac MRI, aided by periodic and predictable movement patterns. There are few methods for abdominal organs, due to the complex and freely moving anatomy of the gut.

There is a significant amount research is currently being conducted on MRI quality enhancement and motion correction tasks using 2D models for 3D reconstruction [176]. The 3D volume images are sliced into multiple 2D image stacks that are individually processed and then combined into a 3D volume output (slice-to-volume) [4]. However, the 2D-based model limits the acquisition of prior information, such as structural information originating from axes parallel to the slices. Therefore, it is challenging for the model to determine whether details are noise or are the structure of the image

when processing each 2D slice [28]. Utilizing a full 3D model alleviates the problem and provides superiority over 2D models [28]. There are a number of challenges associated with 3D models, particularly in terms of stable training and cost. The complexity of 3D data requires expensive memory and parameters, leading to higher computational and time costs.

As opposed to 2D-based, slice-to-volume, ours is 3D-based and extracts more spatial structural information from 3D to learn the mapping between image resolutions. Ours combines prior knowledge of in-plane structure from multiple views to predict anatomically structure and texture of inter-slices, which provides guidance for 3D motion correction. The proposed internal architecture of the generator discriminator not only stabilises the training process for 3D data, but our model enables efficient use to learn features with an appropriate cost and parameter scale.

3.2 Respiratory Motion Correction and Super-Resolution for 3D Abdominal MRI

3.2.1 Overview

Abdominal MRI is critical for diagnosing a wide variety of diseases. However, due to respiratory motion and other organ motions, it is challenging to obtain motion-free and isotropic MRI for clinical diagnosis. Imaging patients with inflammatory bowel disease (IBD) can be especially problematic, owing to involuntary bowel movements and difficulties with long breath-holds during acquisition. Therefore, this paper proposes a deep adversarial super-resolution (SR) reconstruction approach to address the problem of multi-task degradation by utilizing cycle consistency in a staged reconstruction model. We leverage a low-resolution (LR) latent space for motion correction, followed by super-resolution reconstruction, compensating for imaging artefacts caused by respiratory motion and spontaneous bowel movements. This alleviates the need for semantic knowledge about the intestines and paired data. Both are examined through variations of our proposed approach and we compare them to conventional, model-based, and learning-based MC and SR methods. Learned image reconstruction approaches are believed to occasionally hide disease signs. We investigate this hypothesis by evaluating a downstream task, automatically scoring IBD in the area of the terminal ileum on the reconstructed images and show evidence that our method does not suffer a synthetic domain bias.

3.2.2 Introduction

Inflammatory bowel disease (IBD) is a relatively common, but easily overlooked disease. Its insidious clinical presentation [60] can lead to a long delay between the initial causative event and the diagnosis [179]. One of its manifestations, Crohn’s disease, often exhibits symptoms such as abdominal pain, diarrhoea, fatigue, and cramping pain, which can be accompanied by severe complications [174]. Although Crohn’s disease cannot be completely cured, early diagnosis can significantly reduce treatment costs and permanent physical damage [201]. MRI plays a crucial role in diagnosing and monitoring Crohn’s disease. In clinical applications and research, high-resolution (HR) MRI is often preferred over endoscopy as it is non-invasive and visualises more details for the small bowel. MRI is preferred over computed tomography (CT) imaging as it does not use radiation, which is an important consideration in younger patients. Unfortunately, MR acquisition for patients with Crohn’s disease can easily become compromised by respiratory motion.

As a result, many patients’ images are degraded by respiration, involuntary movements and peristalsis. Furthermore, due to technical limitations, it is difficult to acquire HR images in all scan orientations. This limits the assessment of the complete volume in 3D. Given these problems, we aim to develop a novel method that can perform both motion correction (MC) and super-resolution (SR) to improve the quality of 3D IBD MRI and to support accurate interpretation and diagnosis.

Motion can cause multiple issues for MR acquisition. Abdominal MRI scans are usually 2D multi-slice acquisitions [50]. As a result, 3D bowel motion can lead to intra- and inter-plane corruptions [232], e.g., slice misregistration, slice profile effects, and anisotropic spatial resolution. SR can be used to enhance these scans, but conventional methods often struggle with this type of anisotropic data or may unintentionally hide significant imaging findings.

Despite these challenges, MC and SR are crucial because corrupted MR images can lead to inaccurate interpretation and diagnosis [1]. Manual correction or enhancement of these volumes is not feasible.

3.2.3 Degradation Models and Assumptions

The reconstruction problem addressed in this chapter involves multiple, interacting sources of image degradation that arise naturally in abdominal MRI. Unlike conventional super-resolution settings that assume a single, well-defined degradation operator, the observed data in this work are affected by a combination of resolution loss, motion corruption, and acquisition-related artifacts. Clarifying these degradation mechanisms is essential for understanding both the design choices and the scope of the proposed MoCoSR framework.

First, abdominal MRI data are inherently anisotropic due to 2D multi-slice acquisition protocols. Limited through-plane resolution and slice thickness introduce resolution degradation that cannot be adequately modeled by simple isotropic downsampling. This motivates the formulation of super-resolution as a three-dimensional reconstruction problem rather than a purely in-plane enhancement task.

Second, respiratory motion, bowel peristalsis, and involuntary patient movement introduce non-rigid and pseudo-periodic motion during acquisition. These effects lead to inter-slice misalignment and spatial inconsistency across the volume. Importantly, such motion is not directly observable in the reconstructed images and cannot be reliably parameterized using explicit motion fields in clinical practice. As a result, motion corruption is treated as an unknown, latent degradation process rather than an explicitly estimated physical transformation.

Third, acquisition-related degradations such as k-space truncation, noise, and scanner-specific variations further contribute to the complexity of the observed data distribution. In practice, these effects vary across patients and imaging protocols, making it difficult to define a single forward operator that accurately captures all sources of degradation.

In this work, we model these heterogeneous degradations implicitly. Rather than assuming access to paired high-resolution and motion-free data or explicit knowledge of the motion process, we assume that degraded observations and quality images are drawn from related but unpaired distributions. The proposed framework therefore operates under a weak supervision regime, where degradation is addressed through latent-space regularization and cycle consistency constraints,

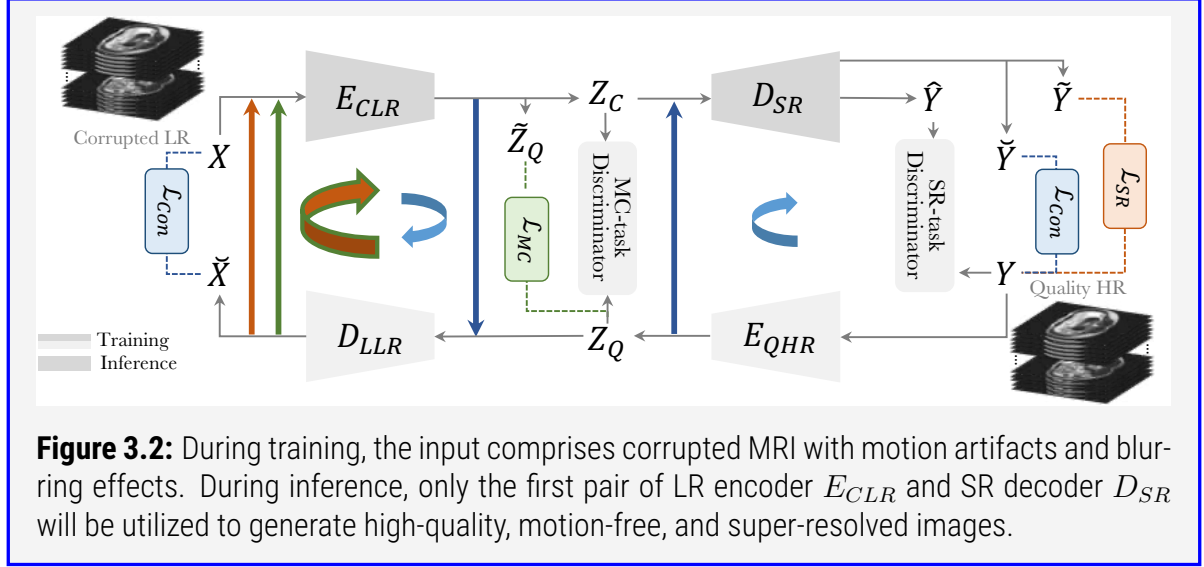
instead of explicit physical modeling.

Contribution: Our method (MoCoSR) alleviates the need for semantic knowledge and manual paired-annotation of individual structures and the requirement for acquiring multiple image stacks from different orientations, *e.g.*, [4].

There are several methodological contributions of our work: (1) First, to account for non-isotropic voxel sizes of abdominal images, we reconstruct spatial resolution from corrupted bowel MR images by enforcing cycle consistency. (2) Second, volumes are corrected by incorporating latent features in the LR domain. The complementary spatial information from unpaired quality images is exploited via cycle regularisation to provide an explicit constraint. Third, we conduct extensive evaluations on 200 subjects from a UK Crohn’s disease study, and a public abdominal MRI dataset with realistic respiratory motion. (3) Experimental evaluation and analysis show that our MoCoSR is able to generate high-quality MR images and performs favourably against other, alternative methods. Furthermore, we explore confidence in the generated data and improvements to the diagnostic process. (4) Experiments with existing models for predicting the degree of small bowel inflammation in Crohn’s disease patients show that MoCoSR can retain diagnostically relevant features and maintain the original HR feature distribution for downstream image analysis tasks.

Related Work: MRI SR. For learning-based MRI super-resolution, [216] discussed the feasibility of learning-based SR methods, where encoder-decoder methods [53, 54, 125] are commonly used to model a variety of complex structures while preserving details. Most single-forward [27, 61, 178] and adversarial methods [45, 244] rely on paired data to learn the mapping and degradation processes, which is not acceptable in real-world scenarios where data are mismatched. [130] utilizes cyclic consistency structures to address unpaired degradation adaptation in brain SR, however abdominal data would be more complicated and suffer from motion corruption. Joint optimization of MC and SR remains challenging because of the high-dimensionality of HR image space, and LR latent space has been introduced in order to alleviate this issue. Recent studies on SR joint with other tasks (*e.g.*, reconstruction, denoising) have demonstrated improvements in the LR space [54, 136, 212]. For this purpose, we utilize a cycle consistency framework to handle unpaired data and joint tasks.

Automated evaluation of IBD. In the field of machine learning and gastrointestinal disease, [140] used random forests to segment diseased intestines, which is the first time that image analysis support has been applied to bowel MRI. However, this technique requires radiologists to label and evaluate diseased bowel segments, and patients’ scan times are long. In [83] residual networks focused on the terminal ileum to detect Crohn’s disease. In this case, quality reconstruction data is extremely important for the detection of relevant structures.



3.2.4 Method

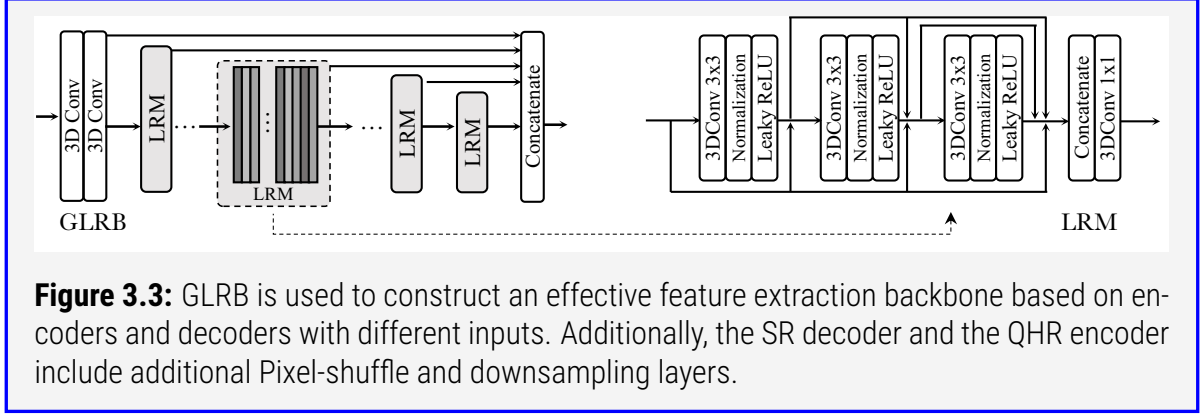
Problem formulation and preliminaries: In 3D SR, the degradation process is modeled with: $I_{LR} = \mathcal{D}(I_{HR}; k_I, \downarrow_s) + n$, $\mathcal{D}()$ represents the downsampling with the blur kernel k_I , scaling factor s , and noise n . In this work, we propose the motion corruption term M , which operates on LR latent space. And our MC-SR model can be refined to

$$I_{LR} = M(\mathcal{D}(I_{HR}; k_I, \downarrow_s); Z) + n \quad (3.1)$$

MoCoSR concept: As shown in Fig. 3.2, our multi-task framework consists of three parts: a pair of corrupted LR (CLR) encoder and SR decoder on corrupted LR input, a pair of quality SR (QLR) encoder and learned LR (LLR) decoder on HR input and two task-specific discriminators. For the first two pairs, individual features are extracted and scaling is applied to provide our network with the ability to handle multiple tasks at different scales. The discriminators are then used to identify features for each scale.

A quality HR volume Y is fed first into the QHR encoder in order to obtain a quality latent feature map Z_Q upon which the corrupted LR can be trained. due to the unpaired scenario. For MC, Z_Q passes to LLR and CLR to gain $\tilde{Z}_Q$ with $\mathcal{L}_{MC}$, where LLR learns the corruption feature and degradation, and CLR encodes the multi-degradation from the corrupted LR input X . For SR, $\tilde{Y}$ is then generated after the SR decoder and ensure the SR purpose with $\mathcal{L}_{SR}$. To ensure training stability, two consistency loss functions $\mathcal{L}_{Con}$ are utilized for each resolution space. The arrows indicate the flow of the tasks. A pair of task-specific discriminators is used to improve the performance of each task-related encoder and decoder.

Loss functions: Rather than aiming to reconstruct motion-free and HR images in high dimensional space with paired data, we propose to regularize in low dimensional latent space to obtain a high



quality LR feature that can be used for upscaling. A $\mathcal{L}_{MC}$ between $\tilde{Z}_Q$ downsampled from QHR and Z_Q cycled after LLR and CLR, defines in an unpaired manner as follows:

$$\mathcal{L}_{MC} = E \left[\left\| Z_Q - \tilde{Z}_Q \right\| \right] \quad (3.2)$$

As the ultimate goal, SR focuses on the recovery and upscaling of detailed high-frequency feature, $\mathcal{L}_{SR}$ is proposed to optimize the reconstruction capability by passing through cycles at various spaces:

$$\mathcal{L}_{SR} = \mathcal{L}_1(Y, \tilde{Y}) \quad (3.3)$$

The dual adversarial $\mathcal{L}_{DAdv}$ is applied to improve the generation capability of the two sets of single generating processes in the cyclic network:

$$\mathcal{L}_{DAdv} = \mathcal{L}_{adv}(\check{X}) + \mathcal{L}_{adv}(\check{Y}) + \mathcal{L}_{adv}(\check{\tilde{Y}}) \quad (3.4)$$

The corresponding two task-specific discriminators $\mathcal{L}_{DMC}$ and $\mathcal{L}_{DSR}$ for discriminating between corrupted and quality images followed are used for the purpose of staged reconstruction Z_Q and $\hat{Y}$ of MC at latent space and SR at spatial space, respectively. Furthermore, a joint cycle consistency loss is used to improve the stability of training in both spaces:

$$\mathcal{L}_{Con} = \mathcal{L}_1(X, \check{X}) + \mathcal{L}_1(Y, \check{Y}) \quad (3.5)$$

For MoCoSR generators, the joint loss function is finally defined as follows:

$$\mathcal{L} = \mathcal{L}_{SR} + \lambda_1 \mathcal{L}_{MC} + \lambda_2 \mathcal{L}_{DAdv} + \lambda_3 \mathcal{L}_{Con} \quad (3.6)$$

Network architecture: In the paired encoder-decoder structure, we developed a 3D Global and Local Residual Blocks (GLRB) with Local Residual Modules (LRM) in Fig. 3.3 based on [76]. The GLRB is designed to extract local features and global structural information at 3D level, and then construct blocks for multi-scale features connected to the output of the first layer. The residual output is then added to the input using a residual connection to obtain a staged output. The model implements the extraction of local features while integrating all previous features through the connected blocks and compression layers. The decoders are employed with the upsampling prior to the convolution layers.

3.2.5 Experiments

In this chapter, the degradation models are designed to reflect common acquisition constraints and motion-related artifacts encountered in routine abdominal MRI, rather than to exactly simulate a specific scanner or pulse sequence. In particular, we distinguish between a reference-quality image, denoted as I_{HR} , and a clinically realistic low-quality observation, denoted as I_{LR} .

The reference images I_{HR} correspond to high-quality volumes with relatively high in-plane resolution and minimal motion, which may be obtained under controlled conditions such as short breath-holds, repeated acquisitions, or careful retrospective selection. In contrast, the degraded observations I_{LR} are intended to represent typical free-breathing 2D multi-slice abdominal MRI acquisitions, which commonly exhibit anisotropic resolution, inter-slice misalignment, and motion-induced artifacts due to respiration, peristalsis, and involuntary patient motion.

In our experiments, spatial resolution degradation is introduced through k-space truncation and downsampling to mimic limited spatial encoding and slice thickness commonly used in clinical protocols. This abstraction reflects the practical trade-off between spatial resolution and acquisition time, rather than a specific vendor-dependent reconstruction pipeline. Additive noise is applied to approximate thermal and system noise in magnitude MR images. Importantly, the resulting low-resolution volumes are anisotropic and exhibit reduced high-frequency content, consistent with real-world abdominal MRI acquisitions.

To account for respiratory and bowel motion, we further introduce a pseudo-periodic multi-factor respiratory motion (PMRM) model. Rather than explicitly modeling physical organ deformation, the PMRM model captures the dominant effects of motion as inter-slice misregistration and slice-wise spatial shifts. This abstraction reflects the fact that abdominal MRI is often acquired as sequential 2D slices, where through-plane motion leads to discontinuities and inconsistencies in the reconstructed 3D volume. The motion operator $M(\cdot)$ in Eq. (2.1) therefore represents a task-oriented latent-space corruption that aggregates multiple motion sources, including respiration and involuntary bowel movement, without assuming access to explicit motion trajectories or navigator signals.

Data degradation: We use $64 \times 64 \times 64$ patches. For downsampling and the degradation asso-

Table 3.1: IBD data set: the upper shows MRI data acquisition parameters. The lower shows the number of patients in each severity level of inflammation.

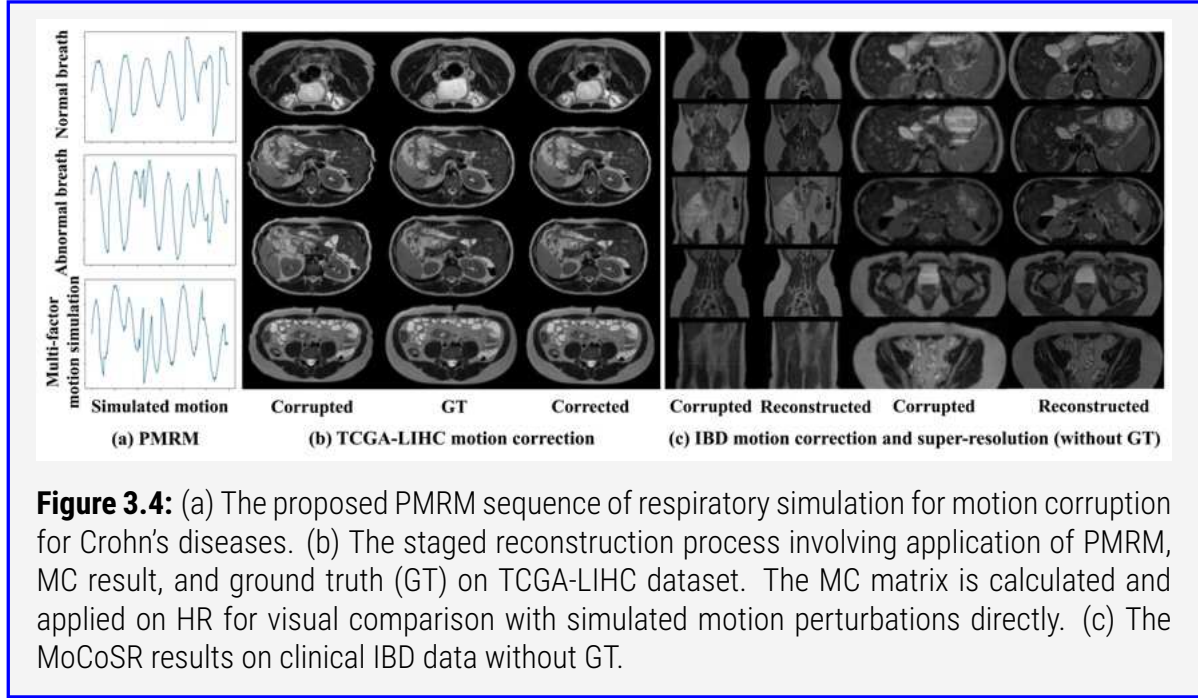
Planes	Sequence	TR [ms]/TE [ms]	Matrix	Slice [mm]	FOV	Time [s]
Axial	T1 FFE (e-THRIVE)	5.9/3.4	512x512x96	3.00	375	20.7x2
Axial Postcon	single-shot T2 TSE	587/120	528x528x72	3.50	375	22.3x2
Coronal	single-shot T2 TSE	554/120	512x512x34	3.00	375	21.1
Inflammation class		Healthy	Mild	Moderate	Severe	
Number of patients		100	49	42	9	

ciated with MRI scanning, (1) Gaussian noise with a standard deviation of 0.25 was added to the image. (2) Truncation at the k-space, retaining only the central region of the data. On top of this, we have developed a motion simulation process to represent the pseudo-periodic multi-factor respiratory motion (PMRM) that occurs during scanning, as shown in Fig. 3.4(a). The simulated motion includes the influence of environmental factors that cause the respiratory intensity and frequency to fluctuate within a certain range. This lead to the presence of inter-slice misalignment in image domains.

TCGA-LIHC data set, abdominal MRI: A total of 237 MR exams are collected from The Cancer Imaging Archive Liver Hepatocellular Carcinoma (TCGA-LIHC). The data contains 908 MRI series from 97 patients. We applied simulated motion to TCGA MRI-Abdomen series to generate the motion-corrupted dataset with respiration-induced shift.

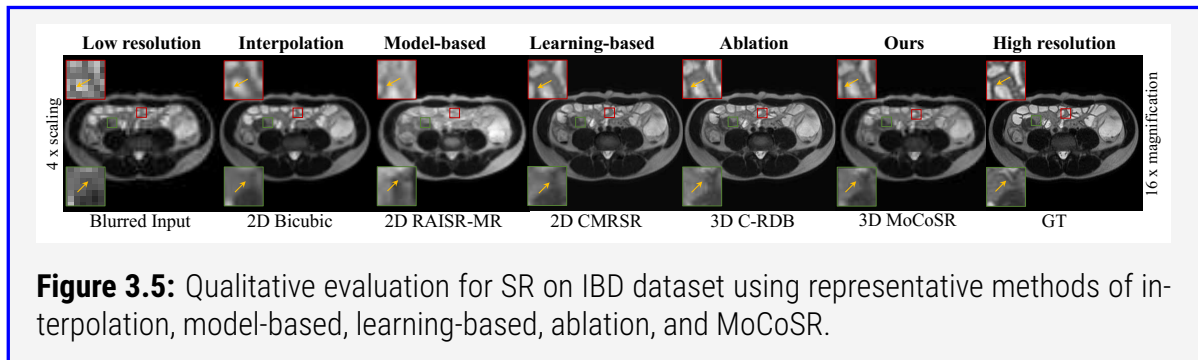
IBD data set, inflammatory bowel disease: MRI sequences obtained include axial T2 images, coronal T2 images and axial postcontrast MRI data on a Philips Achieva 1.5 T MR scanner. Abdominal 2D-acquired images exhibit motion shifts between slices and fibrillation artefacts due to the difficulty of holding one’s breath/body movement and suppressing random organ motion for extended periods. The dataset contains 200 available sample cases with four classes, healthy, mild, moderate, and severe small bowel Crohn’s disease inflammation as shown in Table 3.1. The abnormal Crohn’s disease sample cases, which could contain more than one segment of terminal ileum and small bowel Crohn’s disease, were defined based on the review of clinical endoscopic, histological, and radiological images and by unanimous review by the same two radiologists (this criterion has been used in the recent METRIC trial investigating imaging in Crohn’s disease [200]).

Setting: We compare with interpolation of bicubic and bilinear techniques, rapid and accurate MR image SR (RAISR-MR) with hashing-based learning [172], which we use as the representative of the model-based methods, and published MRI SR with various generative adversarial networks (MRESR [45], CMRSR [244]) as the learning-based representatives. For training, with train/val/test with 70%/10%/20%. Various methods were included in the quantitative evaluation, including single forward WGAN (S-WGAN) and cycle RDB (C-RDB) for ablation experiments on the cycle consistency



framework and the GLRB setting.

Results: The quantitative results are presented in Table 3.2 on TCGA-LIHC and IBD data sets. There is a performance gap between interpolation and GAN-based methods, and the SR method based on learning has a significant mapping advantage for complex gastrointestinal images. MoCoSR achieves the best performance among all evaluation metrics. CMRSR and MRESR cannot guarantee the output quality of mapping from HR back to LR, resulting in poor performance on complex 3D bowel data. Representatives for the qualitative evaluation are shown in Fig. 3.5.

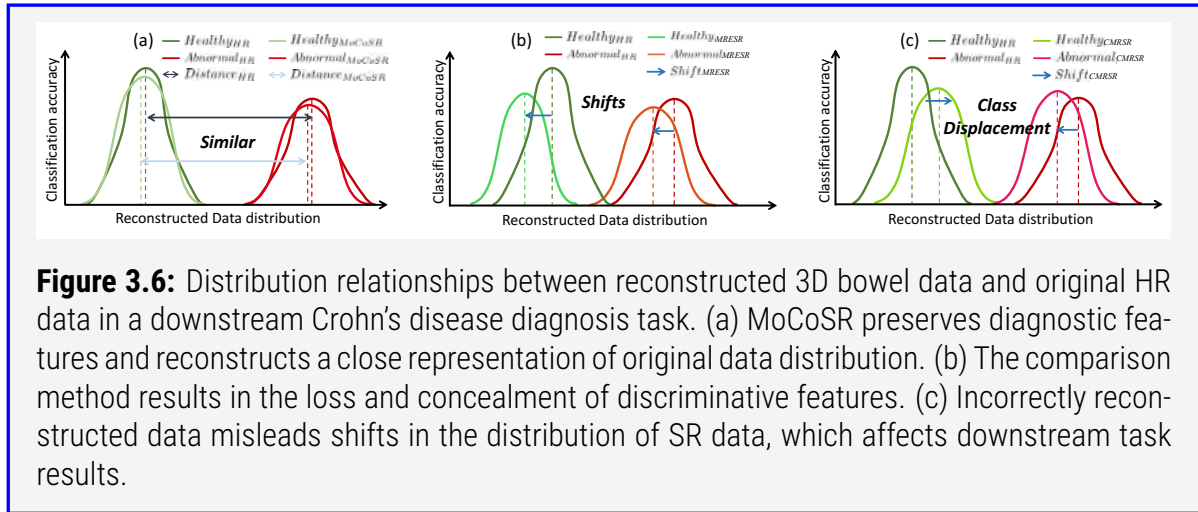


Sensitivity analysis: We evaluate if our method influences downstream analysis using the example of automatic scoring of Crohn's disease with existing deep networks. If features defining Crohn's disease are hidden by the method, this would affect disease scoring. We use a classifier with an attention mechanism similar to [83], trained on HR raw data. Our evaluation is based on the average possibility of normal and abnormal small bowel inflammation on MRI. The degree of

Table 3.2: Quantitative evaluation of extensive methods on TCGA-LIHC and IBD data sets in terms of SSIM and PSNR for 3D abdominal MRI SR

Method	TCGA-LIHC upper abdominal		IBD lower abdominal	
	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$
Bicubic	0.8166	28.53	0.6582	25.30
Bilinear	0.8219	29.15	0.6749	26.27
RAISR-MR [172]	0.8940	32.24	0.8512	31.53
MRESR [45]	0.9042	33.74	0.8529	30.60
CMRSR [244]	0.9185	35.92	0.8662	32.90
S-WGAN	0.8521	30.91	0.7313	28.19
C-RDB	0.9102	35.15	0.8624	32.78
MoCoSR (C-GLRB)	0.9257	35.67	0.8719	33.56

small bowel inflammation on abnormal MRIs was classified by Radiologists as mild, moderate or severe. This outcome was compared against the results of the data constructed from different SR methods.



Complete results including LR degraded image, SR image reconstructed by MRESR, CMRSR, and MoCoSR, are shown in Table 3.3. We tracked and quantified the changes by performing a significance evaluation (t-test) based on p-values < 0.05 . The ideal SR data can achieve classification results as close as possible to HR data with lower requirements. Our method obtains similar small-scale attenuation results on both healthy and abnormal samples. The p-value is larger than 0.05 for MoCoSR, i.e., there is no statistically significant difference between the original and reconstructed data for the prediction results. The results of MRESR are volatile but present an unexpected improvement on healthy samples. CMRSR makes the predicted probability much lower than that of HR.

Discussion: According to the sensitivity analysis and comparison results, our MoCoSR method

Table 3.3: Crohn’s disease classification performance on different 3D data. MoCoSR has a negligible effect on the downstream classification task as shown by high p-values in contrast to LR, MRESR, and CMRSR which produce significantly lower performance.

Data	Classification	Changes in/ Prediction	Significance p-value	CI 95% (Lower,Upper)
HR	Healthy Abnormal	0.784 0.730	N/A	N/A
LR	Healthy Abnormal	-0.19*↓↓ -0.21*↓↓	<0.001 <0.001	(0.13, 0.24) (0.14, 0.25)
MRESR [45]	Healthy Abnormal	-0.11*↓↓ -0.10*↓↓	<0.001 0.002	(0.05, 0.16) (0.04, 0.15)
CMRSR [244]	Healthy Abnormal	+0.03↑ -0.08*↓	0.370 0.012	(-0.08, 0.03) (0.02, 0.13)
MoCoSR (ours)	Healthy Abnormal	-0.01↓ -0.02↓	0.630 0.561	(-0.04, 0.06) (-0.04, 0.07)

shows superior results compared to the forward adversarial reconstruction algorithms and encoder-decoder structures. Combining multi-scale image information in the feature space of different resolution image domains yields better results than inter-domain integration. The cycle consistency network splits the different resolution spaces and latent space, which facilitates the flexibility of the neural network to customize the MC according to the specific purpose and ensures consistency of the corrected data with the unpaired data. Furthermore, although these methods can obtain acceptable SSIM and PSNR, the key features used by the classifier for downstream tasks are potentially lost during the reconstructions. Conversely, the reconstruction result will implicitly cause domain shift. This leads to a distribution shift in the samples, which makes the disease prediction biased as shown in Fig. 3.6. The data generated by ours can reconstruct the results, retain likely all of the diagnostically valuable features, and maintain the original data distribution. The present sensitivity study is limited to the automatic classification from single domain and down-stream task framework, and future extensions will explore model-based and learning segmentation tasks across data domains and acquisitions.

3.2.6 Summary

MoCoSR is a DL-based approach to reconstruct high-quality SR MRI. MoCoSR is evaluated extensively and compared to the various image SR reconstruction algorithms on a public abdominal dataset, simulating different degrees of respiratory motion, and an IBD dataset with inherent motion. MoCoSR demonstrated superior performance. To test if our learned reconstruction preserves clinically relevant features, we tested on a downstream disease scoring method and found no decrease in disease prediction performance with MoCoSR.

Chapter 4

Test-time-Adapted Quantitative Body MRI Super-resolution

Recent advances in deep learning-based super-resolution (SR) techniques have significantly improved MRI quality, particularly for inflammatory bowel disease (IBD) and Crohn's disease, where early and accurate diagnosis is essential. However, as discussed in previous section, SR models often struggle with domain shifts caused by variations in acquisition protocols, imaging devices, and patient-specific motion artifacts. These challenges are further compounded by the reliance on oversimplified degradation models that fail to capture the complex physical processes of MRI acquisition.

To address these limitations, we propose UTAR, an unsupervised test-time adaptation framework that enhances SR model generalization across unseen target domains. UTAR integrates a physics-informed imaging chain degradation model (PIC) and a conditional consistency learning (CCL) strategy, enabling adaptation without access to source-domain data. By leveraging dual-domain degradation modeling and unsupervised adaptation, UTAR mitigates domain shifts, ensuring the preservation of diagnostically relevant features. This makes it a robust solution for deploying SR models in clinical settings, where variations across hospitals and imaging protocols are unavoidable.

4.1 Physics-informed Unsupervised Test-time Adaptation for MRI Super-Resolution

4.1.1 Overview

Deep learning-based super-resolution (SR) usually does not perform well across domains, e.g., different magnetic resonance imaging (MRI) datasets, usually due to two reasons: 1) there is often a mismatch in terms of data distributions between the training data and test data, and 2) the overly simplified image degradation models in popular SR techniques do not model the underlying physical processes well. Hence, SR models often require extensive supervised fine-tuning when transferring to new datasets (target domains). In this paper we propose UTAR, a physics-informed **U**nsupervised **T**est-time domain **A**daptation framework for deep learning-based super-resolution for magnetic **R**esonance imaging. UTAR adapts a pre-trained source-domain SR model to a new target domain (e.g., images from a new hospital) using low-resolution (LR) target-domain images only. This is possible through a new physics-inspired imaging chain degradation model (PIC), which incorporates both spatial and frequency domain (k-space) degradation. Such dual-domain degradation modeling enables us to better capture the variations of LR images that could potentially affect a model's performance across different unseen MRI domains. We further propose a conditional consistency learning (CCL) strategy to encourage prediction consistency, serving as the supervision signal for TTA. Our method is generic and can accommodate arbitrary pre-trained SR networks. Experimental results verify the effectiveness of UTAR in reducing the performance gap with minimum efforts. UTAR outperforms earlier methods based on different degradation-based strategies and existing test time adaptation (TTA) approaches. We expect our method to provide a key step towards the deployment of MRI SR algorithms in clinical applications where significant domain shifts are inevitable.

Why test-time adaptation is necessary for MRI super-resolution. In clinical MRI super-resolution, deployment conditions often differ substantially from the data distributions encountered during training. In contrast to natural image super-resolution, where degradation processes can be approximated by fixed downsampling and noise models, MRI image quality is strongly influenced by acquisition-specific factors such as pulse sequence design, k-space sampling density, reconstruction pipelines, and hardware-dependent system imperfections. These factors vary across hospitals, scanner vendors, field strengths, and clinical protocols, and are rarely fully known or reproducible at deployment time.

As a consequence, even well-trained super-resolution models may exhibit systematic performance degradation when applied to images acquired under unseen conditions. Retraining or supervised fine-tuning is typically impractical in such scenarios, as paired high-resolution ground truth data are unavailable and access to source-domain training data may be restricted due to privacy, gov-

ernance, or storage constraints. This creates a fundamental gap between research settings and real-world clinical deployment.

Test-time adaptation provides a principled mechanism to bridge this gap by allowing models to adjust their internal representations using only target-domain observations available at inference. For MRI super-resolution, this paradigm is particularly appealing because it enables adaptation without modifying acquisition protocols or requiring additional annotations. However, effective test-time adaptation requires meaningful self-supervision signals that reflect the underlying image formation process. This motivates the integration of physics-informed degradation modeling and consistency constraints, which anchor the adaptation process to acquisition-aware transformations rather than purely statistical assumptions.

4.1.2 Related Work

Learning-based super-resolution and reconstruction methods have become increasingly prominent in magnetic resonance imaging (MRI), driven by their ability to recover high-frequency details and suppress noise in ill-posed inverse problems. Early deep learning approaches to MRI super-resolution primarily relied on supervised training with paired low- and high-resolution images, often employing convolutional encoder–decoder architectures or residual networks adapted from natural image super-resolution [54, 61, 125]. While these methods demonstrate strong performance on benchmark datasets, they typically assume simplified and fixed degradation models, such as isotropic blurring combined with additive Gaussian noise. Such assumptions are rarely satisfied in clinical MRI, where image quality is determined by a complex interplay of k-space sampling, pulse sequence design, hardware characteristics, and reconstruction pipelines.

To improve robustness, a line of work has focused on enhancing degradation modeling for super-resolution. Blind or semi-blind approaches attempt to estimate spatial degradation kernels directly from low-resolution images, while others introduce more diverse or anisotropic degradation patterns during training [214, 237]. Although these methods improve generalization to some extent, they remain largely image-domain driven and do not explicitly account for the physical MRI acquisition process, which fundamentally operates in the frequency domain. As a result, they struggle to generalize across scanner vendors, field strengths, pulse sequences, or sampling strategies, where domain shifts manifest not only as spatial blur but also as structured frequency-dependent distortions.

In parallel, physics-informed learning has gained traction in MRI reconstruction, where elements of the forward model—such as Fourier encoding, sampling masks, and data consistency constraints—are incorporated into neural networks to improve stability and interpretability [73, 95]. These approaches have demonstrated clear benefits for undersampled reconstruction and inverse problems with well-defined acquisition operators. However, most physics-informed methods assume access to raw k-space data and detailed acquisition parameters, and are designed for super-

vised reconstruction rather than image-space super-resolution. Their applicability to post-processing super-resolution, especially in source-free or unpaired settings, remains limited. Moreover, they typically assume a fixed forward model during inference and do not explicitly address domain shifts arising from deployment in unseen clinical environments.

Domain adaptation methods have been widely explored to mitigate performance degradation caused by distribution shifts between training and deployment data. Unsupervised domain adaptation approaches align feature representations between source and target domains using adversarial objectives or discrepancy minimization, and have been applied to medical image segmentation and reconstruction [18, 29, 49, 58]. However, these methods generally require simultaneous access to source and target data, which is often infeasible in clinical practice due to privacy constraints, data governance policies, or storage limitations. In addition, the absence of paired low- and high-resolution data across domains further complicates their application to MRI super-resolution.

More recently, test-time adaptation (TTA) has emerged as an alternative paradigm in which models adapt to target-domain data during inference without retraining on source data [15]. In medical imaging, TTA has been investigated primarily for segmentation and reconstruction tasks, using strategies such as entropy minimization, normalization statistics adaptation, or self-supervised consistency objectives [38, 101]. While promising, existing TTA approaches often rely on external priors, source-domain statistics, or task-specific assumptions, and their extension to MRI super-resolution remains under-explored. In particular, most existing methods do not explicitly model acquisition-induced domain shifts, nor do they address the frequency-domain discrepancies that are critical for high-fidelity super-resolution.

Against this background, UTAR is positioned at the intersection of physics-informed modeling and source-free test-time adaptation for MRI super-resolution. By explicitly modeling the MRI imaging chain across both frequency and spatial domains and coupling this with a conditional consistency learning strategy, UTAR addresses domain shifts arising from unseen acquisition protocols without requiring access to source-domain data or high-resolution targets at test time. This combination fills an important gap between existing degradation-based super-resolution methods and test-time adaptation frameworks, offering a principled approach to improving robustness and generalization in realistic clinical deployment scenarios.

4.1.3 Introduction

Magnetic resonance imaging (MRI) is a non-invasive imaging technique used to acquire tomographic images with various image contrasts, especially for soft tissues. The image acquisition is done in k-space, which is the discrete Fourier representation of complex-valued MR radio signals. However, the MRI acquisition is slow and, thus, the k-space is often under-sampled, *e.g.*, to reduce scan time for patient comfort and reduced motion artifacts. Spatial resolution is limited due to a

low signal-to-noise ratio.

To enhance the quality of MR images, deep learning (DL)-based MRI super-resolution (SR) methods are gaining popularity, often outperforming traditional solutions for the ill-defined inverse problem [121, 134, 136]. As a post-processing technique, SR improves the quality and resolution of lower resolution slices or volumes in image space [54, 212]. While existing approaches predominantly rely on Gaussian variants of blur and noise to model image degradation, they usually do not incorporate the MR image acquisition starting from k-space data and do not capture the wide range of possible image contrasts in MRI. Furthermore, existing MR SR or reconstruction methods often show deterioration in practice, when applied to LR images from an unknown acquisition process (domain) due to severe distribution shifts [18, 29, 38, 106].

Contributions: (1) We introduce a physics-informed unsupervised test-time domain adaptation framework for MRI SR (UTAR) that enables a pre-trained source-domain model to effectively adapt to **unseen** data from the target domain during testing, using **LR images from target domain only**. Our method is generic and can be used as a plug-in module in pre-trained SR networks. As key components of UTAR (2) we propose a physical imaging chain degradation model (PIC) for accurate modeling of the degradation process of real-world MRI in SR frameworks. (3) We further propose a conditional consistency learning (CCL) strategy during **test time**, as the supervision signal for TTA. (4) We demonstrate UTAR’s capabilities on public data (IXI brain, fastMRI knee [235]). **Related**

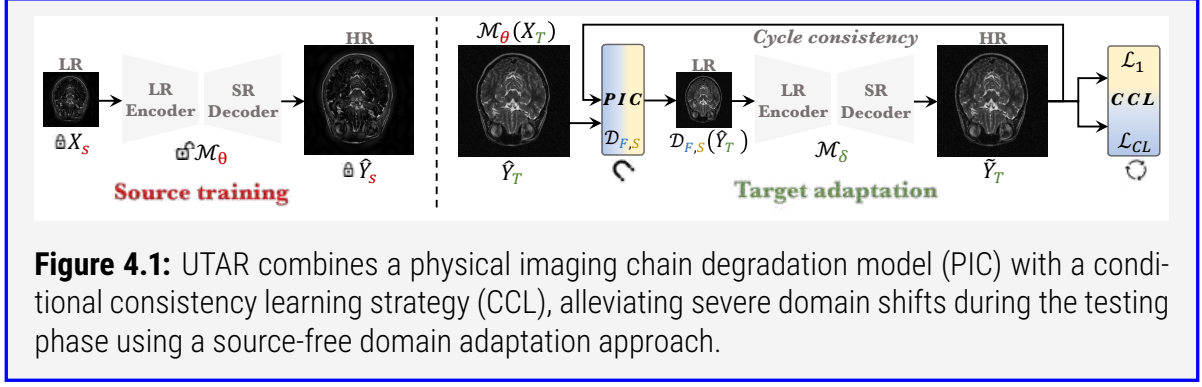
Works: Methods in this area can be divided into two categories:

Degradation modeling: SR methods based on degradation modeling combine spatial augmentation with the designed degradation pattern or conduct intermediate spatial feature transformation according to a blindly estimated kernel [126]. These methods rely on empirical estimates about the degenerating factors [134, 136] or spatial degradation assumptions [214, 237]. However, this may be insufficient for MRI since the underlying point spread function (PSF) depends on how the k-space is sampled and on the consistency within individual measurements.

Domain adaptation for SR: Conventional unsupervised domain adaptation methods require samples from the source and target domains to be available such that domain gaps can be estimated and reduced [18, 49, 58]. These approaches can be combined with a pre-trained SR model for utilizing the transferred information without target-domain HR images [29]. However, accessing source images in a new setting is often impractical due to data sharing constraints. Also, the lack of paired data between source and target domains hinders optimization through direct transfer learning and fine-tuning. These motivate test-time adaptation (TTA) [15, 38], an under-explored problem in MRI SR.

4.1.4 Method

Preliminaries: MRI SR problem settings that use artificial image degradation during training are commonly described through the relationship between the HR image y and LR image x . A degra-



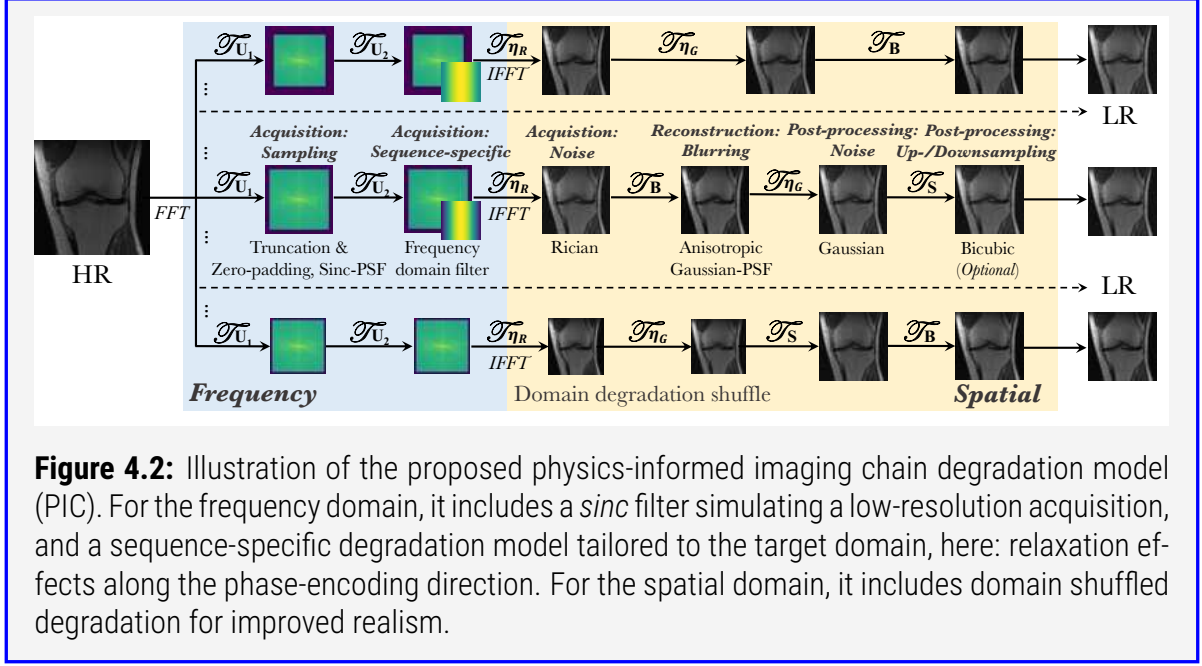
dation model in the source domain S can be defined as: $\mathbf{x}_s = (\mathbf{k}_s * \mathbf{y}_s) \downarrow + \boldsymbol{\eta}_s$, where $\mathbf{x}$ is a degraded image generated from $\mathbf{y}$ via a blurring kernel $\mathbf{k}$, the downsampling operation $\downarrow$ and additive noise perturbation $\boldsymbol{\eta}$. Here $*$ denotes the convolution operation. A pre-trained model $\mathcal{M}_\theta$ for such an ill-posed inverse problem can be affected by distribution shifts in MRI from the source domain S to the target domain T [38]. Here, we do not consider errors introduced by a poorly fitted model, but assume a generic test-time domain adaptation framework for SR without paired data. Existing methods often model adverse factors $\mathcal{T}_i$ as $\mathcal{D}_S$ at source domain that influence model performance in new domains cumulatively via blurring $\mathbf{k}$, and additive Gaussian noise $\boldsymbol{\eta}_G$:

$$\mathbf{x} = \mathcal{D}_S(\mathbf{y}) = (\mathcal{T}_G(\mathbf{k}) * \mathbf{y}) \downarrow + \mathcal{T}_{\eta_G}(\boldsymbol{\eta}_G), \quad (4.1)$$

It is common to model tractable solutions for deconvolution with a Gaussian kernel $\mathcal{G}$, mimicking a sampling scheme with respect to the chosen field of view (FOV) and given PSF of a capturing device. Consequently, a deblurring filter will degrade high-frequency components, which vary from domain to domain.

UTAR: We are interested in TTA for generically pre-trained MRI SR models, where source data is not available during adaptation. Suppose we have a source-domain pretrained model $\mathcal{M}_\theta$ parameterized by θ where training images are from source (*i.e.* s) domain distribution $\{x_s^i, y_s^i\}_{i=1}^{n_s} \in \mathcal{S}_s$. $x_s^i \in \mathcal{X}_s$ and $y_s^i \in \mathcal{Y}_s$ are LR-HR pairs. We aim to update the parameters θ of a given model $\mathcal{M}$: $\theta \rightarrow \delta$ using target domain (*i.e.* t) images only so that $\mathcal{M}_\delta$ lead to higher performance. Here target domain images $\{x_t^i\}_{i=1}^{n_t} \in \mathcal{X}_t$ are LR images only as corresponding HR counterparts $\{y_t^i\}_{i=1}^{n_t} \in \mathcal{Y}_t$ are not available, which is common in practice.

Physical imaging chain degradation model (PIC): Real-world MRI degradation is usually caused by system imperfections and short scan times that limit the amount of k-space that can be acquired. The degradation can be modelled through a PSF, which reflects the specific k-space encoding of the MR sequence. However, modeling of the entire imaging chain is necessary to account for accumulated changes to the PSF along commonly applied reconstruction and post-processing methods. Firstly, we focus on the degradation model in frequency domain (see blue part in Fig 4.2),



which can be described as

$$\mathbf{x} = \mathcal{D}_{\mathcal{F}}(\mathbf{y}) = \mathcal{F}^{-1}\{(\mathcal{F}(\mathbf{y}) \downarrow \cdot \mathcal{T}_U(H))\}, \quad (4.2)$$

with $\mathcal{D}_{\mathcal{F}}$ the frequency domain degradation model applied on 2D or 3D images, $\mathcal{F}$ and $\mathcal{F}^{-1}$ the forward and inverse Fourier transform, $H(u, v) = \mathcal{F}(\mathcal{P}(x, y))$ a k-space filter corresponding to the Fourier transform of the initial PSF, $\mathcal{T}_U$ the degradation in frequency space based on the acquisition matrix size and sequence-specific k-space weightings (e.g., relaxation blurring in turbo spin echo (TSE) sequences [161]), and $\downarrow$ the down-sampling operator corresponding to truncation of the frequency space.

Image degradation in MRI depends on the specific encoding of k-space. In this work, we focus on Cartesian sampling-based k-space for proof of concept. Limited spatial encoding causes a sinc-profile of the PSF $\mathcal{P}$ ($\mathcal{T}_{U_1}$ in Fig 4.2), shown here for a 2D acquisition example: $\mathcal{P}(x_1, x_2) = \mathcal{F}(H(u, v)) = \mathcal{F}(\Pi(u) \cdot \Pi(v)) \propto \text{sinc}(x_1) \cdot \text{sinc}(x_2)$, with $\Pi(u)$ a rectangular function with $\Pi(u) = 1$ if $u \leq u_{\max}$ and $\Pi(u) = 0$ if $u > u_{\max}$. Furthermore, sequence-specific PSF deviations usually occur along the phase-encoding direction. We apply a Gaussian weighting along the phase-encoding direction inspired by the relaxation blurring of long echo trains in TSE imaging [161] ($\mathcal{T}_{U_2}$ in Fig 4.2).

To further improve the realism of artificial degradation, a second phase of degradation is performed in image space (yellow part in Fig 4.2). Bicubic and bilinear interpolation are downsampling schemes ($\mathcal{T}_S$ in Fig 4.2) and the degradation with an anisotropic Gaussian kernel is commonly used for other SR tasks [136, 215]. This phase can pick up the PSF modifications during image reconstruction. In addition to Gaussian noise, we further add Rician noise to the magnitude images ($\mathcal{T}_{\eta_R}$

in Fig 4.2), allowing a better approximation of the noise modeling in MR magnitude images [68]. We shuffle degradation to increase complexity [237].

Conditional consistency learning (CCL): Our CCL strategy can be applied to a generic architecture consisting of a feature encoder $f_s(\cdot) : \mathcal{X}_s \rightarrow \mathbf{R}^D$ and a scaling decoder $h_s(\cdot) : \mathbf{R}^D \rightarrow \hat{\mathcal{Y}}_s$ where D and $\hat{\mathcal{Y}}_s$ are feature dimension and HR prediction. Such a pre-trained model may be capable of reconstructing a reasonable image, but there will be a notable domain gap in several frequency bands. Thus, inspired by the focal frequency loss [94, 224], we propose conditional learning in k-space. The frequency distance $\mathcal{W}$ between the frequency bands of $\hat{Y}$ predicted from $\mathcal{M}_\theta$, and the super-resolved frequency bands of $\tilde{Y}$ predicted from $\mathcal{M}_\delta$ based on a PIC-corrupted LR image, is:

$$\begin{aligned} \mathcal{W}(u, v) &= \|\mathcal{F}(\hat{Y})(u, v) - \mathcal{F}(\tilde{Y})(u, v)\|_2^\gamma \\ \text{with } \|z_1 - z_2\|_2^\gamma &= ((\Re(z_1) - \Re(z_2))^2 + (\Im(z_1) - \Im(z_2))^2)^{\gamma/2}, \end{aligned} \quad (4.3)$$

where $\Re$ and $\Im$ denote the real and imaginary parts, and γ determines the sharpness of the distance function and generally equals to 1. Furthermore, we employ an L1 Loss for the spatial reconstruction between $\hat{Y}$ predicted by $\mathcal{M}_\theta$ and $\tilde{Y}$. A total loss $\mathcal{L}_{CCL}$ can be consequently summarised as:

$$\mathcal{L}_{CCL} = \mathcal{L}_1(\hat{Y}, \tilde{Y}) + \lambda \cdot \mathcal{L}_{CL}(\mathcal{T}_s * \mathcal{F}(\hat{Y}), \mathcal{T}_s * \mathcal{F}(\tilde{Y})), \quad (4.4)$$

where λ is a weighting hyperparameter, and $\mathcal{L}_{CL}$ is defined by $\mathcal{T}$ and $\mathcal{W}$. $\mathcal{T}$ is a conditional constraint for the frequency domain focusing on the partial frequency-bands with the truncation s , corresponding to the scale of the SR.

4.1.5 Experiments

Datasets: Two public datasets focusing on different anatomy are used to investigate UTAR with respect to several domain shifts. **Brain data**¹ contains clinical T1- and T2-weighted scans from three hospitals with different imaging protocols: HH, Guys, and IOP. We selected 184 HH T2 subjects as the source-domain data [train/val/test ratio: 7:1:2], and 37 subjects each from Guys and IOP as two target-domain datasets without degradation, acquisition parameters, datasets, and patient-wise crossovers. Bicubic downsampling was applied to obtain LR data in the source domain and a kernel degradation [17] was applied in target domain. The central 60 slices were selected. **Knee data** [235] contains raw k-space data of multi coil Cartesian 2D TSE acquisitions with proton-density (PD) or PD with fat-suppression (PDFS) contrast. We selected 21 central slices of 584 subjects scanned with the PD sequence as source-domain data [ratio: 7/1/2], and 117 subjects with the PDFS sequence as target-domain data. The fully-sampled k-space data were reconstructed using either

¹<http://brain-development.org/ixi-dataset/>

the full or reduced matrix to obtain HR and LR magnitude images after sum-of-squares reconstruction.

Source training: EDSR [125] serves as a backbone for many SR models [54, 61]. We also use an EDSR backbone

(16 residual blocks, 64 feature channels). The model was trained on the source domain for 150 epochs with L1 loss and the Adam optimizer [lr: 1e-4, PyTorch on NVIDIA Quadro RTX 5000].

Test-time adaptation: We freeze the upscaling layers in order to update the feature extractor [lr: 1e-4], and apply learning rate decay and early stopping.

Baselines for TTA: We consider two baseline ideas: (1) *Repurposed blind SR (BSR) for TTA:* We utilized BSR degradation methods in frequency **SFM** [51] and spatial **PDM** [237] domain. To ensure fairness, training settings as well as the backbone were kept consistent. TTA was performed with an $\mathcal{L}_1$ loss function to isolate the influence of the degradation model. (2) *Native TTA:* We compare to a source-free TTA method **ACT** [38] using external priors [43, 187] and a second TTA method proposed initially for MR reconstruction **CST** [38] with cycle consistency at source-domain. The settings and adaptation strategies of the comparison methods were used directly. SSIM, PSNR are applied for evaluation.

Experiment 1: Quantitative and qualitative comparison. We used the brain T2 data for showing the method’s adaption to dataset shifts due to different acquisition parameters, other MR imaging devices (different vendors and field-strengths) and an unseen degradation. For evaluating the TTA performance on a realistic simulation of LR MR data combined with an unseen image modality, we used the knee dataset. In Table 4.1 PSNR and SSIM scores are reported with respect to the baselines on the target domain dataset. On the knee dataset, compared to CST, we improve the PSNR from 27.50dB to 28.25dB for 4× enlargement. With higher enlargement, both, repurposed BSR and native TTA obtain worse scores, while UTAR still preserves the adaptation capabilities. Visual comparisons are shown in Fig. 4.3. The first two rows show the restored images from the Guys and IOP dataset with 4× enlargement, while the last row is the result for the multi-coil knee dataset with 2× enlargement. The input has artifacts and shows a loss of anatomical details. The comparison methods can improve the performance, but are less effective than UTAR. For high enlargement, UTAR achieves notably better results than other adaptation methods.

Experiment 2: Frequency distribution analysis. We conducted a frequency distribution analysis for the SR reconstruction performance and the adaptation strategies beyond the the common evaluation metrics inspired by [51]. As shown in Fig. 4.4(a), source-domain SR cannot reconstruct the target-domain data effectively despite their similar anatomy (brain dataset), especially for high frequencies (blue). The deterioration can be explained by the model’s tendency to adhere to the source-domain distribution (red). In Fig. 4.4(b), although BSR (modified model + spat., e.g., PDM) and SR TTA (generic model + TTA, e.g., ACT) strategies were capable of capturing the low-frequency

Table 4.1: Quantitative results on the three unseen datasets Guys and IOP of IXI, PDFS of fastMRI under different enlargement scales. → indicates adaptation from source domain to target domain. Note that CST [38] is supervised and has extra access into the source data, UTAR is unsupervised and source-free.

				Brain: HH → Guys				Brain: HH → IOP				Knee: PD → PDFS				
		Source	Degradation	2×		4×		2×		4×		2×		4×		
		-free	Spat. Freq.	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	
BSR	Source [125]		w/o TTA	0.782	25.01	0.700	23.03	0.803	27.02	0.727	25.10	0.833	32.72	0.682	26.51	
	SFM [51]	-	×	✓	0.786	25.15	0.711	23.28	0.809	27.17	0.731	25.18	0.844	33.40	0.701	26.83
	PDM [237]	-	✓	×	0.830	25.83	0.709	22.89	0.851	27.93	0.739	25.06	0.841	33.05	0.698	26.67
TTA	ACT [164]	✓	-	-	0.807	24.50	0.707	22.80	0.826	26.38	0.735	24.95	0.761	28.64	0.685	26.53
	CST [38]	×	-	-	0.827	26.24	0.714	23.70	0.837	28.55	0.746	26.12	0.849	34.24	0.717	27.50
	Ours: UTAR	✓	✓	✓	0.830	26.54	0.773	24.62	0.865	28.27	0.775	26.26	0.847	33.88	0.728	28.25

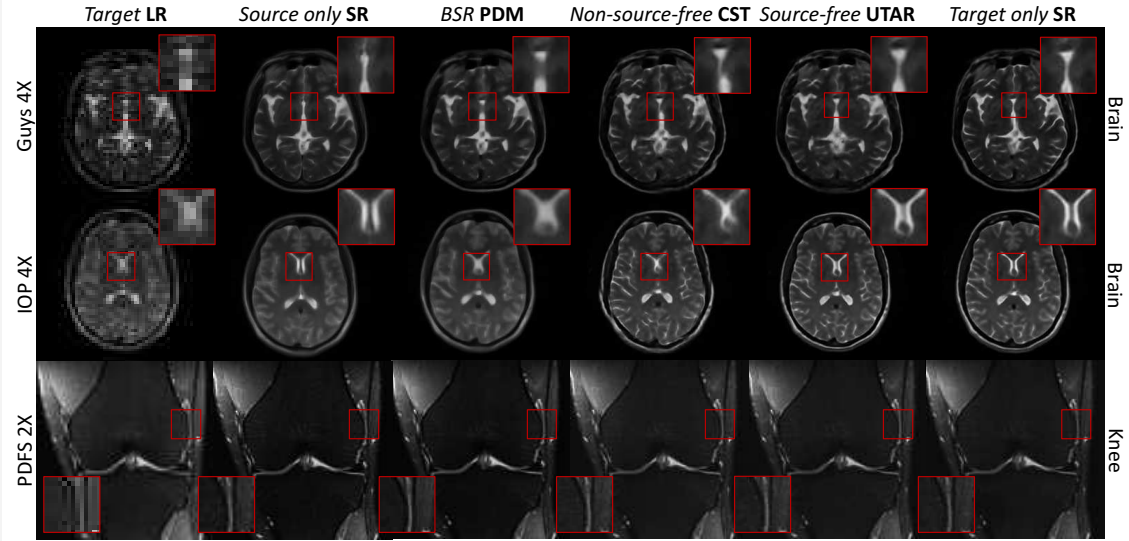


Figure 4.3: Visual comparison of different methods on the three datasets across target and source domain. (a) is the LR target, (b) and (f) are SR methods trained within the same domain, (c)-(e) are representative TTA methods.

details of the target domain for this task (green), they failed to achieve adequate reconstruction for high frequencies (red). When compared with the SR TTA strategy in Fig. 4.4(c), in addition to decreasing the bias of the model itself and narrowing the gap to the target domain (blue), UTAR is also able to further improve the fidelity in both high- and low-frequency reconstruction based on the general SR (green).

Experiment 3: Domain shifts in MRI SR.

We examined the effectiveness of PIC, CCL and UTAR for various domain shifts using a 4× pre-trained model on 5 unseen datasets shown in Table 4.2. UTAR yielded the best results, which indicates that conditional frequency restoration and imaging chain degradation are important factors. PIC outperformed CCL, indicating that an enriched degradation model is necessary to achieve

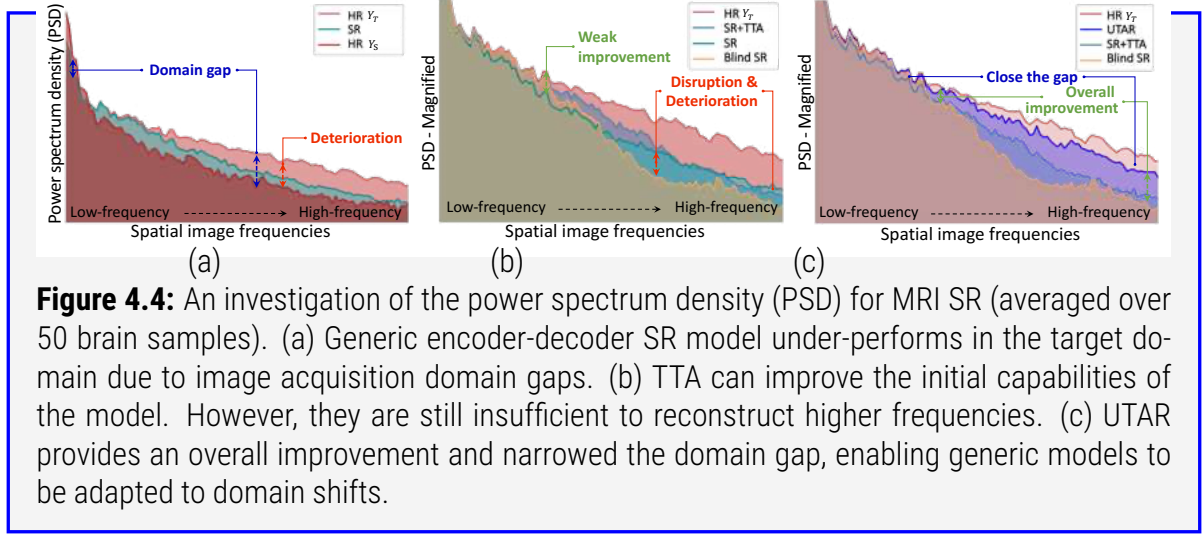


Figure 4.4: An investigation of the power spectrum density (PSD) for MRI SR (averaged over 50 brain samples). (a) Generic encoder-decoder SR model under-performs in the target domain due to image acquisition domain gaps. (b) TTA can improve the initial capabilities of the model. However, they are still insufficient to reconstruct higher frequencies. (c) UTAR provides an overall improvement and narrowed the domain gap, enabling generic models to be adapted to domain shifts.

Table 4.2: PIC, CCL and UTAR for various domain shifts with a $4\times$ SR model.

Training scheme $4\times$		Target-domain unseen datasets with shifts									
Source domain		Acquisition		Acquisition		Modality		Degradation		Anatomy	
		IOP T2		Guys T2		HH T1		HH _{unseen}		Knee	
		SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
HH T2	+ nil	0.908	30.76	0.901	28.85	0.847	30.10	0.740	25.77	0.633	25.46
	+ CCL	0.908	30.88	0.904	29.24	0.865	30.87	0.767	26.52	0.710	27.74
	+ PIC	0.911	31.18	0.906	29.32	0.872	31.01	0.780	26.78	0.721	28.28
	+ UTAR	0.917	31.50	0.912	29.95	0.875	31.12	0.785	26.84	0.743	29.26

source-free adaptation to the target domain for MRI SR TTA. With an adequate posterior degradation distribution like PIC, CCL is able to provide improved reconstruction capabilities, with particular attention to high frequencies, where such collaboration is most beneficial.

Discussion: From our experiments we make the following observations: (1) The feature extraction in MRI SR tends to be fine-tuned to the source domain and requires adaptation. Our method demonstrates a successful adaptation without accessing the source domain, whereas TTA methods in MR reconstruction [38] and segmentation [16, 101] usually rely on external priors or data from the source domain, which might not be available for MRI SR. (2) We show an improved performance compared to methods accessing the source domain, indicating that appropriate modelling of the domain shifts, as possible with the PIC, is essential. The implementation of the PIC in this work is specific to the public datasets used and could be further specialized if more acquisition parameters are known, e.g. the exact sampling pattern of k-space. (3) Our novel and more complex modelling of the entire imaging chain including the specific focus on the restoration of high frequency features seems to be a promising approach to close domain gaps of current SR models. (4) Our highlight is on generalizing the model to the *unseen* domain without compromising data privacy. UTAR adapts a pre-trained model *without re-accessing* the source domain data at *test time*, which is highly relevant in practice.

4.1.6 Summary

In this paper we propose UTAR, a conditional consistency learning (CCL) strategy and physics-inspired imaging chain degradation model (PIC) for generic encoder-decoder MRI SR architectures. UTAR alleviates several possible domain shifts in MRI by fine-tuning the pre-trained models during the inference phase without access to the source domain. We have demonstrated improved domain adaptation performance without extensive re-training while making the model more adaptive. Since we have shown that the degradation model has a major impact on the adaptation performance, we intend to develop an implicit physics-informed SR reconstruction system in the future.

Chapter 5

Generalizable and Stable Diffusion Models for Inverse Problems

Building on advancements in domain adaptation and super-resolution techniques for medical imaging, particularly MRI, prior work on UTAR (Unsupervised Test-time Adaptation for MRI Super-Resolution) emphasized the importance of physics-informed degradation models and conditional consistency learning in addressing domain shifts in deep learning-based super-resolution. While it successfully adapted pre-trained models to unseen target domains without requiring paired data, the focus was primarily on improving the quality of low-resolution MRI images through a dual-domain degradation model.

However, ensuring stability and generalizability in diffusion models, especially when applied to complex inverse problems like MRI reconstruction, remains a critical challenge. This is where the D^3GM (Dynamics-aware SDE Diffusion Generative Model) framework comes in. By leveraging measure-preserving dynamics from Random Dynamical Systems (RDS), it addresses the instability of traditional diffusion models, particularly in cases of severe degradation and domain shifts. The framework not only bridges the gap between high-quality and degraded images but also ensures the diffusion process remains robust to real-world degradations.

The progression from UTAR to D^3GM highlights the increasing need for more theoretically grounded approaches in diffusion models, particularly when tackling challenging inverse problems in medical imaging.

5.1 Background:

In recent years, diffusion models have emerged as a powerful tool for generative modeling, showing impressive results in high-resolution image synthesis. Their success stems from a simple yet elegant idea: gradually turning random noise into structured data through a learned denoising process governed by a stochastic differential equation (SDE). This formulation has also attracted attention in the context of inverse problems—tasks like image reconstruction or super-resolution—where information must be recovered from partial or degraded observations.

Yet despite their promise, diffusion models often struggle when applied to real-world medical data. The assumptions made in many existing methods—such as idealized noise, linear degradation, or perfect data distributions—rarely hold in practice. Medical images, especially those affected by motion, low signal-to-noise ratios, or domain shifts across scanners and sites, present far more complexity. When faced with such conditions, conventional diffusion models can become unstable: their sampling paths drift, reconstructions degrade, and generalization suffers.

Part of the problem lies in how these models treat dynamics. The transformations they learn do not necessarily preserve the structure or distribution of real data. Without constraints to anchor the process, errors can accumulate across sampling steps. This chapter proposes an alternative. By framing diffusion as a measure-preserving stochastic dynamical system, we aim to improve both the theoretical stability and practical robustness of these models. The method we introduce—D3GM—integrates distribution-preserving structure directly into the generative process. We show that this approach not only improves stability but also leads to better generalization across tasks and domains.

5.1.1 Related Work

Diffusion Models for Generative Learning. Diffusion probabilistic models have become a dominant paradigm for generative modeling, driven by their strong empirical performance and theoretical grounding. Early formulations, including denoising diffusion probabilistic models (DDPMs) and score-based generative models, established diffusion as a process of gradually transforming data distributions into tractable noise distributions and learning the corresponding reverse dynamics [81, 190, 193]. Subsequent works refined these formulations through improved noise schedules, variance control, and continuous-time SDE interpretations, enabling more efficient training and sampling [102]. Recent research has also focused on stabilizing diffusion dynamics and accelerating convergence, highlighting the importance of principled stochastic modeling in high-dimensional generative processes.

Diffusion Models for Inverse Problems. Building on their generative capacity, diffusion models have been widely applied to inverse problems, where the goal is to recover latent signals from in-

complete or degraded observations. Representative approaches include posterior sampling methods that incorporate measurement operators into the reverse diffusion process [31, 32], as well as conditional and transitional diffusion frameworks that initialize sampling from corrupted measurements rather than pure noise [135, 142, 231]. While these methods have demonstrated promising results on a range of restoration and reconstruction tasks, they often rely on assumptions such as linear degradation models, well-aligned priors, or stable forward processes. In practice, violations of these assumptions—common in medical imaging and real-world data—can lead to instability, accumulated error, and degraded generalization.

Stability and Generalization in Diffusion Dynamics. An emerging line of work has begun to examine the stability and robustness of diffusion models under distribution shift, imperfect priors, and heterogeneous degradation. Several studies have identified sensitivity to noise schedules, initialization strategies, and forward process design as key factors influencing performance [39, 223]. More recent efforts have explored connections between diffusion models and dynamical systems, including mean-reverting processes and Schrödinger bridges, to improve convergence and controllability [56]. However, most existing approaches focus on local regularization or heuristic stabilization techniques, without explicitly enforcing global distributional consistency throughout the diffusion trajectory.

Measure-Preserving and Dynamical Systems Perspectives. Outside the diffusion literature, measure-preserving dynamics and random dynamical systems (RDS) have long been studied as a principled framework for analyzing stability, recurrence, and invariant measures in stochastic processes. These concepts provide theoretical tools for understanding long-term behavior and robustness under perturbations, yet have received limited attention in the design of modern diffusion-based generative models. In contrast to prior work, our approach explicitly frames diffusion as a measure-preserving stochastic dynamical system, ensuring that the generative process respects invariant distributions and avoids cumulative temporal discrepancies. This perspective enables a unified treatment of stability and generalization, particularly for challenging inverse problems involving complex degradation and domain shift.

5.2 Stability and Generalizability in SDE Diffusion Models with Measure-Preserving Dynamics

5.2.1 Overview

Inverse problems describe the process of estimating the causal factors from a set of measurements or data. Mapping of often incomplete or degraded data to parameters is ill-posed, thus data-driven iterative solutions are required, for example when reconstructing clean images from poor signals. Diffusion models have shown promise as potent generative tools for solving inverse problems due to their superior reconstruction quality and their compatibility with iterative solvers. However, most existing approaches are limited to linear inverse problems represented as Stochastic Differential Equations (SDEs). This simplification falls short of addressing the challenging nature of real-world problems, leading to amplified cumulative errors and biases. We provide an explanation for this gap through the lens of measure-preserving dynamics of Random Dynamical Systems (RDS) with which we analyse Temporal Distribution Discrepancy and thus introduce a theoretical framework based on RDS for SDE diffusion models. We uncover several strategies that inherently enhance the stability and generalizability of diffusion models for inverse problems and introduce a novel score-based diffusion framework, the **D**ynamics-aware **S**DE **D**iffusion **G**enerative **M**odel (D³GM). The *Measure-preserving property* can return the degraded measurement to the original state despite complex degradation with the RDS concept of *stability*. Our extensive experimental results corroborate the effectiveness of D³GM across multiple benchmarks including a prominent application for inverse problems, magnetic resonance imaging.

5.2.2 Introduction

Diffusion probabilistic models [190–192] have demonstrated impressive performance across various image generation tasks, primarily by modeling a diffusion process and then learning an associated reverse process. Among the many commonly used approaches [226], diffusion models that incorporate the concept of score functions [90, 193] can capture the intrinsic random fluctuations of the forward diffusion process, positioning them as a good choice for in-depth analysis. Score-based generative models (SGMs) entail gradually diffusing images towards a noise distribution, and then generating samples by chaining the score functions at decreasing noise levels with score-based sampling approaches. One such example of an SGM with a score-based sampling technique, known as score matching [192], has gained popularity for density estimation [41]. It employs methods such as Langevin dynamics [67, 98, 151] and SDEs [97, 102, 135, 238] to simulate the underlying probability distribution of training samples.

However, vanilla unconditional SGMs can be extended to inverse problems by leveraging an implicit prior distribution, based on the available counterpart measurement, subjected to corruption

and/or noise. To this end *transitional SGMs* enable an iterative recovery of the data from this noisy counterpart, instead of relying on Gaussian white noise as a starting point [34, 56, 128, 135, 142, 231].

Intuitively, leveraging priors and a generative capacity into transitional SGMs offers the possibility of exploring high quality reconstruction and restoration and gaining better performance. However, current transitional SGMs that incorporate priors have largely overlooked the unreliable quality of the prior and its measurement. Empirically, transitional SGMs cannot always be trusted in terms of stability and efficiency, especially in a regime of non-uniformly distributed noise or corrupted signal quality [31, 135, 231]. Hence, the exploitation of transitional learning within SGMs does not come without costs as their advantages vanish in limited data quality settings. Theoretical understanding is notably lacking in this field with the following fundamental open problem: Can we realize reliable transitional diffusion processes in practice for inverse problems with a theoretical guarantee?

While recent works have started to lay down a theoretical foundation for these models, a detailed understanding is still lacking. Current best practice advocates for smaller initialisation values (e.g., noise schedule [56, 231], instead of large values [39, 223] to ensure that the forward dynamics brings the diffusion sufficiently close to a known prior and simple noise distribution. However, a proper choice of the values conditioned on the prior within a theoretical framework should be preferred for a better approximation of the score-matching objective and higher computational efficiency. To fully facilitate the power of reversion and generation of transitional SGMs and to mitigate the influence of low-quality measurements for solving inverse problems, this paper provides a measure-preserving dynamics of random dynamical system (RDS) perspective as a promising way to obtain reliable reversion and generation. Notably, our ‘measure’ is not only the observations (e.g., degraded images), but also represents the invariant probability measure (distribution) of the RDS. This allows to consider the concept of an RDS *stability* and to frame challenging degradation learning within a measure-preserving dynamical system. Thus, we can start from a transitional SGM interpretation of diffusion models and connect RDS to the SDE in transitional SGMs. The pitfalls (e.g., Instability) are discussed in Sect. 5.2.6 and further implications can be found in the Appendix. Transitional SGMs have not been fully explored before, and we provide a theoretical interpretation of a stationary process as a possible solution.

Our D³GM framework is abstracted from transitional SGMs. The key to our framework is a stationary process following measure-preserving dynamics to ensure the stability and generalizability of the diffusion, as well as reducing the influence of accumulated error, distribution bias and degradation discrepancy. Our contributions can be summarised as follows:

- **Temporal Distribution Discrepancy:** We conduct a rigorous theoretical examination of the instability issue of transitional SGMs, measured as Temporal Distribution Discrepancy (i.e., lower bound of modeling error). This analysis sheds light on critical aspects related to stability and

generalizability¹ effectively addressing an unexplored fundamental gap in the understanding of solving challenging inverse problems with SDEs.

- **D³GM Framework:** We propose a solution, D³GM, and an explanation from measure-preserving dynamics of Random Dynamical Systems (RDS). ‘Measure’ includes both measurements (degraded image) and invariant measures (distribution) of RDS, which allows complex degradation learning and enhances restoration and reconstruction accuracy.
- **Thorough Evaluation:** Our contributions are substantiated by extensive validation. We demonstrate the practical benefits of our D³GM framework across various benchmarks, including challenging tasks such the reconstruction of Magnetic Resonance Imaging (MRI) data.

We address the instability of diffusion models for inverse problems under domain shift and concept drift (unknown and heterogeneous degradation). This leads to what we believe is a completely novel view on the theoretical foundation of how the *degradation* process is modelled. The result is an approach that is more in line with the original intention of the theory of diffusion. We chose inverse problems as a relevant application area to demonstrate our ideas but also included a variety of challenging problem settings to explore the generalizability of *D³GM*. To the best of our knowledge, no other method can handle a **diverse range** of challenging tasks like *real-world* de-hazing, *compressed* MRI reconstruction, *blind* MRI super-resolution, etc. with a **unified** underlying theoretical framework. In Tab. 5.1 we illustrate the key differences of *D³GM* compared with SGMs and transitional SGMs.

5.2.3 Preliminaries

SGMs. We will follow the typical construction of the diffusion process $\mathbf{x}(t)$, $t \in [0, T]$ with $\mathbf{x}(t) \in \mathbb{R}^d$. Concretely, we want $\mathbf{x}(0) \sim p_0(\mathbf{x})$, where $p_0 = p_{\text{data}}$, and $\mathbf{x}(T) \sim p_T$, where p_T is a tractable distribution that can be sampled. In this work, we consider the score-based diffusion form of the SDE [193]. Consider the following Itô diffusion process defined by an SDE:

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{W}, \quad (5.1)$$

where $\mathbf{f} : \mathbb{R}^d \mapsto \mathbb{R}^d$ is the drift coefficient of $\mathbf{x}(t)$, $g : \mathbb{R} \mapsto \mathbb{R}$ is the diffusion coefficient coupled with the standard d-dimensional Wiener process $\mathbf{w} \in \mathbb{R}^d$. By carefully choosing $\bar{\mathbf{f}}, g$, one can achieve a spherical Gaussian distribution as $t \rightarrow T$. For the forward SDE in Eq. 5.1 there exists a corresponding reverse-time SDE [9, 193]:

$$d\mathbf{x} = [\mathbf{f}(\mathbf{x}, t) - \underbrace{g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x})}_{\text{score function}}]dt + g(t)d\mathbf{W}, \quad (5.2)$$

¹Generalizability refers to the extent to which out-of-distribution and domain shift impacts the fidelity of the restoration process. Stability in SDE diffusion models is demonstrated by its resilience to degradation beyond its domain and its consistent ability to restore high-quality images.

Table 5.1: Differences between state-of-the-art SDE diffusion-based approaches.

Model	$p(X_0)$	$p(X_1)$	Theorem Foundation	Properties	TDD (Prop.2)	Attractor	Operator Dir.	Inverse problem solver
SGM [191]	p_A	$\mathcal{N}(0, I)$	SDE diffusion	No prior	Unsolvable	No subset attracts	one-sided	
IR-SDE [135]	p_A	$p_B(\cdot X_0, \mu)$	Mean-reverting SDE	Instability	Limited	Gaussian	one-sided	Lin/Non-Lin
I ² SB [128]	δ_a	$p_B(\cdot X_0)$	Schrodinger Bridge	Strict Prior	Limited	No subset attracts	one-sided	Lin/ Non-Lin
D ³ GM (ours)	p_A	$p_B(\cdot X_0, \mu, \tau)$	Measure-Preserving RDS	Stability	Robust	$\mathcal{N}(\mu, \tau^2 \sigma^2 I)$	two-sided	Lin/ Non-Lin/Blind

where dt is the infinitesimal negative time step, and $\mathbf{w}$ is the Brownian motion running backwards. The score function $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ is in general intractable and thus SDE-based diffusion models approximate it by training a time-dependent neural network under a score function [90, 210].

Transitional SGMs. [34, 42, 56, 128, 135, 231] leverage a transitional iterative denoising paradigm for the inverse problems. In inverse problems, such as super-resolution, we have an (nonlinear, partial, and noisy) observation $\mathbf{y}$ of the underlying high-quality signal $\mathbf{x}$. The mapping $\mathbf{x} \mapsto \mathbf{y}$ is many-to-one, posing an ill-posed problem. In this case, a strong prior on $\mathbf{x}$ is needed for finding a realistic solution. Formally, the general form of the *forward* (measurement) model is:

$$\mathbf{y} = \mathcal{A}(\mathbf{x}) + \mathbf{n}, \quad \mathbf{y}, \mathbf{n} \in \mathbb{R}^n, \mathbf{x} \in \mathbb{R}^d, \quad (5.3)$$

where $\mathcal{A}(\cdot): \mathbb{R}^d \mapsto \mathbb{R}^n$ oftentimes $n \ll d$ is the forward measurement operator and $\mathbf{n}$ is the measurement noise, assuming $\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$. While sharing a similar aim of bridging $\mathbf{y}$ and $\mathbf{x}$ in transitional SGMs, different mathematical frameworks have been used: [42] employs Inversion by Direct Iteration; [56, 128, 135] model it as a Mean-reverting SDE.

Transitional SGM has become an increasingly important line of SDE research due to the applicability on images with theoretical guarantees. However, they often perform poorly in real-world scenarios. To provide a theoretical investigation of this gap, we interpret *Transitional SGM* as *Ornstein-Uhlenbeck (OU)* process. This perspective allows us to understand the random fluctuations in image degradation as stochastic processes, providing a foundation to integrate *random dynamical systems (RDS)* with the diffusion process as a natural extension of the SDE framework involving the OU process.

The *Measure-preserving property* is introduced from the perspective of RDS: The distribution can still return to the original state despite severe degradation. Our approach constructs a bridge from measure-preserving dynamical system to transitional SGM through measure-preserving dynamics, and highlights the *Temporal Distribution Discrepancy* in Sect. 5.2.6. Subsequently, we address this issue of instability: by incorporating a measure-preserving strategy into the solution of inverse problems, which is detailed in Sect. 5.2.7. This covers counterpart modeling, bridging a transition from uncertain diffusion modeling to deterministic solutions, yielding significant improvements in both performance and efficiency as demonstrated in Sect. 5.2.8.

²MRI signals are defined on $\mathbb{C}^n$ and $\mathbb{C}^d$. We demonstrate in Sect. 5.2.8 that our approach is applicable to MRI.

5.2.4 Intuition of Measure-Preserving Dynamics in SDE:

Consider the analogy of a stretched rubber band, which naturally seeks to return to its original but does so with a lot of oscillations when released. This elastic behavior parallels the dynamics of the OU process, where deviations from a mean state are counteracted by a restorative force, guiding the system back towards equilibrium (i.e., final state), with random perturbations.

Our process models the noise as a stochastic component that fluctuates around a stationary process and improve the OU process with RDS. Measure-preserving dynamics ensure that while the image undergoes transformations during the denoising process, the overall statistical properties remain consistent (i.e., invariant image features), which cannot be satisfied by vanilla OU processes or previous approaches (Tab. 1).

5.2.5 Mathematical Preliminaries

Consider a probability space $(\Omega, \mathcal{F}, P)$ accompanied by a standard Brownian motion W_t . A stochastic process x_t over the interval $0 \leq t \leq T$ can be formulated by the following stochastic differential equation (SDE):

$$dx_t = b(t, x_t)dt + \sigma(t, x_t)dW_t \quad (5.4)$$

Definition 1 Filtration A collection of sigma-fields,

$$\mathbb{F} := \{\mathcal{F}_t, 0 \leq t \leq T\},$$

is termed a filtration if:

1. $\mathcal{F}_t \subset \mathcal{F}$ is a sub- σ -field for every $t \in [0, T]$;
2. If $0 \leq t_1 < t_2 \leq T$, then $\mathcal{F}_{t_1} \subset \mathcal{F}_{t_2}$.

Here, $\mathcal{F}_t$ represents the information set at time t .

Definition 2 Strong Solution A process x , which is $\mathcal{F}$ -progressively measurable, is considered a strong solution to the SDE given by Eq. 5.1 if:

$$\int_0^T (|b(t, x_t)|^2 + |\sigma(t, x_t)|^2)dt < \infty$$

almost surely. This is captured by:

$$x_t = x_0 + \int_0^t b(s, x_s)ds + \int_0^t \sigma(s, x_s)dW_s, \forall t \in [0, T] \quad (5.5)$$

Definition 3 Lipschitz Continuity For an N -dimensional stochastic process x_t over $t \in [0, \infty)$,

adapted to the filtration $\mathcal{F}$, function $w(x_t)$ exhibits Lipschitz continuity in x if:

1. $w(x_t)$ is $\mathcal{F}$ -measurable with the requisite dimensions.
2. There exists a non-negative constant K such that, for every x_t and x_s ,

$$|w(x_t) - w(x_s)| \leq K|x_t - x_s|.$$

In numerous diffusion models, Lipschitz continuity is inherent, ensuring the existence and uniqueness of the solution to the stochastic process. However, for clarity in network design, the emphasis on Lipschitz continuity ensures the foundation of neural networks remains consistent.

Definition 4 Random Dynamical System A random dynamical system(RDS) consists of a base flow, the "noise", and a cocycle dynamical system on the "physical" phase space, we first discuss one fundamental element of our RDS, the base flow.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, the noise space. Define the base flow $\vartheta : \mathbb{R} \times \Omega \rightarrow \Omega$ as follows: for each "time" $s \in \mathbb{R}$, let $\vartheta_s : \Omega \rightarrow \Omega$ be a measure-preserving measurable function: $\mathbb{P}(E) = \mathbb{P}(\vartheta_s^{-1}(E))$ for all $E \in \mathcal{F}$ and $s \in \mathbb{R}$

Suppose also that

1. $\vartheta_0 = \text{id}_\Omega : \Omega \rightarrow \Omega$, the identity function on Ω ;
2. for all $s, t \in \mathbb{R}$, $\vartheta_s \circ \vartheta_t = \vartheta_{s+t}$.

That is, $\vartheta_s, s \in \mathbb{R}$, forms a group of measure-preserving transformation of the noise $(\Omega, \mathcal{F}, \mathbb{P})$

Now we are ready to define the **random dynamical system(RDS)**.

let (X, d) be a complete separable metric space, the phase space. Let $\varphi : \mathbb{R} \times \Omega \times X \rightarrow X$ be a $(\mathcal{B}(\mathbb{R}) \otimes \mathcal{F} \otimes \mathcal{B}(X), \mathcal{B}(X))$ measurable function such that 1. for all $\omega \in \Omega$, $\varphi(0, \omega) = \text{id}_X : X \rightarrow X$, the identity function on X ; 2. for (almost) all $\omega \in \Omega$, $(t, x) \mapsto \varphi(t, \omega, x)$ is continuous; 3. φ satisfies the (crude) cocycle property: for almost all $\omega \in \Omega$,

$$\varphi(t, \vartheta_s(\omega)) \circ \varphi(s, \omega) = \varphi(t + s, \omega)$$

In the case of random dynamical systems driven by a Wiener process $W : \mathbb{R} \times \Omega \rightarrow X$, the base flow $\vartheta_s : \Omega \rightarrow \Omega$ would be given by

$$W(t, \vartheta_s(\omega)) = W(t + s, \omega) - W(s, \omega).$$

Theorem 1 Existence and Uniqueness If the initial condition $x_0 \in \mathbb{L}^2$ is a random variable that's independent of W and both $\mu(0, x_0)$ and $\sigma(0, x_0) \in \mathbb{H}^2$, then, provided there exists a constant $K > 0$ that satisfies:

$$\begin{aligned} & |b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \\ & \leq K|x - y|, \quad \forall t \in [0, T], x, y \in \mathbb{R}^n \end{aligned} \tag{5.6}$$

(an attribute also recognized as Lipschitz continuity), a unique strong solution to Eq. 5.1 exists in $\mathbb{H}^2$ for every $T > 0$. Additionally:

$$\mathbb{E} \left[\sup_{t \leq T} |x_t|^2 \right] \leq C(1 + \mathbb{E}|x_0|^2)e^{CT} \quad (5.7)$$

holds true, where the constant C depends on both T and K .

5.2.6 Instability Analysis: Transitional SGMs with Corrupted SDE Diffusion

Ornstein-Uhlenbeck (OU) process. An OU process is a common case in transitional SGMs, where x_t is defined using an SDE: $dx_t = -\theta_t x_t dt + \sigma_t dW_t$. W_t is standard Brownian motion. A drift term μ can be introduced:

$$dx_t = \theta_t(\mu - x_t)dt + \sigma_t dW_t, \quad (5.8)$$

where μ denotes state mean, reflecting the expected state of the measurement (e.g., corrupted image [135], noisy speech [219]) over time. θ_t, σ_t are time-dependent parameters. The drift term corrects deviations from the constant μ , effectively pulling the process towards μ ($t \rightarrow \infty$) with *Stability* as opposed to pure noise in Eq. 5.1

Measure-preserving Dynamics in SDE Diffusion. The solution of the above SDE can be represented by a continuous-time random dynamical system φ defined on a complete separable metric space (X, d) . (See precise definition of RDS in Appx. 5.2.5). More generally, we can extend the RDS to a two-sided solution operator with a flow map. The base flow driven by Brownian motion can be written as $W(t, \vartheta_s(\omega)) = W(t + s, \omega) - W(s, \omega)$.

Proposition 1 *After extending the solution of the OU process to RDS, the measure-preserving RDS φ should meet the property $\varphi(t, s; \omega)x = \varphi(t - s, 0; \vartheta_s \omega)x$. However, OU processes with time-varying coefficients usually do not satisfy this property. In this situation, the system breaks the forward-reverse processes, making it difficult to maintain stability.*

Intuition 1. A two-sided measure-preserving random dynamical (MP-RDS) system formulation enables us to use the Poincare recurrence theorem [159], intuitively, with a two-sided MP-RDS φ_t , the Poincare recurrence theorem ensures that the system φ_t starts from terminal condition x_T , run backward in time, will hit a region $(x_0 - \epsilon, x_0 + \epsilon)$ for small ϵ in finite time, where x_0 is the high-quality image.

Example 1. Following Intuition 1. and Prop. 1, suppose that the OU's θ_t follows a cosine schedule, such that $\theta_t = \cos(t)$ for $0 \leq t \leq T$, then for some $0 \leq s \leq t \leq T$, $\varphi(t, s; \omega)x \neq \varphi(t - s, 0; \vartheta_s \omega)x$ because the change of θ_t w.r.t. time is not uniform. The OU-process instability exists due to Temporal Distribution Discrepancy (Prop. 2).

At a high level, Proposition 1 can be extended to show that there exists a compact attracting set at any $-\infty < t < \infty$, and this convention has allowed us to characterize the attractor $K(\omega) = \mathcal{N}(\mu, \sigma^2/2\theta)$. The closed-form distribution for $\mathbf{y}$ can be complex and may not be tractable depending on the particular scenarios of the actual image degradation process μ . The modification of σ and θ is used to regularize the perturbation and attempt to close the distribution. However, these injections might bypass the stationary process.

Instability-Temporal Distribution Discrepancy. Given the process OU $(x_t, \mu; t, \theta)$ with Eq. [5.8], where $x_T \neq x_\infty$ for finite T , indicates that the perturbed state cannot move towards the degraded LQ image and fails in matching the theoretical distribution. This inherent discrepancy further causes bias in the estimation of μ_t , which gradually accumulates into error in the reverse process.

5.2.6.1 Analysis of Proposition 1

Proposition 1 After extending the solution of the OU process to RDS, the measure-preserving flow map of the solution should meet the property $\varphi(t, s; \omega)x = \varphi(t - s, 0; \vartheta_s \omega)x$. However, OU processes with time-varying coefficients are usually not satisfied for this property (can be referred to as time-homogeneity) and thus the stability of the system breaks.

The forward process in SDE notation

$$dx_t = \theta_t(\mu - x_t)dt + \sigma_t dW_t$$

and the solution to the above SDE is given by

$$x_t = \mu + (x_s - \mu)e^{-\bar{\theta}_{s:t}} + \int_s^t \sigma_z e^{-\bar{\theta}_{z:t}} dW_z$$

where $\bar{\theta}_{s:t} = \int_s^t \theta_z dz$.

This solution above can be represented by a continuous random dynamical system (RDS) φ defined on a complete separable metric space (X, d) , where the noise is chosen from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. More details can be found in [36].

More generally, we can extend the RDS to two-sided, infinite time, define a flow map or (solution operator) $\varphi : \mathbb{R} \times \Omega \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ by $\varphi(t, s, |\omega, x_0) := x(t, s | \omega, x_0)$ with $\omega \in \Omega$, $-\infty < s \leq t < \infty$. The base flow driven by Brownian motion can be explicitly written as $W(t, \vartheta_s(\omega)) =$

$$W(t + s, \omega) - W(s, \omega).$$

Now suppose that:

1. The flow map $\vartheta_t, t \in \mathbb{R}$ is a measure-preserving transformations of $(\Omega, \mathcal{F}, P)$, with the property that for all $s < t$ and $x \in X$,

$$\varphi(t, s; \omega)x = \varphi(t - s, 0; \vartheta_s \omega)x, \quad P\text{-a.s.} \quad (5.9)$$

2. (i) $\varphi(t, r; \omega)\varphi(r, s; \omega)x = \varphi(t, s; \omega)x$ for all $s \leq r \leq t$ and $x \in X$;

(ii) $\varphi(t, s; \omega)$ is continuous in X , for all $s \leq t$.

(iii) for all $s < t$ and $x \in X$, the mapping

$$\omega \mapsto \varphi(t, s; \omega)x$$

is measurable from $(\Omega, \mathcal{F})$ to $(X, \mathcal{B}(X))$; and

(iv) for all $t, x \in X$, and P -a.e. ω , the mapping $s \mapsto \varphi(t, s; \omega)x$ is right continuous at any point.

Where $\mathcal{B}(X)$ denotes the σ -algebra generated by X .

Under assumptions (i), (ii), (iii), (iv) and suppose that for P -a.e. ω there exists a compact attracting set $K(\omega)$ at time 0, i.e., such that for all bounded sets $B \subset X$,

$$d(\varphi(0, s; \omega)B, K(\omega)) \rightarrow 0 \quad \text{as} \quad s \rightarrow -\infty$$

We can see that the attractor of this system is defined in the pullback sense, such that time is rewind backward before iterating forward.

Moreover, the reverse process with any starting time t to s is defined as the RDS going backward in time

$$\varphi(s, t | \vartheta_t \omega, x_t)$$

start from the time t realization and run backwards to s .

the above proposition can be extended to show that there exists a compact attracting set at any $-\infty < t < \infty$. and this convention has allowed us to characterize the attractor $K(\omega) = \mathcal{N}(\mu, \lambda^2)$, when the time becomes finite, for example from 0 to T , the random attractors can be abstractly viewed as $\mathcal{N}(\mu + (x_s - \mu)e^{-\bar{\theta}_{0:t}}, \lambda(1 - e^{-\bar{\theta}_{0:t}}))$.

Moreover, an important assumption in the [5.9] is usually not satisfied for OU processes with time-varying coefficients, therefore, we impose that for every t , $\frac{\sigma_t^2}{2\theta_t} = \lambda^2$, where λ is a constant, and will be the asymptotic variance of the forward process. This convention has allowed us to reduce the regularization on two variables σ_t, θ_t to just one variable to satisfy [5.9] and this convention has allowed as to characterize the attractor $K(\omega) = \mathcal{N}(\mu, \lambda^2)$, when the time becomes finite, for example from 0 to T , the random attractors can be abstractly viewed as the Gaussian measure $\mathcal{N}(\mu + (x_s - \mu)e^{-\bar{\theta}_{0:t}}, \lambda(1 - e^{-\bar{\theta}_{0:t}}))$.

5.2.6.2 Proof for reverse-time D³GM process:

There is a one-to-one and onto correspondence between the stochastic differential equation and the Kolmogorov equation for $p(x_t, t | x_s, s), t \geq s$, which describes the evolution of the underlying probability distribution. Consequently, there should be a one-to-one and onto correspondence between a reverse-time equation for $\tilde{x}_t$ and a Kolmogorov equation for $p(x_t, t | x_s, s), s \geq t$

$$dx_t = \theta_t(\mu - x_t)dt + \tau\sigma dW_t$$

We have the corresponding Kolmogorov backward equation given by

$$\begin{aligned} -\frac{\partial p(x_s, s | x_t, t)}{\partial t} &= \theta_t(\mu - x_t) \cdot \frac{\partial p(x_s, s | x_t, t)}{\partial x_t} \\ &\quad + \frac{1}{2} \tau^2 \sigma_t^2 \cdot \frac{\partial^2 p(x_s, s | x_t, t)}{\partial x_t^2}, \end{aligned} \quad (5.10)$$

The unconditioned Kolmogorov forward equation is given by

$$\begin{aligned} -\frac{\partial p(x_s, s | x_t, t)}{\partial t} &= \frac{\partial (\theta_t(\mu - x_t) \cdot p(x_s, s | x_t, t))}{\partial x_t} \\ &\quad - \frac{1}{2} \cdot \frac{\partial^2 (\tau^2 \sigma_t^2 \cdot p(x_s, s | x_t, t))}{\partial x_t^2}, \end{aligned} \quad (5.11)$$

See [9] for more details on Kolmogorov equations.

Bayes rule gives

$$p(x_t, t, x_s, s) = p(x_s, s | x_t, t) p(x_t, t)$$

We plug this result into [5.10] which gives us the Kolmogorov equation

$$\begin{aligned} -\frac{\partial}{\partial t} p(x_t, t, x_s, s) &= \frac{\partial}{\partial x_t} [\bar{f}(x_t, t) p(x_t, t, x_s, s)] \\ &\quad + \frac{1}{2} \frac{\partial^2 [p(x_t, t, x_s, s) \cdot \tau^2 \sigma_t^2]}{\partial x_t^2} \end{aligned} \quad (5.12)$$

and the expression for $\bar{f}$ is given by

$$\begin{aligned} \bar{f}(x_t, t) &= \theta_t(\mu - x_t) - \frac{1}{p(x_t, t)} \frac{\partial}{\partial x_t} [p(x_t, t) \tau^2 \sigma_t^2] \\ &= \theta_t(\mu - x_t) - \tau^2 \sigma_t^2 \log \frac{\partial}{\partial x_t} [p(x_t, t)] \end{aligned} \quad (5.13)$$

Therefore, we have that the reverse process corresponds to the Kolmogorov equation [5.13] is given by

$$dx_t = \theta_t(\mu - x_t) - \tau^2 \sigma_t^2 \log \nabla_x p_t(x_t) + \tau^2 \sigma_t^2 d\bar{W}_t$$

Definition 5 Different from deterministic dynamical systems, random dynamical systems usually

consider a pullback attractor rather than a forward attractor due to the non-autonomousness introduced by the random noise. The pullback attractor (or random global attractor) $\mathcal{A}(\omega)$ for the RDS φ we defined in [7] is a $\mathbb{P}$ -almost surely unique random set such that:

1. $\mathcal{A}(\omega)$ is a random compact set: $\mathcal{A}(\omega) \subseteq X$ is almost surely compact and $\omega \mapsto d(x, \mathcal{A}(\omega))$ is a $(\mathcal{F}, \mathcal{B}(X))$ -measurable function for every $x \in X$
2. $\mathcal{A}(\omega)$ is invariant: for all $\varphi(t, \omega)(\mathcal{A}(\omega)) = \mathcal{A}(\vartheta_t \omega)$ almost surely;
3. $\mathcal{A}(\omega)$ is attractive: for any deterministic bounded set $B \subseteq X$, $\lim_{t \rightarrow +\infty} d(\varphi(t, \vartheta_{-t} \omega)(B), \mathcal{A}(\omega)) = 0$ almost surely.

$\mathcal{B}(X)$ denotes the Borel σ -algebra generated by the space X where the RDS is defined.

Definition 6 Poincare Recurrence Theorem

Let

$$(X, \Sigma, \mu)$$

be a finite measure space and let

$$f : X \rightarrow X$$

be a measure-preserving transformation. then we have that for any $E \in \Sigma$, the set of those points x of E for which there exists $N \in \mathbb{N}$ such that $f^n(x) \notin E$ for all $n > N$ has zero measure.

In other words, almost every point of E returns to E . In fact, almost every point returns infinitely often; i.e. $\mu(\{x \in E : \text{there exists } N \text{ such that } f^n(x) \notin E \text{ for all } n > N\}) = 0$.

5.2.6.3 Proof for Proposition 2

Proposition 2 Given Eq. [5.3] and Eq. [5.8] and assume that the score function is bounded by C in L^2 norm, then the discrepancy between the reference and the retrieved data is, with probability at least $(1 - \delta)$:

$$\begin{aligned} & \|x_0 - \text{OU}(x_0, \mu; T, \theta)\|_2^2 \\ & \geq |((x_0 - \mu)^2 - \sigma_T^2/2\theta_T) e^{-2\bar{\theta}_T} + \sigma_T^2/2\theta_T \\ & - \sigma_{max}^2 \left(C\sigma_{max}^2 + d + 2\sqrt{-d \cdot \log \delta} - 2 \log \delta \right) |, \end{aligned} \quad (5.14)$$

where $x_0, \hat{x}_0$ are the quality reference and sampling data. For a noisy inverse problem scenario, the retrieved data with any finite T always indicates difference depends on $\sigma_t, \mu_t, \lambda, T, \bar{K}$, where $\bar{K}$ is the Lipschitz constant for the reverse process.

we have that the absolute value between the theoretical expectation and actual expectation after T period is given by

$$\|\mu - \mathbb{E}(\hat{x}_T)\| = \|(x_0 - \mu)e^{-\bar{\theta}_T}\| > 0$$

Similarly, the difference between theoretical variance and T-period variance also has a strictly positive difference. where $\bar{\theta}_t = \int_0^t \theta_s ds$.

Therefore, with finite T , the final state of the forward process can only reach a $\hat{x}_T$ rather than the theoretical stationary distribution, which we denote by x_∞ , we denote the retrieved image after T -periods from theoretical stationary distribution by $\hat{x}_0$, x_0 by the ground truth HQ image, $\hat{x}_T$ the true distribution after T iteration,

$$\begin{aligned} \|x_Q - f_{\text{OU}}(x_Q, \mu; t_0,)\| &= \|\hat{x}_T - x_\infty - (\hat{x}_0) - x_\infty\|_2^2 \\ &\geq \|\hat{x}_T - x_\infty\|_2^2 - \|\hat{x}_0 - x_\infty\|_2^2 \end{aligned} \quad (5.15)$$

Inside the norm, the first term is bounded below, since both $\hat{x}_T$ and x_∞ both follow a normal distribution and are independent of each other, the difference between those two random variables, we denote by z_T , that follows a normal distribution

$$\begin{aligned} &\mathcal{N}\left(\mu + (x_0 - \mu)e^{-\bar{\theta}_t} - \mu, \lambda^2(1 - e^{-2\bar{\theta}_t}) + \lambda^2\right) \\ &= \mathcal{N}\left((x_0 - \mu)e^{-\bar{\theta}_t}, \lambda^2(2 - e^{-2\bar{\theta}_t})\right) \end{aligned} \quad (5.16)$$

Therefore, we could rewrite the equality as:

$$\begin{aligned} &\|x_Q - f_{\text{OU}}(x_Q, \mu; t_0)\| \\ &\geq \left\| \|z_T\|_2^2 - \left\| \int_T^0 -\frac{d\sigma_t^2}{dt} \nabla_x \log p_t(x) dt + d\bar{\mathbf{w}}_t \right\|_2^2 \right\|_2^2 \\ &= \left\| (x_0 - \mu)^2 e^{-2\bar{\theta}_t} + \lambda^2(2 - e^{-2\bar{\theta}_t}) - C\sigma_{\max}^4 - \left\| \int_T^0 \sqrt{\frac{d\sigma_t^2}{dt}} d\bar{\mathbf{w}}_t \right\|_2^2 \right\|_2^2 \end{aligned} \quad (5.17)$$

Since we require $\lambda = \frac{\sigma_t^2}{2\bar{\theta}_t}$, we can find a $\sigma_{\max}$ such that $\sigma_t < \sigma_{\max}$ for all t . The last term only concerns the random noise, according to [116], we have that the last term is equivalent to the squared L_2 norm of a random variable from a Wiener process at time $t = 0$, with marginal distribution being $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma_T^2 \mathbf{I})$. The squared L_2 norm of ϵ divided by σ_T^2 is a χ^2 -distribution with d -degrees of freedom, we have the following one-sided tail bound, according to [116]

$$\Pr\left(\|\epsilon\|_2^2 / \sigma^2(t_0) \geq d + 2\sqrt{d \cdot -\log \delta} - 2\log \delta\right) \leq \exp(\log \delta) = \delta$$

Therefore, with probability $1 - \delta$, the HQ image and retrieved image has a lower bound of, (d is the number of dimension for x_0),

$$\begin{aligned} &|((x_0 - \mu)^2 - \lambda^2) e^{-2\bar{\theta}_T} + 2\lambda^2 \\ &- \sigma_{\max}^2 (C\sigma_{\max}^2 + d + 2\sqrt{d \cdot -\log \delta} - 2\log \delta)| \end{aligned}$$

A SDE modeling that has suffered complex degradation is considered 'corrupted':

Example 1. Based on the Prop. The conditional SDE diffusion is composed of three general stages, forward, backward, and sampling. Both time-reversal processes have been declared unstable under the degradation process μ according to the Prop. 2.

The following advantages would be offered by our measure-preserving dynamical system when formulated as SDEs:

Intuition 1. A two-sided measure-preserving random dynamical(MP-RDS) system formulation enables us to use the Poincare recurrence theorem, intuitively, with a two-sided MP-RDS φ_t , the Poincare recurrence theorem ensures that the system φ_t starts from terminal condition x_T , run backwards in time, will hit a region $(x_0 - \epsilon, x_0 + \epsilon)$ for small ϵ in finite time, where x_0 is the high-quality image.

Proposition 2 Given Eq. 5.3 and Eq. 5.8 and assume that the score function is bounded by C in L^2 norm, then the discrepancy between the reference and the retrieved data is, with probability $(1 - \delta)$ at least:

$$\|x_0 - \text{OU}(x_0, \mu; T, \theta)\|_2^2 \geq |((x_0 - \mu)^2 - \sigma_T^2/2\theta_T) e^{-2\bar{\theta}_T} + \sigma_T^2/2\theta_T - \sigma_{\max}^2 (C\sigma_{\max}^2 + d + 2\sqrt{-d \cdot \log \delta} - 2\log \delta)|, \quad (5.18)$$

Intuition 2. Intuitively, Proposition 2 provides a theoretical measurement on how the difference between finite iteration distribution and the asymptotic distribution of the OU-process in L^2 could further enlarge the discrepancy between the retrieved image and the actual HQ image. Discrepancies are typically explained by complex degradation vs. monotonic modeling.

For a noisy inverse problem, the retrieved data with any finite T depends on $\sigma_t, \mu_t, \lambda, T, \bar{K}$ (Lipshitz constant). This proposition also correlates to [142], where the lower bound of the distance between the high quality image and retrieved image in L^2 norm in our model is further enlarged by this discrepancy, which correlates to the term $((x_0 - \mu)^2 - (\sigma_T^2/2\theta_T)) e^{-2\bar{\theta}_T} + \sigma_T^2/2\theta_T$.

Another way to further minimize this bound is through the term $e^{-2\bar{\theta}_T}$ with $[0, T]$ normalized to $[0, 1]$. What we refer to as θ -schedule corresponds to the exact functional form of θ_t , several schedules can be set here, e.g., constant, linear, cosine, and log. At a high level, the discrepancy between the reference and the retrieved data stems from the divergence between the forwarded final state and the low quality image. Eq. 5.18 can be factored into three constituent parts: the data residual, the stationary disturbance, and the random noise. While conditional diffusion generation entails a trade-off between variability and faithfulness [231], the persistent discrepancy within the residual has a significant impact on the generalizability of solving the transitional tasks. This also establishes a connection with SDEdit [142] and CCDF [34]. When fitting inverse problems involving

paired data into diffusion models, while accounting for deviations and degradation, inserting them into Eq. 5.8 directly may not be the most effective strategy.

During sampling and inference with the degraded input y , the discrepancy identified in Prop. 2 intensifies. The complex degradations in y exacerbate the divergence from the expected μ distribution, significantly impacting the accuracy of the restored data $\hat{x}_0$.

Temporal Distribution Discrepancy during Sampling

Theorem 2 Suppose that both the drift $b_t(x)$ and diffusion $\sigma_t(x)$ term of a stochastic process x_t is Lipschitz continuous with some constant K , moreover, $x \in \mathbb{L}^2(\mathbb{F}, \mathbb{R})$ is a solution to the SDE

$$dx_t = b(t, x)dt + \sigma(t, x)dW_t$$

and initial condition b_0, σ_0 then we have that

$$\mathbb{E} [|x_T - x_0|^2] \leq CI_0^2, \quad (5.19)$$

such that

$$I_0^2 := \mathbb{E} \left[\left(\int_0^T |b_0| dt \right)^2 + \int_0^T |\sigma_0|^2 dt \right]$$

and C depends only on T, K , which is the running time and Lipschitz constant

Firstly, we have the following relationship:

$$\begin{aligned} x_T &\leq |x_0| + \int_0^T |b(t, x)| dt + \sup_{0 \leq t \leq T} \left| \int_0^t \sigma(s, x) dW_s \right|, \\ |x_T - x_0| &\leq \int_0^T |b(t, x)| dt + \sup_{0 \leq t \leq T} \left| \int_0^t \sigma(s, x) dW_s \right|. \end{aligned} \quad (5.20)$$

Squaring both sides, taking expectations, and applying the Burkholder-Davis-Gundy inequality, we get:

$$\begin{aligned} &\mathbb{E} [|x_T - x_0|^2] \\ &\leq C\mathbb{E} \left[\left(\int_0^T |b(t, x_t)| dt \right)^2 + \sup_{0 \leq t \leq T} \left(\int_0^t \sigma(s, x_s) dB_s \right)^2 \right] \\ &\leq C\mathbb{E} \left[\left(\int_0^T [|b_0| + |x_t|] dt \right)^2 + \int_0^T |\sigma(t, x_t)|^2 dt \right] \\ &\leq C\mathbb{E} \left[\left(\int_0^T |b_0| dt \right)^2 + \int_0^T [|\sigma_0|^2 + |x_t|^2] dt \right]. \end{aligned} \quad (5.21)$$

Remark: It should be noted that the constant C , which depends on T and K , varies from line to

line.

Next, we show that for any $\varepsilon > 0$, there exists a constant $C_\varepsilon > 0$ such that:

$$\sup_{0 \leq t \leq T} \mathbb{E} [|x_t|^2] \leq \varepsilon \mathbb{E} [|x_T^*|^2] + C_\varepsilon I_0^2. \quad (5.22)$$

Applying Itô's formula, we get:

$$d|x_t|^2 = [2x_t b(t, x_t) + |\sigma(t, x_t)|^2] dt + 2x_t \sigma(t, x_t) dB_t. \quad (5.23)$$

Considering the martingale property of the third term, integrating, and taking expectation on both sides, we have:

$$\begin{aligned} \mathbb{E} [|x_t|^2] &= \mathbb{E} \left[|x_0|^2 + \int_0^t [2x_s b(s, x_s) + |\sigma(s, x_s)|^2] ds \right] \\ &\leq \mathbb{E} \left[|x_0|^2 + \int_0^t [C|x_s|^2 + 2|x_s||b_0| + C|\sigma_0|^2] ds \right] \\ &\leq C \int_0^t \mathbb{E} [|x_s|^2] ds + 2\mathbb{E} \left[x_T \int_0^T |b_0| ds \right] + CI_0^2. \end{aligned} \quad (5.24)$$

Using Gronwall's inequality, we can prove (5.22) with the result above. Substituting (5.22) into (5.21) completes the proof, showing that the distance between x_T and x_0 in the L^2 -sense is bounded by $C(K, T)$.

With theorem 1 in hand, since the OU process is Lipschitz continuous, then the reverse process for the OU process is also Lipschitz continuous.

Now, suppose that the Lipschitz constant for the **reverse process** is given by $\bar{K}$. then we have that in L^2 -norm, the distance between any final state $x_T \in \mathcal{N}(\mu, \lambda)$ and the initial state(HQ image) is bounded by a constant that only depends on time T and Lipschitz constant $\bar{K}$, and the initial condition for the drift and diffusion term, where we denote as $C(\bar{K}, T, \mu, \lambda)$, which will be written as $C(\bar{K}, T)$ for short.

Now, since the time T is finite, where theoretically only when $T \rightarrow \infty$ would x_T converge to the theoretical stationary distribution, thus, if we denote the sample from T time steps by $\hat{x}_T$, and x_T by the sample from the theoretical stationary distribution, then we have $\mathbb{E}[x_T - \hat{x}_T] > \epsilon(T, K, \mu, \lambda)$, note that the K here is the Lipschitz constant for the forward process, and this distance strictly decrease in T .

Suppose that in inference, when the ground truth x_0 is unknown, the distance between ground truth x_0 and sample from the theoretical distribution x_T is bounded from below by

$$||x_0 - x_\infty|| > C(\bar{K}, T)I_0^2 + \epsilon \quad (5.25)$$

Then, we can see that the gap between x_T and $\hat{x}_T$ increases such bound, which is

$$\begin{aligned} \|x_0 - \hat{x}_T\| &\geq \|x_0 - x_\infty - (\hat{x}_T - x_\infty)\| \\ &> C(\bar{K}, T)I_0^2 + \epsilon(T, K, \mu, \lambda) \end{aligned}$$

where T is the time step for the forward process, and ϵ is some strictly positive constant. Now suppose that $\hat{x}_t$ is the solution to the reverse process, which runs for T periods in total, then we have that for $t \in [0, T]$, denote $x_0, \hat{x}_0$ as the original HQ image and the final state of the reverse process, respectively.

$$\begin{aligned} \|x^{Quality} - \text{OU}(x^{Quality}, \mu; t_0, \theta)\|_2^2 &= \\ \|x_0 - \hat{x}_0\| &= \|x_0 - \hat{x}_T - (\hat{x}_0 - \hat{x}_T)\| \geq \\ \| \|x_0 - \hat{x}_T\| - \|(\hat{x}_0 - \hat{x}_T)\| \| &> \\ \epsilon + C(\bar{K}, T)I_0^2 - C(\bar{K}, T)I_0^2 &= \epsilon \end{aligned} \tag{5.26}$$

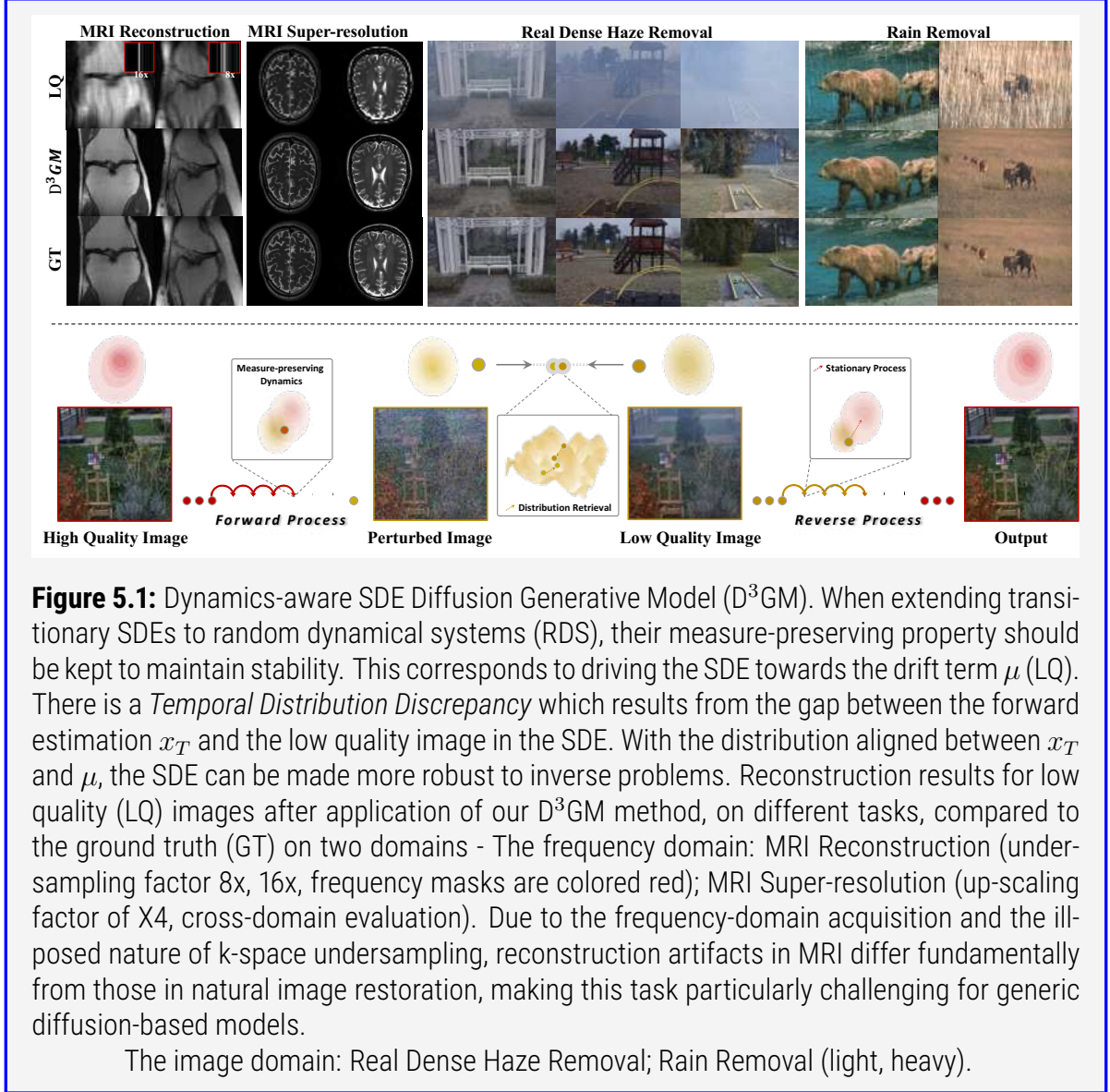
Therefore, the bias created in inference depends on the LQ image μ , stationary variance, Lipschitz constant, and the time steps T .

5.2.7 Towards Stability: Measure-preserving Dynamics in SDE Diffusion

In Sect. 5.2.6, we extrapolate and theorize the Temporal Distribution Discrepancy on μ and x in the diffusion model for the inverse problem. Our key idea is to combine the stationary process to alleviate the Temporal Distribution Discrepancy problem following the measure-preserving dynamics from RDS. Recall that our ‘measure’ is not only the measurements (*i.e.*, degraded image), but also represents the *invariant measure* (distribution) of the RDS.

We begin by describing the forward and reverse processes of the D³GM, which serves as a stable bridge between the quality data and the counterpart measurement. We adapt score-based training methods to estimate this SDE. Following this, we describe the essential constructions for preserving the stationary process in the diffusion model and solving for Temporal Distribution Discrepancy on an orthogonal basis compared to current transitional diffusion models.

Measure-preserving Dynamics with the Stationary Process. Following Prop. 1 in a SDE Diffusion from 0 to T , the corresponding ‘attractors’ (states) can be viewed as $\mathcal{N}(\mu + (x_0 - \mu)e^{-\bar{\theta}_t}, \sigma_t^2(1 - e^{-2\bar{\theta}_t})/2\theta_t)$. We can guide SDE Diffusion towards a stable and robust solution based on the properties of Measure-preserving in RDS. It can be extended to impose that for every t , $\frac{\sigma_t^2}{2\theta_t} = \lambda^2$, where λ is the variance of the designated stationary measure forward process. This convention allows us to reduce the regularization on two variables σ_t, θ_t to just one variable to satisfy the prop-



erty of the measure-preserving dynamics in the asymptotic sense, *i.e.*, $\lim_{t \rightarrow \infty} \varphi(t, s; \omega)x = \lim_{t \rightarrow \infty} \varphi(t - s, 0; \vartheta_s \omega)x$. This convention allows us to characterize the attractor of the system as $K(\omega) = \mathcal{N}(\mu, \lambda^2)$.

The definition and constraint of the attractor are significant; without imposing this constraint, the measure-preserving property cannot be maintained, and the system would degrade into a Coefficient Decoupled SDE (Coe. Dec. SDE), we also analyse this in Fig. 5.3 and Tab. 5.4.

Example 2. When $\frac{\sigma}{\theta} \rightarrow \infty$, the attractor becomes excessively large, reducing the significance of μ . SDEs exhibiting this behavior are defined as Coefficient Decoupled SDEs. In practice, μ demonstrates non-infinite properties as an input image, while the corresponding sigma and theta are indeed unconstrained. In such decoupled forms, the coefficients of the attractor size increases $\frac{\sigma}{\mu}$, diminishing the significance of μ .

5.2.7.1 Implementation details

Model Implementation: Our exploration into mitigating Temporal Distribution Discrepancy in diffusion models employs two neural network architectures, each catering to different dataset complexities.

We utilize the adopted UNet, a staple in DDPM [81] and DDIM [191] frameworks, chosen for its widespread use and strong benchmarking capabilities. By solving the discrepancy through this established structure, we achieve state-of-the-art results on synthetic data, showcasing the potential of improving transanary SDE diffusion models in terms of Temporal Distribution Discrepancy and stationary process. Additionally, the use of this prevalent architecture allows for comprehensive analysis and discussion.

Real-world data with its inherent complexity, such as combined degradations, large resolutions, and extensive interdependencies, requires an architecture beyond the conventional UNet. Our improved model incorporates Squeeze-and-Excitation [141] and NAF [30], explicitly designed to capture intricate feature interrelations. While these models do not seek to innovate the architectural paradigm, it provides a solid baseline that provides better feature extraction ability than UNet performance in demanding scenarios.

Table 5.2: Datasets parameters and split setting. IOP details of the scan parameters are not available.

Datasets	Data setting		Collection		Sequence parameters			
Source domain	Subjects	Slices	Hospitals	Scanner	Repetition time	Echo train length	Matrix size	Receiver coil
HH (IXI Brain)	184	60	Hammersmith	Philips 3T	5725	16	192 x 187	Single
<i>Target domain</i>								
Guys (IXI Brain)	30	60	Guy's Hospital	Philips 1.5T	8178	16	Unnormalized	Single
IOP (IXI Brain)	30	60	Institute of Psychiatry	GE 1.5T		Unknown		Single

Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
WD UNet	27.31	0.8322	0.227
SN UNet	28.63	0.8687	0.144
UNet	31.03	0.9001	0.058

Table 5.3: Evaluation of the Lipschitz continuity.

Additional details: The U-Net we adopted is similar to DDPM as described in [31, 33, 135], where the improved model incorporates Squeeze-and-Excitation [141] to replace the Attention module within NAF [30]. EDSR [125] is employed as the base model for TTA-based comparison methods in MRI super-resolution. In different downstream tasks, we follow the common setting of the latest compared methods: Deraining [135], Real Dehazing [163], MRI Reconstruction [88], and MRI super-resolution [38] and [125]. Below are more specific details about MRI reconstruction and MRI superresolution. For most of the experiments, the training patch-size is set to 128x128 with a batch

Table 5.4: Exemplary deraining results of (A) SGM, (B) transitional SGMs: Coefficient Decoupled SDE, OU SDE, and D³GM.

Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
A SGM	27.27	0.840	0.144
Coef. Dec. SDE	26.18	0.826	0.205
B OU SDE	30.58	0.900	0.051
D ³ GM (ours)	32.41	0.912	0.040

size of 16. We utilize the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, a learning rate of 10^{-4} with a decay strategy. Our models are trained on three RTX 6000 GPUs for about four days, each with 40GB of memory. Random seed is 42. All mathematical variants follow the cosine schedule as per [146]. The variance λ is set to 10 for the OU and stationary processes. In the coef. decoupled SDE, σ is kept decoupled and does not vary with θ . We observed that the adaptability of τ to tasks is contingent upon the task’s corruption for model stability and generalizability.

τ is a hyperparameter as one of the ways to introduce Measure-preserving Dynamics to shape more stable SDE diffusion. It is informed by the deviation between the degraded image and the expected high-quality image. For tasks with moderate intra-domain deviations, τ is set to 2, while for tasks with substantial cross-domain and degradation discrepancies, a larger τ is used. This metric is based on RDS settings of different degradations, rather than learned. A MLP can be used to learn τ as adaptive metric, although it is not comparable with obvious improvements.

SGM, and Transitional SGMs vs. D³GM: We perform qualitative [5.3] and quantitative [5.4] analyses using variants of closely related formulations for Prop. 1 and 2 and evaluate across (A) SGMs and (B) transitional SGMs. (A) uses a common score-based SDE, (B) uses a Coefficient Decoupled SDE (e.g., variance exploding SDE with the drift term μ) according to Prop. 1 and OU SDE, alongside our D³GM.

MRI Reconstruction: We applied 584 proton-density weighted knee MRI scans without fat suppression. These were subsequently partitioned into a training set (420 scans), a validation set (64 scans), and a testing set (100 scans). For each scan, we extracted 20 coronal 2D single-channel complex-valued slices, predominantly from the central region with the uniform size of 320 x 320. Our test set differs from the 200 reported in [88], possibly as a result of the change in the official versions of FastMRI. We subjected all experiments to Cartesian under-sampling masks with undersampling factor 8x and 16x. In the undersampled MRI scans, the acquisition process entails sampling a fractional subset of the Fourier-space (k-space), typically governed by a mask along the dimension of undersampling rate. Undersampling inherently induces aliasing artifacts in the resultant images. Considering the domain deviation, we did not supplement the extra test set into the final 100 samples. Otherwise, we stayed consistent with the details of the paper. For a robust comparison, we benchmarked against a diverse set of deep learning-based state-

of-the-art reconstruction methods, including CNN-based approaches such as D5C5 [182] and DA-GAN [225], the Transformer-based SwinMR [89], diffusion model-inspired DiffuseRecon [155], and the CDiffMR [88]. Quantitative results in Tab. 5.8 exhibit a differing trend from the image domain. The task-specific diffusion models achieve better results and are more capable of capturing the complex degradation that occurs in the frequency domain.

Generally, MRI uses a mask in the phase-encoding direction (the shortest anatomical direction [145]) to model the complex degradation caused by undersampling. In the knee data, based upon the uncertainty of clinical diagnosis (longitudinal artifacts can significantly confuse the diagnosis of meniscal injury), we also masked the frequency direction, which will cause resolution reduction, deterioration of image features, and longitudinal artifacts, more qualitative results can be found in the next section.

MRI Super-resolution: IXI³ contains clinical T1- and T2-weighted scans from three hospitals with different imaging protocols: HH, Guys, and IOP. We selected 184 HH T2 subjects as the source-domain data [train/val/test ratio: 7:1:2], and 30 subjects each from Guys and IOP as two target-domain datasets without degradation, acquisition parameters, datasets, and patient-wise crossovers. The central 60 slices were selected.

We consider two benchmark ideas: (1) Blind SR degradation methods in frequency SFM [51] and spatial PDM [237] domain. (2) Source-free SR adaptation ACT [38] using external priors and test-time adaptation proposed initially for MRI CST [38] with data consistency in the source-domain. We are concerned that existing work on robustness and generalization focuses on medical image segmentation and natural images, and rarely on super-resolution and reconstruction of medical images. Employing the available evaluation criteria in this inevitable problem in practice makes it difficult to reflect the performance of D³GM. Therefore, domain-aware datasets isolation standard for public datasets is designed for the SRR adaptation to cross-domain data rather than same-domain in a source-free manner. The publicly available data were split into several subsets based on hospitals, scanner, acquisition parameters, modality, and anatomy as illustrated in Table 5.2. Explicit reference standards implicitly correspond to various degradation patterns, thus enabling the isolation of natural degradation patterns in source training domain and target-testing domain. In addition to this, different artificial degradation patterns for the subsets in the training and testing domains are employed: K-space truncation downsampling was applied to obtain LR data in the source domain and a kernel degradation [17] was applied in target domain.

We reproduce two types of benchmark ideas on the top of EDSR [125] backbone to achieve the multi-purpose goal: (1) *Repurposed blind SR (BSR) for cross-domain data*: We utilized BSR degradation methods in frequency SFM [51] and spatial PDM [237] domain. (2) *Test-time adaptation (TTA)*: We compare to a source-free TTA method ACT [38] using external priors and a second TTA method proposed initially for MR reconstruction CST [38] with cycle consistency at source-domain. The

³<http://brain-development.org/ixi-dataset/>

settings and adaptation strategies of the comparison methods were used directly.

Lipschitz Continuity for Stationary Process:

We hypothesize that ensuring Lipschitz continuity of the neural network is pivotal for the convergence and stability of the diffusion process, particularly in the context of achieving a stationary process. Theoretically, Lipschitz continuity offers two key benefits: (1) it mitigates the impact of small perturbations in input data or model parameters, thus safeguarding against excessive output variability which could cause numerical instabilities or result in an ill-posed problem, and (2) it guarantees the existence of a unique solution to the diffusion process, underpinning the reliability and convergence of the numerical methods employed to solve these equations.

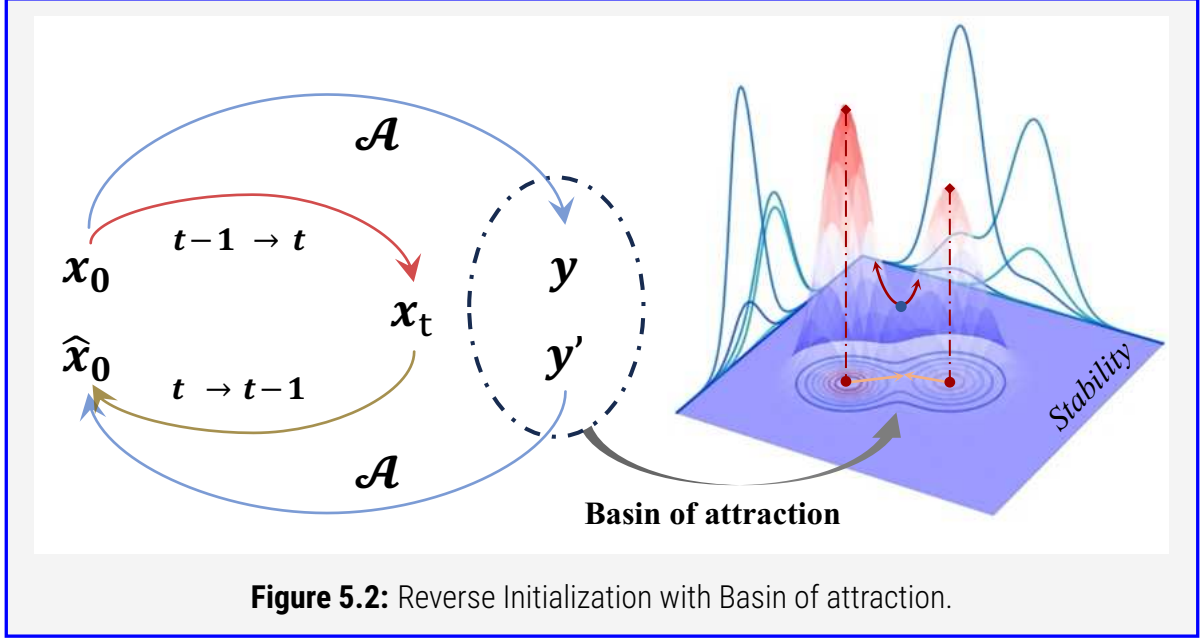
To instantiate these theoretical benefits within our architecture, we integrated spectral normalization (SN) [144] and weight decay (WD) [129] into a U-Net structured score network. In the case of SN, it is achieved by rescaling each layer's weight matrix $W^{(l)}$ by its spectral norm, $\sigma_{max}(W^{(l)})$, to obtain a normalized weight $\tilde{W}^{(l)} = \frac{W^{(l)}}{\sigma_{max}(W^{(l)})}$. By doing so, we intend to control the overall Lipschitz constant of the network for robust score matching within our diffusion model framework. A quantitative result can be seen in Tab. 5.3.

5.2.7.2 Additional Insights

Basin of Attraction in Reverse Sampling: Within our diffusion framework, we establish a forward process with variety to accommodate a wide range of potential corruptions, ensuring the desired final distribution close to the expected distribution. We consider that our diffusion models inherently encompass the forward operator $\mathcal{A}$ within their structure. The transition from high to low-quality images is implicitly encoded in the diffusion pathway, hence the additional forward operator guidance might not contribute supplementary information, which we initially hypothesized would enhance the reverse process.

Thus, the keypoint transfers from the $\mathcal{A}$ to the initial y' (Detailed analysis based on Prop. 2 is provided in Appx. 5.2.6.3). Also we found this problem fall into the Basin of Attraction (BA) in dynamical system as shown in Fig. 5.2. BA can be interpreted as the quality of the attractor of the degraded image in the reverse process here. A BA-guided initialization might be a more effective approach for posterior sampling in transitional SDE diffusion models. By initializing the reverse process closer to the expected solution, we may bypass the initial hurdle of distribution mismatch.

Based on Prop. 2 since the temporal distribution discrepancy always exists as long as the running time T is finite, and we want the final state x_T to be as close as possible to the distribution of x_∞ . Therefore, we introduce τ such that given T , the distribution of x_T follows $\mathcal{N}(\mu(1 - e^{\bar{\theta}_T}) + x_0 e^{\bar{\theta}_T}, \tau^2 \lambda^2 (1 - e^{2\bar{\theta}_T}))$, and x_∞ follows $\mathcal{N}(\mu, \tau^2 \lambda^2)$, with $\tau > 1$, we increase the possibility of a sample $\tilde{x}_T$ from x_T to become closer to the distribution of x_∞ , and thus serves as a plausible initial state for the reverse process. We can control how much to close the distributions, either



by increasing the stiffness τ at the cost of potentially destabilizing the reverse process, or by decreasing τ to further smooth the density functions of both distributions at the cost of more reverse iterations.

By connecting the inverse problem with the analysis above, we clarify the discrepancy in the stationary modeling process from measure-preserving dynamics and thereby improve the generalization of diffusion processes and the accuracy of the reverse process. This is particularly important for accommodating the diversity of degradation states and to ensure accurate sampling.

Table 5.5: $\mu_t(x_t, t)$, $v_t(x_t, t)$ solutions with various θ_t , σ_t .

	$\mu_t(x_t, t)$	$v_t(x_t, t)$
Lin	$e^{\theta \frac{x^2}{2}} \mathbf{x}_0 + \left(1 - e^{\theta \frac{x^2}{2}}\right) \mathbf{y}$	$\tau^2 \lambda^2 \left(1 - e^{\theta x^2}\right) \mathbf{y}$
Log	$e^{\frac{\theta}{k} \log(1+e^{kt})} \mathbf{x}_0 + \left(1 - e^{\frac{\theta}{k} \log(1+e^{kt})}\right) \mathbf{y}$	$\tau^2 \lambda^2 \left(1 - e^{\frac{2\theta}{k} \log(1+e^{kt})}\right) \mathbf{y}$
Cos	$e^{-\theta \left(t - \frac{\sin(\theta t)}{\theta}\right)} \mathbf{x}_0 + \left(1 - e^{-\theta \left(t - \frac{\sin(\theta t)}{\theta}\right)}\right) \mathbf{y}$	$\tau^2 \lambda^2 \left(1 - e^{-2\theta \left(t - \frac{\sin(\theta t)}{\theta}\right)}\right) \mathbf{y}$
Quad	$e^{\theta \frac{x^3}{3}} \mathbf{x}_0 + \left(1 - e^{\theta \frac{x^3}{3}}\right) \mathbf{y}$	$\tau^2 \lambda^2 \left(1 - e^{2\theta \frac{x^3}{3}}\right) \mathbf{y}$

Forward Process. We describe the forward process as: $dx_t = \theta_t(\mu - x_t)dt + \tau\sigma_t dW_t$, parameterized by τ to calibrate the SDE modeling, μ is the state mean. The parameters θ_t and σ_t , both being time-dependent and strictly positive, correspond to the rate of mean reversion and the stochastic volatility, respectively. The selection of θ_t and σ_t offers flexibility in Tab. 5.5 c.p., Sect. 5.2.6.

Considering the trade-off between complexity and effectiveness, Cos has been chosen for both θ_t and σ_t . This aims at capturing complex temporal dynamics in a computationally tractable manner, thereby optimizing the balance between the performance and calculation convenience.

In the forward process, the mean μ_t approaches the low-quality image with $E(x_t) = \mu_t$, while the variance tends toward the stationary variance $\text{var}(x_t) = \frac{\tau^2 \sigma^2}{2\theta}$. Essentially, the forward SDE transitions the high-quality image to a low-quality counterpart infused with Gaussian noise. The discretized SDE for the forward process is $x_{t_i} = x_{t_{i-1}} + \theta_{t_{i-1}}(\mu - x_{t_{i-1}})\Delta t + \tau\sigma_{t_{i-1}}\Delta W_i$. We employ a transition strategy utilizing a varied stationary variance. Additionally, we execute an unconditional update, which operates without the need for matching in the reverse process. These not only allow image corruption but also provides effective adaptability for improvements.

Reverse Process. The reverse process aims at reconstructing the original image by gradually denoising a low quality image. It utilizes the score of the marginal distribution, denoted as $\nabla_x \log \hat{p}_t(x)$, and is governed by:

$$dx_t = [\theta_t(\mu - x_t) - \tau^2 \sigma_t^2 \nabla_x \log \hat{p}_t(x)] dt + \tau \sigma_t d\widehat{W}_t. \quad (5.27)$$

This closely mirrors the forward process and incorporates an additional drift term proportional to the score of the marginal distribution. The ground truth score for this process, necessary for training our generative model, is:

$$\nabla_x \log \hat{p}_t(x | x_0) = -\frac{x_t - \mu_t(x)}{v_t}, \quad (5.28)$$

where $\mu_t(x)$ represents the random attractor of the process at time t , and v_t is the variance. Our training objective is defined as the minimization of the expected discrepancy between the predicted and true scores over the data distribution:

$$\theta^* = \arg \min_{\theta} \mathbb{E}_{t, (\mathbf{x}_0, \mathbf{y}), \mathbf{z}, \mathbf{x}_t} [w \|S_{\theta}(\mathbf{x}_t, \mathbf{y}, t) - \nabla_{\mathbf{x}_t} \log p_{0t}(\mathbf{x}_t | \mathbf{x}_0, \mathbf{y})\|_2^2], \quad (5.29)$$

where $w = -1/\tau^2$ is a time-dependent weighting function, and S_{θ} denotes the score network parameterized by θ which approximates the score of the marginal distribution. The optimization is conducted over the network parameters θ , under the expectation with respect to the time variable t , the initial image $\mathbf{x}_0$, noisy image $\mathbf{x}_t$, and data $\mathbf{y}$.

5.2.8 Evaluation and results

Experimental Settings: We evaluate D³GM on various challenging restoration and reconstruction problems. We initially analyze our method by examining its performance with closely related diffusion formulation variants. Subsequently, we benchmark D³GM against the state-of-the-art techniques in these domains. For comprehensive evaluation across all experiments, we report the PSNR [84] and SSIM [217] for pixel- and structural-level alignment, LPIPS [239] for measuring perceptual variance.

Table 5.6: Quantitative results for Rain100H and Rain100L.(best in bold and second best underlined)

Method	Rain100H			Rain100L		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
JORDER [228]	26.25	0.835	0.197	36.61	0.974	0.028
IRCNN [135]	29.12	0.882	0.153	33.17	0.958	0.068
PreNet [167]	29.46	0.899	0.128	37.48	0.979	0.020
MPRNet [233]	30.41	0.891	0.158	36.40	0.965	0.077
MAXIM [204]	30.81	0.903	0.133	38.06	0.977	0.048
VPB (CD) [248]	30.89	0.885	0.051	38.12	0.968	0.023
IR-SDE (OU) [135]	31.65	<u>0.904</u>	<u>0.047</u>	<u>38.30</u>	<u>0.981</u>	<u>0.014</u>
D ³ GM	32.41	0.912	0.040	38.40	0.982	0.013

Table 5.7: Quantitative results for O-HAZE and Dense-Haze.

Methods	O-HAZE		Dense-Haze	
	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$
DCP [75]	16.78	0.653	12.72	0.442
DehazeNet [24]	17.57	0.770	13.84	0.430
GFN [168]	18.16	0.671	-	-
GDN [131]	18.92	0.672	14.96	0.536
MSBDN [48]	24.36	0.749	15.13	0.555
FFA-Net [162]	22.12	0.770	15.70	0.549
AECR-Net [221]	-	-	15.80	0.466
SGID-PFF [12]	20.96	0.741	12.49	0.517
Restormer [234]	23.58	0.768	15.78	0.548
Dehazer [70]	25.11	0.777	16.62	<u>0.560</u>
MB-TF [163]	<u>25.31</u>	<u>0.782</u>	<u>16.44</u>	0.566
D ³ GM	26.23	0.786	15.85	0.551

5.2.8.1 Stability: Illustrations of the Measure-preserving Dynamics within Diffusion Models

SGM, and Transitional SGMs vs. D³GM:

We perform qualitative and quantitative analyses using variants of closely related formulations for Prop. 1 and 2 and evaluate across (A) SGMs and (B) transitional SGMs. (A) uses a common score-based SDE, (B) uses a Coefficient Decoupled SDE (e.g., variance exploding SDE with the drift term μ) according to Prop. 1 and OU SDE, alongside our D³GM.

Following Tab. 5.1, VPB [248] can be regarded as a Coefficient Decoupled SDE, and IR-SDE [135] as an OU SDE. Fig. 5.3 provides an intuitive visualization of the stability properties analyzed in Prop. 1 and Prop. 2. Standard SGMs and transitional SGMs suffer from unstable or drifting trajectories caused by temporal distribution discrepancy accumulated during the diffusion process. In contrast, the proposed D³GM preserves the invariant measure of the underlying random dynamical system, resulting in trajectories that consistently converge towards the desired attractor. This empirical behavior aligns with our theoretical analysis of measure-preserving dynamics and explains the improved robustness observed in subsequent inverse problem experiments. This highlights the reliance of other techniques, e.g., score-based SDEs, on retrospective measurement consistency corrections.

Simulated Deraining: We evaluated D³GM together with the state-of-the-art deraining strategies: (1) OU SDE method IR-SDE [135], Coefficient Decoupled (CD) VPB [248] and other CNNs [167] [204] [228] [233]. We use two of the most renowned synthetic raining datasets: Rain100H [227] and Rain100L [227]. Rain100H contains 1800 pairs of images with and without rain, along with 100 test pairs. As for Rain100L, it consists of 200 pairs for training and 100 pairs for testing. We present results based on the PSNR, SSIM, and LPIPS metrics.

Quantitative results from the two raining datasets are presented in Tab. 5.6. Based on both dis-

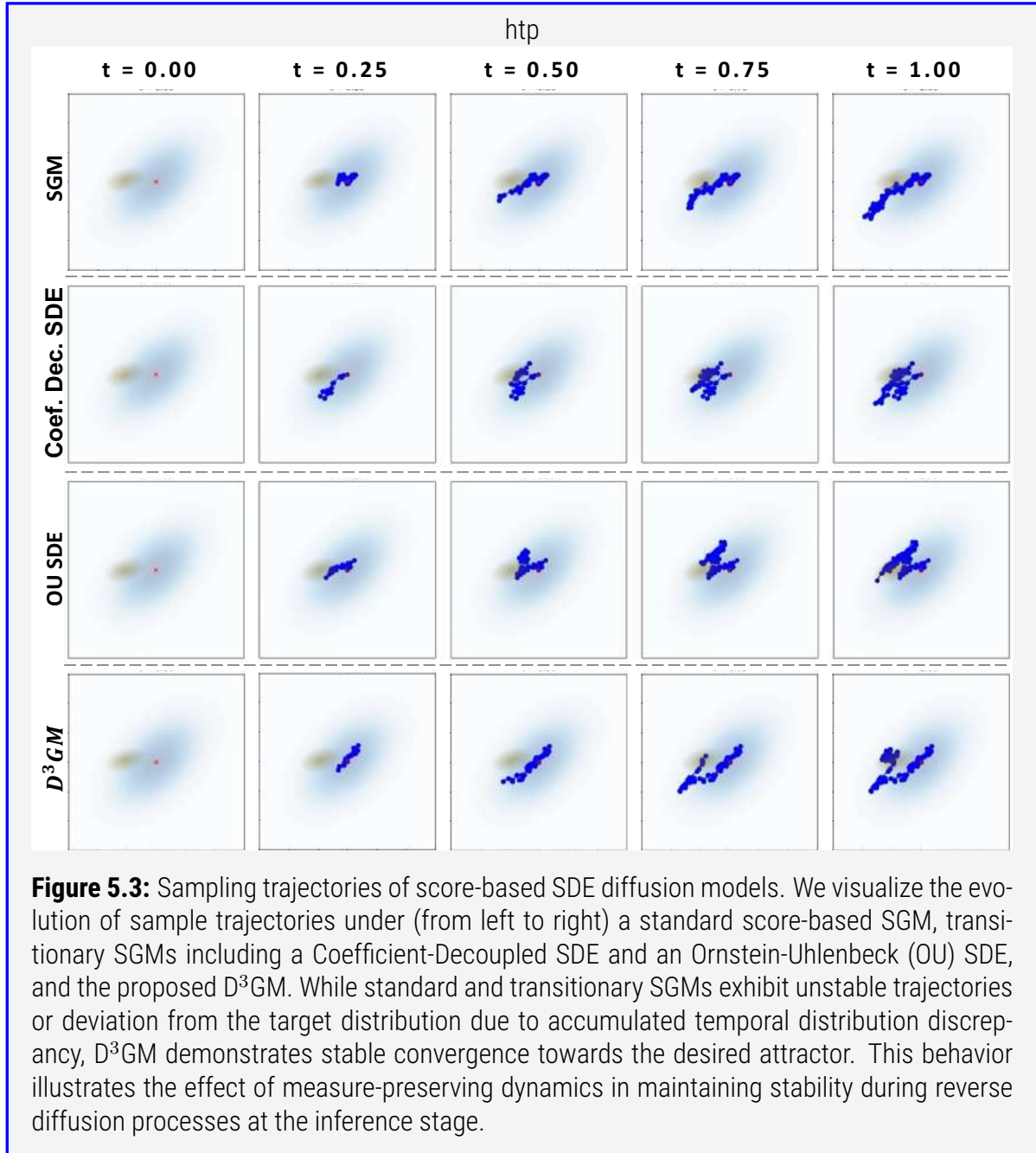




Figure 5.4: O-HAZE and Dense-Haze.

tortion and perceptual metrics, D^3GM is capable of generating the most realistic and high fidelity results as shown in Fig. 5.5.

5.2.8.2 Generalizability: D^3GM for real-world data

Case Study 1: Dehazing. We utilize the real-world datasets O-HAZE [8], Dense-Haze [7], which contain 45 and 55 paired images, respectively. We use the last 5 images of each dataset as the testing set and the rest as the training set following the common split of other methods. Results are shown in Tab. 5.7 and Fig. 5.4. Our work improves results on O-HAZE both quantitatively and qualitatively. Smaller improvements are observed on Dense-Haze. This can be attributed to the severe signal corruption of the Dense-Haze data. A combination with tailored task-specific, Transformer-based methods [70, 163] might lead to further performance gains for such data. Such extensions are beyond the focus of this paper. Qualitatively D^3GM achieves excellent visual results (Fig. 5.1 and Fig. 5.4).

Case Study 2: MRI Reconstruction. MRI data is represented in the complex-valued frequency domain, which is distinctly different from the natural image domain. We utilized the fastMRI dataset [235], containing single-channel, complex-valued MRI samples. For a robust comparison, we benchmarked against a diverse set of deep learning-based state-of-the-art reconstruction methods. Although our method does not have a task-specific design, we still get comparable performance (more details and results are provided in Appx. 5.2.7.1). Fig. 5.1 illustrates our reconstruction results from masked k-space data for 8x and 16x acceleration, *i.e.*, under-sampling for faster data acquisition.

Case Study 3: MRI Super-resolution (SR). IXI⁴ dataset is the largest benchmark considered in our MRI SR evaluation. Clinical MRI T2-weighted (T2w) scans are collected from three hospitals with different imaging protocols: HH, Guys, and IOP. For investigating our cross-domain generalization and robustness, a challenging task for both MRI SR and natural image restoration, we trained on

⁴<http://brain-development.org/ixi-dataset/>

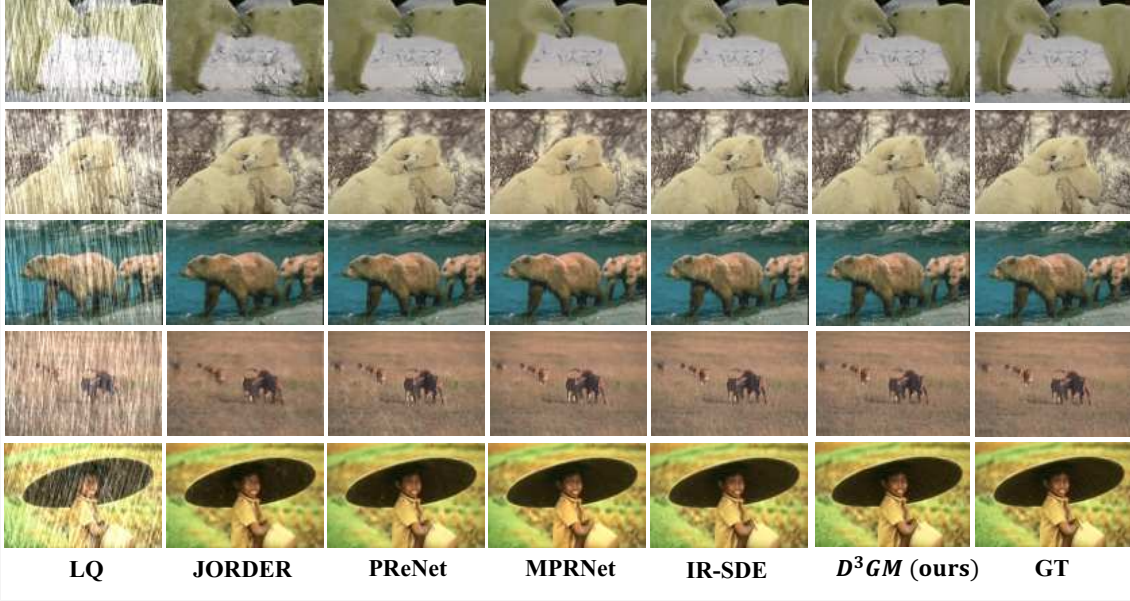


Figure 5.5: Rain100H and Rain100L.

Figure 5.6: Qualitative results for (a) deraining and (b) dehazing.

Table 5.8: Quantitative results for fastMRI dataset with acceleration rates x8 and x16.

Method	Undersampling factor 8x			Undersampling factor 16x		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
ZeroFilling	22.74	0.678	0.504	20.04	0.624	0.580
D5C5 [182]	25.99	0.719	0.291	23.35	0.667	0.412
DAGAN [225]	25.19	0.709	0.262	23.87	0.673	0.317
SwinMR [89]	26.98	0.730	0.254	24.85	0.673	0.327
DiffuseRecon [155]	27.40	0.738	0.286	25.75	0.688	0.362
CDiffMR [88]	27.26	0.744	<u>0.236</u>	25.77	0.707	<u>0.293</u>
D³GM	27.92	<u>0.740</u>	0.175	25.26	<u>0.701</u>	0.153

Table 5.9: Quantitative results for IXI MRI SR on unseen datasets.

X4	HH	Guys		IOP	
		PSNR↑	SSIM↑	PSNR↑	SSIM↑
IXI T2w	EDSR [125]	23.03	0.700	25.10	0.727
	SFM [51]	23.28	0.711	25.18	0.731
	PDM [237]	22.89	0.709	27.93	<u>0.851</u>
	ACT [164]	22.80	0.707	26.38	0.826
	CST [38]	<u>23.70</u>	<u>0.714</u>	<u>28.55</u>	0.837
	D³GM	25.13	0.799	28.60	0.863

HH data with k-space truncation, and tested on Guys and IOP with kernel degradation with an up-scaling factor of X4.

The methods are tested under unseen data conditions, including different acquisition parameters, MRI scanners (different vendors and field-strengths) and unseen degradations. With D³GM we are able to demonstrate varying degrees of improvement, as well as generalizability to the discrepancy within the training domain and across the test domain as shown in Tab. 5.9.

5.2.9 Discussion and limitations

Other works, like VPB (CD) [248], l²SB [128] are based on diffusion bridges assuming that clean and degraded images are already close. Thus, the tractability of the reverse process heavily relies on the validity of the assumed Dirac delta distribution. IR-SDE [135] employs the mean-reverting

SDE theorem based on running the reverse SDE with *instability*. Since unstable errors accumulate in each step, this model will eventually become unable to learn the transformation, e.g., degradation. DPS [31] and CDDDB [32] assume that the degradation process is known, or linear operations are directly used to simulate the degradation process, which limits the generalizability of the method. In contrast, D³GM is built on the theorem of Measure-preserving RDS, which bridges clean and degraded image distributions while taking both degradation and measurements into account. Moreover, D³GM can be extended to a two-sided solution operator (tractability) with a flow map according to Prop. 1.

Limitations in natural images. Even though our results are better than others when the degradation process is very severely corrupted (e.g., real dehazing), the overall quality of the restored image is still limited, which is consistent with Prop. 2. This might be alleviated via guiding the sampling process with priors and enhanced μ , such as posterior sampling or degradation maps on the data manifold, but such approaches are still limited as shown in Tab. 5.10.

Limitations in MRI Reconstruction and Domain-specific Considerations While the proposed D³GM framework demonstrates strong stability and generalizability across a wide range of inverse problems, its performance on MRI reconstruction exhibits comparatively smaller gains than those observed in natural image restoration tasks. This behavior is not unexpected and can be attributed to fundamental differences in imaging physics, degradation mechanisms, and evaluation protocols between MRI and natural images.

Unlike natural images, which are acquired directly in the spatial domain, MRI data are measured in the frequency domain (k-space) through a Fourier encoding process. Reconstruction from under-sampled k-space therefore constitutes a highly ill-posed inverse problem, where aliasing artifacts arise from structured sampling patterns rather than additive noise. As a result, the degradation process in MRI cannot be fully characterized by spatial-domain corruption models commonly used in natural image restoration.

Moreover, MRI measurements are typically complex-valued and affected by additional factors such as coil sensitivity profiles, scanner-specific acquisition protocols, and reconstruction pipelines. These factors introduce domain-specific variability and non-stationarity that further challenge the generalization of data-driven diffusion models. In contrast, many natural image benchmarks assume more homogeneous degradation processes and well-defined forward models.

Another important limitation lies in the definition of ground truth for MRI reconstruction. In practice, the reference images used for evaluation are often obtained from fully sampled or retrospectively reconstructed data, which themselves are approximations rather than absolute physical ground truth. Consequently, commonly used quantitative metrics such as PSNR and SSIM may not fully capture perceptual fidelity or diagnostic relevance, particularly when small discrepancies in k-space can lead to visually subtle but numerically significant differences.

Despite these challenges, D³GM achieves competitive performance on MRI reconstruction and super-resolution tasks without relying on task-specific architectural designs or explicit physical modeling. This highlights the robustness of the proposed measure-preserving diffusion framework, while also indicating that further integration of MRI-specific priors and forward operators may be beneficial for closing the remaining performance gap.

Computational Complexity vs. Performance. Tab. 5.10 highlights that prior work is often tailor-made for a specific subset of tasks and thus also generalisation-limited for challenging environments in practice. D³GM’s focus on a generic robust solutions from an RDS perspective can mitigate this, while maintaining en-par performance with *task-specific* approaches in Tab. 5.11.

Table 5.10: Comparison to tailor-made approaches.

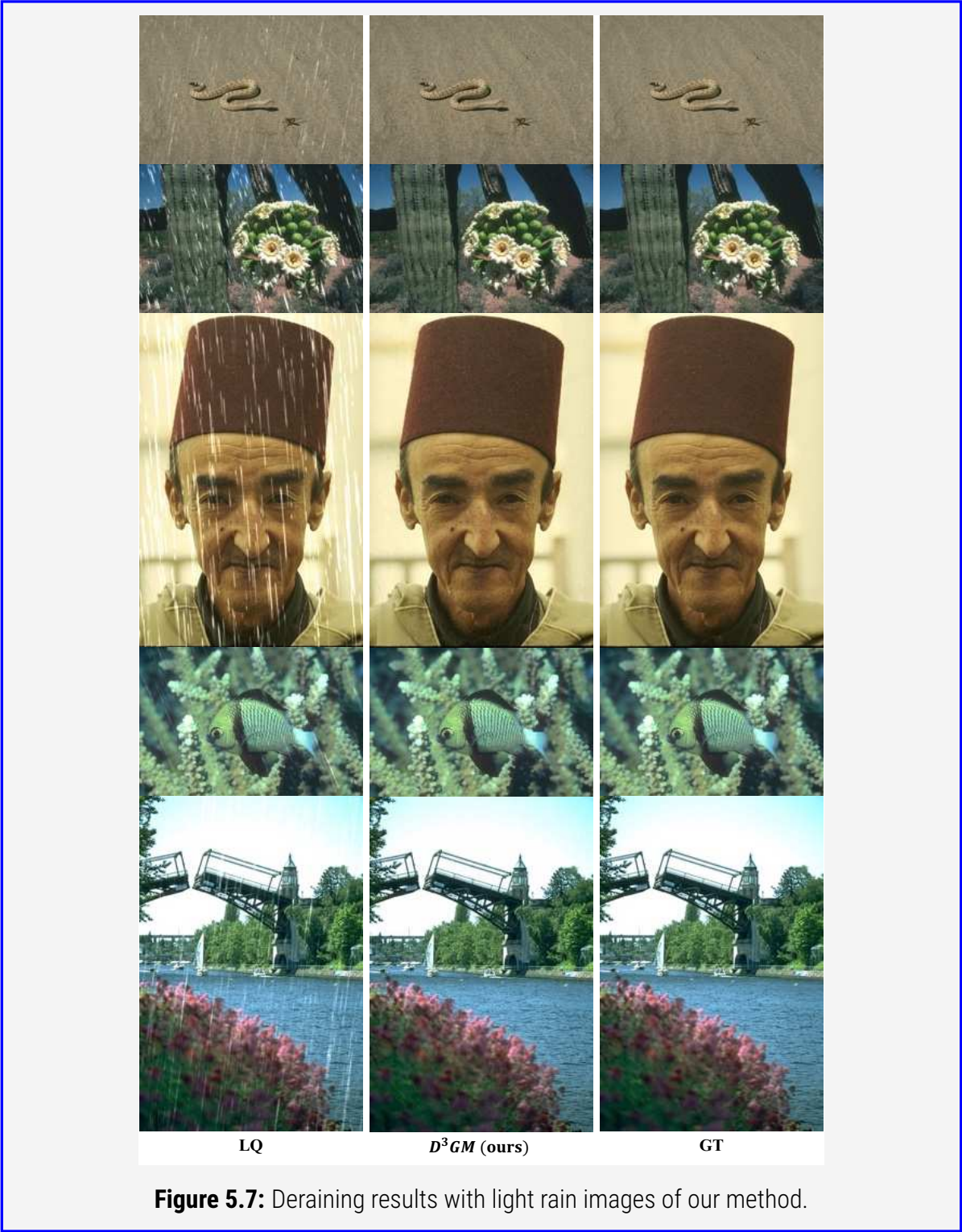
Real Dehazing	Diffu. for Inverse Problems.		Transitionalary SGMs.		
Methods	DPS [31]	CDDb [32]	I ² SB [128]	IR-SDE [135]	D ³ GM
PSNR↑	18.63	21.55	21.51	24.52	26.23
SSIM↑	0.448	0.591	0.583	0.691	0.786

Table 5.11: D³GM vs. deraining works.

Deraining	rain200H		Model Complexity	
Methods	PSNR↑	SSIM↑	Param.	FLOPs
DRSformer [26]	32.17	0.933	33.7M	242.9G
D ³ GM	32.21	0.925	36.5M	104.7G

5.2.10 Summary

The proposed D³GM framework enhances the stability and generalizability of SDE-based diffusion methods for challenging inverse problems. Our approach, grounded in measure-preserving dynamics of random dynamical systems, ensures broad applicability and relevance. We demonstrate D³GM’s effectiveness across various benchmarks, including challenging tasks like MRI reconstruction.



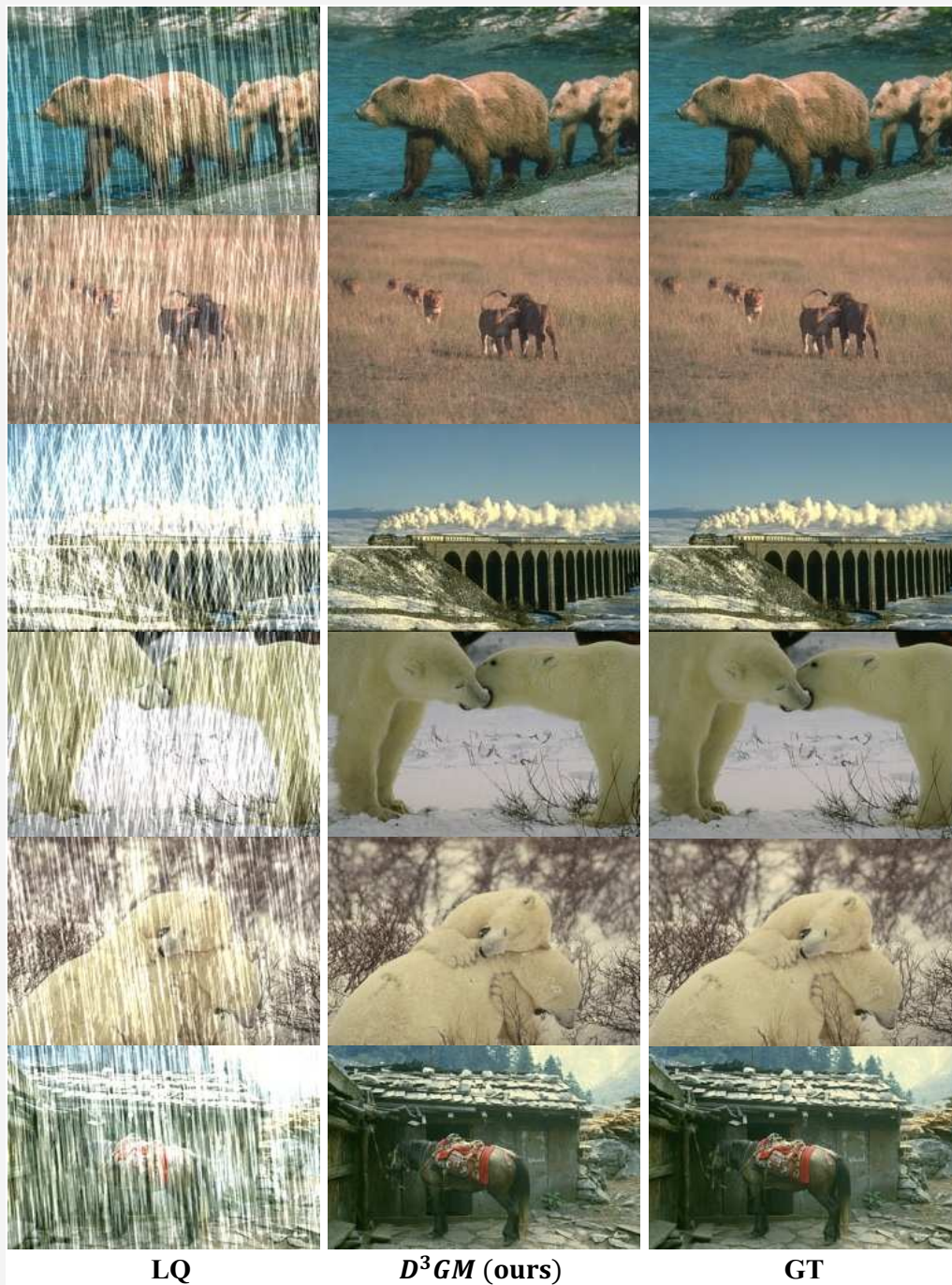
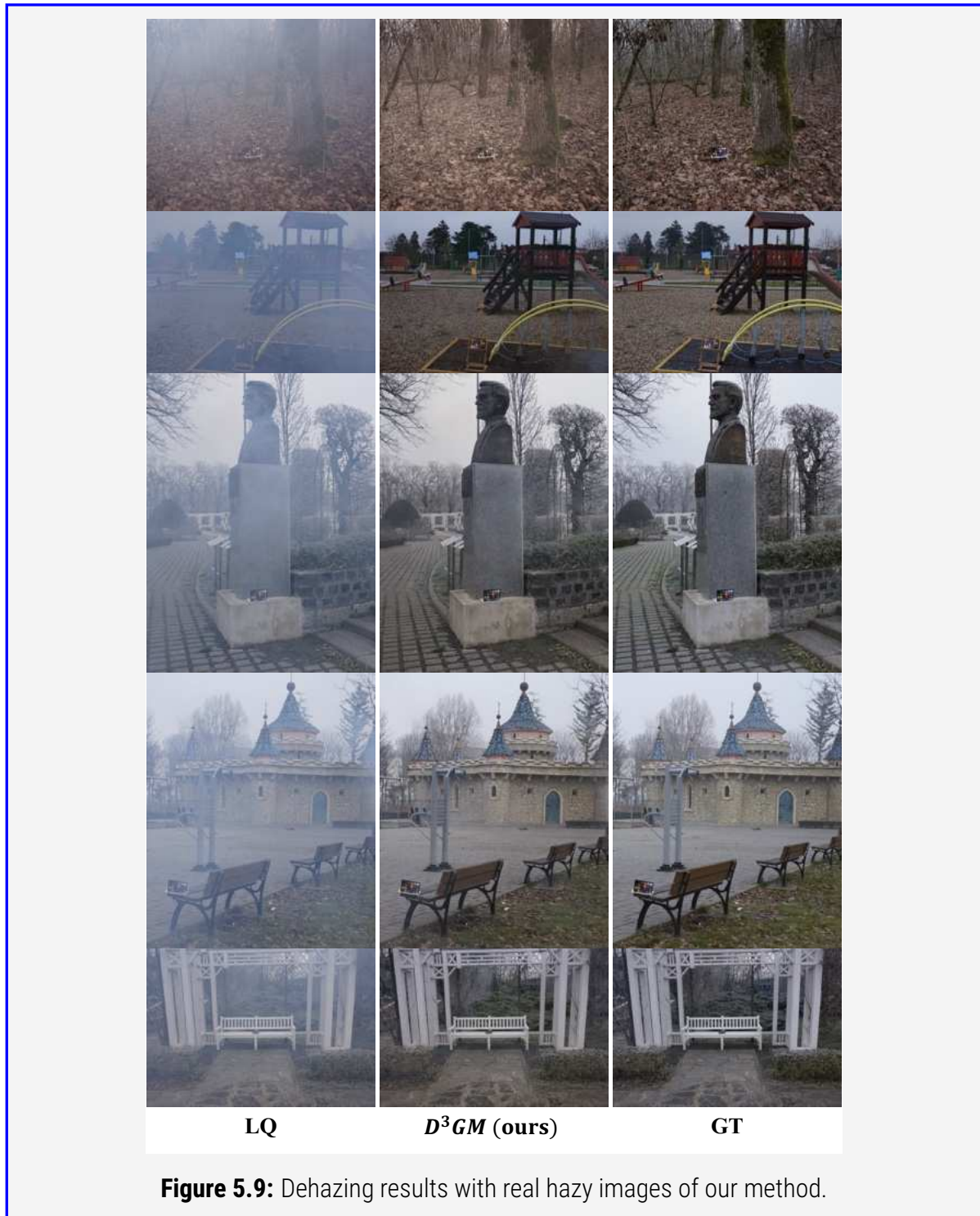


Figure 5.8: Deraining results with heavy rain images of our method.



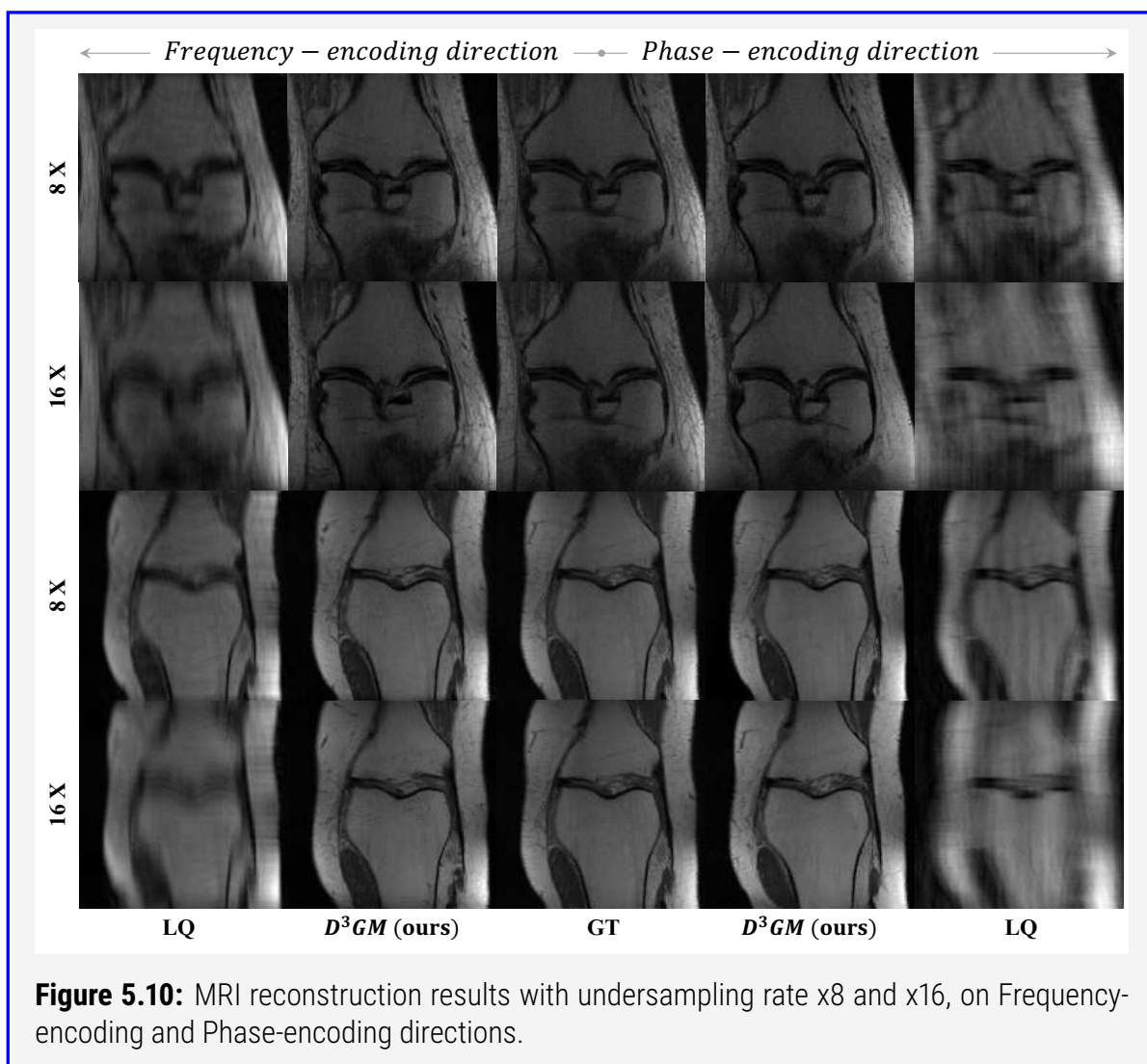
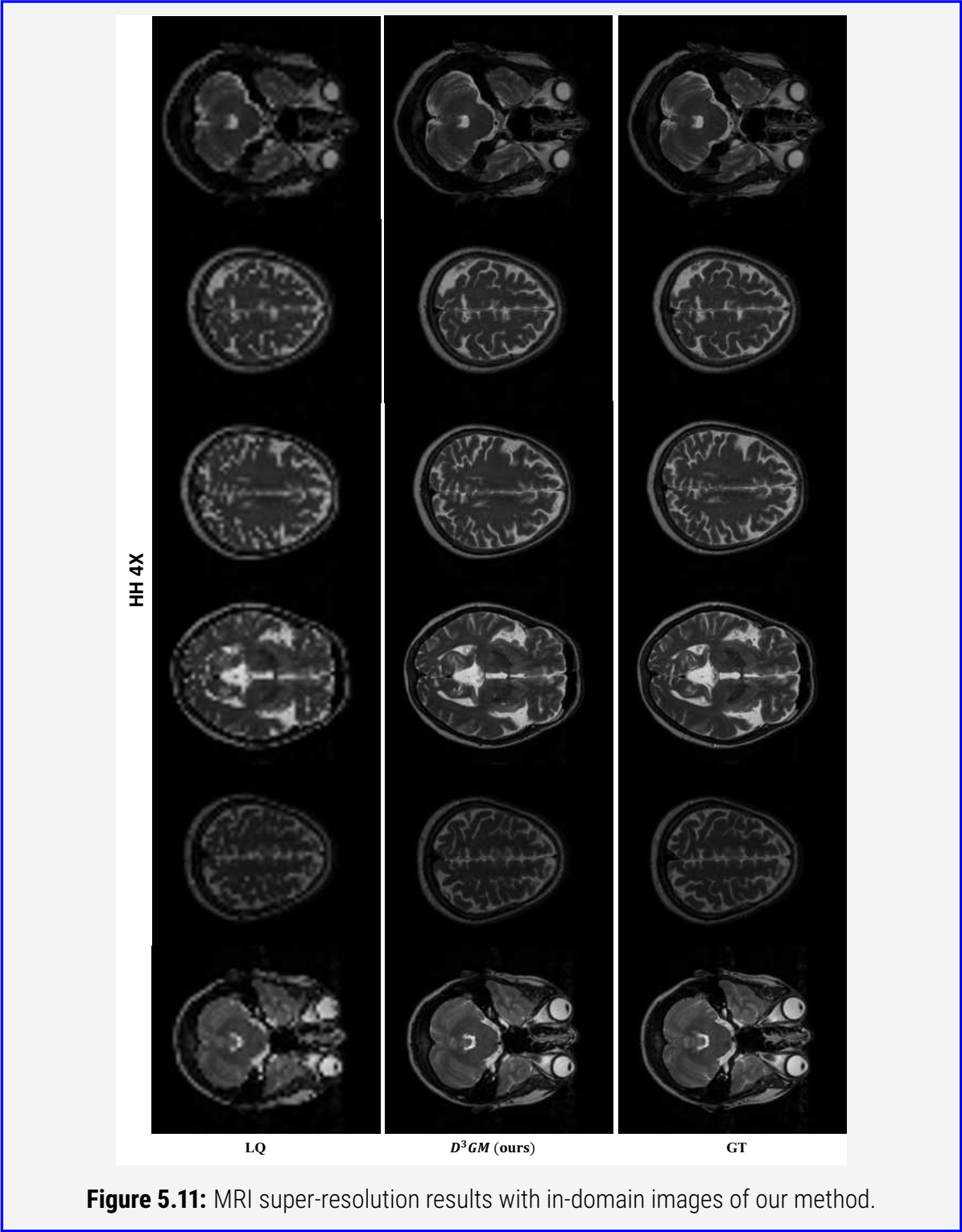


Figure 5.10: MRI reconstruction results with undersampling rate x8 and x16, on Frequency-encoding and Phase-encoding directions.



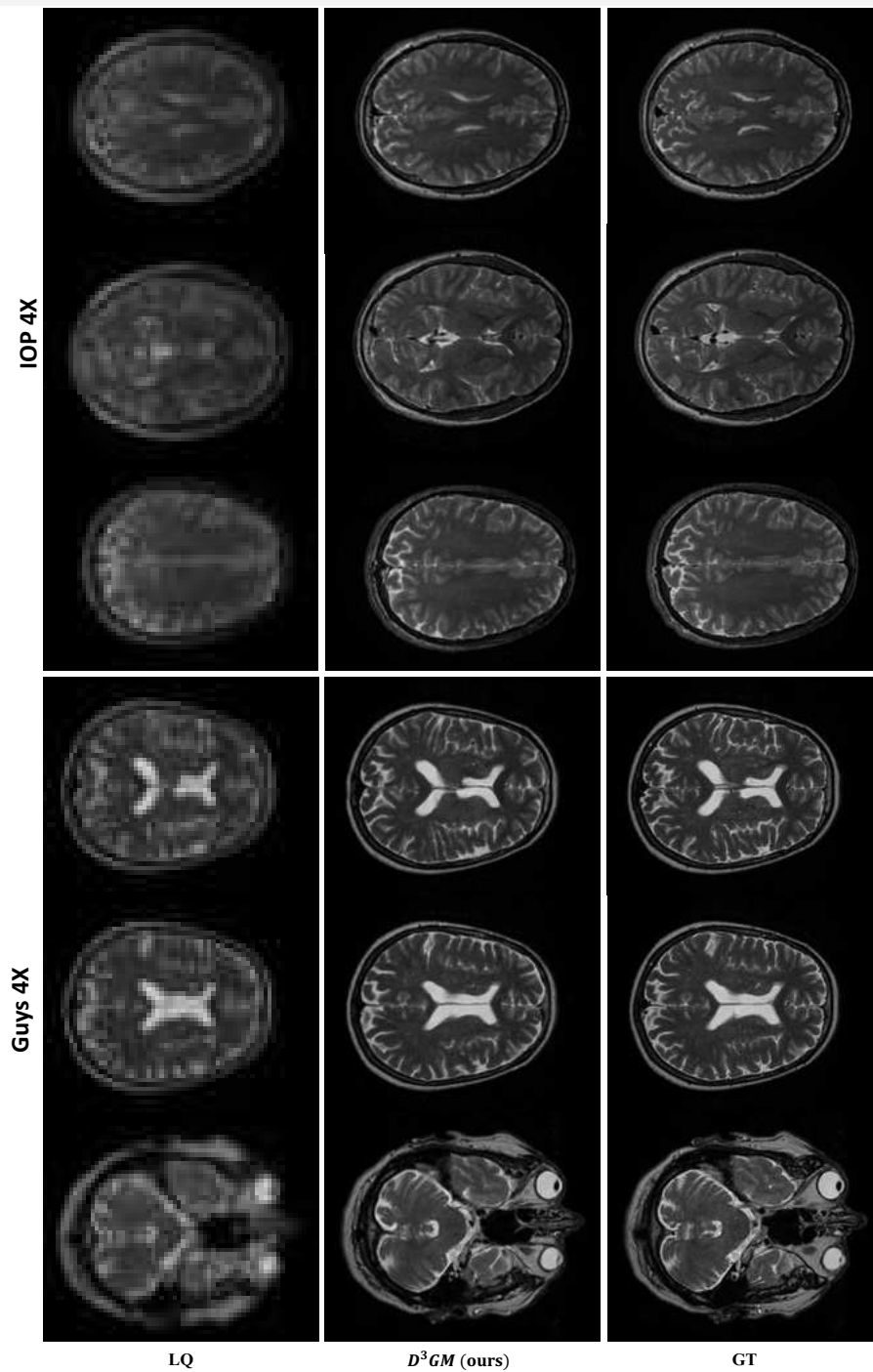


Figure 5.12: MRI super-resolution results with cross-domain (different imaging devices and degradation methods) images of our method.

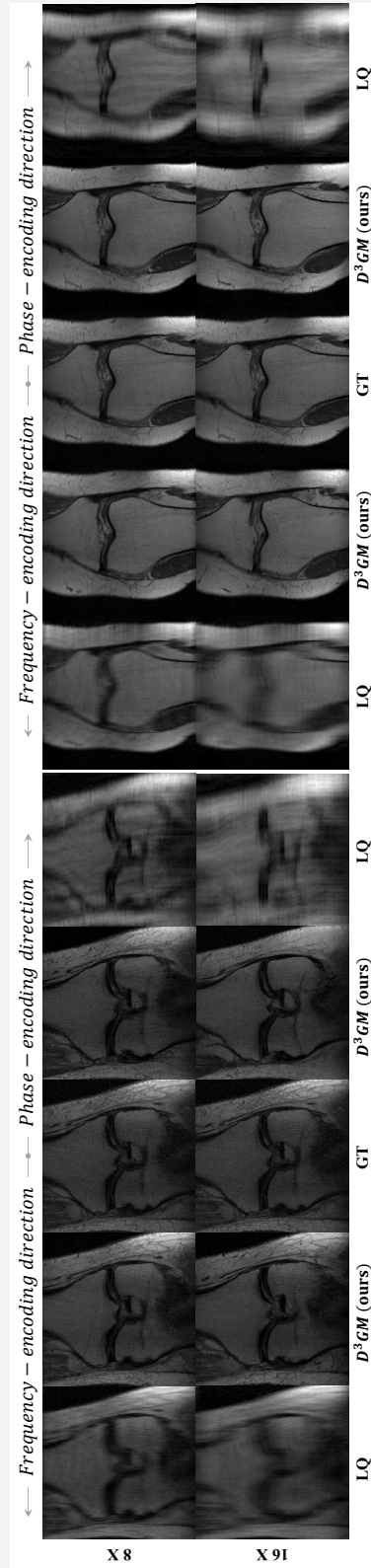


Figure 5.13: MRI reconstruction results with undersampling rate x8 and x16 of our method, on Phase-encoding and Frequency-encoding directions.

Chapter 6

Conclusions

This chapter reflects on the central findings of the thesis and revisits the research questions with a focus on broader implications. Rather than reiterating technical results, it aims to articulate the conceptual contributions of the work and frame them within the evolving landscape of machine learning for medical imaging.

This thesis explores how to build medical image reconstruction models that are not only accurate but generalizable and data-efficient. The core challenge lies in designing systems that can perform well in real-world settings without access to extensive labeled data from every possible domain. To this end, we investigate domain adaptation, domain generalization, self-supervised learning, and uncertainty calibration techniques. Each approach addresses a different aspect of the generalization problem, aiming to improve both the reliability and safety of deep learning in clinical workflows.

Through a series of contributions, this work proposes practical solutions that reduce annotation dependency, extend model adaptability, and enhance interpretability under domain shift. Together, these advances offer a foundation for more trustworthy AI systems in healthcare.

6.1 Thesis Summary

6.1.1 Motion-Corrected Quantitative Body MRI Super-resolution

Abdominal MRI is critical for diagnosing a wide variety of diseases. However, due to respiratory motion and other organ motions, it is challenging to obtain motion-free and isotropic MRI for clinical diagnosis. Imaging patients with inflammatory bowel disease (IBD) can be especially problematic, owing to involuntary bowel movements and difficulties with long breath-holds during acquisition. Therefore, this paper proposes a deep adversarial super-resolution (SR) reconstruction approach to address the problem of multi-task degradation by utilizing cycle consistency in a staged reconstruction model. We leverage a low-resolution (LR) latent space for motion correction, followed by super-resolution reconstruction, compensating for imaging artefacts caused by respiratory motion and spontaneous bowel movements. This alleviates the need for semantic knowledge about the intestines and paired data. Both are examined through variations of our proposed approach and we compare them to conventional, model-based, and learning-based MC and SR methods. Learned image reconstruction approaches are believed to occasionally hide disease signs. We investigate this hypothesis by evaluating a downstream task, automatically scoring IBD in the area of the terminal ileum on the reconstructed images and show evidence that our method does not suffer a synthetic domain bias

6.1.2 Test-time-Adapted Quantitative Body MRI Super-resolution

Deep learning-based super-resolution (SR) usually does not perform well across domains, e.g., different magnetic resonance imaging (MRI) datasets, usually due to two reasons: 1) there is often a mismatch in terms of data distributions between the training data and test data, and 2) the overly simplified image degradation models in popular SR techniques do not model the underlying physical processes well. Hence, SR models often require extensive supervised fine-tuning when transferring to new datasets (target domains). In this paper we propose UTAR, a physics-informed Unsupervised Test-time domain Adaptation framework for deep learning-based super-resolution for magnetic Resonance imaging. UTAR adapts a pre-trained source-domain SR model to a new target domain (e.g., images from a new hospital) using low-resolution (LR) target-domain images only. This is possible through a new physics-inspired imaging chain degradation model (PIC), which incorporates both spatial and frequency domain (k-space) degradation. Such dual-domain degradation modeling enables us to better capture the variations of LR images that could potentially affect a model's performance across different unseen MRI domains. We further propose a conditional consistency learning (CCL) strategy to encourage prediction consistency, serving as the supervision signal for TTA. Our method is generic and can accommodate arbitrary pre-trained SR networks. Experimental results verify the effectiveness of UTAR in reducing the performance gap with minimum efforts. UTAR outperforms earlier methods based on different degradation-based

strategies and existing test time adaptation (TTA) approaches. We expect our method to provide a key step towards the deployment of MRI SR algorithms in clinical applications where significant domain shifts are inevitable.

6.1.3 Generalizable and Stable Diffusion Models for Inverse Problems

Inverse problems describe the process of estimating the causal factors from a set of measurements or data. Mapping of often incomplete or degraded data to parameters is ill-posed, thus data-driven iterative solutions are required, for example when reconstructing clean images from poor signals. Diffusion models have shown promise as potent generative tools for solving inverse problems due to their superior reconstruction quality and their compatibility with iterative solvers. However, most existing approaches are limited to linear inverse problems represented as Stochastic Differential Equations (SDEs). This simplification falls short of addressing the challenging nature of real-world problems, leading to amplified cumulative errors and biases. We provide an explanation for this gap through the lens of measure-preserving dynamics of Random Dynamical Systems (RDS) with which we analyse Temporal Distribution Discrepancy and thus introduce a theoretical framework based on RDS for SDE diffusion models. We uncover several strategies that inherently enhance the stability and generalizability of diffusion models for inverse problems and introduce a novel score-based diffusion framework, the Dynamics-aware SDE Diffusion Generative Model (D3GM). The Measure-preserving property can return the degraded measurement to the original state despite complex degradation with the RDS concept of stability. Our extensive experimental results corroborate the effectiveness of D3GM across multiple benchmarks including a prominent application for inverse problems, magnetic resonance imaging.

6.2 Limitations

6.2.1 Motion Correction and Super-Resolution

A. The MoCoSR approach relies on cycle consistency in latent space, which may not fully model the complex physiological movements in abdominal organs. The method can address common respiratory patterns but may struggle with extreme or irregular motion patterns. B. The generalization capability across different IBD severity levels is variable, with the most severe cases showing comparatively smaller improvements due to larger structural changes and more unpredictable motion artifacts. C. While downstream tasks validate the clinical utility of our approach, comprehensive assessment with radiologist input on a larger multi-center dataset would strengthen clinical validation claims.

6.2.2 Test-time Adaptation Approach

A. The physics-informed degradation model (PIC), while more realistic than conventional approaches, necessarily simplifies certain aspects of MRI acquisition physics due to computational constraints. B. Performance gains diminish with more severe domain shifts, particularly when target-domain acquisition parameters differ significantly from those encountered during training. C. The conditional consistency learning strategy depends on reliable low-frequency information in the degraded images, which may not be guaranteed in highly corrupted or noise-dominated cases.

6.2.3 Dynamics-aware Diffusion Models

A. The theoretical foundations of measure-preserving dynamics, while demonstrating improved stability, rely on assumptions about the data distribution that may not hold for all clinical datasets. B. Computational complexity of D³GM exceeds that of standard diffusion models, creating a practical trade-off between reconstruction quality and inference speed in resource-constrained clinical settings. C. While our approach demonstrates better generalization than existing methods, performance still degrades when presented with extreme out-of-distribution examples or novel artifact types not represented in the training process.

6.3 Further Work

This thesis establishes a foundation for generalizable medical image reconstruction that opens several promising directions for future investigation.

In the immediate term, extending MoCoSR to additional anatomical regions beyond the abdomen would broaden its clinical applicability. Particularly promising areas include cardiac MRI, where similar respiratory and physiological motion challenges exist but with different temporal characteristics and anatomical constraints. Incorporating explicit biomechanical modeling into the latent space representation could further enhance motion correction performance.

For test-time adaptation, promising directions include developing self-supervised strategies to automatically calibrate the PIC model parameters based on initial observations of target domain characteristics, potentially eliminating the need for manual parameter tuning. The conditional consistency learning could be extended to leverage multiple degraded views of the same anatomy, a common scenario in clinical protocols that acquire complementary sequences. Looking further ahead, the D³GM framework provides a theoretical foundation that could be extended beyond MRI to other medical imaging modalities facing similar reconstruction challenges, such as low-dose CT or sparse-view tomographic reconstruction. The measure-preserving dynamics principle could inform the development of training strategies that explicitly optimize for domain generalization, rather than relying on post-hoc adaptation.

A significant open challenge remains in creating unified frameworks that effectively combine the strengths of all three proposed approaches. Future work could explore hierarchical models that integrate motion-awareness, physics-informed priors, and stability guarantees into a single coherent framework, potentially addressing the current limitations while preserving the core advantages of each method. Finally, the clinical integration of these methods represents a crucial next step. Developing interpretability mechanisms that allow radiologists to understand and trust the reconstruction process will be essential for clinical adoption. Collaborative research with clinical partners to design prospective studies measuring the impact on diagnostic accuracy and workflow efficiency would provide valuable validation for the practical utility of these approaches.

Considering recent developments in foundation models for medical imaging, exploring how our domain-specific approaches could be integrated with or enhanced by general-purpose medical vision models represents an exciting direction at the intersection of reconstruction and understanding.

6.3.1 Vision-Language Models for Medical Image Understanding

Beyond pure reconstruction, the emergence of powerful vision-language models (VLMs) offers intriguing possibilities for extending our work. Medical VLMs trained on paired image-report data

could provide semantic guidance for reconstruction tasks. For instance, a model could learn to prioritize the preservation of diagnostically critical features during the reconstruction process by incorporating textual descriptions of relevant pathology. Such models could enable more targeted reconstructions that emphasize clinically relevant regions while appropriately handling less diagnostically important areas.

6.3.2 Safety and Robustness Verification

As reconstruction algorithms become more integral to clinical workflows, ensuring their safety and robustness becomes paramount. Future work should develop formal verification methods for reconstruction guarantees, particularly for applications where subtle artifacts could lead to misdiagnosis. Research into adversarial robustness is especially important, as reconstruction models may be vulnerable to distribution shifts that are imperceptible to humans but potentially compromise diagnostic accuracy. Developing rigorous stress-testing methodologies and uncertainty quantification frameworks will be essential for responsible clinical deployment.

6.3.3 Privacy-Preserving Reconstruction Systems

The sensitive nature of medical data necessitates research into privacy-preserving reconstruction techniques. Federated learning approaches could enable models to be trained across multiple institutions without sharing raw data, potentially accelerating the development of generalizable reconstruction methods. Differential privacy mechanisms could provide formal guarantees around patient data protection while maintaining reconstruction performance. These advances would address significant regulatory and ethical barriers to widespread clinical adoption.

6.3.4 Autonomous Agents for Adaptive Protocols

The development of autonomous agent frameworks represents another promising direction. Intelligent imaging agents could dynamically adjust acquisition parameters based on preliminary reconstruction results, potentially reducing scan time while maintaining diagnostic quality. These agents could learn optimal policies for parameter selection through reinforcement learning, adapting protocols in real-time based on patient-specific factors and the evolving clinical question. This approach could transform current fixed imaging protocols into adaptive, personalized procedures that maximize diagnostic yield while minimizing patient discomfort.

The lack of trust and perceived relevance of AI results, driven by the absence of correctness guarantees and the inability of models to make selective predictions (e.g., acknowledging uncertainty with “I don’t know”), is a critical barrier to AI adoption. Addressing these issues is imperative to unlocking AI’s potential in high-stakes domains, where errors can have severe consequences. Low-stakes fields such as bureaucracy and entertainment have seen widespread adoption of AI, exemplified by the transformative impact of LLMs, precisely because the risks of incorrect predictions are min-

imal. To achieve comparable success in high-stake areas like healthcare, we must prioritize the development of systems that are robust, trustworthy, and capable of making uncertainty-aware decisions. The vision is to create techniques that enable a world of ubiquitous, friendly AI - AI that behaves like a wise and thoughtful human collaborator in high stake scenarios. Such systems would be not only highly accurate but also transparent, openly admitting when something is uncertain or undecidable. Importantly, these AI systems would actively participate in discussions and debates, contributing valuable insights while facilitating collaboration among experts.

This vision is particularly relevant in high-stakes scenarios like medicine, where decisions are often made collectively, such as in tumour boards where specialists evaluate treatment options for cancer patients. AI could serve as a wise mediator in these settings, enhancing discussions by providing relevant, accurate insights while ensuring that no inaccurate or irrelevant information is introduced. By integrating AI as a mediator, we could relatively quickly enable spatially and temporally distributed discussions among experts, ensuring that critical decisions are supported by reliable data and insights. I plan to develop a multi-modal foundation model, Medinator, aimed at transforming high-stakes decision-making and enhancing the efficiency of human collaboration, particularly in contexts such as clinical tumour boards. Tumour boards are multidisciplinary meetings where clinicians, radiologists, pathologists, and other specialists collaboratively assess treatment options for complex cancer cases, leveraging diverse expertise to achieve the best outcomes for patients. However, these discussions are often time-constrained and dependent on the participants' ability to integrate and interpret vast amounts of data efficiently. Traditionally, tumour boards and medical councils involve specialists reviewing cases based on limited information, which can lead to inconsistencies in diagnosis and treatment.

Agent-based techniques are expected to advance this vision by establishing an expertise levelling tool that fuses multimodal data with multi-agent collaboration. One promising approach aims to provide faster and more accurate treatment and diagnosis through a multi-user council interface and a knowledge consolidation system at the model's core. This system acts as an expertise levelling tool, facilitating discussions among domain experts and serving as a clinical expert mediator for collaborative diagnosis based on diverse data points, including images, lab test results, biopsies, written reports, omics, and patient histories.

Beyond expert moderation, agent-based system will also operate in a patient and public involvement mode, explaining diagnostic processes, patient data, and treatment options to affected individuals. It provides a bidirectional communication tool that allows patients to ask questions, ensuring that doctors remain attuned to both the physiological and mental health needs of their patients. In this way, it functions as a moderator not only between patients and doctors but also among experts, fostering a more inclusive and informed approach to healthcare decision-making. By evaluating the effectiveness of this comprehensive, multi-modal approach against traditional workflows, we aim to significantly improve clinical outcomes and enhance the efficiency of com-

plex case discussions. Importantly, participants in these councils do not need to be co-located or present at the same time, as it builds an interface for asynchronous collaboration, accommodating different schedules and time zones. Specialists from London to Edinburgh can collaborate asynchronously, overcoming scheduling and geographical barriers, making healthcare more efficient and accessible.

1. *Multi-Agent Reasoning for Cardiovascular Imaging Phenotype Analysis* [241]: Identifying associations between imaging phenotypes, disease risk factors, and clinical outcomes is essential for understanding disease mechanisms. However, traditional approaches rely on human-driven hypothesis testing and selection of association factors, often overlooking complex, non-linear dependencies among imaging phenotypes and other multi-modal data. To address this, we introduce Multi-agent Exploratory Synergy for the Heart (MESHAgents): a framework that leverages large language models as agents to dynamically elicit, surface, and decide confounders and phenotypes in association studies.

Specifically, we orchestrate a multi-disciplinary team of AI agents, which spontaneously generate and converge on insights through iterative, self-organizing reasoning. The framework dynamically synthesizes statistical correlations with multi-expert consensus, providing an automated pipeline for phenome-wide association studies (PheWAS). We demonstrate the system’s capabilities through a population-based study of imaging phenotypes of the heart and aorta. MESHAgents autonomously uncovered correlations between imaging phenotypes and a wide range of non-imaging factors, identifying additional confounder variables beyond standard demographic factors. Validation on diagnosis tasks reveals that MESHAgents-discovered phenotypes achieve performance comparable to expert-selected phenotypes, with mean AUC differences as small as -0.004 ± 0.010 on disease classification tasks. Notably, the recall score improves for 6 out of 9 disease types. Our framework provides clinically relevant imaging phenotypes with transparent reasoning, offering a scalable alternative to expert-driven methods. [62].

2. *From Self-Check to Consensus: Bayesian Strategic Decoding in Large Language Models* [242]: Large Language Models exhibit logical inconsistency across multi-turn inference processes, undermining correctness in complex inferential tasks. Challenges arise from ensuring that outputs align with both factual correctness and human intent. Approaches like single-agent reflection and multi-agent debate frequently prioritize consistency, but at the expense of accuracy.

To address this problem, we propose a novel game-theoretic consensus mechanism that enables LLMs to self-check their outputs during the decoding stage of output generation. Our method models the decoding process as a multistage Bayesian Decoding Game, where strategic interactions dynamically converge to a consensus on the most reliable outputs without human feedback or additional training. Remarkably, our game design allows smaller

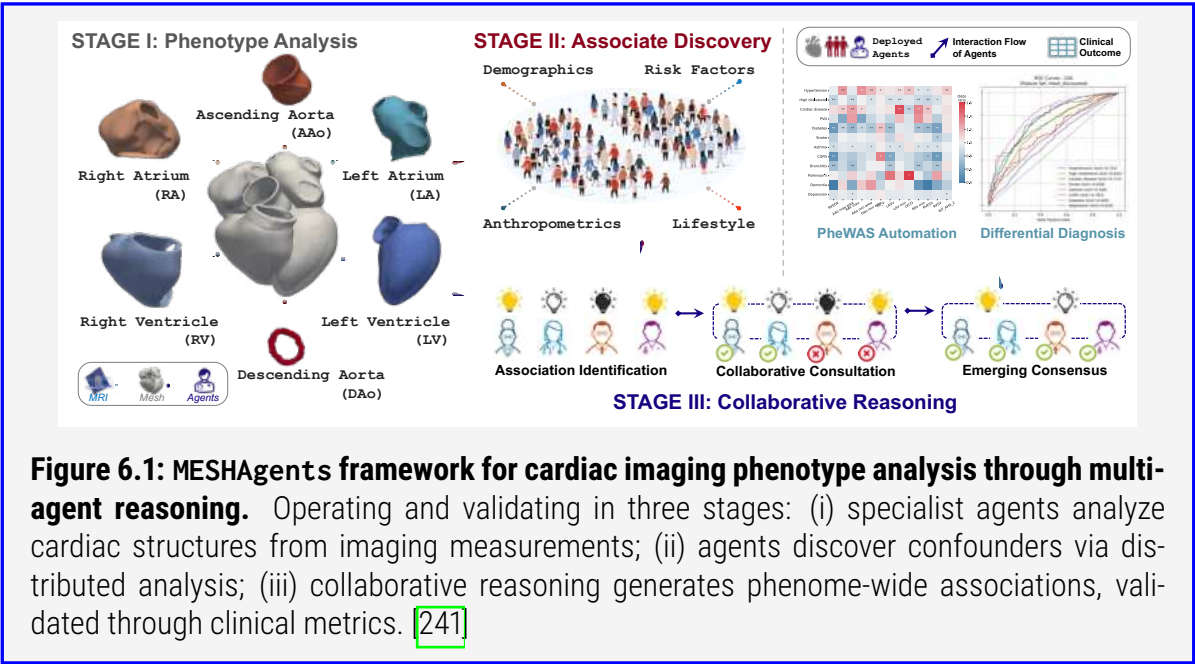


Figure 6.1: MESHAgents framework for cardiac imaging phenotype analysis through multi-agent reasoning. Operating and validating in three stages: (i) specialist agents analyze cardiac structures from imaging measurements; (ii) agents discover confounders via distributed analysis; (iii) collaborative reasoning generates phenome-wide associations, validated through clinical metrics. [241]

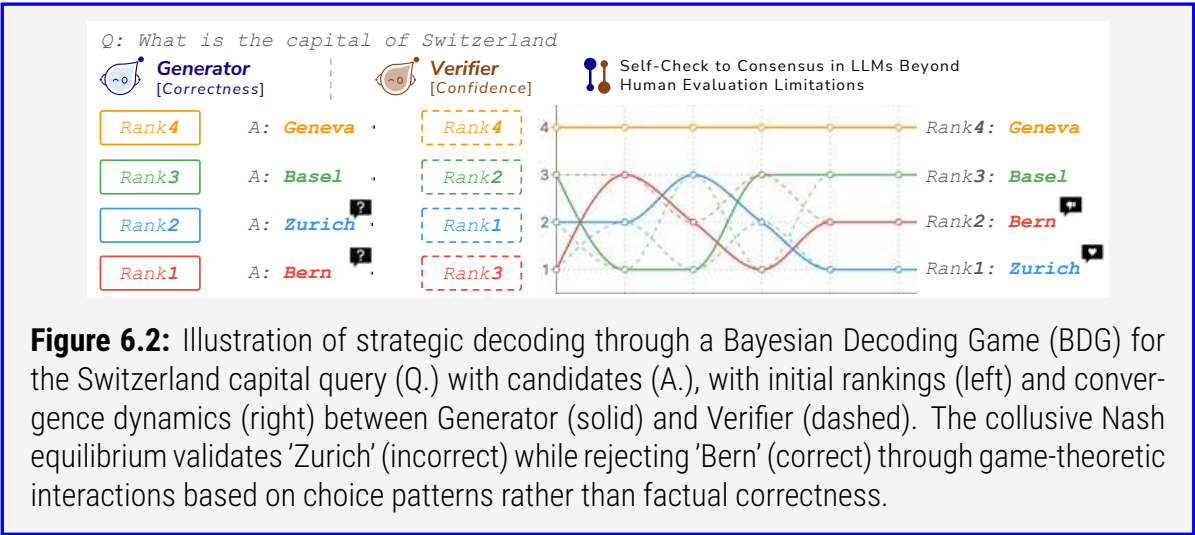
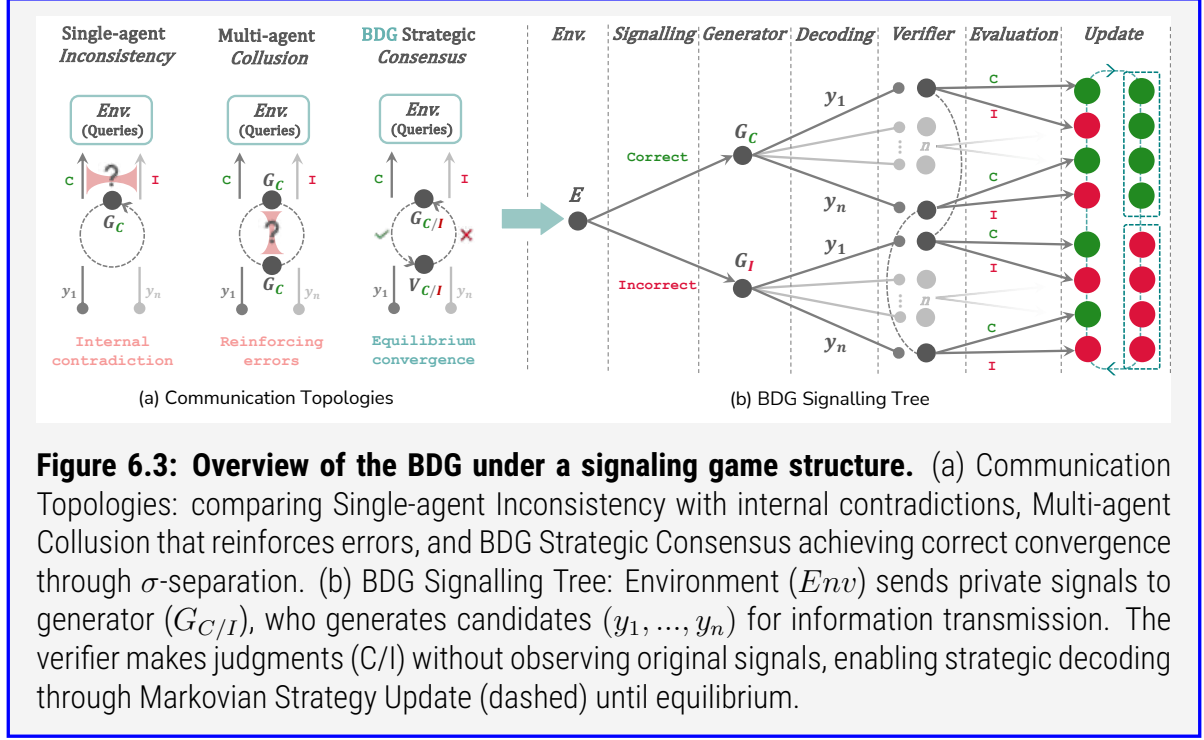


Figure 6.2: Illustration of strategic decoding through a Bayesian Decoding Game (BDG) for the Switzerland capital query (Q.) with candidates (A.), with initial rankings (left) and convergence dynamics (right) between Generator (solid) and Verifier (dashed). The collusive Nash equilibrium validates 'Zurich' (incorrect) while rejecting 'Bern' (correct) through game-theoretic interactions based on choice patterns rather than factual correctness.

models to outperform much larger models through game mechanisms (e.g., 78.1 LLaMA13B vs. 76.6 PaLM540B). As a model-agnostic method, our approach consistently improves even the latest models, enhancing DeepSeek-7B's performance on MMLU by 12.4%. Our framework effectively balances correctness and consistency, demonstrating that properly designed game-theoretic mechanisms can significantly enhance the self-verification capabilities of language models across various tasks and model architectures.

The UK, with its robust healthcare infrastructure, is uniquely positioned to benefit from the Mediator foundation model, which has the potential to drastically change critical care. By integrating extensive datasets—including medical imaging, genomics, clinical records, and real-time monitoring from approximately 200,000 patients - Mediator enables holistic, data-driven decision-making. This will enhance patient outcomes by streamlining multidisciplinary collaboration, improving treat-



ment precision, optimizing resource allocation, and enabling proactive care through predictive analytics. It also bridges gaps in expertise and geographical access, promoting equitable and high-quality healthcare delivery across regions. Mediator supports personalized, inclusive healthcare by acting as a bidirectional communication tool, empowering patients to understand and engage with their care while addressing both physical and mental health needs. For the life sciences and health tech industries, Mediator serves as a transformative AI resource, driving innovation and enabling the development of cutting-edge diagnostic tools, treatments, and precision therapies. By accelerating the journey from discovery to market and providing critical insights for clinical trials and medical device validation, Mediator attracts investment and solidifies the position as a global leader in healthcare innovation and commercialization.

References

- [1] Afaq, A., Faul, D., Chebrolu, V. V., et al. "Pitfalls on PET/MRI". In: *Seminars in Nuclear Medicine*. Vol. 51. Elsevier. 2021, pp. 529–539.
- [2] Aggarwal, H. K., Mani, M. P., and Jacob, M. "MoDL: Model-based deep learning architecture for inverse problems". In: *IEEE transactions on medical imaging* 38.2 (2018), pp. 394–405.
- [3] Alansary, A., Kamnitsas, K., Davidson, A., et al. "Fast fully automatic segmentation of the human placenta from motion corrupted MRI". In: *MICCAI*. Springer. 2016, pp. 589–597.
- [4] Alansary, A., Rajchl, M., McDonagh, S. G., et al. "PVR: patch-to-volume reconstruction for large area motion correction of fetal MRI". In: *IEEE Trans. Med. Imaging* 36.10 (2017), pp. 2031–2044.
- [5] Alom, M. Z., Yakopcic, C., Nasrin, M. S., et al. "Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network". In: *Journal of digital imaging* 32.4 (2019), pp. 605–617.
- [6] Ancona, M., Ceolini, E., Öztireli, C., et al. "Towards better understanding of gradient-based attribution methods for deep neural networks". In: *arXiv preprint arXiv:1711.06104* (2017).
- [7] Ancuti, C. O., Ancuti, C., Sbert, M., et al. "Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images". In: *ICIP*. IEEE. 2019, pp. 1014–1018.
- [8] Ancuti, C. O., Ancuti, C., Timofte, R., et al. "O-haze: a dehazing benchmark with real hazy and haze-free outdoor images". In: *CVPR workshops*. 2018, pp. 754–762.
- [9] Anderson, B. D. "Reverse-time diffusion equation models". In: *Stochastic Processes and their Applications* 12.3 (1982), pp. 313–326.
- [10] Arridge, S., Maass, P., Öktem, O., et al. "Solving inverse problems using data-driven models". In: *Acta Numerica* 28 (2019), pp. 1–174.
- [11] Ashburner, J. and Ridgway, G. R. "Symmetric diffeomorphic modeling of longitudinal structural MRI". In: *Frontiers in neuroscience* 6 (2013), p. 197.
- [12] Bai, H., Pan, J., Xiang, X., et al. "Self-guided image dehazing using progressive feature fusion". In: *TIP* 31 (2022), pp. 1217–1229.
- [13] Balakrishnan, G., Zhao, A., Sabuncu, M. R., et al. "Voxelmorph: a learning framework for deformable medical image registration". In: *IEEE transactions on medical imaging* 38.8 (2019), pp. 1788–1800.
- [14] Basaran, B. D., Zhang, W., Qiao, M., et al. "LesionMix: A lesion-level data augmentation method for medical image segmentation". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2023, pp. 73–83.
- [15] Bateson, M., Dolz, J., Kervadec, H., et al. "Constrained domain adaptation for image segmentation". In: *IEEE Transactions on Medical Imaging* 40.7 (2021), pp. 1875–1887.
- [16] Bateson, M., Lombaert, H., and Ben Ayed, I. "Test-time adaptation with shape moments for image segmentation". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2022, pp. 736–745.
- [17] Bell-Kligler, S., Shocher, A., and Irani, M. "Blind super-resolution kernel estimation using an internal-gan". In: *NeurIPS* 32 (2019).
- [18] Ben-David, S., Blitzer, J., Crammer, K., et al. "A theory of learning from different domains". In: *Machine learning* 79 (2010), pp. 151–175.

- [19] Bengio, Y. et al. "Learning deep architectures for AI". In: *Foundations and trends® in Machine Learning* 2.1 (2009), pp. 1–127.
- [20] Bernstein, M. A., King, K. F., and Zhou, X. J. *Handbook of MRI pulse sequences*. Elsevier, 2004.
- [21] Bishop, C. M. and Nasrabadi, N. M. *Pattern recognition and machine learning*. Vol. 4. 4. Springer, 2006.
- [22] Brown, R. W., Cheng, Y.-C. N., Haacke, E. M., et al. *Magnetic resonance imaging: physical principles and sequence design*. John Wiley & Sons, 2014.
- [23] Cabezas, M., Oliver, A., Lladó, X., et al. "A review of atlas-based segmentation for magnetic resonance brain images". In: *Computer methods and programs in biomedicine* 104.3 (2011), e158–e177.
- [24] Cai, B., Xu, X., Jia, K., et al. "Dehazenet: An end-to-end system for single image haze removal". In: *IEEE TIP* 25.11 (2016), pp. 5187–5198.
- [25] Chang, Q., Yan, Z., Zhou, M., et al. "Mining multi-center heterogeneous medical data with distributed synthetic learning". In: *Nature communications* 14.1 (2023), p. 5510.
- [26] Chen, X., Li, H., Li, M., et al. "Learning a Sparse Transformer Network for Effective Image Deraining". In: *CVPR*. 2023.
- [27] Chen, Y., Shi, F., Christodoulou, A. G., et al. "Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network". In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*. Springer. 2018, pp. 91–99.
- [28] Chen, Y., Xie, Y., Zhou, Z., et al. "Brain MRI super resolution using 3D deep densely connected neural networks". In: *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. 2018, pp. 739–742.
- [29] Chen, Z., Yao, X., Xu, Y., et al. "Unsupervised knowledge transfer for nonblind image deconvolution". In: *Pattern Recognition Letters* 164 (2022), pp. 232–238.
- [30] Chu, X., Chen, L., and Yu, W. "NAFSSR: Stereo image super-resolution using NAFNet". In: *CVPR*. 2022, pp. 1239–1248.
- [31] Chung, H., Kim, J., McCann, M. T., et al. "Diffusion posterior sampling for general noisy inverse problems". In: *ICLR* (2023).
- [32] Chung, H., Kim, J., and Ye, J. C. "Direct Diffusion Bridge using Data Consistency for Inverse Problems". In: *NeurIPS* (2023).
- [33] Chung, H., Ryu, D., McCann, M. T., et al. "Solving 3d inverse problems using pre-trained 2d diffusion models". In: *CVPR*. 2023, pp. 22542–22551.
- [34] Chung, H., Sim, B., and Ye, J. C. "Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction". In: *CVPR*. 2022, pp. 12413–12422.
- [35] Cleary, K. and Peters, T. M. "Image-guided interventions: technology review and clinical applications". In: *Annual review of biomedical engineering* 12.1 (2010), pp. 119–142.
- [36] Crauel, H., Debussche, A., and Flandoli, F. "Random attractors". In: *Journal of Dynamics and Differential Equations* 9 (1997), pp. 307–341.
- [37] Crum, W. R., Hartkens, T., and Hill, D. "Non-rigid image registration: theory and practice". In: *The British journal of radiology* 77.suppl_2 (2004), S140–S153.
- [38] Darestani, M. Z., Liu, J., and Heckel, R. "Test-time training can close the natural distribution shift performance gap in deep learning based compressed sensing". In: *ICML*. PMLR. 2022, pp. 4754–4776.

- [39] De Bortoli, V., Thornton, J., Heng, J., et al. "Diffusion Schrodinger bridge with applications to score-based generative modeling". In: *NeurIPS* 34 (2021), pp. 17695–17709.
- [40] De Vos, B. D., Berendsen, F. F., Viergever, M. A., et al. "A deep learning framework for unsupervised affine and deformable image registration". In: *Medical image analysis* 52 (2019), pp. 128–143.
- [41] Dehnad, K. *Density estimation for statistics and data analysis*. 1987.
- [42] Delbracio, M. and Milanfar, P. "Inversion by direct iteration: An alternative to denoising diffusion for image restoration". In: *TMLR* (2023).
- [43] Deng, J., Dong, W., Socher, R., et al. "Imagenet: A large-scale hierarchical image database". In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255.
- [44] Deshpande, A., Estrela, V. V., and Patavardhan, P. "The DCT-CNN-ResNet50 architecture to classify brain tumors with super-resolution, convolutional neural network, and the ResNet50". In: *Neuroscience Informatics* 1.4 (2021), p. 100013.
- [45] Do, H., Bourdon, P., Helbert, D., et al. "7T MRI super-resolution with Generative Adversarial Network". In: *Electronic Imaging* 2021.18 (2021), pp. 106–1.
- [46] Dombrowski, M., Zhang, W., Cechnicka, S., et al. "Image Generation Diversity Issues and How to Tame Them". In: *arXiv preprint arXiv:2411.16171* (2024).
- [47] Dong, C., Loy, C. C., He, K., et al. "Learning a deep convolutional network for image super-resolution". In: *ECCV*. Springer. 2014, pp. 184–199.
- [48] Dong, H., Pan, J., Xiang, L., et al. "Multi-scale boosted dehazing network with dense feature fusion". In: *CVPR*. 2020, pp. 2157–2167.
- [49] Dou, Q., Ouyang, C., Chen, C., et al. "PnP-AdaNet: Plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation". In: *IEEE Access* 7 (2019), pp. 99065–99076.
- [50] Ebner, M., Chouhan, M., Patel, P. A., et al. "Point-spread-function-aware slice-to-volume registration: application to upper abdominal MRI super-resolution". In: *Reconstruction, Segmentation, and Analysis of Medical Images*. Springer, 2016, pp. 3–13.
- [51] El Helou, M., Zhou, R., and Süssstrunk, S. "Stochastic frequency masking to improve super-resolution and denoising networks". In: *ECCV*. Springer. 2020, pp. 749–766.
- [52] Fang, J., Wang, F., Shen, Y., et al. "Super-resolution compressed sensing for line spectral estimation: An iterative reweighted approach". In: *IEEE Trans. Signal Process.* 64.18 (2016), pp. 4649–4662.
- [53] Feng, C.-M., Fu, H., Yuan, S., et al. "Multi-contrast mri super-resolution via a multi-stage integration network". In: *MICCAI*. Springer. 2021, pp. 140–149.
- [54] Feng, C.-M., Yan, Y., Fu, H., et al. "Task transformer network for joint MRI reconstruction and super-resolution". In: *MICCAI*. Springer. 2021, pp. 307–317.
- [55] François-Lavet, V., Henderson, P., Islam, R., et al. "An introduction to deep reinforcement learning". In: *Foundations and Trends® in Machine Learning* 11.3-4 (2018), pp. 219–354.
- [56] Franzese, G., Rossi, S., Yang, L., et al. "How much is enough? a study on diffusion times in score-based generative models". In: *Entropy* 25.4 (2023), p. 633.
- [57] Frid-Adar, M., Diamant, I., Klang, E., et al. "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification". In: *Neurocomputing* 321 (2018), pp. 321–331.
- [58] Ganin, Y. and Lempitsky, V. "Unsupervised domain adaptation by backpropagation". In: *International conference on machine learning*. PMLR. 2015, pp. 1180–1189.
- [59] Gao, H., Yuan, H., Wang, Z., et al. "Pixel transposed convolutional networks". In: *IEEE transactions on pattern analysis and machine intelligence* 42.5 (2019), pp. 1218–1227.

- [60] Gastrointestinal Unit Medical Services, M. G. H., Andres, P. G., Friedman, L. S., et al. "Epidemiology and the natural course of inflammatory bowel disease". In: *Gastroenterol. Clin. North Am.* 28.2 (1999), pp. 255–281.
- [61] Georgescu, M.-I., Ionescu, R. T., Miron, A.-I., et al. "Multimodal multi-head convolutional attention with various kernel sizes for medical image super-resolution". In: *WACV*. 2023, pp. 2195–2205.
- [62] Glocker, B., Robinson, R., Castro, D. C., et al. "Machine learning with multi-site imaging data: An empirical study on the impact of scanner effects". In: *arXiv preprint arXiv:1910.04597* (2019).
- [63] Godenschweger, F., Kägebein, U., Stucht, D., et al. "Motion correction in MRI of the brain". In: *Physics in medicine & biology* 61.5 (2016), R32.
- [64] Goodfellow, I., Bengio, Y., and Courville, A. *Deep learning*. MIT press, 2016.
- [65] Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. "Generative adversarial nets". In: *Advances in neural information processing systems* 27 (2014).
- [66] Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. "Generative adversarial networks". In: *Communications of the ACM* 63.11 (2020), pp. 139–144.
- [67] Grenander, U. and Miller, M. I. "Representations of knowledge in complex systems". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 56.4 (1994), pp. 549–581.
- [68] Gudbjartsson, H. and Patz, S. "The rician distribution of noisy mri data". In: *Magnetic Resonance in Medicine* 34.6 (Dec. 1995), pp. 910–914. doi: [10.1002/mrm.1910340618](https://doi.org/10.1002/mrm.1910340618).
- [69] Guo, B., Lu, D., Szumel, G., et al. "The impact of scanner domain shift on deep learning performance in medical imaging: an experimental study". In: *arXiv preprint arXiv:2409.04368* (2024).
- [70] Guo, C.-L., Yan, Q., Anwar, S., et al. "Image dehazing transformer with transmission-aware 3d position embedding". In: *CVPR*. 2022, pp. 5812–5820.
- [71] Guo, J., Zhang, W., Sinclair, M., et al. "Pay attention to the atlas: Atlas-guided test-time adaptation method for robust 3d medical image segmentation". In: *arXiv preprint arXiv:2307.00676* (2023).
- [72] Haldar, J. P. and Liang, Z.-P. "Low-rank approximations for dynamic imaging". In: *2011 IEEE international symposium on biomedical imaging: From Nano to macro*. IEEE. 2011, pp. 1052–1055.
- [73] Hammernik, K., Klatzer, T., Kobler, E., et al. "Learning a variational network for reconstruction of accelerated MRI data". In: *Magnetic resonance in medicine* 79.6 (2018), pp. 3055–3071.
- [74] Hammernik, K. and Knoll, F. "Machine learning for image reconstruction". In: *Handbook of medical image computing and computer assisted intervention*. Elsevier, 2020, pp. 25–64.
- [75] He, K., Sun, J., and Tang, X. "Single image haze removal using dark channel prior". In: *IEEE TPAMI* 33.12 (2010), pp. 2341–2353.
- [76] He, K., Zhang, X., Ren, S., et al. "Deep residual learning for image recognition". In: *CVPR*. 2016, pp. 770–778.
- [77] Hedley, M. and Yan, H. "Motion artifact suppression: a review of post-processing techniques". In: *Magnetic resonance imaging* 10.4 (1992), pp. 627–635.
- [78] Heiland, S. "From A as in aliasing to Z as in zipper: artifacts in MRI". In: *Clinical neuroradiology* 18.1 (2008), pp. 25–36.
- [79] Hennig, J. "K-space sampling strategies". In: *European radiology* 9.6 (1999), pp. 1020–1031.
- [80] Hinton, G. E. "Connectionist learning procedures". In: *Machine learning*. Elsevier, 1990, pp. 555–610.
- [81] Ho, J., Jain, A., and Abbeel, P. "Denoising diffusion probabilistic models". In: *NeurIPS* 33 (2020), pp. 6840–6851.

- [82] Hoffman, J., Tzeng, E., Park, T., et al. "Cycada: Cycle-consistent adversarial domain adaptation". In: *International conference on machine learning*. Pmlr. 2018, pp. 1989–1998.
- [83] Holland, R., Patel, U., Lung, P., et al. "Automatic detection of bowel disease with residual networks". In: *MICCAI Workshop*. Springer. 2019, pp. 151–159.
- [84] Hore, A. and Ziou, D. "Image quality metrics: PSNR vs. SSIM". In: *ICPR*. IEEE. 2010, pp. 2366–2369.
- [85] Howard, A. G., Zhu, M., Chen, B., et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications". In: *arXiv preprint arXiv:1704.04861* (2017).
- [86] Hu, J., Shen, L., and Sun, G. "Squeeze-and-excitation networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 7132–7141.
- [87] Hu, J., Zhang, C., Zhou, K., et al. "Chest X-ray diagnostic quality assessment: How much is pixel-wise supervision needed?" In: *IEEE Transactions on Medical Imaging* 41.7 (2022), pp. 1711–1723.
- [88] Huang, J., Aviles-Rivero, A. I., Schönlieb, C.-B., et al. "CDiffMR: Can We Replace the Gaussian Noise with K-Space Undersampling for Fast MRI?" In: *MICCAI*. Springer. 2023, pp. 3–12.
- [89] Huang, J., Fang, Y., Wu, Y., et al. "Swin transformer for fast MRI". In: *Neurocomputing* 493 (2022), pp. 281–304.
- [90] Hyvärinen, A. and Dayan, P. "Estimation of non-normalized statistical models by score matching." In: *Journal of Machine Learning Research* 6.4 (2005).
- [91] Ioffe, S. and Szegedy, C. "Batch normalization: Accelerating deep network training by reducing internal covariate shift". In: *International conference on machine learning*. pmlr. 2015, pp. 448–456.
- [92] Isensee, F., Jaeger, P. F., Kohl, S. A., et al. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In: *Nature methods* 18.2 (2021), pp. 203–211.
- [93] Jenkinson, M., Bannister, P., Brady, M., et al. "Improved optimization for the robust and accurate linear registration and motion correction of brain images". In: *Neuroimage* 17.2 (2002), pp. 825–841.
- [94] Jiang, L., Dai, B., Wu, W., et al. "Focal frequency loss for image reconstruction and synthesis". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 13919–13929.
- [95] Jin, K. H., Um, J.-Y., Lee, D., et al. "MRI artifact correction using sparse+ low-rank decomposition of annihilating filter-based hankel matrix". In: *Magn. Reson. Med.* 78.1 (2017), pp. 327–340.
- [96] Jing, L. and Tian, Y. "Self-supervised visual feature learning with deep neural networks: A survey". In: *IEEE transactions on pattern analysis and machine intelligence* 43.11 (2020), pp. 4037–4058.
- [97] Jolicoeur-Martineau, A., Li, K., Piché-Taillefer, R., et al. "Gotta go fast when generating data with score-based models". In: *arXiv preprint arXiv:2105.14080* (2021).
- [98] Jolicoeur-Martineau, A., Piché-Taillefer, R., Combes, R. T. d., et al. "Adversarial score matching and improved sampling for image generation". In: *arXiv preprint arXiv:2009.05475* (2020).
- [99] Kamnitsas, K. "Advancing efficiency and robustness of neural networks for imaging". PhD thesis. Imperial College London, 2020.
- [100] Kamnitsas, K., Baumgartner, C., Ledig, C., et al. "Unsupervised domain adaptation in brain lesion segmentation with adversarial networks". In: *Information Processing in Medical Imaging: 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings* 25. Springer. 2017, pp. 597–609.

- [101] Karani, N., Erdil, E., Chaitanya, K., et al. "Test-time adaptable neural networks for robust medical image segmentation". In: *Medical Image Analysis* 68 (2021), p. 101907.
- [102] Karras, T., Aittala, M., Aila, T., et al. "Elucidating the design space of diffusion-based generative models". In: *arXiv preprint arXiv:2206.00364* (2022).
- [103] Kazeminia, S., Baur, C., Kuijper, A., et al. "GANs for medical image analysis". In: *Artificial Intelligence in Medicine* 109 (2020), p. 101938.
- [104] Kazerouni, A., Aghdam, E. K., Heidari, M., et al. "Diffusion models in medical imaging: A comprehensive survey". In: *Medical image analysis* 88 (2023), p. 102846.
- [105] Keraudren, K., Kainz, B., Oktay, O., et al. "Automated localization of fetal organs in MRI using random forests with steerable features". In: *MICCAI*. Springer. 2015, pp. 620–627.
- [106] Kilim, O., Olar, A., Joó, T., et al. "Physical imaging parameter variation drives domain shift". In: *Scientific Reports* 12.1 (2022), p. 21302.
- [107] Kingma, D. P. and Ba, J. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).
- [108] Kingma, D. P., Welling, M., et al. *Auto-encoding variational bayes*. 2013.
- [109] Kipf, T. N. and Welling, M. "Semi-supervised classification with graph convolutional networks". In: *arXiv preprint arXiv:1609.02907* (2016).
- [110] Kiran, B. R., Sobh, I., Talpaert, V., et al. "Deep reinforcement learning for autonomous driving: A survey". In: *IEEE transactions on intelligent transportation systems* 23.6 (2021), pp. 4909–4926.
- [111] Kodinariya, T. M., Makwana, P. R., et al. "Review on determining number of Cluster in K-Means Clustering". In: *International Journal* 1.6 (2013), pp. 90–95.
- [112] Koh, D.-M., Collins, D. J., and Orton, M. R. "Intravoxel incoherent motion in body diffusion-weighted MRI: reality and challenges". In: *AJR* 196.6 (2011), pp. 1351–1361.
- [113] Krizhevsky, A., Sutskever, I., and Hinton, G. E. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems* 25 (2012).
- [114] Küstner, T., Schwartz, M., Martirosian, P., et al. "MR-based respiratory and cardiac motion correction for PET imaging". In: *Med. Image Anal.* 42 (2017), pp. 129–144.
- [115] Lambert, B., Forbes, F., Doyle, S., et al. "Trustworthy clinical AI solutions: A unified review of uncertainty quantification in Deep Learning models for medical image analysis." In: *Artif. Intell. Medicine* 150 (2024), p. 102830.
- [116] Laurent, B. and Massart, P. "Adaptive estimation of a quadratic functional by model selection". In: *Annals of statistics* (2000), pp. 1302–1338.
- [117] LeCun, Y., Boser, B., Denker, J., et al. "Handwritten digit recognition with a back-propagation network". In: *Advances in neural information processing systems* 2 (1989).
- [118] Li, L., Ma, Q., Li, Z., et al. "Fetal cortex segmentation with topology and thickness loss constraints". In: *Workshop on the Ethical and Philosophical Issues in Medical Imaging*. Springer. 2022, pp. 123–133.
- [119] Li, L., Ma, Q., Ouyang, C., et al. "Robust segmentation via topology violation detection and feature synthesis". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2023, pp. 67–77.
- [120] Li, L., Wang, H., Baugh, M., et al. "Universal Topology Refinement for Medical Image Segmentation with Polynomial Feature Synthesis". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2024, pp. 670–680.
- [121] Li, Y., Sixou, B., and Peyrin, F. "A review of the deep learning methods for medical images super resolution problems". In: *IRBM* 42.2 (2021), pp. 120–133.
- [122] Li, Z., Zhang, W., Cechnicka, S., et al. "Data-Efficient Generation for Dataset Distillation". In: *arXiv preprint arXiv:2409.03929* (2024).

- [123] Liang, J., Homayounfar, N., Ma, W.-C., et al. "Polytransform: Deep polygon transformer for instance segmentation". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 9131–9140.
- [124] Lichtenstein, G. R., Shahabi, A., Seabury, S. A., et al. "Lifetime economic burden of Crohn's disease and ulcerative colitis by age at diagnosis". In: *Clinical Gastroenterology and Hepatology* 18.4 (2020), pp. 889–897.
- [125] Lim, B., Son, S., Kim, H., et al. "Enhanced deep residual networks for single image super-resolution". In: *CVPR workshops*. 2017, pp. 136–144.
- [126] Liu, A., Liu, Y., Gu, J., et al. "Blind image super-resolution: A survey and beyond". In: *TPAMI* (2022).
- [127] Liu, C., Cheng, S., Chen, C., et al. "M-flag: Medical vision-language pre-training with frozen language models and latent space geometry optimization". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2023, pp. 637–647.
- [128] Liu, G.-H., Vahdat, A., Huang, D.-A., et al. "I2SB: Image-to-Image Schröding Bridge". In: *ICML* (2023).
- [129] Liu, H.-T. D., Williams, F., Jacobson, A., et al. "Learning smooth neural functions via lipschitz regularization". In: *ACM SIGGRAPH 2022 Conference Proceedings*. 2022, pp. 1–13.
- [130] Liu, J., Li, H., Huang, T., et al. "Unsupervised Representation Learning for 3D MRI Super Resolution with Degradation Adaptation". In: *arXiv preprint arXiv:2205.06891* (2022).
- [131] Liu, X., Ma, Y., Shi, Z., et al. "Griddehazenet: Attention-based multi-scale network for image dehazing". In: *ICCV*. 2019, pp. 7314–7323.
- [132] Liu, Y., Chen, A., Shi, H., et al. "CT synthesis from MRI using multi-cycle GAN for head-and-neck radiation therapy". In: *CMIG* 91 (2021), p. 101953.
- [133] Lowe, D. G. "Distinctive image features from scale-invariant keypoints". In: *International journal of computer vision* 60 (2004), pp. 91–110.
- [134] Luo, Z., Huang, Y., Li, S., et al. "End-to-end alternating optimization for blind super resolution". In: *arXiv preprint arXiv:2105.06878* (2021).
- [135] Luo, Z., Gustafsson, F. K., Zhao, Z., et al. "Image Restoration with Mean-Reverting Stochastic Differential Equations". In: *ICML* (2023).
- [136] Luo, Z., Huang, H., Yu, L., et al. "Deep constrained least squares for blind image super-resolution". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 17642–17652.
- [137] Lustig, M., Donoho, D., and Pauly, J. M. "Sparse MRI: The application of compressed sensing for rapid MR imaging". In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 58.6 (2007), pp. 1182–1195.
- [138] Lv, J., Zhu, J., and Yang, G. "Which GAN? A comparative study of generative adversarial network-based fast MRI reconstruction". In: *Philos. Trans. Royal Soc.* 379.2200 (2021), p. 20200203.
- [139] Maclaren, J., Herbst, M., Speck, O., et al. "Prospective motion correction in brain imaging: a review". In: *Magn. Reson. Med.* 69.3 (2013), pp. 621–636.
- [140] Mahapatra, D., Schöffler, P. J., Tielbeek, J. A., et al. "Automatic detection and segmentation of Crohn's disease tissues from abdominal MRI". In: *IEEE Trans. Med. Imaging* 32.12 (2013), pp. 2332–2347.
- [141] Mei, K., Jiang, A., Li, J., et al. "An effective single-image super-resolution model using squeeze-and-excitation networks". In: *ICONIP*. Springer. 2018, pp. 542–553.
- [142] Meng, C., He, Y., Song, Y., et al. "Sdedit: Guided image synthesis and editing with stochastic differential equations". In: *ICLR* (2022).

- [143] Milletari, F., Navab, N., and Ahmadi, S.-A. "V-net: Fully convolutional neural networks for volumetric medical image segmentation". In: *2016 fourth international conference on 3D vision (3DV)*. Ieee. 2016, pp. 565–571.
- [144] Miyato, T., Kataoka, T., Koyama, M., et al. "Spectral normalization for generative adversarial networks". In: *ICLR* (2018).
- [145] Nguyen, J. C., De Smet, A. A., Graf, B. K., et al. "MR imaging-based diagnosis and classification of meniscal tears". In: *Radiographics* 34.4 (2014), pp. 981–999.
- [146] Nichol, A. Q. and Dhariwal, P. "Improved denoising diffusion probabilistic models". In: *ICML*. PMLR. 2021, pp. 8162–8171.
- [147] Oakden-Rayner, L., Dunnmon, J., Carneiro, G., et al. "Hidden stratification causes clinically meaningful failures in machine learning for medical imaging". In: *Proceedings of the ACM conference on health, inference, and learning*. 2020, pp. 151–159.
- [148] Oktay, O., Ferrante, E., Kamnitsas, K., et al. "Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation". In: *IEEE transactions on medical imaging* 37.2 (2017), pp. 384–395.
- [149] Oktay, O., Schlemper, J., Folgoc, L. L., et al. "Attention u-net: Learning where to look for the pancreas". In: *arXiv preprint arXiv:1804.03999* (2018).
- [150] Ouyang, C. "Generalizable deep learning based medical image segmentation". PhD thesis. Imperial College London, 2023.
- [151] Parisi, G. "Correlation functions and computer simulations". In: *Nuclear Physics B* 180.3 (1981), pp. 378–384.
- [152] Paschal, C. B. and Morris, H. D. "K-space in the clinic". In: *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 19.2 (2004), pp. 145–159.
- [153] Pawlowski, N., Coelho de Castro, D., and Glocker, B. "Deep structural causal models for tractable counterfactual inference". In: *Advances in neural information processing systems* 33 (2020), pp. 857–869.
- [154] Peled, S. and Yeshurun, Y. "Superresolution in MRI—perhaps sometimes". In: *Magn. Reson. Med.* 48.2 (2002), pp. 409–409.
- [155] Peng, C., Guo, P., Zhou, S. K., et al. "Towards performant and reliable undersampled MR reconstruction via diffusion model sampling". In: *MICCAI*. Springer. 2022, pp. 623–633.
- [156] Peng, J. and Wang, Y. "Medical image segmentation with limited supervision: a review of deep network models". In: *Ieee Access* 9 (2021), pp. 36827–36851.
- [157] Pihl-Lesnovska, K., Hjortswang, H., Ek, A.-C., et al. "Patients' perspective of factors influencing quality of life while living with Crohn disease". In: *Gastroenterology Nursing* 33.1 (2010), pp. 37–44.
- [158] Plenge, E., Poot, D. H., Bernsen, M., et al. "Super-resolution methods in MRI: can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time?" In: *Magn. Reson. Med.* 68.6 (2012), pp. 1983–1993.
- [159] Poincaré, H. "Sur le problème des trois corps et les équations de la dynamique". In: *Acta mathematica* 13.1 (1890), A3–A270.
- [160] Purkait, P. and Chanda, B. "Super resolution image reconstruction through Bregman iteration using morphologic regularization". In: *IEEE Trans. Image Process.* 21.9 (2012), pp. 4029–4039.
- [161] Qin, Q. "Point spread functions of the T2 decay in k-space trajectories with long echo train". In: *Magnetic Resonance Imaging* 30.8 (Oct. 2012), pp. 1134–1142. doi: [10.1016/j.mri.2012.04.017](https://doi.org/10.1016/j.mri.2012.04.017)

- [162] Qin, X., Wang, Z., Bai, Y., et al. "FFA-Net: Feature fusion attention network for single image dehazing". In: *AAAI*. 2020.
- [163] Qiu, Y., Zhang, K., Wang, C., et al. "MB-TaylorFormer: Multi-branch Efficient Transformer Expanded by Taylor Formula for Image Dehazing". In: *CVPR*. 2023, pp. 12802–12813.
- [164] Rad, M. S., Yu, T., Bozorgtabar, B., et al. "Test-Time Adaptation for Super-Resolution: You Only Need to Overfit on a Few More Images". In: *ICCV*. 2021, pp. 1845–1854.
- [165] Redmon, J., Divvala, S., Girshick, R., et al. "You only look once: Unified, real-time object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 779–788.
- [166] Reigada, L. C., Hoogendoorn, C. J., Walsh, L. C., et al. "Anxiety symptoms and disease severity in children and adolescents with Crohn disease". In: *J. Pediatr. Gastroenterol. Nutr.* 60.1 (2015), pp. 30–35.
- [167] Ren, D., Zuo, W., Hu, Q., et al. "Progressive image deraining networks: A better and simpler baseline". In: *CVPR*. 2019, pp. 3937–3946.
- [168] Ren, W., Ma, L., Zhang, J., et al. "Gated fusion network for single image dehazing". In: *CVPR*. 2018, pp. 3253–3261.
- [169] Rezende, D. and Mohamed, S. "Variational inference with normalizing flows". In: *International conference on machine learning*. PMLR. 2015, pp. 1530–1538.
- [170] Rimola, J., Rodríguez, S., García-Bosch, O., et al. "Magnetic resonance for assessment of disease activity and severity in ileocolonic Crohn's disease". In: *Gut* 58.8 (2009), pp. 1113–1120.
- [171] Robinson, R. D. "Reliable Machine Learning for Medical Imaging Data through Automated Quality Control and Data Harmonization". In: (July 2020).
- [172] Romano, Y., Isidoro, J., and Milanfar, P. "RAISR: rapid and accurate image super resolution". In: *IEEE Trans. Comput. Imaging* 3.1 (2016), pp. 110–125.
- [173] Ronneberger, O., Fischer, P., and Brox, T. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.
- [174] Rosen, M. J., Dhawan, A., and Saeed, S. A. "Inflammatory bowel disease in children and adolescents". In: *JAMA pediatrics* 169.11 (2015), pp. 1053–1060.
- [175] Rosenblatt, F. "The perceptron: a probabilistic model for information storage and organization in the brain." In: *Psychological review* 65.6 (1958), p. 386.
- [176] Rueckert, D., Glocker, B., and Kainz, B. *Learning clinically useful information from images: past, present and future*. 2016.
- [177] Rumelhart, D. E., Hinton, G. E., Williams, R. J., et al. *Learning internal representations by error propagation*. 1985.
- [178] Sánchez, I. and Vilaplana, V. "Brain MRI super-resolution using 3D generative adversarial networks". In: *arXiv preprint arXiv:1812.11440* (2018).
- [179] Sandler, R. and Eisen, G. "Epidemiology of inflammatory bowel disease In: Kirsner, editor". In: *Inflammatory bowel disease. 5th ed. Philadelphia: WB Saunders* (2000), p. 96.
- [180] Scarselli, F., Gori, M., Tsoi, A. C., et al. "The graph neural network model". In: *IEEE transactions on neural networks* 20.1 (2008), pp. 61–80.
- [181] Schlegl, T., Seeböck, P., Waldstein, S. M., et al. "f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks". In: *Medical image analysis* 54 (2019), pp. 30–44.
- [182] Schlemper, J., Caballero, J., Hajnal, J. V., et al. "A deep cascade of convolutional neural networks for MR image reconstruction". In: *IPMI*. Springer. 2017, pp. 647–658.

- [183] Schlemper, J., Oktay, O., Schaap, M., et al. "Attention gated networks: Learning to leverage salient regions in medical images". In: *Medical image analysis* 53 (2019), pp. 197–207.
- [184] Shi, L., Lu, Y., Dvornek, N., et al. "Automatic Inter-Frame Patient Motion Correction for Dynamic Cardiac PET Using Deep Learning". In: *IEEE Trans. Med. Imaging* 40.12 (2021), pp. 3293–3304. doi: [10.1109/TMI.2021.3082578](https://doi.org/10.1109/TMI.2021.3082578).
- [185] Shi, W., Caballero, J., Huszár, F., et al. "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network". In: *CVPR*. 2016, pp. 1874–1883.
- [186] Sim, B., Oh, G., Kim, J., et al. "Optimal transport driven CycleGAN for unsupervised learning in inverse problems". In: *SIAM Journal on Imaging Sciences* 13.4 (2020), pp. 2281–2306.
- [187] Simonyan, K. and Zisserman, A. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).
- [188] Siu, W.-C. and Hung, K.-W. "Review of image interpolation and super-resolution". In: *APSIPA*. IEEE. 2012, pp. 1–10.
- [189] Smith, T. B. "MRI artifacts and correction strategies". In: *Imaging in Medicine* 2.4 (2010), p. 445.
- [190] Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., et al. "Deep unsupervised learning using nonequilibrium thermodynamics". In: *ICML*. PMLR. 2015, pp. 2256–2265.
- [191] Song, J., Meng, C., and Ermon, S. "Denoising Diffusion Implicit Models". In: *NeurIPS* 32 (2019).
- [192] Song, Y. and Ermon, S. "Generative modeling by estimating gradients of the data distribution". In: *NeurIPS* 32 (2019).
- [193] Song, Y., Sohl-Dickstein, J., Kingma, D. P., et al. "Score-Based Generative Modeling through Stochastic Differential Equations". In: *ICLR*. 2020.
- [194] Spieker, V., Eichhorn, H., Hammernik, K., et al. "Deep learning for retrospective motion correction in MRI: a comprehensive review". In: *IEEE Transactions on Medical Imaging* 43.2 (2023), pp. 846–859.
- [195] Stelter, J., Weiss, K., Meineke, J., et al. "Subspace modeling enables time-efficient B1-corrected volumetric water-specific relaxometry with a long echo train CASPR trajectory". In: ().
- [196] Stelter, J., Weiss, K., Steinhilber, L., et al. "Simultaneous whole-liver water T1 and T2 mapping with isotropic resolution during free-breathing". In: *NMR in Biomedicine* 37.12 (2024), e5216.
- [197] Szegedy, C., Liu, W., Jia, Y., et al. "Going deeper with convolutions". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1–9.
- [198] Tan, M. and Le, Q. "Efficientnet: Rethinking model scaling for convolutional neural networks". In: *International conference on machine learning*. PMLR. 2019, pp. 6105–6114.
- [199] Tarroni, G., Oktay, O., Sinclair, M., et al. "A comprehensive approach for learning-based fully-automated inter-slice motion correction for short-axis cine cardiac MR image stacks". In: *MICCAI*. Springer. 2018, pp. 268–276.
- [200] Taylor, S. A., Mallett, S., Bhatnagar, G., et al. "Diagnostic accuracy of magnetic resonance enterography and small bowel ultrasound for the extent and activity of newly diagnosed and relapsed Crohn's disease (METRIC): a multicentre trial". In: *Lancet Gastroenterol. Hepatol.* 3.8 (2018), pp. 548–558.
- [201] Tielbeek, J. A., Makanyanga, J. C., Bipat, S., et al. "Grading Crohn disease activity with MRI: interobserver variability of MRI features, MRI scoring of severity, and correlation with Crohn disease endoscopic index of severity". In: *American Journal of Roentgenology* 201.6 (2013), pp. 1220–1228.
- [202] Torrents-Barrena, J., Piella, G., Masoller, N., et al. "Segmentation and classification in MRI and US fetal imaging: recent trends and future prospects". In: *Med. Image Anal.* 51 (2019), pp. 61–88.

- [203] Tschandl, P., Rinner, C., Apalla, Z., et al. "Human-computer collaboration for skin cancer recognition". In: *Nature medicine* 26.8 (2020), pp. 1229–1234.
- [204] Tu, Z., Talebi, H., Zhang, H., et al. "Maxim: Multi-axis mlp for image processing". In: *CVPR*. 2022, pp. 5769–5780.
- [205] Ulyanov, D., Vedaldi, A., and Lempitsky, V. "Instance normalization: The missing ingredient for fast stylization". In: *arXiv preprint arXiv:1607.08022* (2016).
- [206] Van de Walle, R., Lemahieu, I., and Achten, E. "Magnetic resonance imaging and the reduction of motion artifacts: review of the principles". In: *Technology and Health Care* 5.6 (1997), pp. 419–435.
- [207] Vapnik, V. "Principles of risk minimization for learning theory". In: *Advances in neural information processing systems* 4 (1991).
- [208] Venkatakrishnan, S. V., Bouman, C. A., and Wohlberg, B. "Plug-and-play priors for model based reconstruction". In: *2013 IEEE global conference on signal and information processing*. IEEE. 2013, pp. 945–948.
- [209] Viergever, M. A., Maintz, J. A., Klein, S., et al. *A survey of medical image registration—under review*. 2016.
- [210] Vincent, P. "A connection between score matching and denoising autoencoders". In: *Neural computation* 23.7 (2011), pp. 1661–1674.
- [211] Wang, D., Shelhamer, E., Liu, S., et al. "Tent: Fully test-time adaptation by entropy minimization". In: *arXiv preprint arXiv:2006.10726* (2020).
- [212] Wang, S., Qin, C., Savioli, N., et al. "Joint Motion Correction and Super Resolution for Cardiac Segmentation via Latent Optimisation". In: *MICCAI*. Springer. 2021, pp. 14–24.
- [213] Wang, W. and Shen, J. "Deep visual attention prediction". In: *IEEE Transactions on Image Processing* 27.5 (2017), pp. 2368–2378.
- [214] Wang, X., Xie, L., Dong, C., et al. "Real-esrgan: Training real-world blind super-resolution with pure synthetic data". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 1905–1914.
- [215] Wang, X., Xie, L., Dong, C., et al. "Realesrgan: Training real-world blind super-resolution with pure synthetic data supplementary material". In: *Computer Vision Foundation open access* (2022).
- [216] Wang, Z., Chen, J., and Hoi, S. C. H. "Deep Learning for Image Super-Resolution: A Survey". In: *IEEE Trans. Pattern Anal. Mach. Intell.* 43.10 (2021), pp. 3365–3387. doi: [10.1109/TPAMI.2020.2982166](https://doi.org/10.1109/TPAMI.2020.2982166)
- [217] Wang, Z., Bovik, A. C., Sheikh, H. R., et al. "Image quality assessment: from error visibility to structural similarity". In: *IEEE TIP* 13.4 (2004), pp. 600–612.
- [218] Webber, G. and Reader, A. J. "Diffusion models for medical image reconstruction". In: *BJR—Artificial Intelligence* 1.1 (2024), ubae013.
- [219] Welker, S., Richter, J., and Gerkmann, T. "Speech enhancement with score-based generative models in the complex STFT domain". In: *arXiv preprint arXiv:2203.17004* (2022).
- [220] Wood, M. L. and Henkelman, R. M. "MR image artifacts from periodic motion". In: *Medical physics* 12.2 (1985), pp. 143–151.
- [221] Wu, H., Qu, Y., Lin, S., et al. "Contrastive learning for compact single image dehazing". In: *CVPR*. 2021, pp. 10551–10560.
- [222] Wu, Y. and He, K. "Group normalization". In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 3–19.
- [223] Xiao, Z., Kreis, K., and Vahdat, A. "Tackling the Generative Learning Trilemma with Denoising Diffusion GANs". In: *ICLR*. 2021.

- [224] Xie, J., Li, W., Zhan, X., et al. "Masked frequency modeling for self-supervised visual pre-training". In: *arXiv preprint arXiv:2206.07706* (2022).
- [225] Yang, G., Yu, S., Dong, H., et al. "DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction". In: *TMI* 37.6 (2017), pp. 1310–1321.
- [226] Yang, L., Zhang, Z., Song, Y., et al. "Diffusion models: A comprehensive survey of methods and applications". In: *ACM Computing Surveys* (2022).
- [227] Yang, W., Tan, R. T., Feng, J., et al. "Deep joint rain detection and removal from a single image". In: *CVPR*. 2017, pp. 1357–1366.
- [228] Yang, W., Tan, R. T., Feng, J., et al. "Joint rain detection and removal from a single image with contextualized deep networks". In: *IEEE TPAMI* 42.6 (2019), pp. 1377–1393.
- [229] Yoon, J., Jordon, J., and Schaar, M. "Gain: Missing data imputation using generative adversarial nets". In: *International conference on machine learning*. PMLR. 2018, pp. 5689–5698.
- [230] You, H., Cheng, Y., Cheng, T., et al. "Bayesian cycle-consistent generative adversarial networks via marginalizing latent sampling". In: *IEEE Trans. Neural Netw. Learn.* 32.10 (2020), pp. 4389–4403.
- [231] Yue, Z., Wang, J., and Loy, C. C. "ResShift: Efficient Diffusion Model for Image Super-resolution by Residual Shifting". In: *NeurIPS* (2023).
- [232] Zaitsev, M., Maclaren, J., and Herbst, M. "Motion artifacts in MRI: A complex problem with many partial solutions". In: *Magn. Reson. Imaging* 42.4 (2015), pp. 887–901.
- [233] Zamir, S. W., Arora, A., Khan, S., et al. "Multi-stage progressive image restoration". In: *CVPR*. 2021, pp. 14821–14831.
- [234] Zamir, S. W., Arora, A., Khan, S., et al. "Restormer: Efficient transformer for high-resolution image restoration". In: *CVPR*. 2022, pp. 5728–5739.
- [235] Zbontar, J., Knoll, F., Sriram, A., et al. "fastMRI: An open dataset and benchmarks for accelerated MRI". In: *arXiv preprint arXiv:1811.08839* (2018).
- [236] Zech, J. R., Badgeley, M. A., Liu, M., et al. "Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study". In: *PLoS medicine* 15.11 (2018), e1002683.
- [237] Zhang, K., Liang, J., Van Gool, L., et al. "Designing a practical degradation model for deep blind image super-resolution". In: *CVPR*. 2021, pp. 4791–4800.
- [238] Zhang, Q. and Chen, Y. "Fast sampling of diffusion models with exponential integrator". In: *ICLR* (2022).
- [239] Zhang, R., Isola, P., Efros, A. A., et al. "The unreasonable effectiveness of deep features as a perceptual metric". In: *CVPR*. 2018, pp. 586–595.
- [240] Zhang, W., Basaran, B., Meng, Q., et al. "MoCoSR: Respiratory Motion Correction and Super-Resolution for 3D Abdominal MRI". In: *MICCAI*. Springer. 2023, pp. 121–131.
- [241] Zhang, W., Qiao, M., Zang, C., et al. "Multi-agent reasoning for cardiovascular imaging phenotype analysis". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2025, pp. 429–439.
- [242] Zhang, W., Zang, C., and Kainz, B. "From Self-Check to Consensus: Bayesian Strategic Decoding in Large Language Models". In: *The Thirty-ninth Annual Conference on Neural Information Processing Systems*.
- [243] Zhang, W., Zang, C., Li, L., et al. "Stability and Generalizability in SDE Diffusion Models with Measure-Preserving Dynamics". In: *Advances in Neural Information Processing Systems* 37 (2025), pp. 81606–81644.
- [244] Zhao, M., Wei, Y., and Wong, K. K. "A Generative Adversarial Network technique for high-quality super-resolution reconstruction of cardiac magnetic resonance images". In: *Magn. Reson. Imaging* 85 (2022), pp. 153–160.

- [245] Zhao, M., Gönczi, L., Lakatos, P. L., et al. "The burden of inflammatory bowel disease in Europe in 2020". In: *Journal of Crohn's and Colitis* 15.9 (2021), pp. 1573–1587.
- [246] Zhao, X., Zhang, Y., Zhang, T., et al. "Channel Splitting Network for Single MR Image Super-Resolution". In: *IEEE Trans. Image Process.* 28.11 (2019), pp. 5649–5662. doi: [10.1109/TIP.2019.2921882](https://doi.org/10.1109/TIP.2019.2921882)
- [247] Zheng, Y., John, M., Liao, R., et al. "Automatic aorta segmentation and valve landmark detection in C-arm CT: application to aortic valve implantation". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010: 13th International Conference, Beijing, China, September 20-24, 2010, Proceedings, Part I* 13. Springer. 2010, pp. 476–483.
- [248] Zhou, L., Lou, A., Khanna, S., et al. "Denoising diffusion bridge models". In: *arXiv preprint arXiv:2309.16948* (2023).
- [249] Zhou, X., Koltun, V., and Krähenbühl, P. "Probabilistic two-stage detection". In: *arXiv preprint arXiv:2103.07461* (2021).
- [250] Zhu, J.-Y., Park, T., Isola, P., et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks". In: *CVPR*. 2017, pp. 2223–2232.