Unsupervised Motion-Compensated Decomposition for Cardiac MRI Reconstruction via Neural Representation

Xuanyu Tian[1,2], Lixuan Chen[3], Qing Wu[1], Xiao Wang[1], Jie Feng[4], Yuyao Zhang[1], Hongjiang Wei[4*]

[1] School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China
[2] Lingang Laboratory, Shanghai 200031, China
[3] Electrical and Computer Engineering, University of Michigan, MI 48105, United States
[4] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200127, China
{tianxy, wuqing, zhangyy8}@shanghaitech.edu.cn, chenlx@umich.edu, {jiefeng, hongjiang.wei}@sjtu.edu.cn

Abstract

Cardiac magnetic resonance (CMR) imaging is widely used to characterize cardiac morphology and function. To accelerate CMR imaging, various methods have been proposed to recover high-quality spatiotemporal CMR images from highly undersampled k-t space data. However, current CMR reconstruction techniques either fail to achieve satisfactory image quality or are restricted by the scarcity of ground truth data, leading to limited applicability in clinical scenarios. In this work, we proposed MoCo-INR, a new unsupervised method that integrates implicit neural representations (INR) with the conventional motion-compensated (MoCo) framework. Using the explicit motion modeling and the continuous prior of INRs, our MoCo-INR can produce accurate cardiac motion decomposition and high-quality CMR reconstruction. Moreover, we present a new INR network architecture tailored to the CMR problem, which can greatly stabilize model optimization. Experiments on retrospective (*i.e.*, simulated) datasets demonstrate the superiority of MoCo-INR over state-of-the-art methods, achieving fast convergence and fine-detailed reconstructions at ultra-high acceleration factors (*e.g.*, 20× in VISTA sampling). In addition, evaluations on prospective (*i.e.*, real-acquired) free-breathing CMR scans highlight its clinical practicality for real-time imaging. Several ablation studies also confirm the effectiveness of critical components of MoCo-INR.

Code — https://github.com/MeijiTian/MoCo-INR

1 Introduction

Magnetic resonance (MR) imaging offers unparalleled soft tissue contrast and, as a non-invasive modality, serves as a versatile tool for evaluating cardiac function (Pennell 2010). However, the long acquisition time makes it difficult to capture cardiac motion accurately. Scanning undersampled k-space data for each short temporal frame is an effective strategy to accelerate cardiac MR (CMR) acquisition. Nevertheless, reconstructing artifact-free, dynamic MR images from undersampled k-t space (*i.e.*, spatiotemporal) data poses a challenging ill-posed problem due to the violation of the Nyquist–Shannon sampling theorem (Nyquist 1928).

Many studies have proposed exploiting the inherent spatial and temporal correlations in the image sequences to alleviate the ill-posedness of the CMR problem (Oscanoa et al. 2023). A classical strategy is to incorporate a low-rank prior into the compressed sensing (CS) framework (Lingala et al. 2011; Zhao et al. 2011; Otazo, Candès, and Sodickson 2015), where the dynamic image sequence is decomposed into low-rank and sparse components. The integration of the low-rank prior effectively utilizes the spatiotemporal redundancy of dynamic images, thus enhancing CMR results.

In contrast, motion-compensated (MoCo) methods (Pan et al. 2022, 2024; Morales et al. 2019) explicitly decouple frame-specific deformations from a single canonical image, allowing all temporal frames to share a common canonical spatial representation. Thus, this decoupled modeling can achieve better CMR reconstructions for highly undersampled acquisitions. However, most existing MoCo approaches rely on fully sampled cine CMR data for supervised training. Although effective, cine CMR acquisition requires breath-holding, which increases acquisition cost and restricts the practicality and generalizability of these methods in real-time free-breathing scenarios.

As an unsupervised learning paradigm, implicit neural representation (INR) has shown great promise in dynamic medical reconstruction (Reed et al. 2021; Huang et al. 2023; Kunz, Ruschke, and Heckel 2024; Catalán et al. 2025; Feng et al. 2025), where the image sequences are formulated as a continuous function of spatial–temporal coordinates. Thanks to the learning basis of neural networks towards low-frequency signals (Xu et al. 2019; Rahaman et al. 2019), INR implicitly captures the spatiotemporal redundancy of dynamic images, which produces improved reconstructions. However, when applied to extremely ill-posed inverse problems, the continuous prior of INR is often insufficient and requires additional regularization priors, such as image-domain prior (Kazerouni et al. 2024; Tian et al. 2025), low-rank models (Feng et al. 2025), denoisers (Iskender et al. 2025), or generative priors (Du et al. 2024), to enhance reconstruction quality and stability.

With the achievements of INR combined with MoCo scheme in 4D scene reconstruction (Pumarola et al. 2021; Park et al. 2021), several studies have extended this framework to 4D CT (Zhang et al. 2023) and time-resolved MRI

[*] Corresponding author.

reconstruction (Shao et al. 2024, 2025; Chen et al. 2025). These approaches effectively capture respiratory motion, which is relatively simpler than cardiac motion, but often struggle to represent high-frequency details. Due to the high-frequency and fine detail of cardiac motion, adopting INR to achieve accurate cardiac motion decomposition from undersampled data is non-trivial. Meanwhile, INR is known for slow optimization, limiting its clinical practicality. Hash-grid encoding (Müller et al. 2022) has been proposed to accelerate convergence and enhance high-frequency representation. However, its inherent discrete feature representation compromises the continuity of INR, leading to inconsistencies in continuous space and unstable optimization in dynamic reconstruction.

In this work, we propose **MoCo-INR**, a novel unsupervised CMR reconstruction method. Our key idea is to introduce unsupervised INRs into the MoCo framework, enabling accurate cardiac motion decomposition and the recovery of high-frequency image details. Conceptually, we explicitly decompose dynamic CMR sequences into time-varying deformations and a shared canonical image, both modeled as continuous functions parameterized by two INR networks. Benefiting from the continuous priors of INRs and the explicit motion decomposition, we effectively solve the highly ill-posed CMR inverse problem in an unsupervised manner. Moreover, we present a new INR network architecture tailored to the CMR problem, which consists of a coarse-to-fine hash encoding strategy and a CNN-based decoder. Compared to existing INR architectures, our proposed design achieves more stable optimization and produces CMR images with fine anatomical details.

We evaluate the proposed MoCo-INR on both retrospective cine CMR reconstruction under various acquisition schemes and prospective free-breathing CMR reconstruction. The results demonstrate that MoCo-INR outperforms state-of-the-art (SOTA) unsupervised methods, delivering both faster convergence and higher-fidelity reconstructions, particularly under ultra-high acceleration factors (20× for Cartesian and 69× for non-Cartesian). In addition, extensive ablation studies validate the effectiveness the several key components of our MoCo-INR.

The main contributions as summarized as follows:

- We introduce the INR to the MoCo framework, enabling accurate cardiac motion decomposition and fine-detailed reconstruction in an unsupervised manner.
- We propose a novel INR network architecture tailored to the CMR problem, which can greatly stabilize model optimization.
- We perform extensive experiments confirming the superiority of our unsupervised MoCo-INR in fast convergence and robustness with various CMR acquisitions.

2 Related Work

2.1 Motion-Compensated Approaches for CMR

To leverage motion information, motion-compensated (MoCo) methods are introduced into CMR reconstruction to further improve performance. MoCo methods (Batchelor et al. 2005; Qi et al. 2021; Hammernik et al. 2021; Munoz et al. 2022; Zou et al. 2022; Pan et al. 2024) explicitly decompose dynamic images into a canonical (or template) image and a sequence of canonical-to-observation displacement vector fields (DVFs), which can effectively exploit the spatio-temporal redundancies. The reconstruction task is reformulated as two sub-problems: motion estimation and canonical image reconstruction, which are typically solved iteratively or within a joint optimization framework. Thus, accurate estimation of cardiac motion is crucial to both the canonical image quality and the final reconstruction performance. With the emergence of deep learning (DL), supervised MoCo methods (Qi et al. 2021; Hammernik et al. 2021; Pan et al. 2024) have gained importance in motion estimation due to their promising performance. However, these methods often suffer from performance degradation when the acquisition settings change (*e.g.*, different sampling patterns and accelerator factors) deviate from the training data. Moreover, the long acquisition time of MRI makes it difficult to obtain high-quality ground-truth labels. These issues pose substantial obstacles to the practical application of supervised methods in clinical settings. Current unsupervised MoCo methods have mainly focused on respiratory motion (Munoz et al. 2022; Zou et al. 2022) but have rarely explored cardiac motion (Kettelkamp et al. 2023), which is more complex and requires higher temporal resolution.

2.2 INR for Dynamic MRI Reconstruction

Recently, many INR-based dynamic MRI reconstruction methods (Huang et al. 2023; Catalán et al. 2025; Feng et al. 2025) have been proposed, formulating dynamic MRI image sequences as spatial–temporal functions represented in either the image domain or in k-space. While INR effectively exploits spatial–temporal correlations to constrain the reconstruction, existing methods are known for their slow convergence, often requiring hours to reconstruct a single slice (Kunz, Ruschke, and Heckel 2024), particularly when modeling high-frequency features. The current SOTA approach (Feng et al. 2025) adopts hash-grid encoding to accelerate convergence but still relies on additional low-rank and sparsity constraints to ensure consistent reconstruction quality. However, explicit motion-compensated representations remain unexplored in improving the robustness and efficiency of INR-based optimization in real-time cardiac MRI reconstruction.

3 Preliminaries

3.1 Forward Model of Dynamic CMR

The forward physical model of dynamic CMR acquisition can be expressed as:

$$\boldsymbol{y}_{t,c} = \boldsymbol{M}_t \mathcal{T} \boldsymbol{S}_c \boldsymbol{x}_t + \boldsymbol{n}_{t,c}, \tag{1}$$

where $\boldsymbol{x}_t \in \mathbb{C}^N$ is the image at any timestemp $t = 1, \ldots, T$ and $\boldsymbol{S}_c \in \mathbb{C}^{N \times N}$ represents the c^{th} coil sensitivity map. $\mathcal{T}$ is the Fourier transform operator, $\boldsymbol{M}_t \in \mathbb{R}^{M \times N}$ is the binary diagonal undersampling matrix, $\boldsymbol{n}_{t,c} \in \mathbb{C}^M$ is the system noise assuming Gaussian distribution, and $\boldsymbol{y}_{t,c} \in \mathbb{C}^M$ is the acquired k-space measurement.

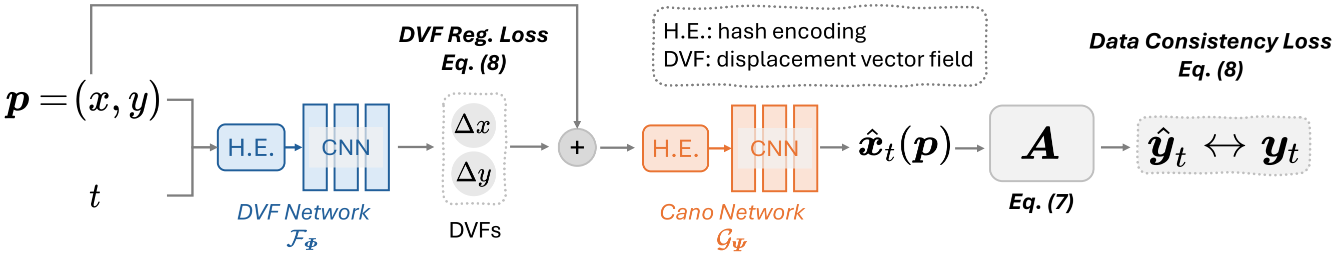
$p = (x, y)$
t
H.E.
CNN
DVF Network
$\mathcal{F}_{\boldsymbol{\Phi}}$
DVF Reg. Loss
Eq. (8)
Δx
Δy
DVFs
H.E.: hash encoding
DVF: displacement vector field
+
H.E.
CNN
Cano Network
$\mathcal{G}_{\boldsymbol{\Psi}}$
$\hat{\boldsymbol{x}}_t(\boldsymbol{p})$
$\boldsymbol{A}$
Eq. (7)
Data Consistency Loss
Eq. (8)
$\hat{\boldsymbol{y}}_t \leftrightarrow \boldsymbol{y}_t$

Figure 1: Overview of the proposed MoCo-INR framework. Given any spatial coordinate $\boldsymbol{p} = (x, y)$ in the physical space and temporal coordinate t, our deformation network predicts the corresponding time-varying displacement vector field (DVF). Adding this displacement to the spatial coordinate in physical space yields the associated location in the canonical space. Then, the canonical network maps these warped coordinates to the dynamic image $\boldsymbol{x}_t$. Finally, the two networks are jointly optimized by minimizing the data-consistency loss (Eq. 8) and DVF regularization loss (Eq. 8).

3.2 Motion-Compensated Representation

To fully exploit the spatiotemporal redundancy in the image sequence $\{\boldsymbol{x}_t\}_{t=1}^T$ and alleviate the ill-posedness of the CMR inverse problem, motion-compensated (MoCo) representation decouples each frame $\boldsymbol{x}_t$ into a shared canonical image $\boldsymbol{x}_{\text{cano}}$ and a corresponding displacement vector field (DVF) $\boldsymbol{u}_t$. Formally, this can be expressed as:

$$\boldsymbol{x}_t = \mathcal{W}(\boldsymbol{x}_{\text{cano}}, \boldsymbol{u}_t), \tag{2}$$

where $\mathcal{W}$ denotes the image warping operator. The DVF defines, for each voxel of the frame $\boldsymbol{x}_t$, the offset $(\Delta x, \Delta y)_t$ between the canonical space and its physical location.

Conventional MoCo approaches for dynamic MRI represent the canonical image $\boldsymbol{x}_{\text{cano}}$ as a discrete matrix. The warping operator is then used with the DVFs $\{\boldsymbol{u}_t\}_{t=1}^T$ to interpolate this matrix and generate the image sequence $\{\boldsymbol{x}_t\}_{t=1}^T$. Although effective, discrete interpolation may lose high-frequency details and thus limits reconstruction quality.

4 Proposed Method

Our goal is to reconstruct artifact-free CMR images with high spatiotemporal resolution from highly undersampled k-t space data in *an unsupervised way*. To this end, we propose MoCo-INR, a new unsupervised CMR method that first integrates INR into the MoCo framework. Thanks to the continuous representation enabled by INR, MoCo-INR can achieve accurate estimation of cardiac motion and image reconstructions with preserved high-frequency details.

4.1 Continuous Representations of DVFs and Canonical Image

To accurately recover both DVFs $\{\boldsymbol{u}_t\}_{t=1}^T$ and the shared canonical image $\boldsymbol{x}_{\text{cano}}$, we leverage INR to formulate them in a continuous form, instead of discrete matrices as in traditional MoCo-based methods. Specifically, the DVFs are defined as a single continuous function $\boldsymbol{f}$ of spatial-temporal coordinate, as below:

$$\boldsymbol{f}: \quad (\boldsymbol{p}, t) \in \mathbb{R}^3 \mapsto \boldsymbol{u}_t(\boldsymbol{p}) = (\Delta x, \Delta y) \in \mathbb{R}^2, \tag{3}$$

where $(\boldsymbol{p}, t)$ denotes any spatial-temporal coordinate in the physical space, and $\boldsymbol{u}_t(\boldsymbol{p})$ is the displacement vector to map $\boldsymbol{p}$ into the canonical space. While the complex-valued canonical image $\boldsymbol{x}_{\text{cano}}$ is formulated as a continuous function $\boldsymbol{g}$ of spatial coordinate, as below:

$$\boldsymbol{g}: \quad \tilde{\boldsymbol{p}} \in \mathbb{R}^2 \mapsto \boldsymbol{x}_{\text{cano}}(\tilde{\boldsymbol{p}}) = a(\tilde{\boldsymbol{p}}) + jb(\tilde{\boldsymbol{p}}) \in \mathbb{C}, \tag{4}$$

where $\tilde{\boldsymbol{p}}$ represents any coordinate in a 2D canonical space, and $\boldsymbol{x}_{\text{cano}}(\tilde{\boldsymbol{p}})$ is the corresponding complex-valued intensity.

MoCo-INR uses a DVF network $\mathcal{F}_{\boldsymbol{\Phi}}$ and a canonical network $\mathcal{G}_{\boldsymbol{\Psi}}$ to approximate the two functions, respectively. Technically, $\mathcal{F}_{\boldsymbol{\Phi}}$ takes spatial-temporal coordinates as input and outputs the DVF estimations (*i.e.*, $\boldsymbol{u}_t(\boldsymbol{p}) = \mathcal{F}_{\boldsymbol{\Phi}}(\boldsymbol{p}, t)$), while $\mathcal{G}_{\boldsymbol{\Psi}}$ takes spatial coordinates as input and predicts the real and imaginary parts of the canonical image (*i.e.*, $[a(\tilde{\boldsymbol{p}}), b(\tilde{\boldsymbol{p}})] = \mathcal{G}_{\boldsymbol{\Psi}}(\tilde{\boldsymbol{p}})$). Benefiting from the learning bias toward low-frequency continuous signals (Xu et al. 2019; Rahaman et al. 2019), the continuous functions $\boldsymbol{f}$ and $\boldsymbol{g}$ can be well approximated, enabling high-quality reconstructions of both the DVFs and canonical image.

4.2 Model Optimization

Fig. 1 shows the workflow of MoCo-INR, where we jointly optimize the DVF $\mathcal{F}_{\boldsymbol{\Phi}}$ and the canonical network $\mathcal{G}_{\boldsymbol{\Psi}}$.

Prediction of CMR Image. Given the acquired k-space data $\boldsymbol{y}_t$ at any timestemp $t = 1, \ldots, T$, we first feed the spatial-temporal coordinate $(\boldsymbol{p}, t)$, defined in the physical space, into the network $\mathcal{F}_{\boldsymbol{\Phi}}$ to predict the corresponding DVF $\boldsymbol{u}_t(\boldsymbol{p})$. This DVF is then used to transform the spatial coordinate from the physical space to the canonical space. Formally, this process can be expressed as:

$$\tilde{\boldsymbol{p}} = \boldsymbol{p} + \boldsymbol{u}_t(\boldsymbol{p}), \quad \text{with} \quad \boldsymbol{u}_t(\boldsymbol{p}) = \mathcal{F}_{\boldsymbol{\Phi}}(\boldsymbol{p}, t). \tag{5}$$

Then, the canonical network $\mathcal{G}_{\boldsymbol{\Psi}}$ takes the warped coordinates $\tilde{\boldsymbol{p}}$ as input and estimates the corresponding dynamic image as follows:

$$\hat{\boldsymbol{x}}_t(\tilde{\boldsymbol{p}}) = \mathcal{G}_{\boldsymbol{\Psi}}(\tilde{\boldsymbol{p}}). \tag{6}$$

Differentiable Forward Model. According to the forward acquisition of dynamic CMR (Eq. 1), we generate the k-space data estimations $\hat{\boldsymbol{y}}_t$ from the predicted CMR image $\hat{\boldsymbol{x}}_t$, which is defined as follows:

$$\boldsymbol{A}: \quad \hat{\boldsymbol{y}}_t = \boldsymbol{M}_t \mathcal{T} \boldsymbol{S}_c \hat{\boldsymbol{x}}_t, \tag{7}$$

where the operators $\boldsymbol{M}_t$, $\mathcal{T}$, and $\boldsymbol{S}_c$ depend on the CMR acquisition protocols and are known. More importantly, the forward model $\boldsymbol{A}$ is differentiable, allowing the use of gradient descent-based backpropagation algorithms (*e.g.*, Adam) for network optimization.

Loss Function. Finally, the DVF network $\mathcal{F}_{\boldsymbol{\Phi}}$ and the canonical network $\mathcal{G}_{\boldsymbol{\Psi}}$ are jointly optimized by minimizing the following loss function $\mathcal{L}$ as below:

$$\begin{aligned} \mathcal{L} &= \underbrace{\left\|\hat{\boldsymbol{y}}_t - \boldsymbol{y}_t\right\|_1}_{\mathcal{L}_{\mathrm{DC}}} + \mathcal{L}_{\mathrm{DVF}}, \\ \text{with} \quad \mathcal{L}_{\mathrm{DVF}} &= \left\|\boldsymbol{u}_t\right\|_1 + \left\|\nabla \boldsymbol{u}_t\right\|_1 + \left\|\nabla^2 \boldsymbol{u}_t\right\|_1, \end{aligned} \tag{8}$$

where $\mathcal{L}_{\mathrm{DC}}$ represents the data consistency term that minimizes the distance between the acquired and estimated k-space data. $\mathcal{L}_{\mathrm{DVF}}$ is a sparsity and smoothness regularization term that enforces plausible DVF estimations and further stabilizes the joint network optimization. Its effectiveness is explored in the following experiments.

CMR Image Reconstruction. After model optimization, the high-quality CMR sequence $\{\boldsymbol{x}_t^*\}_{t=1}^T$ can be directly reconstructed by feeding all spatial-temporal coordinates $(\boldsymbol{p}, t)$ into the well-trained MoCo-INR model.

4.3 INR Network Designed for CMR Problem

Traditional INR networks often consist of a coordinate encoding module, such as Hash encoding (Müller et al. 2022), and an MLP decoder. These encoding modules can greatly enhance the network's ability to capture high-frequency signals, improving detail recovery. However, applying existing INR networks in CMR often yields unsatisfactory performance due to the problem's ill-posedness and strong nonlinearity. To address this, we propose a novel coarse-to-fine hash encoding and a CNN-based decoder to achieve reliable DVF estimation and detailed image reconstruction.

Coarse-to-Fine Hash Encoding. Hash encoding (Müller et al. 2022) is a cutting-edge encoding strategy. It maps low-dimensional coordinates $\boldsymbol{p}$ into high-dimensional features $\gamma(\boldsymbol{p}) = \gamma_1(\boldsymbol{p}) \oplus \cdots \oplus \gamma_l(\boldsymbol{p}) \oplus \cdots \oplus \gamma_L(\boldsymbol{p}) \in \mathbb{R}^{LF}$, where each $\gamma_l(\boldsymbol{p}) \in \mathbb{R}^F$ denotes an F-dimensional feature at the l-th resolution level. The low-resolution features (*e.g.*, γ_1) capture low-frequency global structures, while the high-resolution ones (*e.g.*, γ_L) model high-frequency local details. A recent study on MRI reconstruction (Wu et al. 2025) demonstrated that the global structures are more crucial for rigid motion correction. Inspired by this observation, we propose a novel coarse-to-fine scheme for the CMR problem. As shown in Fig. 2, the optimization starts by learning the low-frequency features to capture global motion. Then, the higher-frequency features are progressively optimized to refine fine-scale motion details. This coarse-to-fine hash encoding can enhance reliable DVF estimations, thus enabling improved CMR reconstructions.

CNN-based Decoder. Existing INR networks typically use an MLP as decoder to transform the encoded features into target signals. Although effective, the voxel-wise mapping of MLP-based decoders struggles to capture the spatial continuity of images (Mihajlovic et al. 2024). Moreover, the powerful fitting capability introduced by the encoding may further lead to overfitting to undersampled data, resulting in high-frequency artifacts. To address these issues, we introduce a three-layer convolutional neural network (CNN) to replace the conventional MLP decoder. Owing to the inductive bias of CNNs toward local structures, the continuous functions $\boldsymbol{f}$ and $\boldsymbol{g}$, which represent the DVFs and the canonical image, can be better approximated, thereby improving the quality of reconstructed CMR images.

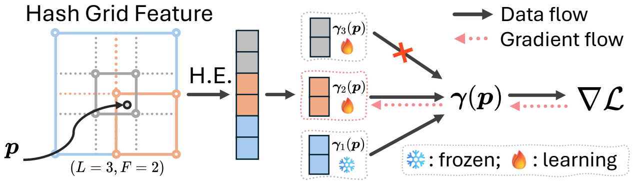

Figure 2: Illustration of the proposed coarse-to-fine hash encoding strategy. Given any input coordinate $\boldsymbol{p}$, the low-frequency feature (*i.e.*, γ_1) is learned first and then frozen. As the optimization proceeds, higher-frequency features (*i.e.*, γ_2 and γ_3) are progressively optimized.

5 Experimental Settings

5.1 Retrospective Reconstruction Study

Dataset. We used the fully sampled cardiac cine dataset from the public OCMR dataset (Chen et al. 2020). All scans were acquired using prospective ECG gating and breath-holding. For this study, we selected 11 slices, including five long-axis (LAX) views and six short-axis (SAX) views, acquired on a 1.5T clinical MRI scanner (Magnetom Aera, Siemens Healthineers).

Simulation Process. The original data were cropped into a square shape and resized to a size of 208×208. To evaluate reconstruction performance under different sampling strategies, we simulated both Cartesian and non-Cartesian undersampling. **Cartesian**: we adopted VISTA pattern with two accelelation factor (AF) of 12 and 20. **Non-Cartesian**: we adopted golden-angle (GA) radial pattern combined with the NUFFT operator (Fessler and Sutton 2003), using 8 and 3 spokes per frame, corresponding to AF of 26 and 69.3.

5.2 Prospective Reconstruction Study

We used prospectively acquired real-time CMR data from the public OCMR dataset (Chen et al. 2020), under **free-breathing** conditions with VISTA sampling mask and the acceleration factor of AF=9. Ten slices of SAX view were selected for the study, each with an in-plane resolution of $2.08 \times 2.08\text{mm}^2$ and a slice thickness of 8mm. Each slice comprises 65 temporal frames, corresponding to a temporal resolution of 38.4 ms ($\approx$26 Hz).

5.3 Methods in Comparison & Metrics

Methods in Comparison. We compare our method with five representative unsupervised methods: (1) Compressed-sensing (CS) based: ℓ_1**-Wavelet** (Lustig, Donoho, and Pauly

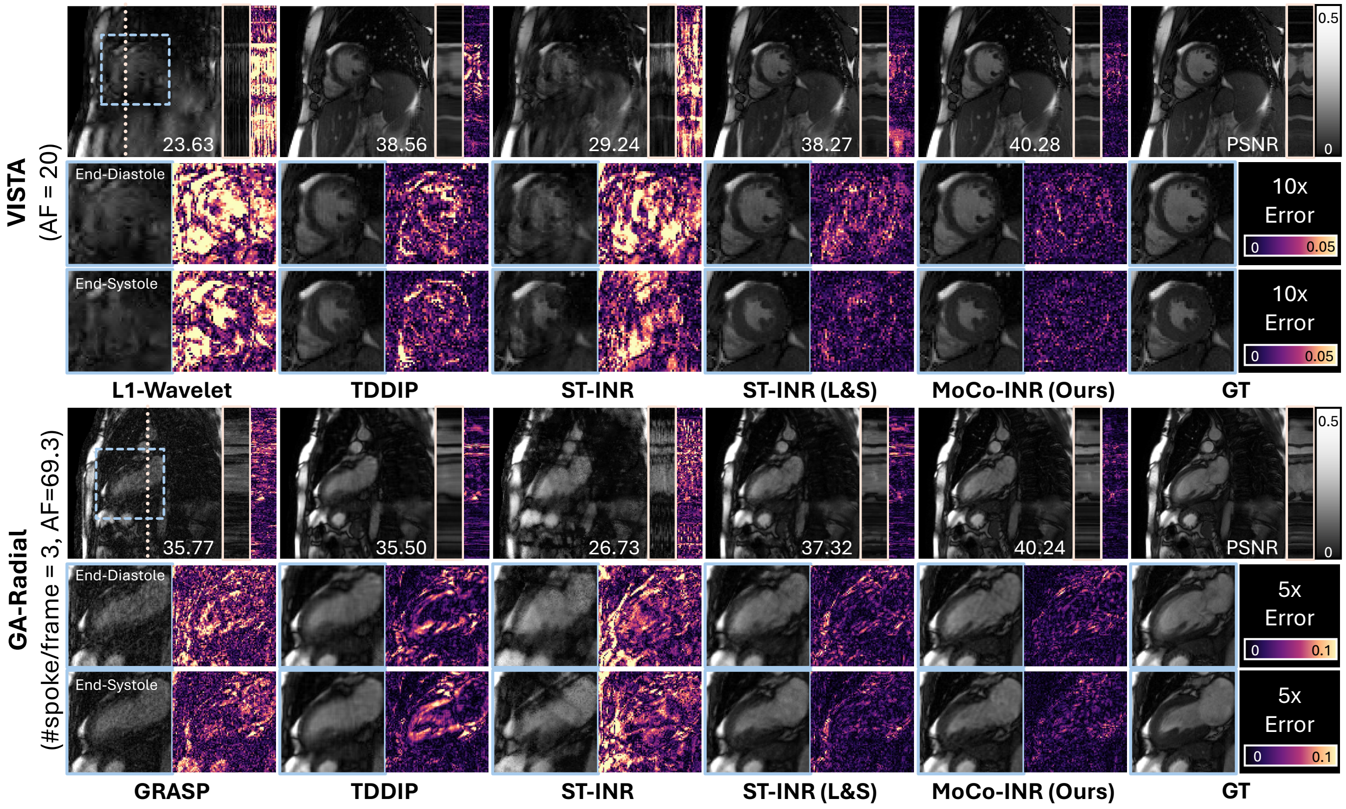

Figure 3: Qualitative results of retrospective reconstructions obtained with the compared methods. The figure displays the reconstructed image, its profile line over time (the y-t plane), and the corresponding error map. The selected y-axis is indicated by a white dashed line, and zoom-in boxes highlight regions of interest at the end-diastole (ED) and end-systole (ES) phases. The upper part shows results for SAX slice acquired using a VISTA sampling pattern with an acceleration factor of AF=20. The bottom part shows results for LAX slice acquired using a golden-angle radial sampling pattern with 3 spokes.

Sampling	AF	Metric	ℓ_1-Wavelet	TDDIP	ST-INR	ST-INR (L&S)	MoCo-INR
VISTA	12×	PSNR	28.00±1.98***	38.05±2.99***	36.31±2.62***	41.35±2.60***	**42.25±2.64**
		SSIM	0.734±0.038***	0.943±0.025***	0.934±0.020***	**0.972±0.012**$^{\blacktriangledown}$	0.971±0.013
		nRMSE (ROI)	0.450±0.088***	0.206±0.037***	0.150±0.022***	0.109±0.024***	**0.093±0.017**
	20×	PSNR	23.82±1.69***	36.58±2.70***	31.24±3.00***	36.26±2.94***	**39.53±2.58**
		SSIM	0.576±0.039***	0.929±0.026***	0.843±0.042***	0.937±0.021***	**0.957±0.017**
		nRMSE (ROI)	0.658±0.044***	0.217±0.037***	0.229±0.030***	0.158±0.024***	**0.125±0.022**

Sampling	AF	Metric	GRASP	TDDIP	ST-INR	ST-INR (L&S)	MoCo-INR
GA Radial	26.0×	PSNR	32.14±3.33***	34.10±1.97***	30.96±2.04***	38.85±2.86**	**40.33±2.48**
		SSIM	0.886±0.037***	0.895±0.022***	0.812±0.028***	0.956±0.016*	**0.960±0.016**
		nRMSE (ROI)	0.253±0.056***	0.227±0.033***	0.219±0.040***	0.118±0.014*	**0.109±0.012**
	69.3×	PSNR	26.24±2.24***	33.62±2.10***	26.58±1.82***	33.92±3.13***	**37.75±2.53**
		SSIM	0.717±0.038***	0.883±0.028***	0.682±0.039***	0.910±0.034**	**0.940±0.024**
		nRMSE (ROI)	0.422±0.118***	0.238±0.034***	0.338±0.084***	0.196±0.040*	**0.165±0.022**

The best and second performances are highlighted in **bold** and underline. Statistical significant differences compared with our MoCo-INR are marked (*** p <0.001; ** p <0.01; * p <0.05; and $^{\blacktriangledown}$ p ≥0.05, not significant).

Table 1: Quantitative results (PSNR (dB)/SSIM/nRMSE) of the compared methods under a Cartesian sampling pattern (VISTA) and a non-Cartesian sampling pattern (GA Radial) under two acceleration factors, respectively.

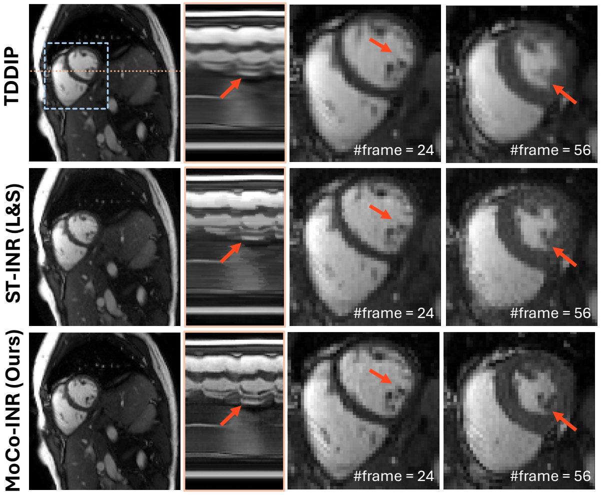

Figure 4: Qualitative results of prospective reconstruction under free-breathing scans.

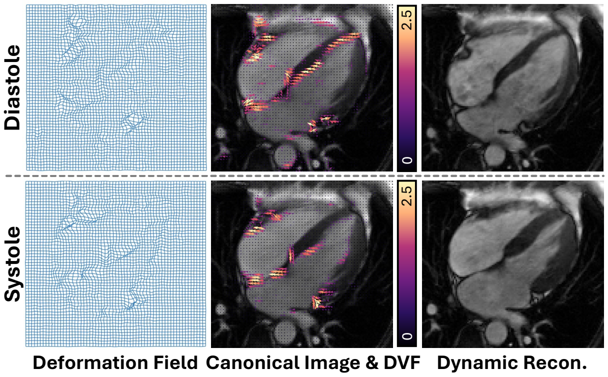

Figure 5: Visualization of the estimated DVFs and canonical image of MoCo-INR at the diastolic and systolic phases.

Study	Sampling	TDDIP	ST-INR (L&S)	MoCo-INR
Retro.	VISTA	3.2	<u>1.5</u>	**1.3**
	GA Radial	23.3	<u>10.9</u>	**5.5**
Prosp.	VISTA	19.3	<u>6.7</u>	**3.4**

Table 2: Runtime (in minutes) comparisons for the unsupervised DL-based methods.

2007) for Cartesian sampling; **GRASP** (Feng et al. 2014) for golden-angle radial sampling; (2) DIP-based: **Time-Depend DIP (TDDIP)** (Yoo et al. 2021); (3) INR-based: Feng et al. (2025) proposed an INR-based dynamic MRI reconstruction method that incorporates hash encoding with additional low-rank and sparsity (L&S) constraints; we refer to this approach as **ST-INR (L&S)**. To evaluate the effectiveness of these additional constraints, we also include a variant without the (L&S) constraints, denoted simply as **ST-INR**.

Evaluation Metrics. For the reconstructed CMR images, we employ peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) as quantitative evaluation metrics. To specifically evaluate the accurate reconstruction of cardiac anatomy and its temporal dynamics, we segment the cardiac region and compute the normalized root-mean-square error (nRMSE) within it. To quantify reconstruction efficiency, we present the runtime for DL-based methods to achieve the reported optimal performance.

5.4 Implementation Details

In MoCo-INR, the DVF network $\mathcal{F}_{\Phi}$ adopts hash encoding set of $N_{\min} = 2$, $T = 2^{21}$, $L = 10$, $F = 4$ and $b = 2$, while the canonical network $\mathcal{G}_{\Psi}$ is set as follows $N_{\min} = 2$, $T = 2^{21}$, $L = 12$, $F = 8$ and $b = 2$. Both networks employ lightweight CNN decoders composed of three convolutional layers. The first two convolutional layers are followed by nonlinear activation functions, with 64 filters of size of 3, and the final convolutional layer outputs without activation. Due to the space constraint, we introduce the other implemental details of MoCo-INR and baselines in the supplemental materials.

6 Results

6.1 Retrospective Reconstruction Results

Table 1 compares the performance of our MoCo-INR with baselines. Under ultra-high acceleration factors (20× for Cartesian and 69× for non-Cartesian), MoCo-INR attains the highest PSNR/SSIM values, with improvements that are statistically significant (p <0.001). The results highlight its robustness to severe undersampling and demonstrate its suitability for challenging reconstruction scenarios. Moreover, MoCo-INR consistently yields the lowest ROI nRMSE across every sampling pattern and acceleration setting, confirming its superior ability to preserve both the dynamic motion and the anatomical detail of the cardiac region.

Fig. 3 shows the qualitative results of reconstruction. The ℓ_1-Wavelet method fails to recover images at high acceleration factors, exhibiting severe blurring and aliasing artifacts. TDDIP insufficiently captures temporal dynamics. Particularly, under golden-angle radial sampling, the cardiac anatomy appears nearly identical between ED and ES phases. ST-INR introduces numerous artifacts due to its lack of explicit regularization, while ST-INR (L&S) mitigates these artifacts yet still suffers from noticeable noise and indistinct tissue boundaries. In contrast, the proposed MoCo-INR exploits shared spatial information across temporal frames within a motion-compensation framework and employs a CNN decoder that robustly processes hash-grid features, thereby enabling accurate cardiac motion tracking and high-fidelity anatomical reconstruction.

6.2 Prospective Reconstruction Results

We further evaluate the proposed MoCo-INR and the compared methods on prospectively undersampled data. The visual comparison is illustrated in Fig. 4, consistent with the observations from the retrospective study. TDDIP exhibits over-smoothing in both spatial and temporal dimensions, and the zoom-in views reveal anatomically implausi-

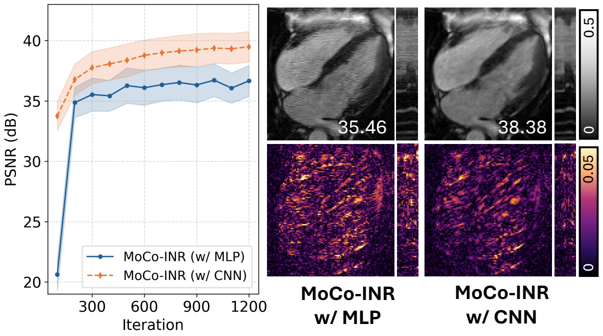

Figure 6: Performance curves and qualitative results for VISTA sampling with AF=20 of MoCo-INR with either an MLP decoder or a CNN decoder.

ble structures. Compared with ST-INR (L&S), the proposed MoCo-INR yields sharper tissue detail with significantly reduced artifacts, as highlighted by the red arrows. Notably, the intensity profile shows that, although ST-INR (L&S) suffers from spatial noise, it fails to capture temporal detail, whereas MoCo-INR successfully resolves both large-scale cardiac motion and subtle intramural deformations.

6.3 Evaluation of DVFs and Canonical Image

Fig. 5 shows the estimated DVFs alongside the canonical image. The quiver plots (second column) illustrate vector patterns consistent with myocardial relaxation and chamber enlargement during diastole, and with myocardial contraction and ventricular volume reduction during systole. The learned DVFs are consistent with the known biomechanics of the cardiac cycle, demonstrating our method accurately capture cardiac motion.

6.4 Evaluation of Runtime

Fast runtime is essential for clinical applicability. Table 2 reports the runtime on a single NVIDIA RTX 4090 GPU, showing that the proposed MoCo-INR achieves fast reconstruction for both retrospective and prospective studies. By explicitly decomposing deformation and canonical image content, MoCo-INR enables faster convergence with fewer optimization steps compared to ST-INR (L&S). This significantly reduces the computational cost, particularly for non-Cartesian sampling where the NUFFT operator is inherently slow.

6.5 Ablation Studies

Effectiveness of CNN-Based INR Network. Fig. 6 compares the performance of MoCo-INR using an MLP decoder versus a CNN decoder. The performance curves show that the CNN-based decoder consistently outperforms the MLP decoder and provides a more stable optimization process. In the reconstructed MR images, the MLP decoder introduces noticeable high-frequency artifacts, whereas the CNN decoder produces smoother and more accurate results, as further highlighted in the error maps.

Model	PSNR	SSIM
w/o $\mathcal{L}_{\text{DVF}}$	34.42±2.73***	0.895±0.031***
w/o Coarse2fine	35.51±2.84***	0.926±0.026***
Full	**37.75±2.53**	**0.940±0.024**

Table 3: Quantitative comparisons on retrospective study using MoCo-INR, evaluated without key components under golden-angle radial sampling with AF=69.3.

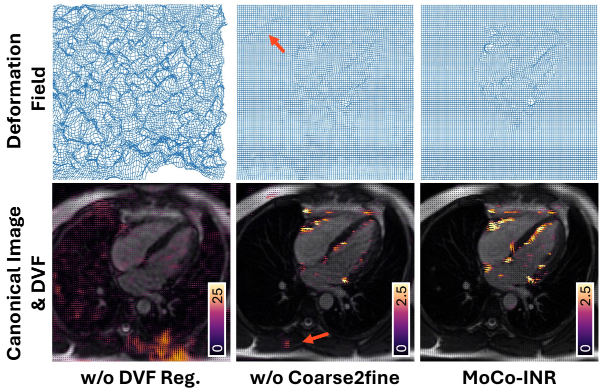

Figure 7: Qualitative comparison showing the influence of DVF regularization and the coarse-to-fine hash-encoding learning strategy of MoCo-INR on DVF estimation and canonical image reconstruction.

Influence of DVF Regularization and Coarse2fine Hash Encoding Learning. Table 3 demonstrates the effectiveness of DVF regularization and the coarse-to-fine learning strategy, showing a significant degradation in reconstruction performance when these components are removed. Fig. 7 further illustrates their influence. Without DVF regularization, the estimated DVF is largely incorrect. When the coarse-to-fine learning strategy is not applied, the DVF estimation is relatively reasonable but still exhibits abnormal motion in static regions (highlighted by orange arrows). In contrast, MoCo-INR with the proposed coarse-to-fine learning accurately captures plausible cardiac motion.

7 Conclusion & Discussion

This work introduces MoCo-INR, a novel unsupervised motion-compensated framework for cardiac MR reconstruction. Experimental results show that MoCo-INR achieves superior performance under ultra-high acceleration factors acquisitions and is capable of accurately reconstructing real-time free-breathing scans. Benefiting from the flexibility of unsupervised nature and fast convergence, MoCo-INR is well-suited for a variety of acquisition conditions encountered in clinical practice. Despite these promising results, several challenges remain. Future work will focus on extending MoCo-INR to high-resolution 3D spatial–temporal reconstructions and addressing limitations of motion compensation when intensity changes occur, such as in dynamic contrast-enhanced (DCE) MRI.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 62571328.

References

Batchelor, P.; Atkinson, D.; Irarrazaval, P.; Hill, D.; Hajnal, J.; and Larkman, D. 2005. Matrix description of general motion correction applied to multishot images. *Magnetic Resonance in Medicine*, 54(5): 1273–1280.

Catalán, T.; Courdurier, M.; Osses, A.; Fotaki, A.; Botnar, R.; Sahli-Costabal, F.; and Prieto, C. 2025. Unsupervised reconstruction of accelerated cardiac cine MRI using neural fields. *Computers in Biology and Medicine*, 185: 109467.

Chen, C.; Liu, Y.; Schniter, P.; Tong, M.; Zareba, K.; Simonetti, O.; Potter, L.; and Ahmad, R. 2020. OCMR (v1. 0)–open-access multi-coil k-space dataset for cardiovascular magnetic resonance imaging.

Chen, L.; Balter, J. M.; Shen, L.; and Park, J. J. 2025. Single-Spoke Motion-Compensated Dynamic 3D MRI Reconstruction via Neural Representation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 513–522. Springer.

Du, C.; Lin, X.; Wu, Q.; Tian, X.; Su, Y.; Luo, Z.; Zheng, R.; Chen, Y.; Wei, H.; Zhou, S. K.; et al. 2024. DPER: Diffusion prior driven neural representation for limited angle and sparse view CT reconstruction. *arXiv preprint arXiv:2404.17890*.

Feng, J.; Feng, R.; Wu, Q.; Shen, X.; Chen, L.; Li, X.; Feng, L.; Chen, J.; Zhang, Z.; Liu, C.; Zhang, Y.; and Wei, H. 2025. Spatiotemporal Implicit Neural Representation for Unsupervised Dynamic MRI Reconstruction. *IEEE Transactions on Medical Imaging*, 44(5): 2143–2156.

Feng, L.; Grimm, R.; Block, K. T.; Chandarana, H.; Kim, S.; Xu, J.; Axel, L.; Sodickson, D. K.; and Otazo, R. 2014. Golden-angle radial sparse parallel MRI: Combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI. *Magnetic Resonance in Medicine*, 72(3): 707–717.

Fessler, J. A.; and Sutton, B. P. 2003. Nonuniform fast Fourier transforms using min-max interpolation. *IEEE transactions on signal processing*, 51(2): 560–574.

Hammernik, K.; Pan, J.; Rueckert, D.; and Küstner, T. 2021. Motion-Guided Physics-Based Learning for Cardiac MRI Reconstruction. In *2021 55th Asilomar Conference on Signals, Systems, and Computers*, 900–907.

Huang, W.; Li, H. B.; Pan, J.; Cruz, G.; Rueckert, D.; and Hammernik, K. 2023. Neural implicit k-space for binning-free non-cartesian cardiac MR imaging. In *International Conference on Information Processing in Medical Imaging*, 548–560. Springer.

Iskender, B.; Nakarmi, S.; Daphalapurkar, N.; Klasky, M. L.; and Bresler, Y. 2025. Rsr-nf: Neural field regularization by static restoration priors for dynamic imaging. *arXiv preprint arXiv:2503.10015*.

Kazerouni, A.; Azad, R.; Hosseini, A.; Merhof, D.; and Bagci, U. 2024. Incode: Implicit neural conditioning with prior knowledge embeddings. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1298–1307.

Kettelkamp, J.; Romanin, L.; Piccini, D.; Priya, S.; and Jacob, M. 2023. Motion Compensated Unsupervised Deep Learning for 5D MRI. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 419–427. Springer.

Kunz, J. F.; Ruschke, S.; and Heckel, R. 2024. Implicit neural networks with fourier-feature inputs for free-breathing cardiac MRI reconstruction. *IEEE Transactions on Computational Imaging*, 10: 1280–1289.

Lingala, S. G.; Hu, Y.; DiBella, E.; and Jacob, M. 2011. Accelerated dynamic MRI exploiting sparsity and low-rank structure: kt SLR. *IEEE Transactions on Medical Imaging*, 30(5): 1042–1054.

Lustig, M.; Donoho, D.; and Pauly, J. M. 2007. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6): 1182–1195.

Mihajlovic, M.; Prokudin, S.; Pollefeys, M.; and Tang, S. 2024. ResFields: Residual Neural Fields for Spatiotemporal Signals. In *The Twelfth International Conference on Learning Representations*.

Morales, M. A.; Izquierdo-Garcia, D.; Aganj, I.; Kalpathy-Cramer, J.; Rosen, B. R.; and Catana, C. 2019. Implementation and Validation of a Three-dimensional Cardiac Motion Estimation Network. *Radiology: Artificial Intelligence*, 1(4): e180080. PMID: 32076659.

Müller, T.; Evans, A.; Schied, C.; and Keller, A. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, 41(4): 1–15.

Munoz, C.; Qi, H.; Cruz, G.; Küstner, T.; Botnar, R. M.; and Prieto, C. 2022. Self-supervised learning-based diffeomorphic non-rigid motion estimation for fast motion-compensated coronary MR angiography. *Magnetic Resonance Imaging*, 85: 10–18.

Nyquist, H. 1928. Certain Topics in Telegraph Transmission Theory. *Transactions of the American Institute of Electrical Engineers*, 47(2): 617–644.

Oscanoa, J. A.; Middione, M. J.; Alkan, C.; Yurt, M.; Loecher, M.; Vasanawala, S. S.; and Ennis, D. B. 2023. Deep learning-based reconstruction for cardiac MRI: a review. *Bioengineering*, 10(3): 334.

Otazo, R.; Candès, E.; and Sodickson, D. K. 2015. Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components. *Magnetic Resonance in Medicine*, 73(3): 1125–1136.

Pan, J.; Huang, W.; Rueckert, D.; Küstner, T.; and Hammernik, K. 2024. Motion-Compensated MR CINE Reconstruction With Reconstruction-Driven Motion Estimation. *IEEE Transactions on Medical Imaging*, 43(7): 2420–2433.

Pan, J.; Rueckert, D.; Küstner, T.; and Hammernik, K. 2022. Learning-based and unrolled motion-compensated reconstruction for cardiac MR CINE imaging. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 686–696. Springer.

Park, K.; Sinha, U.; Barron, J. T.; Bouaziz, S.; Goldman, D. B.; Seitz, S. M.; and Martin-Brualla, R. 2021. Nerfies: Deformable neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5865–5874.

Pennell, D. J. 2010. Cardiovascular Magnetic Resonance. *Circulation*, 121(5): 692–705.

Pumarola, A.; Corona, E.; Pons-Moll, G.; and Moreno-Noguer, F. 2021. D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10318–10327.

Qi, H.; Hajhosseiny, R.; Cruz, G.; Kuestner, T.; Kunze, K.; Neji, R.; Botnar, R.; and Prieto, C. 2021. End-to-end deep learning nonrigid motion-corrected reconstruction for highly accelerated free-breathing coronary MRA. *Magnetic Resonance in Medicine*, 86(4): 1983–1996.

Rahaman, N.; Baratin, A.; Arpit, D.; Draxler, F.; Lin, M.; Hamprecht, F.; Bengio, Y.; and Courville, A. 2019. On the spectral bias of neural networks. In *International Conference on Machine Learning*, 5301–5310. PMLR.

Reed, A. W.; Kim, H.; Anirudh, R.; Mohan, K. A.; Champley, K.; Kang, J.; and Jayasuriya, S. 2021. Dynamic ct reconstruction from limited views with implicit neural representations and parametric motion fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2258–2268.

Shao, H.-C.; Mengke, T.; Deng, J.; and Zhang, Y. 2024. 3D cine-magnetic resonance imaging using spatial and temporal implicit neural representation learning (STINR-MR). *Physics in Medicine & Biology*, 69(9): 095007.

Shao, H.-C.; Qian, X.; Xu, G.; Wu, C.; Otazo, R.; Deng, J.; and Zhang, Y. 2025. A dynamic reconstruction and motion estimation framework for cardiorespiratory motion-resolved real-time volumetric MR imaging (DREME-MR).

Tian, X.; Chen, L.; Wu, Q.; Du, C.; Shi, J.; Wei, H.; and Zhang, Y. 2025. Unsupervised Self-Prior Embedding Neural Representation for Iterative Sparse-View CT Reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 7383–7391.

Wu, Q.; Du, C.; Tian, X.; Yu, J.; Zhang, Y.; and Wei, H. 2025. Moner: Motion Correction in Undersampled Radial MRI with Unsupervised Neural Representation. In *The Thirteenth International Conference on Learning Representations*.

Xu, Z.-Q. J.; Zhang, Y.; Luo, T.; Xiao, Y.; and Ma, Z. 2019. Frequency principle: Fourier analysis sheds light on deep neural networks.

Yoo, J.; Jin, K. H.; Gupta, H.; Yerly, J.; Stuber, M.; and Unser, M. 2021. Time-Dependent Deep Image Prior for Dynamic MRI. *IEEE Transactions on Medical Imaging*, 40(12): 3337–3348.

Zhang, Y.; Shao, H.-C.; Pan, T.; and Mengke, T. 2023. Dynamic cone-beam CT reconstruction using spatial and temporal implicit neural representation learning (STINR). *Physics in Medicine & Biology*, 68(4): 045005.

Zhao, B.; Haldar, J. P.; Christodoulou, A. G.; and Liang, Z.-P. 2011. Further development of image reconstruction from highly undersampled (k, t)-space data with joint partial separability and sparsity constraints. In *IEEE International Symposium on Biomedical Imaging*, 1593–1596.

Zou, Q.; Torres, L.; Fain, S.; and Jacob, M. 2022. Dynamic Imaging Using Motion-Compensated Smoothness Regularization on Manifolds (MOCO-STORM). In *IEEE International Symposium on Biomedical Imaging*, 1–4.