Zero-shot Implicit Neural Manifold Representation (INMR) for Ultra-high Temporal Resolution Dynamic MRI

Jie Feng[1], Rui Luo[2], Tian Zeng[1], Xin Shen[1], Haikun Qi[2], Yuyao Zhang[3], Dong Liang[4*], Hongjiang Wei[1*]

[1] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China
[2] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai, China
[3] School of Information Science and Technology, ShanghaiTech University, Shanghai, China
[4] Research Center for Medical AI, Shenzhen Institute of Advanced Technology, Chinese Academic of Sciences, Shenzhen, China
{jiefeng, 2334519828, xin-shen, hongjiang.wei}@sjtu.edu.cn,
{luorui2023, qihk, zhangyy8}@shanghaitech.edu.cn, dong.liang@siat.ac.cn

Abstract

Capturing accurate dynamic information of moving organs is essential for functional assessment using non-invasive imaging modalities. Achieving high temporal resolution visualization of physiological processes remains a critical challenge in dynamic magnetic resonance imaging (MRI) when reconstructing from extremely limited acquisitions. We introduce an unsupervised zero-shot reconstruction framework combining Implicit Neural Representation (INR) with manifold learning, capable of reconstructing dynamic MRI data at unprecedented temporal resolutions (less than 10 ms per frame for 2D imaging, less than 400 ms per frame for 3D imaging). The framework employs learnable low-dimensional manifold vectors to autonomously capture motion in real time directly from undersampled data, and dynamically condition coordinate-based spatial representations to generate high-fidelity image sequences. Through a novel spatiotemporal coarse-to-fine (C2F) optimization strategy, our method outperforms current state-of-the-art (SOTA) techniques across multiple imaging scenarios, including cardiac, speech and dynamic-contrast-enhanced (DCE) abdominal MRI, demonstrating robust performance under challenging motion patterns and contrast dynamics. The learned manifolds additionally provide intuitive visualization of motion and contrast evolution during imaging. These advances indicate strong clinical potential for applications requiring extreme temporal resolution while maintaining both anatomical and temporal fidelity.

Code —
https://github.com/AMRI-Lab/INMR_for_DynamicMRI

Introduction

High-temporal-resolution visualization of dynamic physiological processes via MRI remains a fundamental pursuit in both research and clinical practice. Achieving diagnostically meaningful spatial resolution while improving temporal resolution is particularly crucial, as it enables reduced motion artifacts and more precise characterization of rapid physiological dynamics (Nayak et al. 2022). However, the physical limitation of long proton relaxation time has restricted the temporal resolution in basic fully-sampled dynamic MRI to hundreds of milliseconds in 2D imaging or several tens of seconds in 3D imaging (Uecker et al. 2010), which is not enough for capturing motion like heartbeat or respiratory.

To overcome temporal resolution limitations in dynamic MRI, various reconstruction strategies for highly-undersampled $\boldsymbol{k}$-space data have been developed. Conventional compressed sensing (CS) methods, incorporating both data consistency and temporal regularization, can achieve temporal resolutions of 30-55 ms for 2D in-vivo imaging (Uecker et al. 2010) and several seconds for 3D imaging (Feng et al. 2014). Subsequent advances through low-rank matrix completion and subspace modeling have further enhanced temporal resolution capabilities. With spatiotemporal low-rank regularization (Lingala et al. 2011), explicit spatiotemporal low-rank structure (Zhao et al. 2012; Feng et al. 2020) or low-dimensional manifold embedding of temporal dynamics (Poddar and Jacob 2016; Nakarmi et al. 2017; Shetty et al. 2020; Ahmed et al. 2020), the complexity of the CS-based optimization task is highly reduced, inducing improved temporal fidelity performances at the temporal resolution of 40 ms for 2D imaging (Ahmed et al. 2020) or sub-second for 3D imaging (Chen et al. 2024). However, they typically require additional calibration scans or sophisticated extraction of inherent data redundancies to estimate the temporal priors for reconstruction calibration, which may also introduce additional errors that degrade the reconstruction quality.

Recently, supervised deep learning (DL) has achieved remarkable success in dynamic MRI reconstruction (Wang et al. 2016; Han et al. 2018; Qin et al. 2019; Qi et al. 2021; Huang et al. 2021; Wang et al. 2023). However, its reliance on fully sampled training data significantly limits its ap-

*Corresponding authors.

plicability to ultrahigh-temporal-resolution imaging, where acquiring reliable fully sampled datasets is fundamentally infeasible. Alternatively, unsupervised zero-shot Deep Image Prior (DIP) approaches have been explored for dynamic cardiac imaging (Yoo et al. 2021; Zou et al. 2021, 2022). These approaches leverage untrained convolutional neural networks (CNN) as implicit regularizers. Nevertheless, their performances for ultrahigh resolution remains constrained due to the discrete nature of CNN (Wu et al. 2023).

Implicit Neural Representation (INR) has emerged as an innovative and powerful paradigm for signal representation, where multi-layer perceptions (MLPs) learn continuous functions from coordinates to the corresponding signal intensities (Mildenhall et al. 2020; Müller et al. 2022; Feng et al. 2024). The inherent continuity of INR provides strong implicit regularization, enabling recent breakthroughs in unsupervised dynamic MRI reconstruction for solving ill-posed inverse problems (Huang et al. 2023; Spieker et al. 2023; Al-Haj Hemidi et al. 2024; Kunz, Ruschke, and Heckel 2024; Huang et al. 2024; Feng et al. 2025; Catalán et al. 2025). Notably, studies such as (Kunz, Ruschke, and Heckel 2024; Huang et al. 2024) have demonstrated the potential of INR for achieving approximately 20 ms in 2D free-breathing cardiac MRI. However, current spatiotemporal coordinate-based INR approaches exclusively rely on network optimization to capture inter-frame dependencies, requiring a single network to fully capture temporal dynamics. This design results in significant computational overhead and clinically impractical reconstruction times.

In this work, we propose a novel unsupervised zero-shot method, Implicit Neural Manifold Representation (INMR), for ultra-high temporal resolution dynamic MRI. To compactly encode temporal information while preserving implicit continuity across dynamic states, INMR employs an MLP-based implicit function that takes spatial coordinates and learnable low-dimensional manifold vectors as inputs. These manifold vectors and the INMR function are jointly optimized using a spatiotemporal coarse-to-fine (C2F) strategy, which avoids suboptimal convergence in highly ill-posed scenarios and eliminating the need for prior temporal calibration. We extensively evaluated INMR across diverse imaging scenarios, including periodic motion (free-breathing cardiac MRI), aperiodic motion (speech articulation MRI), and dynamic contrast variations (DCE abdominal MRI). Results demonstrate that INMR outperforms unsupervised state-of-the-art (SOTA) methods, delivering artifact-free images with high temporal fidelity at temporal resolutions below 10 ms for 2D and 400 ms for 3D imaging. Visualization of the optimized manifold vectors, through principle component analysis or covariance matrices, confirms their ability to effectively capture motion and contrast dynamics. Ablation studies further validate the effectiveness and robustness of INMR.

The main contributions of this work are summarized as follows:

- We propose INMR, a novel zero-shot unsupervised neural representation method for ultra-high temporal resolution dynamic MRI, achieving unprecedented temporal resolutions (<10 ms per 2D frame, <400 ms per 3D volume). INMR demonstrates superior convergence speed and reconstruction fidelity compared to SOTA unsupervised methods across diverse imaging scenarios.
- We introduce a novel framework that employs learnable low-dimensional manifold vectors to condition coordinate-based neural representation, enabling effective capturing motion and contrast dynamics without requiring prior calibration data.
- We develop a spatiotemporal C2F optimization strategy that efficiently guides INMR training, ensuring robust convergence while avoiding suboptimal solutions.

Background

Inverse problem in dynamic MRI reconstruction

The relationship between the acquired $\boldsymbol{k}$-data and the dynamic MRI sequences can be formally described by the following model:

$$\mathbf{D} = F_u\mathbf{d} + \mathbf{n}, \tag{1}$$

where $\mathbf{d} \in \mathbb{C}^{N\times T}$ is the discretized spatiotemporal image matrix for dynamic MRI sequence, $\mathbf{D}$ is the measured ($\boldsymbol{k}$, t)-space data. F_u is the forward MRI undersampling acquisition operator with Fourier Transform, to simulate the physical acquisition of dynamic MRI. Where $\mathbf{n}$ represents zero-mean acquisition noise.

Given the fact that recovering dynamic image sequence from the highly-undersampled ($\boldsymbol{k}$, t)-space data is an ill-posed problem, we usually try to find the optimal $\mathbf{d}$ with an iterative optimization problem:

$$\arg\min_{\mathbf{d}} \frac{1}{2}\|F_u\mathbf{d} - \mathbf{D}\|_2^2 + \mathcal{R}(\mathbf{d}). \tag{2}$$

Here, $\mathcal{R}(\mathbf{d})$ represents the regularization term of the optimization framework for better convergence, such as temporal Total Variation (TV), etc.

Manifold learning for dynamic MRI reconstruction

Recently, the manifold learning strategy has been utilized for dynamic MRI reconstruction (Zou et al. 2021; Ke et al. 2021; Zou et al. 2022), where a smooth manifold of low-dimensional vectors $\mathbf{z} = [\mathbf{z}_1, \mathbf{z}_2, ...\mathbf{z}_T] \in \mathbb{R}^{L\times T}(L \ll N)$ is introduced to represent the temporal frames in the dynamic MRI series. These approaches also employ a CNN-based nonlinear mapping function $\mathcal{G}_\theta$, parameterized by θ, to project manifold vectors into image space:

$$d_t = \mathcal{G}_\theta(\mathbf{z}_t). \tag{3}$$

In unsupervised manifold learning pipelines (Zou et al. 2021, 2022), both the manifold representation $\mathbf{z}$ and the mapping function $\mathcal{G}_\theta$ are jointly optimized for each data respectively with the optimization framework of Eq. 2.

Spatiotemporal INR(stINR)-based dynamic MRI reconstruction

Based on the definition of INR, the intensities of the dynamic MR images d are represented with the continuous function f_θ of the spatial coordinates $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_N]$ and the temporal coordinates $\mathbf{t} = [1, 2, ...T]$ in the previous

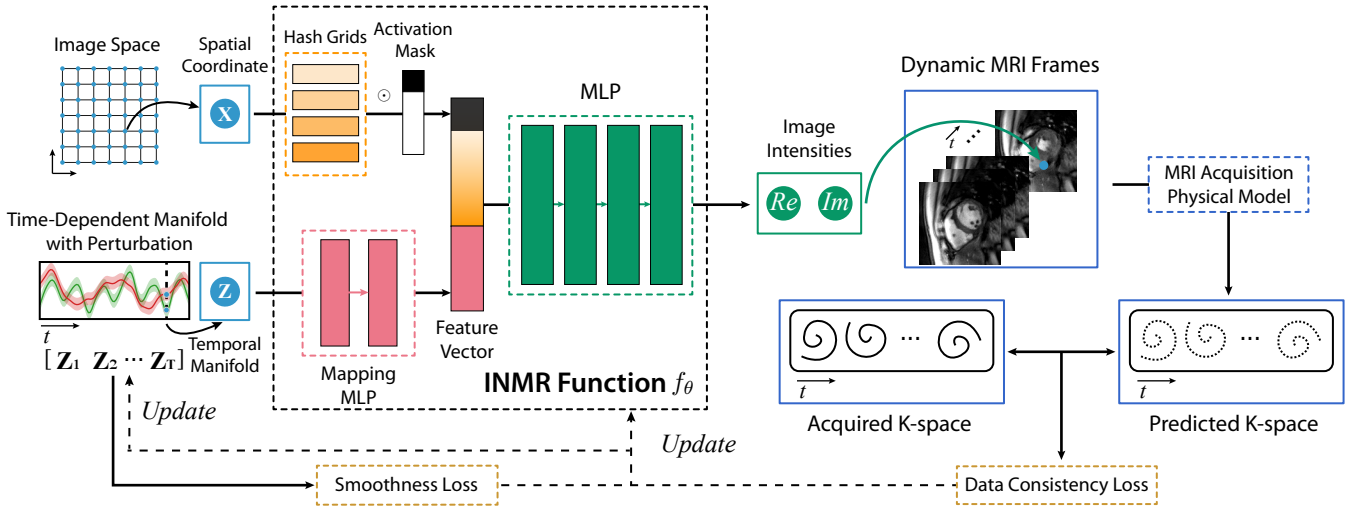

Figure 1: The detailed architecture of INMR. Both the INMR function and the temporal manifold were jointly updated with Data Consistency (DC) loss and Smoothness loss for zero-shot unsupervised reconstruction. A C2F optimization strategy was enabled with both progressive activation masks for spatial features and progressive perturbation for temporal manifold.

stINR works (Feng et al. 2025; Kunz, Ruschke, and Heckel 2024; Catalán et al. 2025):

$$\mathbf{d}_\theta = \begin{bmatrix} f_\theta(\mathbf{x}_1, 1) & \cdots & f_\theta(\mathbf{x}_1, T) \\ \vdots & \ddots & \vdots \\ f_\theta(\mathbf{x}_N, 1) & \cdots & f_\theta(\mathbf{x}_N, T) \end{bmatrix}. \tag{4}$$

Here, f_θ is the MLP-based representation function parameterized by θ. Then Eq. 2 can be transformed into a fitting problem for optimal parameters θ using gradient-descent-based optimizers from the undersampled $\boldsymbol{k}$-data alone.

Method

Implicit Neural Manifold Representation (INMR)

Building upon principles from manifold learning and INR, we introduce a novel Implicit Neural Manifold Representation (INMR) method for dynamic MRI reconstruction. Our approach models dynamic MRI sequences using continuous functions that simultaneously encode fixed spatial coordinates $\mathbf{x}$ and learnable time-dependent manifold vectors $\mathbf{z}$. Thus, the image sequence $\mathbf{d}$ is rewritten as a modified Casorati matrix representation:

$$\mathbf{d}_{\theta,\mathbf{z}} = \begin{bmatrix} f_\theta(\mathbf{x}_1, \mathbf{z}_1) & \cdots & f_\theta(\mathbf{x}_1, \mathbf{z}_T) \\ \vdots & \ddots & \vdots \\ f_\theta(\mathbf{x}_N, \mathbf{z}_1) & \cdots & f_\theta(\mathbf{x}_N, \mathbf{z}_T) \end{bmatrix}. \tag{5}$$

In contrast to the prior stINR-based works employing fixed temporal indices as input, the input of learnable low-dimensional manifold in INMR could encode more compact temporal features, such as complex motion patterns and contrast dynamics. The implicit continuity of the representation function could then generalize across these temporal states rather than only adjacent frames, enabling robust reconstruction at ultra-high temporal resolution without any prior calibration.

Thus the optimization procedure in Eq. 2 can be written as:

$$\arg\min_{\theta,\mathbf{z}} \frac{1}{2}\|F_u \mathbf{d}_{\theta,\mathbf{z}} - \mathbf{D}\|_2^2 + \mathcal{R}(\mathbf{z}), \tag{6}$$

for the optimal solution of both the parameters θ of the function and the manifold vectors $\mathbf{z}$ in INMR.

Detailed framework and implementation of INMR

Fig. 1 illustrates the overall framework of INMR. In INMR, we separate the inputs - the spatial coordinates and temporal manifolds - into two paths, and then combine the features from both paths with an MLP to output the real and imaginary parts of the image intensities. For the spatial path, to enhance high-frequency spatial representation, we employ hash encoding (Müller et al. 2022). This learnable encoding scheme utilizes multi-resolution hash grids that progressively store spatial features, achieving an optimal trade-off between computational efficiency and representation continuity. As for the temporal path, a tiny mapping MLP is adopted to map the manifolds into a comparable dimension of the output of the spatial path, due to the smoothness of the temporal information.

Spatiotemporal C2F optimization strategy To improve both the continuity of the mapping function and the accuracy of the learnt manifolds, a spatiotemporal C2F optimization strategy is introduced in our framework.

As mentioned in the previous works (Li et al. 2023; Wang et al. 2024), optimizing grids at all resolutions of hash encoding simultaneously frequently results in non-smooth local optima. Such suboptimal spatial encoding not only disrupts the continuity of mapping functions but also adversely affects the temporal manifold optimization process. Thus, we implement a spatial C2F optimization strategy (Feng et al. 2025; Wu et al. 2025) wherein finer grid levels are sequentially activated throughout the training procedure.

Similar convergence failures in temporal manifold optimization may also lead to discontinuous solutions. Drawing inspiration from the reparameterization trick in Variational Autoencoders (VAE) (Kingma and Welling 2019), we enforce manifold continuity during optimization through a parameterized Gaussian distribution:

$$\mathbf{z}_t = \boldsymbol{\mu}_t + \boldsymbol{\Sigma}_t \boldsymbol{\epsilon}, \tag{7}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ represents isotropic Gaussian perturbation. This strategy enables Monte-Carlo optimization of the manifold vector expectations $\boldsymbol{\mu}_t$ at each temporal point t. The covariance matrix $\boldsymbol{\Sigma} = \sigma\mathbf{I}$ is progressively annealed according to our C2F paradigm, with the deterministic manifold expectations $\boldsymbol{\mu}_t$ being exclusively utilized upon convergence. Finally, after the completion of the optimization process, we utilize exclusively the deterministic manifold expectations $\mathbf{z}_t = \boldsymbol{\mu}_t$ as the input for INMR.

Loss functions Following Eq. 6, two loss functions are included in this framework for unsupervisedly learning the parameters of the representation function and the manifold vectors:

$$\mathcal{L}_{Total} = \mathcal{L}_{DC} + \lambda \mathcal{L}_{Smooth}. \tag{8}$$

Here, λ serves as a regularization hyperparameter to tune the effect of smooth loss. The further details about these two losses can be found in the supplementary material.

Experiments

Datasets

We applied the proposed method on three different high-temporal-resolution MRI application scenarios including free-breathing cardiac MRI, speech MRI and DCE abdominal MRI, covering a general range from multiple periodic motion, non-periodic motion and mix of periodic motion and contrast dynamics. The general information of the datasets is listed below.

Free-breathing cardiac MRI data A numerical phantom was simulated with quasi-periodic respiratory and irregular heartbeat to emulate the clinical settings. A 2D golden-angle radial sampling scheme (Feng 2022) was implemented across 160 temporal frames, with each radial spoke representing an acquisition time of 3 ms and 2 spokes per frame.

This study additionally incorporated in-vivo free-breathing cardiac MRI data acquired from three healthy volunteers under institutional review board approval. For each temporal frame, data from a single spiral arm (acquisition window of 8 ms) (Luo et al. 2025) were utilized, with a total of 600 frames reconstructed per data. Due to the page limit, the results were included in the supplementary material.

Speech MRI data An open-source speech MRI dataset from (Lim et al. 2021) was included in this study. It provided 2D spiral GRE $\boldsymbol{k}$-data of the vocal tract with different stimulus tasks. From this dataset, we reconstructed imaging data from 10 participants performing 4 distinct speech tasks each. The reconstruction utilized single-spiral-arm data per temporal frame (6 ms/frame), yielding a total of 3000 frames for evaluation.

Abdominal DCE MRI data Two abdominal DCE MRI data of the healthy volunteers were downloaded from the website of CA2R for this study (Feng et al. 2018). Each 3D volumetric dataset comprised 48 slices reconstructed from 800 stack-of-stars (SOS) radial spokes (Feng et al. 2018), with temporal frames generated by binning 2 spokes per frame (temporal resolution of around 345.6 ms).

Compared methods

In this work, we compared our framework with 5 methods: NUFFT, GRASP (Feng et al. 2014), GRASP-Pro (Feng et al. 2020), Time-Depend DIP (TDDIP) (Yoo et al. 2021) and Fourier-feature MLP (FMLP) (Kunz, Ruschke, and Heckel 2024). NUFFT directly zero-filled the acquired $\boldsymbol{k}$-space, demonstrating the reconstruction challenges at ultrahigh temporal resolution. GRASP is a CS-based reconstruction method employing temporal TV as the regularizer. GRASP-Pro represents an advanced CS-based method that incorporates a subspace model specifically designed for highly accelerated dynamic MRI. TDDIP is a SOTA, scan-specific unsupervised reconstruction method based on DIP. FMLP is a SOTA INR-based method designed for high-temporal-resolution free-breathing cardiac MRI, and employs fixed temporal indices as network inputs.

Evaluation metrics

Both qualitative visual evaluation and quantitative evaluation were included. For the simulated free-breathing cardiac data, frame-wise average quantitative metrics including Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) were calculated. For the rest in-vivo data, due to the lack of reference images, temporal profiles were visualized to analyze the temporal blurring. Specifically, due to the significance of the enhancement quantification in DCE MRI, temporal region of interest (ROI) intensity analyses were conducted with the reference of NUFFT results. Although strongly contaminated by the streaking artifacts, the ROI-averaged intensities of the NUFFT results were still suitable to assess temporal fidelity (Chen et al. 2024).

Also, in order to evaluate the effectiveness and interpretability of the optimized manifold vectors, we visualized the first several components of the manifold after principal component analysis (PCA). They were compared with the temporal profiles to prove the correspondence to the temporal motion/contrast information. As for the liver dataset, the projection of the self-navigator (the 1D inverse Fourier projection from the central $\boldsymbol{k}$-space points of the SOS acquisition) was used to extract respiratory motion (Feng et al. 2018) for comparison.

Results

Results on free-breathing cardiac data

In Fig. 2, we visualized the comparison results on the simulated free-breathing cardiac data with the acceleration factor (AF) of 128. According to the figure, our proposed method appears to have the best spatial details and the most accurate temporal fidelity among the compared methods at

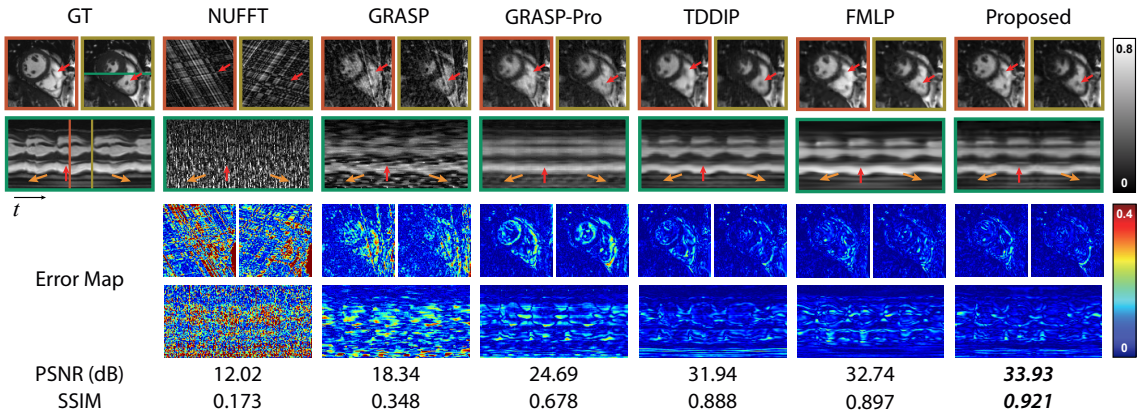

Figure 2: The comparison results on the simulated free-breathing cardiac data with a simulated resolution of 6 ms/frame. The red arrows highlight the physiological structures where the proposed method demonstrates the best performance, while the orange arrows emphasize the area where the superior performance of the proposed method capturing the respiratory motion.

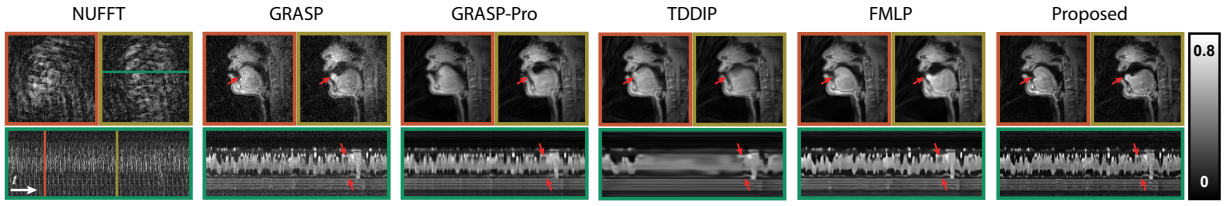

Figure 3: The comparison results on the speech data. The reconstruction results of 1 spiral arm per frame (6 ms per frame) from the free speech task are visualized. The red arrows highlight the physiological structures where the proposed method demonstrates the least motion blurring.

such high AF. In comparison, NUFFT results do not contain enough cardiac structures, while GRASP results strongly suffer from streaking artifacts. GRASP-Pro can reduce the effects of artifacts spatially, but the temporal motion is not correctly preserved. Both TDDIP and FMLP show comparable results towards the proposed method, but they do not provide sharp spatiotemporal physiological structures as indicated by the red arrows. Furthermore, TDDIP cannot match the proposed method in capturing the respiratory motion as pointed out by the orange arrows. The quantitative metrics support our observation, where our proposed method leads the comparison with the highest PSNR of 33.93 dB and the highest SSIM of 0.921.

Results on speech data

We visualized the speech MRI reconstruction results from the acquisition of 6 ms per frame in Fig. 3 for free talk task and in supplementary material for repeated stimulus task. In both tasks, NUFFT fails to provide any meaningful structures due to severe artifacts, the results from GRASP are highly affected by noise and TDDIP totally fails on reconstructing reasonable motion. As for GRASP-Pro, FMLP and the proposed method, all of them produce spatially clean results, but as pointed out by the red arrows, neither GRASP-Pro nor FMLP can prevent motion blurring as the proposed method.

Results on free-breathing 3D DCE abdominal data

For the DCE liver dataset, the visual comparisons are demonstrated in Fig. 4a and 4b. According to these figures, the anatomical details of the vessels are clearly visible in the reconstructed images of the proposed method. However, severe streaking artifacts and noises can be observed in the reconstructed images from NUFFT and GRASP. The results from GRASP-Pro demonstrate worse artifact suppression and motion blurring level than the proposed method, and it cannot capture respiratory motion as accurately as the proposed method in the temporal profiles. As for TDDIP and FMLP, extreme oversmoothness and inaccurate respiratory motion are witnessed. The ROI signal intensity curves in Fig. 4c suggest a better temporal fidelity of the proposed method against GRASP-Pro, which is the only available golden standard at such extreme AF.

Fig. 5 presents the PC curves of the manifold, comparing them with respiratory motion and contrast dynamics. The first PC of the optimized manifold shows strong agreement with respiratory motion derived from self-navigator projections, even during irregular breathing patterns. Additionally, the second and third PCs align with the wash-in/wash-out transitions of both the aorta and portal vein, particularly at key turning points. These findings highlight the optimized manifold's ability to extract clinically relevant features from complex temporal dynamics.

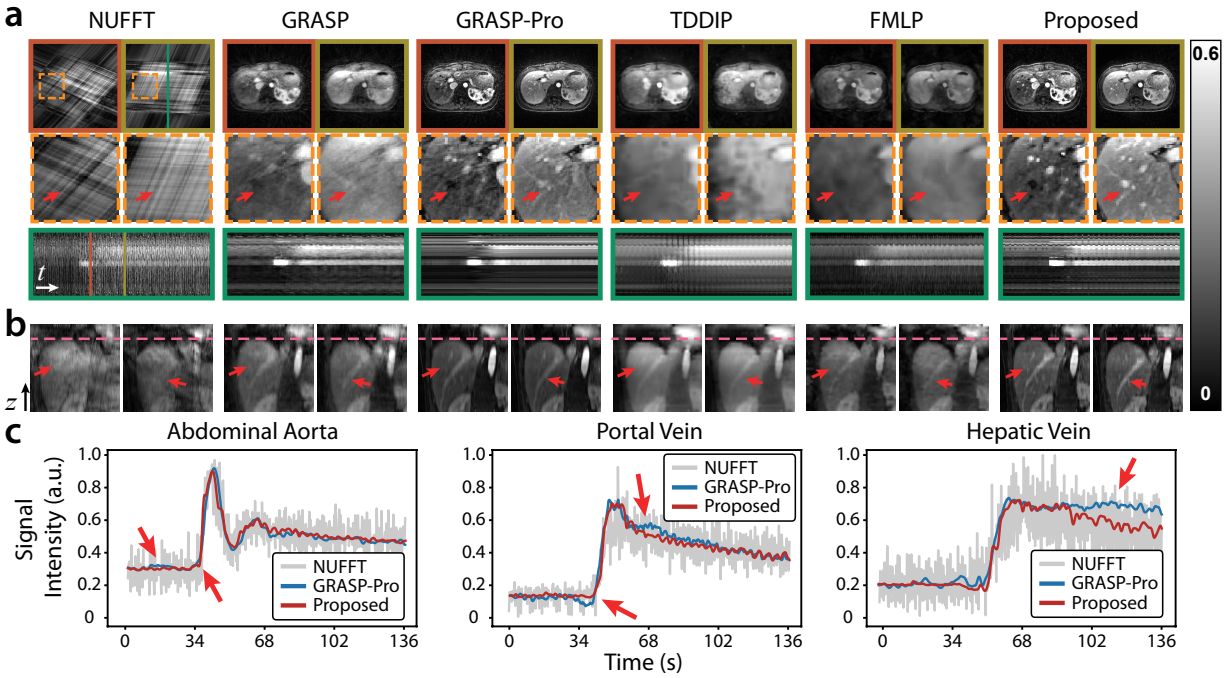

Figure 4: **a**. The comparison results with acquisitions of 345.6 ms per frame on the 3D DCE abdominal data. The red arrows highlight the vessels where the proposed method demonstrates the least motion blurring and artifacts. **b**. The representative stacked z-direction profiles of the compared methods. The pink dashed lines help highlight the respiratory motion between two frames. **c**. The ROI intensity curves of NUFFT, GRASP-Pro and the proposed results. The red arrows highlight where INMR shows better temporal fidelity than GRASP-Pro.

Reconstruction time

Table 1 demonstrates the mean time consumptions of each compared method on all the datasets. For all the datasets, INMR takes significantly shorter time than the compared unsupervised DL-based learning methods TDDIP and FMLP.

Ablation studies

Effectiveness of manifold learning In contrast to stINR frameworks, the key innovation of INMR is to model the temporal dynamics with learnable manifold vectors rather than fixed temporal indices. To evaluate the effectiveness of the manifold strategy, we conduct an experiment on the simulated cardiac dataset. For stINR implementation, we adopted the architecture from (Feng et al. 2025) with the identical hash encoding configurations. All the evaluations were limited to 40 frames from one cardiac cycle, due to stINR's GPU memory constraints. For INMR, the models were trained with both data of one cardiac cycle and 4 cardiac cycles (160 frames) for thorough comparisons.

Table 2 compares the quantitative performance of the two strategies. With the same training data from single cardiac cycle, INMR outperforms stINR in PSNR by 2 dB but suffers a slight 0.05 decrease in SSIM. However, when trained on data from 4 cycles, INMR achieves superior results, with a 3.8 dB increase in PSNR and a 0.008 improvement in SSIM, while using less than half the GPU memory of stINR. These results demonstrate that, although INMR requires sufficient training samples to optimize the manifold, it delivers more consistent and high-quality denoising than stINR with lower computational overhead. This efficiency makes INMR particularly advantageous for long acquisition sequences.

Effectiveness of C2F optimization strategy The spatiotemporal C2F optimization strategy plays a vital role in guiding proper convergence of the INMR framework. To evaluate its effectiveness, ablation studies were conducted to evaluate the framework with and without both spatial and temporal C2F components. As shown in Table 3, both the spatial and temporal C2F significantly improve the reconstruction performance and robustness of INMR.

Configuration of manifold dimension To investigate the influence of manifold dimension number, we evaluated the performance of INMR under different dimension configurations ranging from 2 to 16. Table 4 shows the quantitative results on the simulated cardiac dataset. When the dimension is less than 4, the reconstruction performance is highly degraded, due to the limited capability to capture accurate temporal dynamics. The results of both 8-dim and 16-dim are comparable, but the 16-dim configuration introduces temporal discontinuities in the temporal profiles. A similar trend was witnessed in the rest datasets, thus we set 8-dim manifold vectors for all the experiments.

Configuration of regularization hyperparameter λ To analyze the influence of smooth regularization parameter λ, we evaluated the performance of INMR with λ value of 0, 10, 20 on the simulated cardiac dataset. Table 5 demon-

Category	Method	Simulated cardiac	In-vivo cardiac	Speech	DCE abdomen (per slice)
CS-based	GRASP	62.35	127.87	415.73	111.16
	GRASP-Pro	106.83	227.65	906.58	190.54
Unsupervised DL-based	TDDIP	1488.50	4441.22	2423.11	2582.44
	FMLP	2446.17	5215.32	3163.74	3255.45
	INMR	**219.11**	**489.56**	**1381.33**	**519.37**

Table 1: Reconstruction time comparison of different methods on each dataset, in the unit of seconds.

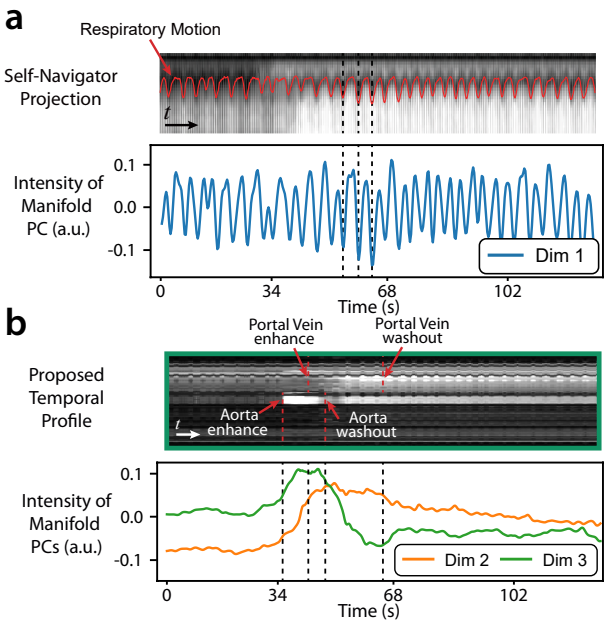

Figure 5: **a**. The respiratory motion extracted from the self-navigator projection ***vs*** the first PC. **b**. The temporal profiles of the proposed results ***vs*** the second and third PCs. Two representative respiratory cycles and the turning points of contrast transition are visualized with the dashed lines.

Temporal modeling	PSNR (dB)	SSIM	GPU memory usage (MB)
stINR (1 cycle)	29.78	0.913	22874
INMR (1 cycle)	31.77	0.861	7868
INMR (4 cycles)	**33.54**	**0.921**	10914

Table 2: Quantitative comparisons and GPU memory usages of different INR temporal modelings on the simulated cardiac dataset.

strates the quantitative results, and it indicates the optimal reconstruction quality at λ=10. For the rest datasets, due to the lack of ground truth, λ was grid-searched on validation data of each dataset across exponential ranges, selecting the minimum λ values that achieved artifact-free motion characterization. The specific values are 2 for both the in-vivo cardiac and DCE datasets, and 0.5 for the speech dataset.

Spatial C2F	Temporal C2F	PSNR (dB)	SSIM
×	×	33.14±1.03	0.908±0.012
×	✓	33.36±0.83	0.910±0.012
✓	×	33.67±0.72	0.919±0.010
✓	✓	**33.93±0.70**	**0.921±0.010**

Table 3: Impact of spatiotemporal C2F optimization on the simulated cardiac dataset.

Manifold dimension	PSNR (dB)	SSIM
2	31.82±1.00	0.845±0.022
4	32.97±0.64	0.893±0.012
8	**33.93±0.70**	**0.921±0.010**
16	33.62±0.73	0.915±0.012

Table 4: Quantitative comparisons of different manifold dimensions on the simulated cardiac dataset.

λ	PSNR (dB)	SSIM
0	29.91±1.15	0.889±0.029
10	**33.93±0.70**	**0.921±0.010**
20	33.24±0.64	0.909±0.011

Table 5: The influence of smooth regularization parameter λ on the simulated cardiac dataset.

Discussion and Conclusion

In this work, we proposed INMR, a zero-shot unsupervised deep learning method designed for ultrafast dynamic MRI that eliminates the need for prior calibration scans or labeled data. By incorporating learnable manifold vectors and a spatiotemporal C2F strategy, INMR outperforms SOTA methods in both artifacts suppression and motion preservation, particularly for irregular physiological motions. Our analysis of the manifold vectors reveals their potential for tracking physiological processes. While INMR demonstrates promising results, future work could further enhance INMR by integrating prior physiological signals (e.g., ECG signal or injection signal functions) to guide the optimization process.

Acknowledgments

This study is supported by the National Natural Science Foundation of China (62471296) and the National Key Research and Development Program of China (2024YFC2421100).

References

Ahmed, A. H.; Zhou, R.; Yang, Y.; Nagpal, P.; Salerno, M.; and Jacob, M. 2020. Free-Breathing and Ungated Dynamic MRI Using Navigator-Less Spiral SToRM. *IEEE Transactions on Medical Imaging*, 39(12): 3933–3943.

Al-Haj Hemidi, Z.; Vogt, N.; Quillien, L.; Weihsbach, C.; Heinrich, M. P.; and Oster, J. 2024. CineJENSE: Simultaneous Cine MRI Image Reconstruction and Sensitivity Map Estimation Using Neural Representations. In *Statistical Atlases and Computational Models of the Heart. Regular and CMRxRecon Challenge Papers*, 467–478.

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, J.; Huang, C.; Shanbhogue, K.; Xia, D.; Bruno, M.; Huang, Y.; Block, K. T.; Chandarana, H.; and Feng, L. 2024. DCE-MRI of the liver with sub-second temporal resolution using GRASP-Pro with navi-stack-of-stars sampling. *NMR in Biomedicine*, e5262.

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. 2022. Golden-Angle Radial MRI: Basics, Advances, and Applications. *Journal of Magnetic Resonance Imaging*, 56(1): 45–62.

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.

Feng, L.; Huang, C.; Shanbhogue, K.; Sodickson, D. K.; Chandarana, H.; and Otazo, R. 2018. RACER-GRASP: Respiratory-weighted, aortic contrast enhancement-guided and coil-unstreaking golden-angle radial sparse MRI. *Magnetic Resonance in Medicine*, 80(1): 77–89.

Feng, L.; Wen, Q.; Huang, C.; Tong, A.; Liu, F.; and Chandarana, H. 2020. GRASP-Pro: imProving GRASP DCE-MRI through self-calibrating subspace-modeling and contrast phase automation. *Magnetic Resonance in Medicine*, 83(1): 94–108.

Feng, R.; Wu, Q.; Feng, J.; She, H.; Liu, C.; Zhang, Y.; and Wei, H. 2024. IMJENSE: Scan-Specific Implicit Representation for Joint Coil Sensitivity and Image Estimation in Parallel MRI. *IEEE Transactions on Medical Imaging*, 43(4): 1539–1553.

Han, Y.; Yoo, J.; Kim, H. H.; Shin, H. J.; Sung, K.; and Ye, J. C. 2018. Deep learning with domain adaptation for accelerated projection-reconstruction MR. *Magnetic Resonance in Medicine*, 80(3): 1189–1205.

Huang, W.; Ke, Z.; Cui, Z.-X.; Cheng, J.; Qiu, Z.; Jia, S.; Ying, L.; Zhu, Y.; and Liang, D. 2021. Deep low-Rank plus sparse network for dynamic MR imaging. *Medical Image Analysis*, 73: 102190.

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 *Proceedings of the International Conference on Information Processing in Medical Imaging (IPMI)*, 548–560.

Huang, W.; Spieker, V.; Xu, S.; Cruz, G.; Prieto, C.; Schnabel, J.; Hammernik, K.; Kuestner, T.; and Rueckert, D. 2024. Subspace Implicit Neural Representations for Real-Time Cardiac Cine MR Imaging. arXiv:2412.12742.

Ke, Z.; Cui, Z.-X.; Huang, W.; Cheng, J.; Jia, S.; Ying, L.; Zhu, Y.; and Liang, D. 2021. Deep Manifold Learning for Dynamic MR Imaging. *IEEE Transactions on Computational Imaging*, 7: 1314–1327.

Kingma, D. P.; and Welling, M. 2019. An Introduction to Variational Autoencoders. *Foundations and Trends in Machine Learning*, 12(4): 307–392.

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.

Li, Z.; Müller, T.; Evans, A.; Taylor, R. H.; Unberath, M.; Liu, M.-Y.; and Lin, C.-H. 2023. Neuralangelo: High-fidelity neural surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8456–8465.

Lim, Y.; Toutios, A.; Bliesener, Y.; Tian, Y.; Lingala, S. G.; Vaz, C.; Sorensen, T.; Oh, M.; Harper, S.; Chen, W.; et al. 2021. A multispeaker dataset of raw and reconstructed speech production real-time MRI video and 3D volumetric images. *Scientific data*, 8(1): 187.

Lingala, S. G.; Hu, Y.; DiBella, E.; and Jacob, M. 2011. Accelerated Dynamic MRI Exploiting Sparsity and Low-Rank Structure: k-t SLR. *IEEE Transactions on Medical Imaging*, 30(5): 1042–1054.

Luo, R.; Huang, H.; Miao, Q.; Xu, J.; Hu, P.; and Qi, H. 2025. A Graphical Method for Designing Time-Optimal Non-Cartesian Gradient Waveforms. arXiv:2507.21625.

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 405–421.

Müller, T.; Evans, A.; Schied, C.; and Keller, A. 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Transactions on Graphics*, 41(4): 102:1–102:15.

Nakarmi, U.; Wang, Y.; Lyu, J.; Liang, D.; and Ying, L. 2017. A Kernel-Based Low-Rank (KLR) Model for Low-Dimensional Manifold Recovery in Highly Accelerated Dynamic MRI. *IEEE Transactions on Medical Imaging*, 36(11): 2297–2307.

Nayak, K. S.; Lim, Y.; Campbell-Washburn, A. E.; and Steeden, J. 2022. Real-Time Magnetic Resonance Imaging. *Journal of Magnetic Resonance Imaging*, 55(1): 81–99.

Poddar, S.; and Jacob, M. 2016. Dynamic MRI Using SmooThness Regularization on Manifolds (STORM). *IEEE Transactions on Medical Imaging*, 35(4): 1106–1115.

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.

Qin, C.; Schlemper, J.; Caballero, J.; Price, A. N.; Hajnal, J. V.; and Rueckert, D. 2019. Convolutional Recurrent Neural Networks for Dynamic MR Image Reconstruction. *IEEE Transactions on Medical Imaging*, 38(1): 280–290.

Shetty, G. N.; Slavakis, K.; Bose, A.; Nakarmi, U.; Scutari, G.; and Ying, L. 2020. Bi-Linear Modeling of Data Manifolds for Dynamic-MRI Recovery. *IEEE Transactions on Medical Imaging*, 39(3): 688–702.

Spieker, V.; Huang, W.; Eichhorn, H.; Stelter, J.; Weiss, K.; Zimmer, V. A.; Braren, R. F.; Karampinos, D. C.; Hammernik, K.; and Schnabel, J. A. 2023. ICoNIK: Generating Respiratory-Resolved Abdominal MR Reconstructions Using Neural Implicit Representations in k-Space. In *Deep Generative Models: Third MICCAI Workshop, DGM4MICCAI 2023*, 183–192.

Uecker, M.; Zhang, S.; Voit, D.; Karaus, A.; Merboldt, K.-D.; and Frahm, J. 2010. Real-time MRI at a resolution of 20 ms. *NMR in Biomedicine*, 23(8): 986–994.

Wang, H.; Yu, T.; Yang, T.; Qiao, H.; and Dai, Q. 2024. Neural physical simulation with multi-resolution hash grid encoding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 5410–5418.

Wang, S.; Su, Z.; Ying, L.; Peng, X.; Zhu, S.; Liang, F.; Feng, D.; and Liang, D. 2016. Accelerating magnetic resonance imaging via deep learning. In *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)*, 514–517.

Wang, Z.; She, H.; Zhang, Y.; and Du, Y. P. 2023. Parallel non-Cartesian spatial-temporal dictionary learning neural networks (stDLNN) for accelerating 4D-MRI. *Medical Image Analysis*, 84: 102701.

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 *International Conference on Representation Learning*, volume 2025, 70680–70701.

Wu, Q.; Feng, R.; Wei, H.; Yu, J.; and Zhang, Y. 2023. Self-Supervised Coordinate Projection Network for Sparse-View Computed Tomography. *IEEE Transactions on Computational Imaging*, 9: 517–529.

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.

Zhao, B.; Haldar, J. P.; Christodoulou, A. G.; and Liang, Z.-P. 2012. Image reconstruction from highly undersampled (k, t)-space data with joint partial separability and sparsity constraints. *IEEE Transactions on Medical Imaging*, 31(9): 1809–1820.

Zou, Q.; Ahmed, A. H.; Nagpal, P.; Kruger, S.; and Jacob, M. 2021. Dynamic Imaging Using a Deep Generative SToRM (Gen-SToRM) Model. *IEEE Transactions on Medical Imaging*, 40(11): 3102–3112.

Zou, Q.; Ahmed, A. H.; Nagpal, P.; Priya, S.; Schulte, R. F.; and Jacob, M. 2022. Variational Manifold Learning From Incomplete Data: Application to Multislice Dynamic MRI. *IEEE Transactions on Medical Imaging*, 41(12): 3552–3561.