Prior as Geometry, Not Generator: Weak-Diffusion Test-Time Reconstruction for Dynamic MRI

Mengzhe Du[1,2*] **Congcong Liu**[1*] **Baoqing Li**[3] **Jinglong Tao**[1,4] **Wenwu He**[5,6†]
Zhuo-Xu Cui[1†]
[1]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
[2]University of Chinese Academy of Sciences, Beijing, China
[3]School of Mathematics and Computational Science, Xiangtan University, Xiangtan, China
[4]Software College of Northeastern University, Shenyang, China
[5]School of Computing and Data Science, Fujian University of Technology, Fuzhou, China
[6]Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fuzhou, China
{mz.du, cc.liu}@siat.ac.cn, 202631510171@smail.xtu.edu.cn
taojl@mails.neu.edu.cn, hwwhbb@163.com, zx.cui@siat.ac.cn

Abstract

Dynamic MRI reconstruction from highly undersampled k-space is an ill-posed inverse problem that requires strong measurement consistency, fine spatial and temporal recovery. Diffusion priors provide useful population-level regularization, but task-matched spatiotemporal diffusion models are costly to train, while generic pretrained priors may be mismatched to protocols, vendors, or anatomies, causing hallucination or over-smoothing. Untrained test-time priors like Time-Dependent Deep Image Prior (TD-DIP) adapt to each scan and naturally couple with data consistency, but can drift toward aliasing artifacts and overfit noise. We propose *Geometry-Gated TD-DIP* for this *weak-diffusion* regime, where diffusion is used primarily as a local geometric reference rather than an unconstrained generative prior. Each TD-DIP correction is split by the diffusion denoiser's Jacobian into a prior-sensitive component and a prior-residual component. A self-calibrated gate regulates the residual contribution to balance scan-specific recovery against prior mismatch and noise amplification. We establish conditional reconstruction error bounds and characterize the effects of geometric gating and diffusion-prior mismatch on reconstruction fidelity. Experiments on cine MRI show consistent gains over diffusion-only, TD-DIP-only, and additive hybrid baselines, especially at high acceleration and under out-of-distribution diffusion priors. The code is publicly available at: https://github.com/MAiTL-Group/weak-diffusion-mri.

1 Introduction

Dynamic magnetic resonance imaging (MRI) is indispensable for capturing time-varying physiology such as cardiac motion and tissue perfusion [1]. To achieve functionally adequate temporal resolution, acquisitions are often aggressively undersampled in k-space, which increases non-identifiability and makes reconstruction highly sensitive to measurement noise and model mismatch [2]. Beyond spatial fidelity, dynamic reconstruction must also preserve *temporal coherence*: small frame-wise inconsistencies readily appear as flickering, motion blurring, or unstable textures.

Score-based diffusion models have emerged as powerful priors for inverse problems, with reconstruction algorithms combining learned image statistics with measurement consistency [3–6]. In

*Mengzhe Du and Congcong Liu contributed equally.
†Corresponding authors: Wenwu He (hwwhbb@163.com), Zhuo-Xu Cui (zx.cui@siat.ac.cn)

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

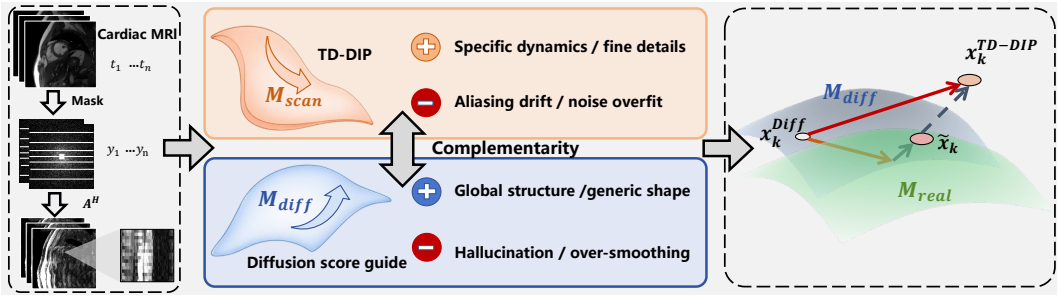


Figure 1: Motivation. Generic diffusion provides weak global geometry, while TD-DIP captures scan-specific temporal dynamics.

MRI, exploiting image correlations and the structure of k-space acquisition can improve reconstruction [7, 8]; dynamic imaging further requires temporal coherence to be preserved alongside fine spatial detail. Training such task-matched diffusion priors is computationally prohibitive and requires fully sampled dynamic datasets that are scarce. More importantly, generic pretrained diffusion models can be biased under protocol and vendor shifts, which often manifests as hallucinated structures or overly conservative smoothing when used as a dominant reconstruction driver.

At the other extreme, untrained scan-adaptive priors offer robustness to distribution shift. Deep Image Prior (DIP) [9] and variants like Time-Dependent DIP (TD-DIP) [10], fit a per-scan implicit representation and promote temporal coherence through shared parameters and time encodings, requiring no external training data. Yet TD-DIP lacks a population-level geometric reference and can drift toward undersampling artifacts or overfit high-frequency noise in later optimization [11]. A common reaction is to combine diffusion and TD-DIP via additive fusion or by imposing diffusion as a rigid constraint, but such couplings provide limited control over *which* correction directions should follow the prior. When diffusion is mismatched, naive fusion can import prior errors, while hard constraints can suppress subject-specific details needed for strict data consistency.

We address this tension by formalizing a *weak diffusion* regime: diffusion is potentially out-of-domain and only partially reliable, and its guidance is coupled to scan-specific measurement fitting at every reverse step. We propose *Geometry-Gated TD-DIP*, which uses diffusion as a soft anchor and a local sensitivity reference. At each reverse step, a pretrained score model provides a diffusion anchor and a local sensitivity operator via the Jacobian of the Tweedie denoiser. In parallel, TD-DIP produces a physics-driven, scan-adaptive correction by fitting the measurements with a soft penalty toward the diffusion anchor. We split this correction into a prior-sensitive component and a prior-residual component carrying candidate scan-specific innovation. A self-calibrated gate regulates the residual contribution across reverse steps to balance adaptation and noise control. This directional coupling goes beyond averaging two reconstructions: it regulates how strongly scan-specific corrections follow the diffusion reference.

Contributions.

- **Weak diffusion as geometry under prior mismatch and domain shift:** we use a generic pretrained diffusion model as a local sensitivity reference under domain mismatch, while TD-DIP remains the primary physics-driven and temporally coherent reconstruction engine.
- **Geometry-gated coupling:** we split TD-DIP corrections into prior-sensitive and prior-residual components with a self-calibrated gate, improving on additive fusion under mismatch.
- **Mismatch-aware theory and robustness:** we derive a conditional recovery bound, an exact gated-error identity, and a full one-step prior-perturbation bound, while empirically evaluating fidelity and temporal consistency under high acceleration and out-of-distribution diffusion priors.

2 Related Work

Dynamic MRI: from explicit models to learned priors. Dynamic MRI reconstruction traditionally relies on explicit spatiotemporal structure to compensate for aggressive undersampling. Classical approaches exploit k–t correlations and physics-based constraints (e.g., k–t methods) [12],

flexible dynamic imaging with golden-angle radial sampling and sparsity/parallel imaging coupling (e.g., GRASP) [13], and low-rank plus sparse decompositions (L+S) that separate background from dynamics [14]. Recent signal models further increase expressiveness beyond global low-rank assumptions, e.g., spatially varying temporal bases and autoregressive (k,t)-space formulations [15]. These model-driven methods are interpretable and robust to limited training data, but may struggle when the assumed model is violated (complex motion, spatially heterogeneous dynamics) or under extremely high acceleration.

Model-based deep learning and physics-guided unrolling. Deep learning improved reconstruction quality by embedding data-consistency (DC) into trainable architectures. Early CNN-based inverse solvers [16] led to unrolled optimization networks that alternate DC with learned regularizers, e.g., variational networks [17], ADMM-inspired unrolling [18], and MoDL-style model-based deep learning [19]. For dynamic cine, learned low-rank operators and sparse+low-rank unrolling explicitly revisit classical priors in a learnable form [20]. While unrolled networks offer fast inference and strong average-case fidelity, their performance can degrade under sampling/protocol shifts, and supervised training typically requires fully sampled references that are scarce in dynamic settings.

Self-supervision and scan-adaptation. To reduce dependence on fully sampled ground truth, self-supervised strategies train physics-guided networks by splitting acquired measurements into disjoint DC and loss subsets (e.g., SSDU) [21]. Going further, zero-shot self-supervised learning adapts a network to a *single* scan by partitioning that scan's measurements into training/loss/validation sets to enable early stopping and mitigate overfitting [22]. These methods enhance data efficiency and robustness relative to purely supervised pipelines, but under very high acceleration the self-supervised signal can be weak, and careful split design and stopping criteria become critical.

Untrained priors and implicit spatiotemporal representations. Untrained reconstruction uses architectural bias as a prior without external datasets. Deep Image Prior (DIP) [9] and untrained MRI reconstruction [23] demonstrate surprising regularization capacity, yet are known to drift toward noise/overfitting as optimization proceeds [11]. For dynamic MRI, TD-DIP introduces time-conditioned parameter sharing and temporal encodings to enforce coherence while fitting undersampled k-space [10], with subsequent dynamic DIP variants expanding applicability to free-breathing and multi-protocol settings [24, 25]. In parallel, implicit neural representations (INRs) formulate dynamic images as continuous functions of spatiotemporal coordinates learned directly from sparse (k, t)-space data, often combined with explicit low-rank/sparsity regularization [26]. Recent unsupervised dynamic MRI variants further incorporate explicit motion decomposition or joint sensitivity estimation, e.g., MoCo-INR for motion-compensated cardiac reconstruction [27] and IMJ-PLUS for free-breathing reconstruction with time-dependent coil sensitivities [28]. Scan-specific priors are attractive under distribution shift, but their lack of a population-level geometric reference can make them vulnerable to undersampling artifacts at extreme acceleration unless paired with additional guidance.

Diffusion/score priors for inverse problems: projection, constraints, and acceleration. Score-based generative modeling [3] and denoising diffusion models [4] have become powerful priors for inverse problems by iteratively denoising while injecting measurement consistency. General frameworks include diffusion posterior sampling under noisy measurements [29], manifold-constraint interpretations that regulate updates in feasible sets [5], and decomposed samplers that reduce computational burden for large-scale inverse problems [6]. Related lines such as diffusion null-space modeling emphasize updating only unconstrained components to enable zero-shot restoration [30]. More recent works couple diffusion with test-time DIP/INR adaptation. DDIP/D3IP [31] adapts an OOD diffusion prior to the test instance for 3D inverse problems, while DPER [32] and DiffINR [33] combine neural representations with diffusion priors through alternating or posterior-sampling style reconstruction. A recurring limitation in medical imaging is that when diffusion dominates the reconstruction trajectory, prior mismatch may induce hallucinations or conservative over-smoothing, motivating couplings that more carefully control *which directions* inherit the prior.

Diffusion for MRI: MRI-specific designs, domains, and spatiotemporal challenges. MRI reconstruction has inspired diffusion variants that operate in different domains (image-space, k-space, latent-space) and incorporate MRI physics more explicitly [34, 35]. Efficiency and determinism

considerations also motivate designs that restrict diffusion to uncertain subspaces (e.g., emphasizing high-frequency components while treating low-frequency data deterministically) [36]. For spatiotemporal inverse problems, training full video diffusion models is expensive; an alternative is to leverage image diffusion models across time with batch-consistent strategies to promote temporal coherence [37]. Despite these advances, robustness under protocol/vendor/domain shift remains a key bottleneck, especially when diffusion priors are pretrained on static or out-of-domain corpora.

Our positioning: weak diffusion as local geometry for scan-adaptive TD-DIP. The above literature reveals a persistent tension in dynamic MRI: diffusion priors provide population-level structure but can be unreliable under mismatch, whereas scan-adaptive priors (TD-DIP/INR) preserve temporal coherence yet may drift toward undersampling artifacts. Existing couplings often adapt the diffusion prior to the test instance [31], alternate between INR reconstruction and diffusion regularization [32], or use INR-guided posterior sampling where diffusion remains a dominant generative prior [33]. In contrast, our method targets the *weak diffusion* regime by using diffusion as a *local geometric reference* within a scan-adaptive reverse sampler, rather than relying on an uncorrected generative trajectory. By splitting each physics-driven TD-DIP correction into prior-sensitive and prior-residual components and adaptively gating the residual, our framework regulates dependence on the frozen diffusion operator while retaining scan-adaptive temporal structure.

3 Method

We propose *Geometry-Gated TD-DIP*, a dynamic MRI reconstruction framework tailored to the *weak diffusion* regime. The design is intentionally asymmetric: a scan-adaptive time-dependent Deep Image Prior (TD-DIP) promotes *data consistency* and *temporal coherence* through per-scan optimization, while a pretrained (possibly mismatched) diffusion prior contributes (i) a *soft anchor* and (ii) a *local sensitivity reference*. This avoids relying on task-matched spatiotemporal diffusion training, yet leverages diffusion's population-level structural bias in a controlled, mismatch-tolerant way.

3.1 Problem Setup

Let $\mathbf{x}_0 \in \mathbb{C}^d$ denote the unknown dynamic image sequence (all frames stacked), and let $\mathbf{y} \in \mathbb{C}^m$ be the undersampled multi-coil k-space measurements. The forward model is

$$\mathbf{y} \;=\; \mathcal{A}(\mathbf{x}_0) + \boldsymbol{\eta}, \qquad \boldsymbol{\eta} \sim \mathcal{CN}\big(\mathbf{0},\, \sigma_y^2 \mathbf{I}\big), \tag{1}$$

where $\mathcal{A}$ is the multi-coil encoding operator (applied frame-wise and stacked over time), and $\mathcal{CN}$ denotes circularly symmetric complex Gaussian noise. For compatibility with real-valued score networks, we represent complex-valued images by stacking real/imaginary parts; all variables may be viewed in $\mathbb{R}^{2d}$ without changing the formulation.

Under aggressive undersampling, reconstruction is severely ill-posed. We combine a score-derived quadratic anchor with a scan-specific image-set constraint induced by TD-DIP.

3.2 Weak Diffusion as a Soft Local Anchor

We assume access to a pretrained *2D* score-based diffusion prior through a score function $\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}, \sigma) \approx \nabla_{\mathbf{x}} \log p_\sigma(\mathbf{x})$. Since this prior may be trained on limited, synthetic, or out-of-domain data, we treat it as *weak* and potentially biased.

Let $\{\mathbf{x}_k\}_{k=1}^K$ denote the reverse diffusion states with noise levels $\{\sigma_k\}_{k=1}^K$. At step k, we compute the Tweedie denoised estimate (the *anchor*)

$$\hat{\mathbf{x}}_{0,k} \;\triangleq\; \mathcal{D}_{\boldsymbol{\theta},k}(\mathbf{x}_k) \;=\; \mathbf{x}_k + \sigma_k^2\, \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}_k, \sigma_k). \tag{2}$$

We use this estimate as the center of a quadratic anchor with step-dependent strength $\mu_k > 0$ (Eq. (5)). A weak anchor allows scan-specific adaptation to a potentially mismatched diffusion reference.

3.3 TD-DIP as a Scan-Adaptive Temporal Image Set

We represent the dynamic sequence via a time-conditioned Deep Image Prior (TD-DIP). Let $t_n \in [0, 1]$ be the normalized time for frame n, and let $\gamma(t)$ be a fixed temporal encoding (e.g., Fourier

features). A generator $\mathcal{G}_{\boldsymbol{\phi}}$ produces each frame by

$$\mathbf{x}_{\boldsymbol{\phi}}^{(n)} = \mathcal{G}_{\boldsymbol{\phi}}(t_n) = \mathcal{H}_{\boldsymbol{\phi}}\big(\mathcal{M}_{\boldsymbol{\phi}}(\mathbf{z}, \gamma(t_n))\big),\ \mathbf{x}_{\boldsymbol{\phi}} = \{\mathbf{x}_{\boldsymbol{\phi}}^{(n)}\}_{n=1}^{N_t}, \tag{3}$$

where $\mathbf{z}$ is a fixed random input, $\mathcal{M}_{\boldsymbol{\phi}}$ is a mapping network (MapNet), and $\mathcal{H}_{\boldsymbol{\phi}}$ is a decoder. All parameters $\boldsymbol{\phi}$ are optimized *per scan* (no external training).

Temporal regularization is induced implicitly: shared parameters across time and smooth structure in $\gamma(t)$ constrain the sequence to a representable image set

$$\mathcal{M}_{\mathrm{DIP}} \triangleq \{\mathbf{x}_{\boldsymbol{\phi}} \mid \boldsymbol{\phi} \in \mathbb{R}^p\}. \tag{4}$$

In our framework, $\mathcal{M}_{\mathrm{DIP}}$ is the *primary* scan-adaptive constraint that stabilizes the reconstruction and promotes temporal coherence, compensating for the deliberately weak diffusion anchor.

3.4 Per-Step Score-Anchored Fitting via Two-Stage Optimization

At each reverse step k, we combine the measurement likelihood and the weak diffusion anchor into the local energy

$$\mathcal{L}_k(\mathbf{x}) \triangleq \frac{1}{2\sigma_y^2}\big\|\mathcal{A}(\mathbf{x}) - \mathbf{y}\big\|_2^2 + \frac{\mu_k}{2}\big\|\mathbf{x} - \hat{\mathbf{x}}_{0,k}\big\|_2^2. \tag{5}$$

We compute a scan-adaptive estimate by approximately minimizing this score-anchored objective over the TD-DIP image set:

$$\boldsymbol{\phi}_k^\star \in \arg\min_{\boldsymbol{\phi}}\ \mathcal{L}_k(\mathbf{x}_{\boldsymbol{\phi}}), \qquad \mathbf{x}_k^{\mathrm{DIP}} \triangleq \mathbf{x}_{\boldsymbol{\phi}_k} \in \mathcal{M}_{\mathrm{DIP}}, \tag{6}$$

where $\boldsymbol{\phi}_k$ denotes the finite-step fit.

Two-stage solver (practical). To stabilize optimization under high noise and severe undersampling, we warm-start with image-domain alignment and then enforce data consistency:

$$\text{(Stage A)} \quad \boldsymbol{\phi}_k^A \in \arg\min_{\boldsymbol{\phi}}\ \big\|\mathbf{x}_{\boldsymbol{\phi}} - \hat{\mathbf{x}}_{0,k}\big\|_2^2, \tag{7}$$

$$\text{(Stage B)} \quad \boldsymbol{\phi}_k: \quad \text{approximately minimize } \mathcal{L}_k(\mathbf{x}_{\boldsymbol{\phi}}) \quad \text{from } \boldsymbol{\phi}_k^A. \tag{8}$$

Stage A initializes the fit near the diffusion anchor; Stage B balances measurement fit and anchor proximity within $\mathcal{M}_{\mathrm{DIP}}$.

We define the physics-driven TD-DIP correction

$$\mathbf{d}_k \triangleq \mathbf{x}_k^{\mathrm{DIP}} - \hat{\mathbf{x}}_{0,k}. \tag{9}$$

3.5 Diffusion-Prior Sensitivity and Residual Decomposition

We use the denoiser's local sensitivity to split the TD-DIP correction before fusion.

Local sensitivity operator. We define the denoiser Jacobian with respect to the current reverse state as

$$\mathbf{J}_{\boldsymbol{\theta},k} \triangleq \nabla_{\mathbf{x}_k} \mathcal{D}_{\boldsymbol{\theta},k}(\mathbf{x}_k). \tag{10}$$

The local sensitivity operator $\mathbf{J}_{\boldsymbol{\theta},k}$ splits the TD-DIP correction $\mathbf{d}_k$ into prior-sensitive and prior-residual components:

$$\mathbf{d}_k^{\mathrm{sens}} \triangleq \mathbf{J}_{\boldsymbol{\theta},k}\mathbf{d}_k, \qquad \mathbf{d}_k^{\mathrm{res}} \triangleq (\mathbf{I} - \mathbf{J}_{\boldsymbol{\theta},k})\mathbf{d}_k. \tag{11}$$

The two components need not be orthogonal. These components describe the denoiser's local response and the remaining correction, without a prescribed signal/noise interpretation. We compute them using Jacobian–vector products without forming the Jacobian explicitly. The 2D denoiser acts frame-wise, while shared TD-DIP parameters couple frames over time.

3.6 Adaptive Gating from Residual Fraction and Alignment

The residual branch carries candidate scan-specific innovation, whose contribution is regulated by an adaptive gate.

We define two normalized cues: the *residual fraction* $m_k \in [0,1]$ and the *directional alignment* $a_k \in [-1,1]$:

$$m_k \triangleq \frac{\|\mathbf{d}_k^{\text{res}}\|_2}{\|\mathbf{d}_k^{\text{sens}}\|_2 + \|\mathbf{d}_k^{\text{res}}\|_2 + \varepsilon},\ a_k \triangleq \frac{\text{Re}\langle \mathbf{d}_k, \mathbf{d}_k^{\text{sens}} \rangle}{\|\mathbf{d}_k\|_2 \|\mathbf{d}_k^{\text{sens}}\|_2 + \varepsilon}. \tag{12}$$

We then compute a content-aware gate $\hat{\lambda}_k$ via a non-linear decision function:

$$\hat{\lambda}_k \triangleq \lambda_{\min} + (\lambda_{\max} - \lambda_{\min}) \cdot \sigma\big(\beta(m_k - \tau_m) - \gamma(a_k - \tau_a)\big), \tag{13}$$

where $\sigma(\cdot)$ is the sigmoid function and τ_m, τ_a are scalar thresholds. The gate increases the residual weight when m_k is high and a_k is low. These cues summarize the correction's disagreement with the prior's local response. We smooth the gate over reverse steps using an exponential moving average (EMA), $\lambda_k \leftarrow \rho\lambda_{k+1} + (1-\rho)\hat{\lambda}_k$ for the descending loop $k = K, \ldots, 1$, initialized by $\lambda_{K+1} = \hat{\lambda}_K$.

Finally, the geometry-gated clean estimate is formed as:

$$\tilde{\mathbf{x}}_{0,k} \ \triangleq\ \hat{\mathbf{x}}_{0,k} + \mathbf{d}_k^{\text{sens}} + \lambda_k\, \mathbf{d}_k^{\text{res}}. \tag{14}$$

The gated correction is then passed to the reverse sampler.

3.7 Sampler-Agnostic Reverse Diffusion Update

Given $\tilde{\mathbf{x}}_{0,k}$, we perform a standard sampler update (e.g., DDIM or predictor–corrector) by replacing the usual clean estimate with our gated one:

$$\mathbf{x}_{k-1} \ =\ \text{SamplerStep}(\mathbf{x}_k,\, \tilde{\mathbf{x}}_{0,k},\, \sigma_k,\, \sigma_{k-1})\,. \tag{15}$$

Hence the framework is sampler-agnostic; the novelty lies in constructing $\tilde{\mathbf{x}}_{0,k}$ via image-set-restricted TD-DIP fitting and diffusion-induced geometric gating.

3.8 Overall Reconstruction Algorithm

Algorithm 1 summarizes the full procedure.The framework diagram of our proposed method is shown in the appendix C.

Theoretical analysis is provided in Appendix H, with detailed proofs in Appendix I.

4 Experiments

Experiment design around *weak diffusion*. Our experiments test whether a diffusion prior trained without cardiac cine data can still improve dynamic MRI reconstruction when used only as a weak geometric anchor. We evaluate four aspects: aggressive undersampling, cross-domain generalization, prospective real-world applicability, and controlled studies of geometric coupling and prior mismatch.

4.1 Datasets and Undersampling Protocols

We evaluate the proposed framework on three MRI datasets and use an additional synthetic dataset for prior-mismatch analysis, spanning diverse acquisition settings and domain shifts: **(i)** an in-house cardiac cine dataset, **(ii)** CMRxRecon2025 [38], **(iii)** OCMR [39], and **(iv)** a synthetic moving-ellipses dataset used to stress-test prior mismatch through diffusion-prior training. Retrospective undersampling is simulated using both Cartesian trajectories (Random and VISTA) and a non-Cartesian Golden-Angle Radial trajectory, with aggressive acceleration factors spanning $8\times$ to $38.4\times$. Complete specifications of data acquisition, preprocessing, and simulation settings are provided in Appendix A.1.

Algorithm 1 Geometry-Gated TD-DIP for Dynamic MRI

Input: measurements $\mathbf{y}$, operator $\mathcal{A}$, score $\mathbf{s}_{\boldsymbol{\theta}}$, TD-DIP generator $\mathcal{G}_{\boldsymbol{\phi}}$, schedule $\{\sigma_k\}_{k=1}^{K}$, anchor strengths $\{\mu_k\}_{k=1}^{K}$.
Initialize: $\mathbf{x}_K \sim \mathcal{N}(\mathcal{A}^{\mathsf{H}}\mathbf{y},\, \sigma_K^2\mathbf{I})$; initialize TD-DIP parameters $\boldsymbol{\phi}$.
for $k = K, \ldots, 1$ **do**
  // weak diffusion anchor
  $\hat{\mathbf{x}}_{0,k} \leftarrow \mathbf{x}_k + \sigma_k^2\, \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}_k, \sigma_k)$ (Eq. 2)
  // TD-DIP two-stage score-anchored fitting
  $\boldsymbol{\phi}_k^A \leftarrow \arg\min_{\boldsymbol{\phi}} \|\mathbf{x}_{\boldsymbol{\phi}} - \hat{\mathbf{x}}_{0,k}\|_2^2$ (Eq. 7)
  $\boldsymbol{\phi}_k \leftarrow$ finite-step minimization of $\mathcal{L}_k(\mathbf{x}_{\boldsymbol{\phi}})$ **from** $\boldsymbol{\phi}_k^A$ (Eq. 8)
  $\mathbf{x}_k^{\mathrm{DIP}} \leftarrow \mathbf{x}_{\boldsymbol{\phi}_k}; \quad \mathbf{d}_k \leftarrow \mathbf{x}_k^{\mathrm{DIP}} - \hat{\mathbf{x}}_{0,k}$ (Eq. 9)
  // prior-sensitive/residual split via Jacobian–vector product
  $\mathbf{d}_k^{\mathrm{sens}} \leftarrow \mathbf{J}_{\boldsymbol{\theta},k}\mathbf{d}_k$ (Eq. 11)
  $\mathbf{d}_k^{\mathrm{res}} \leftarrow \mathbf{d}_k - \mathbf{d}_k^{\mathrm{sens}}$
  // self-calibrated geometric gating
  Compute residual fraction and directional alignment m_k, a_k (Eq. 12)
  Update adaptive gate λ_k via sigmoid decision and EMA (Eq. 13)
  $\tilde{\mathbf{x}}_{0,k} \leftarrow \hat{\mathbf{x}}_{0,k} + \mathbf{d}_k^{\mathrm{sens}} + \lambda_k \mathbf{d}_k^{\mathrm{res}}$ (Eq. 14)
  // sampler step
  $\mathbf{x}_{k-1} \leftarrow \mathrm{SamplerStep}(\mathbf{x}_k, \tilde{\mathbf{x}}_{0,k}, \sigma_k, \sigma_{k-1})$ (Eq. 15)
end for
Return: reconstructed dynamic sequence $\tilde{\mathbf{x}}_{0,1}$.

4.2 Implementation and Baselines

We compare against five representative baselines covering distinct reconstruction paradigms: **L+S** (Low-Rank + Sparse) [14], a classic optimization-based compressed sensing method that decomposes the dynamic sequence into a low-rank background and a sparse dynamic component; **MoDL** [19], a supervised unrolled network that alternates a learned CNN denoiser with physics-based conjugate-gradient updates (trained on the remaining 800 samples from the in-house cardiac cine dataset); **TD-DIP** [10], an unsupervised scan-specific implicit neural representation serving as a zero-shot baseline for temporal coherence; **HashINR** [26], an advanced implicit representation with multi-resolution hash encoding and low-rank regularization (with the reported tTVLR configuration); and **SVI-Diffusion** [37], a standard diffusion posterior sampling approach based on a score prior. Unless stated otherwise, all methods use the same undersampling operators and evaluation protocol.

Table 1: Quantitative comparison of reconstruction performance (PSNR/SSIM) across undersampling patterns and acceleration factors. Best results are **bolded**; second-best are underlined.

Pattern	Acc	Baseline[1]	TD-DIP	HashINR _tTVLR	MoDL	SVI-Diffusion	Linear Fusion	Ours
		PSNR / SSIM[2]	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM
Random Cartesian	8×	41.515 / 96.48	40.034 / 94.25	**46.654** / 98.34	41.175 / 96.48	35.403 / 84.88	40.479 / 95.60	45.889 / **98.41**
	10×	39.968 / 95.83	39.223 / 93.43	42.891 / 97.20	39.858 / 96.03	34.693 / 84.57	39.107 / 94.33	**44.230 / 98.00**
	12×	39.187 / 95.16	38.926 / 92.80	42.685 / 97.04	38.781 / 95.19	32.748 / 80.77	37.601 / 92.47	**44.026 / 97.80**
VISTA	12×	35.688 / 92.39	41.446 / 95.67	**48.053 / 98.62**	42.893 / 97.12	31.466 / 74.91	33.498 / 84.21	45.769 / 98.27
	16×	32.112 / 86.23	40.462 / 94.58	42.387 / 96.87	40.214 / 95.76	28.421 / 68.20	31.915 / 80.88	**43.933 / 97.60**
	20×	28.421 / 78.35	39.620 / 93.60	39.580 / 95.23	37.391 / 93.10	26.475 / 64.79	30.981 / 78.50	**43.036 / 97.19**
	24×	26.337 / 72.54	38.733 / 92.74	36.182 / 92.24	35.717 / 91.96	24.312 / 61.08	28.889 / 73.91	**42.129 / 96.82**
Golden Radial	14.8×	27.707 / 56.75	43.276 / 95.33	**47.081 / 98.50**	34.912 / 88.65	40.849 / 93.31	40.218 / 92.84	44.811 / 94.57
	24×	24.920 / 45.86	42.531 / 95.09	**44.868 / 97.79**	31.226 / 82.72	37.338 / 91.19	39.199 / 92.23	44.182 / 94.56
	38.4×	23.117 / 37.30	41.523 / 94.63	42.214 / **96.54**	28.037 / 76.13	33.406 / 87.12	37.546 / 90.86	**43.469** / 94.51

[1] Baseline represents **L+S** for Random/VISTA, and **NUFFT** for Radial masks.
[2] SSIM values are scaled by 10^2 for readability. Standard deviations for all entries in Table 1 are reported in Appendix D.3.

4.3 Comparative Evaluation on Cardiac Cine MRI

We evaluate robustness under aggressive undersampling on the test split of the in-house cardiac cine dataset. Retrospective undersampling is simulated using Random Cartesian and VISTA masks (AF

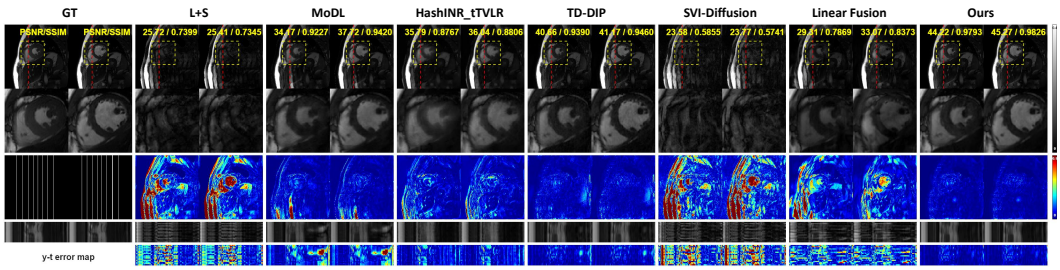


Figure 2: **Visual comparison on the Cardiac Cine dataset under VISTA $24\times$ undersampling.** The columns compare full-FOV reconstructions, zoomed ROIs, spatial error maps, and y-t temporal profiles during systole and diastole. Our method shows the best artifact suppression and structural preservation at high acceleration.

$12\times$–$24\times$), as well as Golden-Angle Radial sampling with 13, 8, and 5 spokes (AF up to $\approx 38.4\times$). Reconstruction quality is measured by PSNR and SSIM against the baselines in Sec. 4.2.

We additionally include a **Linear Fusion** baseline to test whether naive additive coupling is sufficient; its exact definition is provided in Appendix D.1. Quantitative results are summarized in Table 1.

4.4 Additional Evaluation Protocols

Cross-vendor generalization. To evaluate distribution shift, we test on the CMRxRecon2025 dataset using the same baseline set and retrospectively undersample it with a VISTA mask at $8\times$ acceleration to probe robustness to scanner-vendor and acquisition-protocol differences at the same 3.0 T field strength.

Prospective reference-free evaluation. To assess applicability in real-world free-breathing scenarios, we evaluate on the prospective OCMR dataset (estimated AF $\approx 12\times$). Since fully sampled ground truth is unavailable, we adopt a reference-free protocol combining SSDU scores [21], second-order temporal difference energy, motion-compensation consistency, and exploratory atrial area-curve analysis from 3D Slicer segmentations. Full metric definitions are provided in Appendix E.

Ablation protocol. We evaluate the coupling on the in-house cardiac cine dataset under VISTA $12\times$, comparing diffusion-only MCG, TD-DIP alone, naive diffusion+TD-DIP MCG coupling, a matched hybrid without decomposition/gating, and the full method. Results are reported in Sec. 5.4.

Radiologist evaluation. Two radiologists independently scored seven anonymized methods on five randomly selected VISTA $24\times$ cases for overall image quality, structural detail, temporal consistency, and artifact suppression. The scoring protocol is provided in Appendix D.4, with results in Sec. 5.3.

5 Results

5.1 Quantitative Performance Analysis

Table 1 summarizes the quantitative performance (PSNR/SSIM) across three sampling trajectories with acceleration factors ranging from $8\times$ to $38.4\times$. Across the ten reported settings, our method achieves the highest PSNR in six and the highest SSIM in six, with consistent advantages over MoDL and TD-DIP in PSNR. Notably, at extreme acceleration rates (e.g., Golden-Angle Radial $38.4\times$), our method attains the highest PSNR (43.469 dB), although HashINR has the highest SSIM; MoDL suffers a marked performance drop under this setting. The gap over Linear Fusion supports the benefit of Jacobian-based coupling over direct averaging. In the blinded reader study, our method attains the highest mean score in all four categories, including temporal consistency (4.30 versus 3.80 for TD-DIP; Sec. 5.3).

5.2 Generalization to Unseen Protocols and Prospective Scans

On the cross-vendor CMRxRecon dataset, the supervised MoDL baseline degrades noticeably under scanner-vendor and protocol shift at 3.0 T, while HashINR remains more robust but oversmooths

Table 2: Blinded reader-averaged scores on five VISTA $24\times$ cases (mean $\pm$ standard deviation across cases). Higher is better.

Method	Overall	Detail	Temporal	Artifact suppression
L+S	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
MoDL	2.90 ± 0.42	2.80 ± 0.27	3.00 ± 0.00	3.10 ± 0.22
HashINR$_{\mathrm{tTVLR}}$	1.90 ± 0.22	2.40 ± 0.22	2.00 ± 0.00	2.00 ± 0.00
TD-DIP	3.10 ± 0.96	3.10 ± 0.96	3.80 ± 0.45	3.10 ± 0.96
SVI-Diffusion	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
Linear Fusion	1.50 ± 0.00	1.50 ± 0.00	1.00 ± 0.00	1.50 ± 0.00
Ours	$\mathbf{3.60 \pm 0.55}$	$\mathbf{3.60 \pm 0.55}$	$\mathbf{4.30 \pm 0.27}$	$\mathbf{3.20 \pm 0.67}$
Friedman $\chi^2(6)$	28.74	28.12	29.91	27.90
p-value	< 0.001	< 0.001	< 0.001	< 0.001

Table 3: Ablation study of key components (VISTA $12\times$).

Components			Metrics	
Diffusion	TD-DIP	Coupling	PSNR (dB)	SSIM ($\times 10^{-2}$)
✓	-	MCG	34.25 ± 1.18	84.17 ± 3.07
-	✓	-	41.45 ± 2.25	95.67 ± 0.88
✓	✓	MCG	43.92 ± 1.57	97.74 ± 0.69
✓	✓	No decomposition/gate	45.05 ± 1.88	98.14 ± 0.70
✓	✓	**Ours**	$\mathbf{45.77 \pm 1.92}$	$\mathbf{98.27 \pm 0.68}$

high-frequency textures. In contrast, our geometric constraint preserves structural fidelity without retraining, demonstrating promising cross-vendor transfer on the tested subset (Fig. 8). The same trend appears in prospective OCMR, where purely implicit methods such as HashINR suffer from motion blur and TD-DIP misses fine structures, whereas our method yields sharper myocardial boundaries and lower noise (Fig. 9).

5.3 Blinded Radiologist Evaluation

Table 2 reports blinded reader-averaged scores on five VISTA $24\times$ cases for overall image quality, structural detail, temporal consistency, and artifact suppression. The full method receives the highest mean score in all four categories. The reader study assesses reconstruction quality rather than diagnostic accuracy. Detailed scoring and statistical procedures are provided in Appendix D.4.

5.4 Ablation Study on Sensitivity/Residual Coupling

We analyze the contribution of each component based on the configurations defined in Sec. 4.4. Quantitative results are presented in Table 3.

Table 3 reports the gain from combining diffusion and TD-DIP and the additional improvement from the proposed sensitivity/residual gate.

TD-DIP alone and naive diffusion+TD-DIP MCG coupling reach 41.45 and 43.92 dB, respectively. The matched hybrid without decomposition/gating reaches 45.05 dB, compared with 45.77 dB for the full method. The additional 0.72 dB improvement supports the benefit of the proposed coupling beyond combining diffusion and TD-DIP.

5.5 Validation of the "Weak Diffusion" Hypothesis

To empirically validate the central premise that our geometric gating mechanism reduces dependency on task-specific priors, we evaluated reconstruction performance on the Cardiac Cine dataset (VISTA $12\times$) using three distinct diffusion models with varying degrees of domain alignment. We compared a **Matched Prior** trained on 800 augmented in-house cine training samples (a domain-specific reference) against a **Synthetic Geometric Prior** trained solely on 200 samples of moving ellipses (an anatomy-mismatched prior based on synthetic shapes) and a standard **Pre-trained Prior** (representing the practical generic use case). Qualitative comparisons are shown in Fig. 3.

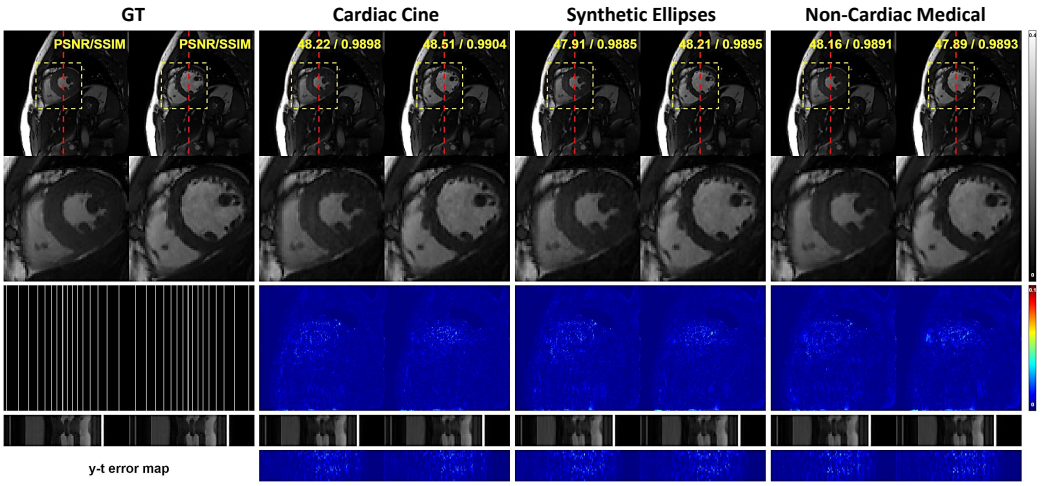


Figure 3: **Validation of robustness against prior mismatch (VISTA** $12\times$**).** We compare reconstruction performance using three diffusion priors with varying degrees of domain knowledge: **Cine** (fully matched domain data), **Ellipses** (synthetic geometric shapes only), and **Pre-trained** (generic weights). **Top Row:** Reconstructed frames with PSNR/SSIM scores. **Middle Row:** Zoomed-in views showing fine anatomical details. **Bottom Rows:** $y - t$ temporal profiles (indicated by the red dashed line) and corresponding error maps. **Observation:** In the illustrated example, the *Ellipses* and *Cine* priors yield similar reconstructions (PSNR gap < 0.4 dB). Anatomical detail is retained across the three tested priors, including the synthetic ellipse prior without cardiac training data.

In the illustrated comparison, all three priors produce similar reconstructions, with the synthetic ellipses prior within 0.4 dB of the matched cine prior. This example shows that detailed reconstruction can be maintained with an anatomy-mismatched diffusion prior; additional qualitative discussion is provided in Appendix D.6.

6 Conclusion

We introduced a geometry-guided reconstruction framework that leverages generic "weak" diffusion models as local references for scan-specific TD-DIP optimization. A Jacobian-based prior-sensitive/residual split combines population-level structural guidance with scan-specific dynamics through adaptive gating. The analysis provides local recovery and operator-mismatch bounds, while reconstruction experiments and radiologist scores support the method under aggressive acceleration and prior mismatch. These results suggest that high-quality dynamic MRI reconstruction need not rely on costly task-matched diffusion pretraining.

Acknowledgments

This work is in part supported by the Shenzhen Science and Technology Program (Grant No.JCYJ20240813155840052), and the National Natural Science Foundation of China (62476268) and the Fujian Province Natural Science Foundation of China under Grant (2024J01158).

The authors thank Pengfei Zhao (The Affiliated Hospital of Inner Mongolia Medical University) and Jun Chai (Inner Mongolia Cardiovascular and Cerebrovascular Hospital) for their radiological assessment of the reconstructed cardiac cine MRI sequences.

Competing Interests Disclosure

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Aaron D Curtis and Hai-Ling M Cheng. Primer and historical review on rapid cardiac CINE MRI. *Journal of Magnetic Resonance Imaging*, 55(2):373–388, 2022.

[2] Onat Dalmaz, Arjun D Desai, Reinhard Heckel, Tolga Cukur, Akshay S Chaudhari, and Brian Hargreaves. Efficient noise calculation in deep learning-based MRI reconstructions. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=br7fTbnd16.

[3] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

[4] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf.

[5] Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, and Jong Chul Ye. Improving diffusion models for inverse problems using manifold constraints. *Advances in Neural Information Processing Systems*, 35:25683–25696, 2022.

[6] Hyungjin Chung, Suhyeon Lee, and Jong Chul Ye. Decomposed diffusion sampler for accelerating large-scale inverse problems. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=DsEhqQtfAG.

[7] Zheng Zhong, Kanghyun Ryu, Jonathan Mao, Kaibao Sun, Guangyu Dan, Shreyas S Vasanawala, and Xiaohong Joe Zhou. Accelerating high b-value diffusion-weighted MRI using a convolutional recurrent neural network (CRNN-DWI). *Bioengineering*, 10(7):864, 2023.

[8] Jiazhen Wang, Zhihao Shi, Xiang Gu, Yan Yang, and Jian Sun. Diffusion-based arbitrary-scale magnetic resonance image super-resolution via progressive k-space reconstruction and denoising. *Medical Image Analysis*, page 103814, 2025.

[9] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018.

[10] Jaejun Yoo, Kyong Hwan Jin, Harshit Gupta, Jerome Yerly, Matthias Stuber, and Michael Unser. Time-dependent deep image prior for dynamic MRI. *IEEE Transactions on Medical Imaging*, 40(12):3337–3348, 2021.

[11] Yeonsik Jo, Se Young Chun, and Jonghyun Choi. Rethinking deep image prior for denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5087–5096, 2021.

[12] Hong Jung, Jong Chul Ye, and Eung Yeop Kim. Improved k–t BLAST and k–t SENSE using FOCUSS. *Physics in Medicine & Biology*, 52(11):3201, 2007.

[13] Li Feng, Robert Grimm, Kai Tobias Block, Hersh Chandarana, Sungheon Kim, Jian Xu, Leon Axel, Daniel K Sodickson, and Ricardo Otazo. 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, 2014.

[14] Ricardo Otazo, Emmanuel Candes, and Daniel K Sodickson. 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, 2015.

[15] Rodrigo A. Lobos, Xiaokai Wang, Rex T. L. Fung, Yongli He, David Frey, Dinank Gupta, Zhongming Liu, Jeffrey A. Fessler, and Douglas C. Noll. Spatiotemporal maps for dynamic mri reconstruction, 2026. URL https://arxiv.org/abs/2507.14429.

[16] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017.

[17] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P Recht, Daniel K Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, 2018.

[18] Yan Yang, Jian Sun, Huibin Li, Zongben Xu, et al. Deep ADMM-net for compressive sensing MRI. *Advances in Neural Information Processing Systems*, 29, 2016.

[19] Hemant K Aggarwal, Merry P Mani, and Mathews Jacob. MoDL: Model-based deep learning architecture for inverse problems. *IEEE Transactions on Medical Imaging*, 38(2):394–405, 2018.

[20] Ziwen Ke, Wenqi Huang, Zhuo-Xu Cui, Jing Cheng, Sen Jia, Haifeng Wang, Xin Liu, Hairong Zheng, Leslie Ying, Yanjie Zhu, and Dong Liang. Learned low-rank priors in dynamic mr imaging. *IEEE Transactions on Medical Imaging*, 40(12):3698–3710, 2021. doi: 10.1109/TMI.2021.3096218.

[21] Burhaneddin Yaman, Seyed Amir Hossein Hosseini, Steen Moeller, Jutta Ellermann, Kâmil Uğurbil, and Mehmet Akçakaya. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magnetic resonance in medicine*, 84(6):3172–3191, 2020.

[22] Burhaneddin Yaman, Seyed Amir Hossein Hosseini, and Mehmet Akçakaya. Zero-shot self-supervised learning for mri reconstruction, 2023. URL `https://arxiv.org/abs/2102.07737`.

[23] Mohammad Zalbagi Darestani and Reinhard Heckel. Accelerated MRI with untrained neural networks. *IEEE Transactions on Computational Imaging*, 7:724–733, 2021.

[24] Marc Vornehm, Chong Chen, Muhammad A Sultan, Syed M Arshad, Yuchi Han, Florian Knoll, and Rizwan Ahmad. Multi-dynamic deep image prior for cardiac MRI. *Magnetic Resonance in Medicine*, 94(6):2668–2679, 2025.

[25] Kalina P Slavkova, Julie C DiCarlo, Viraj Wadhwa, Sidharth Kumar, Chengyue Wu, John Virostko, Thomas E Yankeelov, and Jonathan I Tamir. An untrained deep learning method for reconstructing dynamic MR images from accelerated model-based data. *Magnetic Resonance in Medicine*, 89(4):1617–1633, 2023.

[26] Jie Feng, Ruimin Feng, Qing Wu, Xin Shen, Lixuan Chen, Xin Li, Li Feng, Jingjia Chen, Zhiyong Zhang, Chunlei Liu, et al. Spatiotemporal implicit neural representation for unsupervised dynamic MRI reconstruction. *IEEE Transactions on Medical Imaging*, 2025.

[27] Xuanyu Tian, Lixuan Chen, Qing Wu, Xiao Wang, Jie Feng, Yuyao Zhang, and Hongjiang Wei. Unsupervised motion-compensated decomposition for cardiac mri reconstruction via neural representation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(11):9529–9537, Mar. 2026. doi: 10.1609/aaai.v40i11.37914. URL `https://ojs.aaai.org/index.php/AAAI/article/view/37914`.

[28] Xin Shen, Jie Feng, Zhenghao Li, Qing Zou, Yuyao Zhang, and Hongjiang Wei. Free-breathing dynamic mri reconstruction via joint time-dependent coil sensitivity estimation using implicit neural representation. *Medical Image Analysis*, 108:103847, 2026. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2025.103847. URL `https://www.sciencedirect.com/science/article/pii/S1361841525003937`.

[29] Hyungjin Chung, Jeongsol Kim, Michael T. Mccann, Marc L. Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems, 2024. URL `https://arxiv.org/abs/2209.14687`.

[30] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *The Eleventh International Conference on Learning Representations*, 2023. URL `https://openreview.net/forum?id=mRieQgMtNTQ`.

[31] Hyungjin Chung and Jong Chul Ye. Deep diffusion image prior for efficient ood adaptation in 3d inverse problems. In Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol, editors, *Computer Vision – ECCV 2024*, pages 432–455, Cham, 2025. Springer Nature Switzerland. ISBN 978-3-031-73226-3.

[32] Chenhe Du, Xiyue Lin, Qing Wu, Xuanyu Tian, Ying Su, Zhe Luo, Rui Zheng, Yang Chen, Hongjiang Wei, S. Kevin Zhou, Jingyi Yu, and Yuyao Zhang. Dper: Diffusion prior driven neural representation for limited angle and sparse view ct reconstruction, 2024. URL https://arxiv.org/abs/2404.17890.

[33] Jiayue Chu, Chenhe Du, Xiyue Lin, Xiaoqun Zhang, Lihui Wang, Yuyao Zhang, and Hongjiang Wei. Highly accelerated mri via implicit neural representation guided posterior sampling of diffusion models. *Medical Image Analysis*, 100:103398, 2025. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2024.103398. URL https://www.sciencedirect.com/science/article/pii/S1361841524003232.

[34] Zhuo-Xu Cui, Congcong Liu, Xiaohong Fan, Chentao Cao, Jing Cheng, Qingyong Zhu, Yuanyuan Liu, Sen Jia, Haifeng Wang, Yanjie Zhu, et al. Physics-informed DeepMRI: k-space interpolation meets heat diffusion. *IEEE Transactions on Medical Imaging*, 43(10): 3503–3520, 2024.

[35] Yujuan Lu, Xin Xie, Shaoyu Wang, and Qiegen Liu. Latent-k-space of refinement diffusion model for accelerated MRI reconstruction. *Biomedical Physics & Engineering Express*, 11(5): 055012, 2025.

[36] Chentao Cao, Zhuo-Xu Cui, Yue Wang, Shaonan Liu, Taijin Chen, Hairong Zheng, Dong Liang, and Yanjie Zhu. High-frequency space diffusion model for accelerated mri. *IEEE Transactions on Medical Imaging*, 43(5):1853–1865, 2024. doi: 10.1109/TMI.2024.3351702.

[37] Taesung Kwon and Jong Chul Ye. Solving video inverse problems using image diffusion models. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=TRWxFUzK9K.

[38] Zi Wang, Fanwen Wang, Chen Qin, Jun Lyu, Cheng Ouyang, Shuo Wang, Yan Li, Mengyao Yu, Haoyu Zhang, Kunyuan Guo, et al. CMRxRecon2024: A multimodality, multiview k-space dataset boosting universal machine learning for accelerated cardiac MRI. *Radiology: Artificial Intelligence*, 7(2):e240443, 2025.

[39] Chong Chen, Yingmin Liu, Philip Schniter, Matthew Tong, Karolina Zareba, Orlando Simonetti, Lee Potter, and Rizwan Ahmad. OCMR (v1.0)–open-access multi-coil k-space dataset for cardiovascular magnetic resonance imaging, 2020. URL https://arxiv.org/abs/2008.03410.

[40] Martin Uecker, Peng Lai, Mark J Murphy, Patrick Virtue, Michael Elad, John M Pauly, Shreyas S Vasanawala, and Michael Lustig. ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. *Magnetic Resonance in Medicine*, 71(3):990–1001, 2014.

[41] Zi Wang, Mingkai Huang, Zhang Shi, Hongjie Hu, Lan Lan, Hui Zhang, Yan Li, Xi Hu, Qing Lu, Zongming Zhu, et al. Enabling ultra-fast cardiovascular imaging across heterogeneous clinical environments with a generalist foundation model and multimodal database. *arXiv preprint arXiv:2512.21652*, 2025.

[42] Chengyan Wang, Jun Lyu, Shuo Wang, Chen Qin, Kunyuan Guo, Xinyu Zhang, Xiaotong Yu, Yan Li, Fanwen Wang, Jianhua Jin, et al. CMRxRecon: A publicly available k-space dataset and benchmark to advance deep learning for cardiac MRI. *Scientific Data*, 11(1):687, 2024.

[43] Jun Lyu, Chen Qin, Shuo Wang, Fanwen Wang, Yan Li, Zi Wang, Kunyuan Guo, Cheng Ouyang, Michael Tänzer, Meng Liu, et al. The state-of-the-art in cardiac MRI reconstruction: Results of the CMRxRecon challenge in MICCAI 2023. *Medical Image Analysis*, 101:103485, 2025.

[44] Chengyan Wang, Yan Li, Jun Lv, Jianhua Jin, Xumei Hu, Xutong Kuang, Weibo Chen, and He Wang. Recommendation for cardiac magnetic resonance imaging-based phenotypic study: imaging part. *Phenomics*, 1(4):151–170, 2021.

[45] Kang Wang, Chen Qin, Zhang Shi, Haoran Wang, Xiwen Zhang, Chen Chen, Cheng Ouyang, Chengliang Dai, Yuanhan Mo, Chenchen Dai, et al. Extreme cardiac MRI analysis under respiratory motion: Results of the CMRxMotion challenge. *Medical Image Analysis*, page 103883, 2025.

[46] Jonas Adler, Holger Kohr, and Ozan Öktem. Operator discretization library (odl), January 2017. URL `https://doi.org/10.5281/zenodo.249479`.

[47] Rizwan Ahmad, Hui Xue, Shivraman Giri, Yu Ding, Jason Craft, and Orlando P Simonetti. Variable density incoherent spatiotemporal acquisition (VISTA) for highly accelerated cardiac MRI. *Magnetic Resonance in Medicine*, 74(5):1266–1278, 2015.

[48] Li Feng. Golden-angle radial MRI: basics, advances, and applications. *Journal of Magnetic Resonance Imaging*, 56(1):45–62, 2022.

[49] Pierre-Antoine Comby, Guillaume Daval-Frérot, Caini Pan, Asma Tanabene, Léna Oudjman, Matteo Cencini, Philippe Ciuciu, and Chaithya GR. MRI-NUFFT: Doing non-cartesian MRI has never been easier. *Journal of Open Source Software*, 10(108):7743, 2025.

[50] Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, et al. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*, 30(9):1323–1341, 2012.

Appendix Outline

Appendix A: Detailed Experimental Setup
- **A.1** Dataset Specifications
- **A.2** Retrospective Undersampling Protocols

Appendix B: Pretrained Diffusion Model

Appendix C: Framework

Appendix D: Supplementary Experiment results
- **D.1** Linear Fusion Baseline
- **D.2** Runtime and Memory Overhead
- **D.3** Statistical Variability of Main Results
- **D.4** Radiologist Evaluation Protocol
- **D.5** Local TD-DIP Measurement Conditioning
- **D.6** Robustness to Prior Mismatch

Appendix E: Detailed Evaluation Protocols and Results about OCMR dataset
- **E.1** Quantitative Reference-Free Metrics
- **E.2** Exploratory Atrial Area Analysis

Appendix F: Broader Impacts

Appendix G: Limitations

Appendix H: Theoretical Analysis
- **H.1** Stage-B as Image-Set-Restricted Surrogate Minimization
- **H.2** Local Measurement Conditioning and Recovery
- **H.3** Diffusion-Operator Mismatch and Gate Attenuation
- **H.4** Multi-step error propagation through the sampler

Appendix I: Proofs for Theoretical Analysis
- **I.1** Proof of Proposition H.1
- **I.2** Proof of Theorem H.3
- **I.3** Proof of Proposition H.4
- **I.4** Proof of Corollary H.5
- **I.5** Proof of Proposition H.7
- **I.6** Proof of Proposition H.8
- **I.7** Proof of Proposition H.9
- **I.8** Derivation of Eq. (34)
- **I.9** Derivation of Eq. (35)
- **I.10** Proof of Theorem H.11

A Detailed Experimental Setup

A.1 Dataset Specifications

A.1.1 Cardiac Cine Dataset (In-house)

We utilized a proprietary fully sampled cardiac cine dataset acquired from 29 healthy subjects using a 3.0 T scanner (MAGNETOM Trio, Siemens Healthcare, Erlangen, Germany) with a 20-channel receiver coil array. The data was retrospectively ECG-gated and acquired under breath-hold conditions. The study adhered to IRB protocols with informed consent.

- **Acquisition Protocol:** Retrospective ECG-gated segmented bSSFP sequence.
- **Dimensions:** Field of view (FOV) 330×330 mm, matrix size 256×256, slice thickness 6 mm.
- **Timing Parameters:** TR/TE = 3.0/1.5 ms; temporal resolution 40.0 ms (≈ 25 phases per cycle).
- **Preprocessing & Augmentation:** Coil sensitivity maps were estimated using the ESPIRiT algorithm [40]. To expand the dataset scale, we implemented a sliding window data augmentation strategy. A 3D bounding box of size $192 \times 192 \times 18$ ($x \times y \times t$) was moved with strides of 25, 25, and 7 along the spatial and temporal axes, respectively, generating **800 samples for training and 118 samples for testing**.

A.1.2 CMRxRecon2025 Subset

To benchmark robustness against diverse scanning conditions, we selected a subset from the CMRxRecon2025 challenge [41, 42, 38, 43–45]. We specifically focused on data acquired on a 3.0 T UIH scanner to evaluate generalization to unseen scanner vendors.

- **Subject Details:** 39 samples from 5 volunteers.
- **Dimensions:** FOV 720×320 mm, reconstruction matrix 336×298, slice thickness 8 mm.
- **Timing Parameters:** TR/TE = 2.84/1.30 ms.
- **Hardware:** 28-channel receiver coil array. The data contains 12 temporal cardiac phases with a readout oversampling factor of 2.

A.1.3 OCMR Dataset

The OCMR dataset [39] offers high heterogeneity, acquired from three Siemens scanners (Prisma 3T, Avanto 1.5T, and Sola 1.5T).

- **Fully Sampled Set:** 53 cine series (81 slices) used for quantitative benchmarking against reference standards.
- **Prospective Undersampled Set:** 212 real-time cine series (842 slices). Unlike retrospective data, these acquisitions were performed under free-breathing conditions, capturing realistic physiological noise and acceleration artifacts. This subset allows for qualitative assessment of reconstruction performance in real-world scenarios.

A.1.4 Simulated Dynamic Ellipses Dataset

To rigorously validate the "Weak Diffusion" hypothesis, we constructed a simulated dynamic MRI dataset using the Operator Discretization Library (ODL) [46]. The simulation pipeline mimics the physics of multi-coil dynamic acquisition:

Phantom Generation: Each dynamic sequence consists of $N_t = 18$ temporal frames with a spatial resolution of 192×192 pixels. The geometric parameters—including the number of ellipses, their positions, major/minor axes, rotation angles, and intensity values—were sampled randomly for each frame to simulate rapid morphological changes and complex dynamics. All images were normalized to the range $[0, 1]$.

Physics Simulation: To simulate parallel imaging, we incorporated coil sensitivity maps (CSMs). We modeled a scanner with $N_c = 20$ receiver coils. The sensitivity profile for each coil was generated using a Gaussian distribution centered at uniform angular intervals around the field of view, modulated by a smooth phase variation to reflect realistic electromagnetic properties.

K-space Generation: Let $\mathbf{x}_t$ denote the ground truth image at frame t, and $\mathbf{S}_i$ denote the complex sensitivity map of the i-th coil. The corresponding k-space measurement $\mathbf{y}_{t,i}$ was generated via the forward model:

$$\mathbf{y}_{t,i} = \mathcal{F}(\mathbf{S}_i \cdot \mathbf{x}_t), \quad i = 1, \ldots, N_c, \quad t = 1, \ldots, N_t \tag{16}$$

where $\mathcal{F}$ represents the centered 2D Fast Fourier Transform (FFT). The final dataset comprises 200×18 dynamic samples with data dimensions of $(N_c \times H \times W) = (20 \times 192 \times 192)$.

A.2 Retrospective Undersampling Protocols

To systematically evaluate the robustness and generalization capability of our framework, we applied retrospective undersampling masks to the fully sampled datasets (In-house and CMRxRecon) using both Cartesian and non-Cartesian trajectories.

Cartesian Sampling: For Cartesian experiments, we maintained full sampling along the frequency encoding direction (k_x) while undersampling the phase encoding direction (k_y) for each temporal frame. To verify performance across different sampling distributions, we employed two distinct masking strategies: (1) standard random Cartesian masks [12], and (2) Variable-Density Incoherent Spatial-Temporal Acquisition (VISTA) masks [47], which provide quasi-random spatiotemporal sampling patterns.

Non-Cartesian (Radial) Sampling: To demonstrate the versatility of our method beyond rectilinear grids, we implemented a 2D golden-angle radial acquisition scheme [48]. The radial k-space data were synthesized by transforming the fully sampled image series into the frequency domain using the Multi-coil Non-Uniform Fast Fourier Transform (NUFFT) operator [49]. The trajectory angles were determined based on the Fibonacci sequence. We evaluated the method under three aggressive acceleration regimes, corresponding to 13, 8, and 5 spokes per frame, which yield effective acceleration factors (AF) of approximately $14.8\times$, $24\times$, and $38.4\times$, respectively.

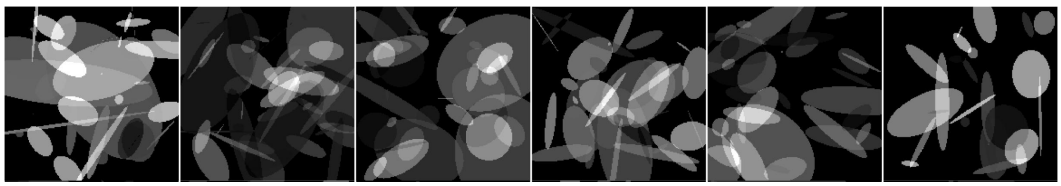


Figure 4: **Representative sample from the simulated dynamic Ellipses dataset.** The images were generated using the ODL library, where geometric parameters (position, orientation, and axes) were randomized for each frame to simulate complex dynamic changes.

B Pretrained Diffusion Model

Model and training setup. We adopt a variance-exploding (VE) score-based diffusion model [3] with an NCSN++ backbone [3]. The VE noise schedule is parameterized by $(\epsilon_1, \epsilon_L, L) = (0.01,\ 378,\ 1000)$, and we maintain an exponential moving average (EMA) of model weights with decay 0.999. Pretraining is performed with Adam (learning rate 2×10^{-4}) for 100 epochs on an in-house *multi-anatomy* MR dataset. Importantly, this pretraining corpus contains **no cardiac cine or cardiac mapping sequences**; it is deliberately constructed to be broad rather than task-matched.

Why this prior is intentionally "weak" for dynamic cardiac MRI. Our goal is to study the regime where diffusion provides a *generic* image prior that is not specialized to the downstream acquisition protocol. By pretraining on diverse anatomies/contrasts and evaluating on in-house cardiac cine and external cardiac datasets, we induce a realistic *task shift* (domain mismatch) that weakens the direct anatomical specificity of the diffusion prior. This setting aligns with the paper theme *When Diffusion Is Weak*: reconstruction quality should not hinge on having a highly matched diffusion model, but instead be stabilized by scan-specific geometric constraints (TD-DIP) and our geometry-gated coupling.

Table 4: **Anatomical composition of the in-house MR pretraining dataset.** The diffusion prior is pretrained on multi-anatomy data and *does not include cardiac cine/mapping*, intentionally creating a task shift for dynamic cardiac MRI experiments.

Region	Knee	Head	Spine	Hip	Ankle	Pelvis	Brachial-plexus	Shoulder	Thigh	Abdomen	Wrist	Sacroiliac	Calves
#Slices	25,690	12,570	10,892	8,870	7,051	3,985	1,590	1,471	669	530	445	306	156

Total: 74,225 slices (sum over regions). **Volunteers:** 2,205. **Contrasts:** T_1, T_2, PD, STIR. **Field strengths:** 1.5T/3T/5T.

Pretraining data. The pretraining dataset consists of 74,225 slices from 2,205 volunteers, covering four contrasts (T_1, T_2, PD, STIR), 13 anatomical regions, and three field strengths (1.5T/3T/5T). The anatomical composition is summarized in Table 4.

C Framework Details

Figure 5 provides a detailed schematic overview of the proposed Geometry-Gated TD-DIP framework.

Pre-trained Diffusion Prior. We utilize a variance-exploding (VE) score-based diffusion model with an NCSN++ backbone. The noise schedule is parameterized by a geometric sequence with $\sigma_{\min} = 0.01$ and $\sigma_{\max} = 378$. This model is pre-trained on a generic multi-anatomy dataset and remains frozen during the reconstruction process.

TD-DIP Generator. The scan-specific Time-Dependent Deep Image Prior (TD-DIP) is implemented as a generator network $\mathcal{G}_\phi$, consisting of: (1) A *Mapping Network* (MLP) with 3 layers and a hidden dimension of 512, mapping the temporal coordinate z_t to a style vector; and (2) A *Convolutional Decoder* with 10 layers and 128 channels, ensuring sufficient capacity to capture dynamic details. The network parameters ϕ are optimized using the Adam optimizer with a learning rate of 1×10^{-4}.

Hyperparameters. We use a single configuration across the reported reconstruction experiments. The configuration covers Cardiac Cine, CMR, and OCMR, Cartesian and radial sampling, and acceleration factors from $8\times$ to $38.4\times$.

Gate parameters. The parameters control the residual-weight range, operating point, and transition slopes:

- **Conservative Injection Bound ($\lambda_{\max} = 0.5$):** We set the maximum residual weight to 0.5, limiting residual injection while retaining dependence on the denoiser sensitivity operator.
- **Decision Boundaries ($\tau_m = 0.4, \tau_a = 0.0$):** These thresholds set the operating point for the residual fraction and directional alignment.
- **Switching Sensitivity ($\beta = 5.0, \gamma = 2.0$):** These slope parameters control how rapidly the residual weight changes with residual fraction and directional alignment.

Adaptive residual weighting. The gate adapts the residual weight to the correction's residual fraction and directional alignment at each reverse step.

Computing Infrastructure. All experiments were implemented in PyTorch. Unless otherwise noted, training and reconstruction runs were conducted on a single NVIDIA RTX 3090 GPU. For the dedicated runtime and peak-memory benchmark in Appendix D.2, we re-ran the main competing methods on a single NVIDIA A800 GPU under the same golden-angle radial NUFFT setup to enable a controlled comparison.

D Supplementary Experiment Results

D.1 Linear Fusion Baseline

To test whether diffusion and TD-DIP can be combined by naive ensembling, we include a **Linear Fusion** baseline obtained by replacing the fusion step in Algorithm 1 with

$$\tilde{\mathbf{x}}_{0,k} \triangleq 0.5\hat{\mathbf{x}}_{0,k} + 0.5\mathbf{x}_k^{\mathrm{DIP}}.$$

This baseline directly averages the diffusion and TD-DIP estimates without geometric selection, and serves as a control for whether additive coupling alone can explain the gains of the proposed method.

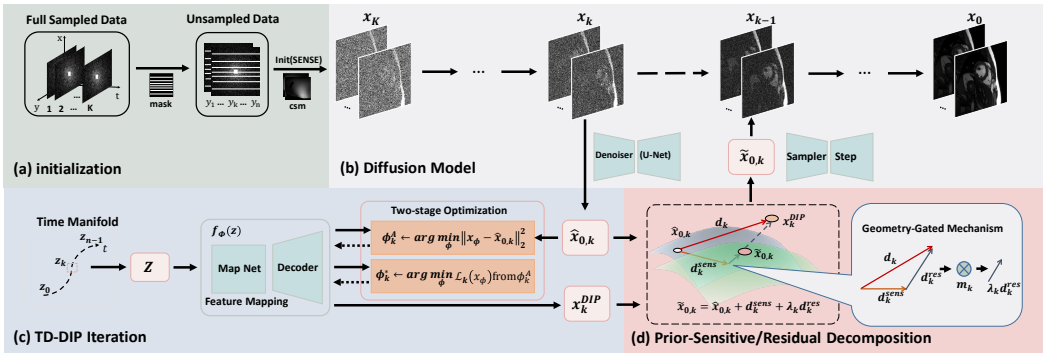


Figure 5: **Schematic overview of the proposed Geometry-Gated TD-DIP framework.** The reconstruction proceeds through an iterative reverse diffusion process coupled with scan-specific optimization. **(a) Initialization:** The process starts with undersampled k-space data and an initial ESPIRiT[40] reconstruction. **(b) Diffusion Model:** At each timestep k, the pre-trained (weak) diffusion model predicts a clean estimate $\hat{x}_{0,k}$ from the noisy state x_k, acting as a local *geometric anchor*. **(c) TD-DIP Iteration:** A Time-Dependent Deep Image Prior (TD-DIP) network $f_\phi(z)$ is optimized per-step using a *Two-stage Optimization* strategy: first aligning with the diffusion anchor (Stage 1), then minimizing $\mathcal{L}_k$ (Stage 2) to yield the scan-specific estimate x_k^{DIP}. **(d) Prior-Sensitive/Residual Split:** The correction d_k is split into $d_k^{\text{sens}} = J_{\theta,k} d_k$ and $d_k^{\text{res}} = (I - J_{\theta,k}) d_k$. A self-calibrated gate λ_k, driven by the residual fraction and directional alignment, forms $\tilde{x}_{0,k} = \hat{x}_{0,k} + d_k^{\text{sens}} + \lambda_k d_k^{\text{res}}$.

Table 5: **Runtime-memory trade-off under golden-angle radial sampling (AF $38.4\times$) on the cardiac cine dataset.** All methods are benchmarked on a single NVIDIA A800 GPU with the same NUFFT-based forward model. Quarter-step guidance reduces runtime to 472 s per sample, nearly halving the time of HashINR_tTVLR, while incurring only a 1.6 dB PSNR drop relative to the full-step variant.

Method	Time (s) ↓	Peak Alloc (MiB) ↓	PSNR (dB) ↑
TD-DIP	1071.981	**1911**	41.523
HashINR_tTVLR	890.854	19377	42.214
Ours (Full steps)	1336.922	6183	**43.469**
Ours (Half steps)	758.625	6183	42.959
Ours (Quarter steps)	**472.228**	6183	41.879

D.2 Runtime and Memory Overhead

We report single-sample reconstruction time and peak GPU memory under the same golden-angle radial NUFFT setup on the in-house cardiac cine dataset. All methods in this benchmark were re-evaluated on a single NVIDIA A800 GPU, and the "HashINR_tTVLR" entry corresponds to the strongest HashINR baseline used throughout the paper. Table 5 compares the full-step solver with reduced-step variants that lower the number of TD-DIP inner iterations to trade a small drop in reconstruction quality for substantial wall-clock savings.

As shown in Table 5, the full-step version prioritizes quality, whereas the reduced-step variants substantially improve efficiency without changing the memory footprint. In particular, the quarter-step configuration is about $1.9\times$ faster than HashINR_tTVLR and uses roughly one-third of its peak memory, while remaining competitive in reconstruction quality. This supports our claim that the proposed coupling can be made practically efficient once the TD-DIP refinement budget is trimmed.

D.3 Statistical Variability of Main Results

Reporting protocol. To keep the main quantitative comparison table compact, Table 6 reports the standard deviation associated with each PSNR/SSIM entry in Table 1. Each cell is formatted as

Table 6: **Standard deviations corresponding to the main quantitative results in Table 1.** Each entry is reported as $\pm\sigma_{\mathrm{PSNR}}/\pm\sigma_{\mathrm{SSIM}}$, where SSIM is scaled by 10^2 for readability.

Pattern	Acc	Baseline[1]	TD-DIP	HashINR +tTVLR	MoDL	SVI-Diffusion	Linear Fusion	Ours
		±PSNR / ±SSIM[2]	±PSNR / ±SSIM	±PSNR / ±SSIM	±PSNR / ±SSIM	±PSNR / ±SSIM	±PSNR / ±SSIM	±PSNR / ±SSIM
Random Cartesian	8×	±1.642 / ±1.08	±1.673 / ±0.67	±1.976 / ±0.49	±1.490 / ±0.87	±1.445 / ±2.79	±1.146 / ±0.89	±1.641 / ±0.55
	10×	±1.858 / ±1.09	±1.538 / ±0.73	±1.880 / ±0.66	±1.915 / ±0.96	±1.429 / ±2.68	±1.128 / ±1.05	±1.523 / ±0.60
	12×	±1.688 / ±1.18	±1.712 / ±0.86	±1.899 / ±0.72	±1.987 / ±1.12	±1.258 / ±2.49	±1.055 / ±1.21	±1.638 / ±0.71
VISTA	12×	±2.156 / ±2.48	±2.246 / ±0.88	±2.126 / ±0.35	±1.639 / ±0.76	±1.454 / ±2.62	±1.232 / ±1.11	±1.924 / ±0.68
	16×	±1.896 / ±3.53	±2.082 / ±0.97	±1.838 / ±0.70	±1.918 / ±1.23	±1.215 / ±2.77	±1.138 / ±1.49	±1.907 / ±0.83
	20×	±1.604 / ±3.71	±2.134 / ±1.17	±1.686 / ±0.88	±1.624 / ±1.48	±1.274 / ±2.80	±1.082 / ±1.90	±1.873 / ±0.94
	24×	±1.503 / ±3.83	±2.018 / ±1.20	±1.548 / ±1.20	±1.665 / ±1.75	±1.280 / ±2.62	±1.187 / ±3.13	±1.848 / ±1.01
Golden Radial	14.8×	±0.786 / ±3.10	±1.442 / ±0.44	±1.171 / ±0.32	±1.082 / ±1.67	±1.264 / ±1.45	±1.070 / ±1.25	±1.579 / ±1.51
	24×	±0.722 / ±2.93	±1.304 / ±0.60	±1.259 / ±0.45	±1.268 / ±2.20	±1.328 / ±1.60	±1.068 / ±1.26	±1.623 / ±1.54
	38.4×	±0.666 / ±2.97	±1.353 / ±0.78	±1.249 / ±0.62	±1.313 / ±2.97	±1.329 / ±1.73	±1.008 / ±1.27	±1.663 / ±1.58

[1] Baseline represents **L+S** for Random/VISTA, and **NUFFT** for Radial masks.
[2] SSIM values are scaled by 10^2 for readability.

Table 7: Restricted conditioning on sampled local TD-DIP image subspaces. Brackets report interquartile ranges.

AF	Stage	$\sigma_{\min}(\mathcal{A}\mathbf{Q}_k)$ [IQR]	Condition number [IQR]
12×	Early	0.282 [0.204, 0.300]	1.28 [1.21, 1.75]
12×	Mid	0.261 [0.158, 0.290]	1.41 [1.26, 2.38]
12×	Late	0.259 [0.130, 0.285]	1.41 [1.27, 3.21]
24×	Early	0.207 [0.162, 0.221]	1.33 [1.25, 1.67]
24×	Mid	0.191 [0.112, 0.212]	1.52 [1.34, 2.68]
24×	Late	0.190 [0.098, 0.214]	1.51 [1.34, 3.26]

$\pm\sigma_{\mathrm{PSNR}}/\pm\sigma_{\mathrm{SSIM}}$, with SSIM scaled by 10^2 to match the notation in the main paper. The statistics are computed over the corresponding evaluation set for each undersampling pattern and acceleration factor. The in-house standard deviations describe variability across augmented cine crops.

D.4 Radiologist Evaluation Protocol

Two radiologists with more than three years of experience independently scored seven anonymized reconstruction methods on five randomly selected VISTA $24\times$ cases. They rated overall image quality, structural detail, temporal consistency, and artifact suppression on a five-point scale. Reader scores were averaged within each case; Table 2 reports the mean and standard deviation across cases and a Friedman test across paired methods.

Reader-averaged results are reported in Sec. 5.3 and Table 2.

D.5 Local TD-DIP Measurement Conditioning

We evaluated local measurement gain at early, middle, and late reconstruction stages under VISTA $12\times$ and $24\times$. At each state, 32 parameter directions were mapped to image variations using TD-DIP Jacobian–vector products and orthonormalized to form $\mathbf{Q}_k$, with $\mathbf{Q}_k^{\mathsf{H}}\mathbf{Q}_k = \mathbf{I}$. We computed the smallest singular value and condition number of $\mathcal{A}\mathbf{Q}_k$.

The sampled subspaces have positive restricted gains throughout the tested trajectories, with smaller gains at $24\times$. These measurements characterize conditioning on sampled local TD-DIP subspaces along the tested trajectories.

D.6 Robustness to Prior Mismatch

Detailed setting. We compare three diffusion priors with increasing domain mismatch: a **Matched Prior** trained on 800 augmented in-house cine training samples, a **Synthetic Geometric Prior** trained on 200 moving-ellipse sequences, and a generic **Pre-trained Prior**. The goal is to test whether reconstruction quality remains stable when the diffusion model provides only coarse geometric information rather than anatomy-specific knowledge.

Additional qualitative discussion. As visualized in Fig. 3, the reconstruction results remain highly consistent across all three priors in the illustrated example. The illustrative performance gap between the *Matched Prior* (48.22 dB) and the *Synthetic Geometric Prior* (47.91 dB) is 0.31 dB, and the

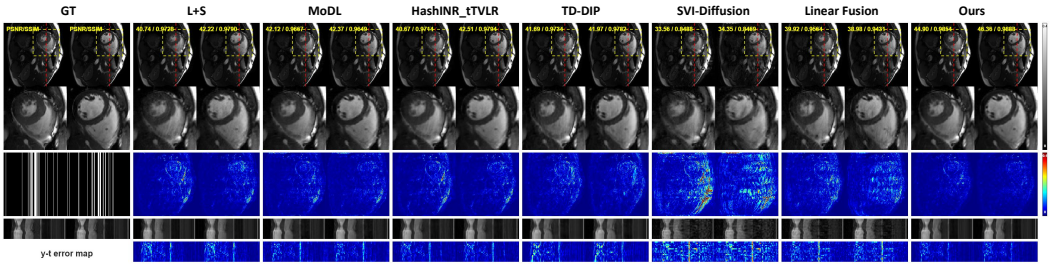


Figure 6: **Visual comparison of reconstruction performance on the Cardiac Cine dataset (Cartesian Undersampling).** The results were evaluated in the undersampling acceleration setting of Random Cartesian $12\times$: during systole (left) and diastole (right). The layout is organized as follows: The **first row** displays the full Field of View (FOV) reconstructions with PSNR/SSIM metrics annotated; The **second row** provides magnified Regions of Interest (ROIs) focusing on the cardiac structures; The **third row** shows the sampling mask (leftmost) and the corresponding spatial error maps for each method; The **fourth and fifth rows** visualize the temporal evolution ($y - t$ profiles) and their corresponding temporal error maps, respectively.

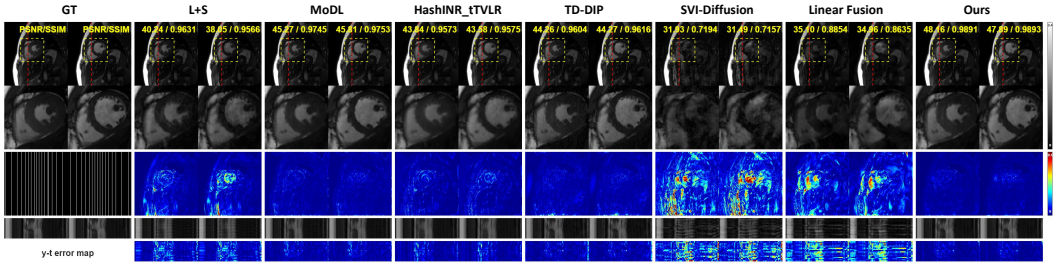


Figure 7: **Visual comparison of reconstruction performance on the Cardiac Cine dataset (VISTA Undersampling).** The results were evaluated in the undersampling acceleration setting of VISTA $12\times$: during systole (left) and diastole (right). The layout is organized as follows: The **first row** displays the full Field of View (FOV) reconstructions with PSNR/SSIM metrics annotated; The **second row** provides magnified Regions of Interest (ROIs) focusing on the cardiac structures; The **third row** shows the sampling mask (leftmost) and the corresponding spatial error maps for each method; The **fourth and fifth rows** visualize the temporal evolution ($y - t$ profiles) and their corresponding temporal error maps, respectively.

model still recovers anatomical structures such as the papillary muscles and right ventricular wall despite their absence from the ellipse training data. The illustrated comparison shows that detailed reconstruction can be maintained with an anatomy-mismatched diffusion prior.

Figures 6 and 7 present qualitative comparisons on the Cardiac Cine dataset under Random Cartesian $12\times$ and VISTA $12\times$ undersampling, respectively. As shown in the results, the conventional L+S method suffers from severe residual aliasing artifacts, particularly evident in the VISTA reconstruction. The deep learning baselines, MoDL and HashINR, tend to produce overly smoothed textures, leading to a loss of fine anatomical details. While TD-DIP captures temporal dynamics, it exhibits slight overfitting to high-frequency noise and undersampling artifacts. Notably, SVI-Diffusion performs poorly due to the lack of temporal constraints, showing heavy artifacts under the Cartesian mask and failing to reconstruct meaningful structures under the VISTA mask. The Linear Fusion baseline also yields results significantly inferior to ours, confirming that naive averaging is insufficient. Our reconstruction combines artifact suppression with visible structural detail, while the $y - t$ profiles show temporal continuity in the illustrated examples.

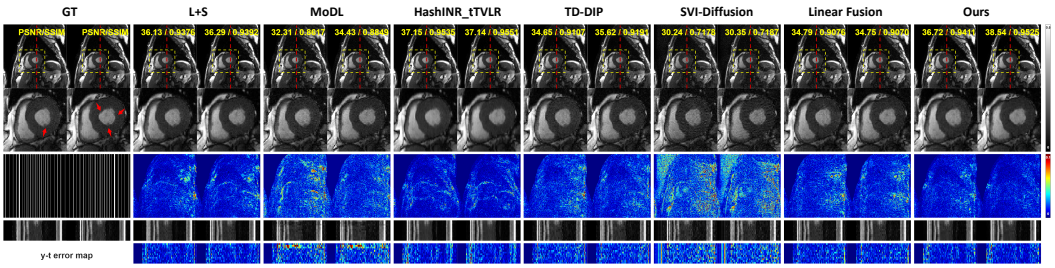


Figure 8: **Reconstruction results on the cross-vendor CMR dataset (Generalization Test).** The results were evaluated in the undersampling acceleration setting of VISTA 8×. The layout is organized as follows: The **first row** displays the full Field of View (FOV) reconstructions with PSNR/SSIM metrics annotated; The **second row** provides magnified Regions of Interest (ROIs) focusing on the cardiac structures; The **third row** shows the sampling mask (leftmost) and the corresponding spatial error maps for each method; The **fourth and fifth rows** visualize the temporal evolution ($y-t$ profiles) and their corresponding temporal error maps, respectively. The red arrows highlight structural details preserved by our reconstruction in the illustrated regions.

Table 8: **Reconstruction performance (PSNR/SSIM) of CMR. The second-best results are underlined.**

Pattern	Acc	L+S	TD-DIP	HashINR +tTVLR	MoDL	SVI-Diffusion	Linear Fusion	Ours
		PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM	PSNR / SSIM
VISTA	8×	41.755 / 96.92	39.540 / 96.22	41.761 / **97.69**	35.198 / 91.84	29.987 / 47.92	40.848 / 96.63	**41.917** / 97.49

E Detailed Evaluation Protocols and Results about OCMR dataset

We evaluate prospective OCMR reconstructions using measurement consistency, temporal coherence, and exploratory atrial area analysis, since fully sampled ground truth is unavailable.

E.1 Quantitative Reference-Free Metrics

We employed three specific indicators to objectively quantify reconstruction quality. The quantitative results comparing our method against baselines are summarized in Table 9.

1. Self-Supervised via Data Undersampling (SSDU) Score Following the principle of self-supervision [21], we partition the acquired non-Cartesian k-space indices Ω into two disjoint sets: a training set Ω_{dc} used for data consistency during reconstruction, and a validation set Ω_{val} used strictly for evaluation. To ensure a robust evaluation, we adopted a stratified partitioning strategy:

- **ACS Region:** All auto-calibration signal (ACS) lines (central low-frequency k-space) were assigned to Ω_{dc} to maintain coil sensitivity estimation stability.
- **Outer k-space:** For high-frequency peripheral lines, we randomly selected 20% as Ω_{val}, with the remaining assigned to Ω_{dc}. This partitioning was kept consistent across all temporal frames to avoid sampling bias.

The SSDU score quantifies the normalized mean squared error (NMSE) on the unseen validation indices:

$$\mathcal{S}_{SSDU} = \frac{\|\mathcal{P}_{\Omega_{val}}(\mathcal{A}(\hat{\mathbf{x}}) - \mathbf{y})\|_2^2}{\|\mathcal{P}_{\Omega_{val}}(\mathbf{y})\|_2^2}, \tag{17}$$

where $\hat{x}$ is the reconstructed image series, y is the acquired multi-coil k-space, $\mathcal{A}$ is the forward operator (involving coil sensitivities and Non-Uniform FFT), and $\mathcal{P}_{\Omega_{val}}$ is the projection operator selecting indices in Ω_{val}.

2. Temporal Consistency Metrics Dynamic MRI requires preserving smooth physiological motion while suppressing flickering artifacts. We employed two metrics to quantify this property:

Table 9: Quantitative comparison of different reconstruction methods on the prospective OCMR dataset. The best results are highlighted in **bold**. Second-best are underlined. All metrics are lower-is-better (↓). Note that TD-DIP achieves the lowest T2 energy due to its strong implicit temporal smoothing, but this comes at the cost of detail loss (see qualitative results).

Method	**SSDU (ratio_all mean)** ($\times 10^{-1}$)	**T2 Energy** (E_{t2}) ($\times 10^{-4}$)	**MC Consistency** (E_{mc}) ($\times 10^{-5}$)
Linear Fusion	4.440	15.605	5.153
L+S	**1.388**	13.049	4.483
MoDL	1.662	4.918	3.353
SVI-Diffusion	12.018	9.219	4.723
HashINR_tTVLR	1.551	3.115	2.145
TD-DIP	1.515	**0.729**	1.232
Ours	1.407	0.853	**1.215**

(a) Second-Order Temporal Difference Energy (E_{t2}): This metric measures the smoothness of the temporal trajectory and is highly sensitive to non-smooth artifacts such as flashing or jittering. It is defined as the energy of the second derivative of the image sequence along the time axis:

$$E_{t2} = \frac{1}{N_t} \sum_t \|\hat{\mathbf{x}}_{t+1} - 2\hat{\mathbf{x}}_t + \hat{\mathbf{x}}_{t-1}\|_2^2 , \tag{18}$$

where N_t is the number of frames. A lower E_{t2} indicates smoother temporal transitions.

(b) Motion Compensation Consistency (E_{mc}): To distinguish between valid physiological motion and artifactual changes, we evaluated the consistency between a frame and its motion-compensated neighbor. We utilized an optical flow algorithm to estimate the dense flow field $\mathcal{W}_{t \leftarrow t+1}$ between adjacent frames. The consistency error is calculated as:

$$E_{mc} = \frac{1}{N_t - 1} \sum_t \|\hat{\mathbf{x}}_t - \mathcal{W}_{t \leftarrow t+1}(\hat{\mathbf{x}}_{t+1})\|_1 , \tag{19}$$

where $\mathcal{W}$ denotes the warping operation based on the estimated flow. A lower E_{mc} suggests that temporal variations are physically explainable by motion rather than random noise.

E.2 Exploratory Atrial Area Analysis

To characterize temporal variation in reconstructed atrial boundaries, segmentation of the Left Atrium (LA) and Right Atrium (RA) was performed using **3D Slicer** [50] software. Based on the time-resolved 2D area curves $A(t)$, we calculated:

- **Maximum atrial area ($A_{\max}$):** The largest segmented 2D atrial area within the observed sequence.
- **Minimum atrial area ($A_{\min}$):** The smallest segmented 2D atrial area within the observed sequence.
- **Area-change fraction (F_A):** A descriptive area-based measure, calculated as

$$F_A = \frac{A_{\max} - A_{\min}}{A_{\max}} \times 100\%. \tag{20}$$

  It is not equivalent to a volumetric ejection fraction.
- **Temporal Area Curves:** The segmented area $A(t)$ was plotted over time to inspect the continuity and timing of the estimated atrial motion.

The segmentation results and the corresponding area-based metric computations are shown in Fig. 10 and Table 10. In the illustrated sequence, our reconstruction yields visually cleaner atrial boundaries, smoother area–time curves, and a left atrial area-change fraction of 59.18%. Independent reference segmentations would be needed to assess quantitative accuracy.

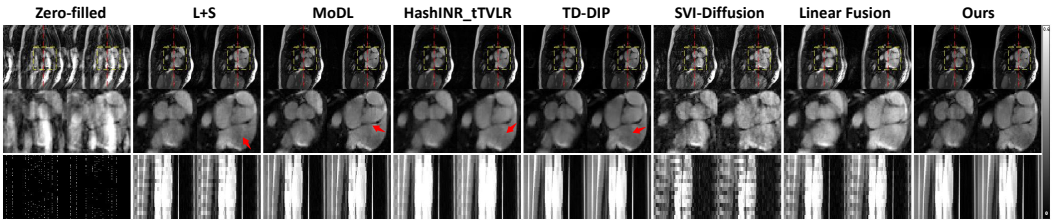


Figure 9: **Visual comparison on prospective OCMR data.** Representative reconstructions from a free-breathing acquisition. The proposed Geometry-Gated TD-DIP (rightmost) suppresses background noise and aliasing while preserving sharp anatomical boundaries in the zoomed-in ROI, compared with baselines that exhibit residual artifacts (MoDL, Linear Fusion) or over-smoothing (TD-DIP). The zero-filled reconstruction reveals residual aliasing in the region marked by the red arrow.

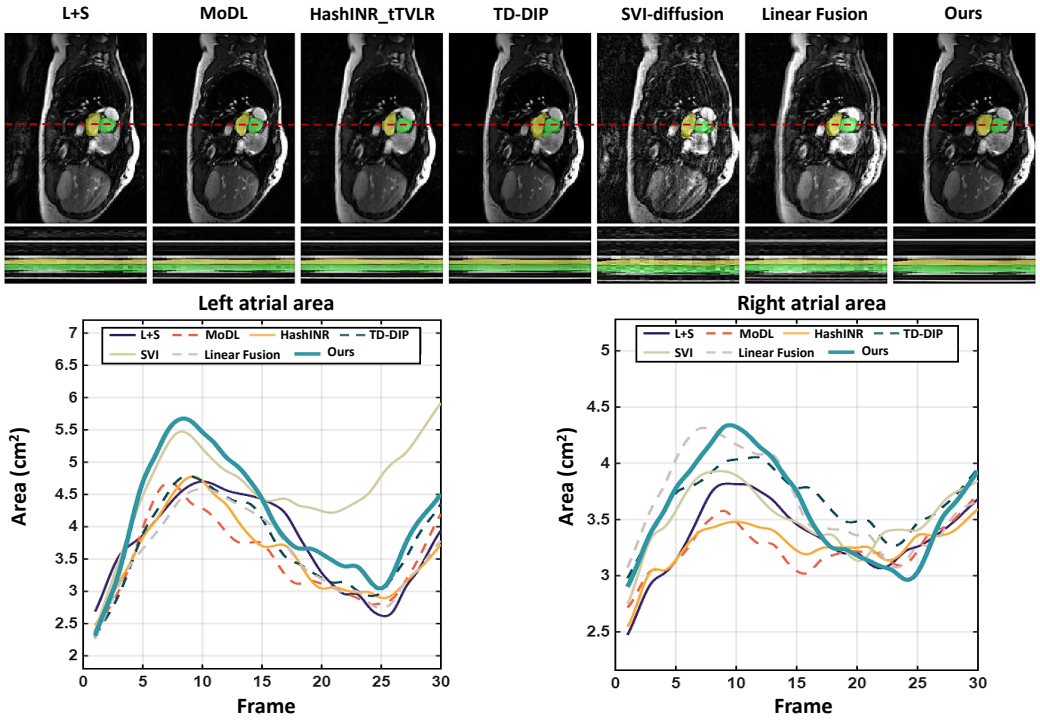


Figure 10: **Atrial segmentations (top row) and corresponding area–time curves (bottom row).** In this example, our reconstruction yields cleaner boundaries and smoother curves with less frame-to-frame jitter.

F Broader Impacts

On the positive side, more robust accelerated MRI reconstruction can improve scan efficiency and reduce the burden of acquiring fully sampled dynamic data. On the negative side, any reconstruction method may still introduce artifacts or suppress subtle structures under severe undersampling, which could propagate misleading visual cues if outputs are used without appropriate human oversight.

Table 10: Exploratory 2D atrial area-change fractions across reconstruction methods.

Metric	L+S	MoDL	HashINR	TD-DIP	SVI-Diffusion	LF	Ours
LA F_A (%)	44.34	49.23	48.21	52.18	59.52	50.68	59.18
RA F_A (%)	35.22	26.98	29.09	26.54	30.02	28.77	33.11

G Limitations

A primary limitation of our proposed framework is the computational cost associated with test-time optimization. In the matched runtime benchmark, the full method takes 1336.922 s per sample, compared with 1071.981 s for TD-DIP (Appendix D.2). While this approach is computationally more intensive than fast supervised networks like MoDL [19], high latency remains a common and inherent challenge for untrained, scan-specific methods that rely on extensive iterations to converge. This trade-off accompanies the observed robustness against distribution shifts without requiring large-scale task-matched dynamic cine training data. Future work will focus on accelerating scan-specific optimization. The reader study covers five cases and two readers; broader validation remains necessary. The prospective atrial area analysis lacks independent reference segmentations.

H Theoretical Analysis

Our analysis follows the reconstruction pipeline: local measurement conditioning yields a recovery bound for Stage-B fitting, the gate controls sensitivity to diffusion-operator mismatch, and a sampler recursion propagates per-step perturbations.

Notation. Let $\hat{\mathbf{x}}_{0,k}$ be the diffusion anchor (Eq. (2)) and $\mathcal{L}_k(\mathbf{x})$ the local energy (Eq. (5)). Let $\mathbf{x}_k^{\mathrm{DIP}} = \mathbf{x}_{\boldsymbol{\phi}_k} \in \mathcal{M}_{\mathrm{DIP}}$ be the Stage-B output (Eq. (6)), and $\mathbf{d}_k = \mathbf{x}_k^{\mathrm{DIP}} - \hat{\mathbf{x}}_{0,k}$ (Eq. (9)). Define the *gated operator*

$$\mathbf{G}_{\boldsymbol{\theta},k}(\lambda) \;\triangleq\; \lambda\,\mathbf{I} + (1-\lambda)\,\mathbf{J}_{\boldsymbol{\theta},k}, \qquad \lambda \in [0,1], \tag{21}$$

where $\mathbf{J}_{\boldsymbol{\theta},k} = \nabla_{\mathbf{x}_k} \mathcal{D}_{\boldsymbol{\theta},k}(\mathbf{x}_k)$ (Eq. (10)). Then the guided clean estimate used by the sampler (Eq. (14)) can be written compactly as

$$\tilde{\mathbf{x}}_{0,k} \;=\; \hat{\mathbf{x}}_{0,k} + \mathbf{G}_{\boldsymbol{\theta},k}(\lambda_k)\,\mathbf{d}_k \;=\; \hat{\mathbf{x}}_{0,k} + \mathbf{d}_k^{\mathrm{sens}} + \lambda_k \mathbf{d}_k^{\mathrm{res}}. \tag{22}$$

Equivalently, $\tilde{\mathbf{x}}_{0,k} = \mathbf{x}_k^{\mathrm{DIP}} - (1-\lambda_k)\mathbf{d}_k^{\mathrm{res}}$. This form connects the guided estimate to the Stage-B fit.

H.1 Stage-B as Image-Set-Restricted Surrogate Minimization

Proposition H.1 (Score-anchored local surrogate)**.** *The measurement term and quadratic anchor define the local objective $\mathcal{L}_k$. Its ideal image-set-restricted solution is*

$$\mathbf{x}_k^\star \;\in\; \arg\min_{\mathbf{x}\in\mathcal{M}_{\mathrm{DIP}}} \;\mathcal{L}_k(\mathbf{x}), \qquad \mathbf{x}_k^{\mathrm{DIP}} \approx \mathbf{x}_k^\star. \tag{23}$$

Appendix I.1 describes the construction of this chosen score-anchored surrogate.

H.2 Local Measurement Conditioning and Recovery

A weak diffusion anchor leaves recovery dependent on the measurements over the local TD-DIP image set.

Assumption H.2 (Local restricted measurement gain)**.** For a neighborhood $\mathcal{V}_k$ of the Stage-B solution parameters, define $\mathcal{M}_k = \{\mathbf{x}_{\boldsymbol{\phi}} : \boldsymbol{\phi} \in \mathcal{V}_k\}$. There exists $\alpha_k > 0$ such that

$$\|\mathcal{A}(\mathbf{u}-\mathbf{v})\|_2 \geq \alpha_k \|\mathbf{u}-\mathbf{v}\|_2, \qquad \mathbf{u},\mathbf{v} \in \mathcal{M}_k. \tag{24}$$

This condition measures local image-space identifiability. Appendix D.5 reports restricted gains along reconstruction trajectories.

Theorem H.3 (Local recovery under weak diffusion)**.** *Assume Assumption H.2. Let* $\bar{\mathbf{x}}_k, \mathbf{x}_k^{\mathrm{DIP}} \in \mathcal{M}_k$ *satisfy*

$$\mathcal{L}_k(\mathbf{x}_k^{\mathrm{DIP}}) \le \mathcal{L}_k(\bar{\mathbf{x}}_k) + \delta_k^{\mathrm{opt}}, \qquad \delta_k^{\mathrm{opt}} \ge 0.$$

Define $e_{M,k} = \|\bar{\mathbf{x}}_k - \mathbf{x}_0\|_2$, $r_{M,k} = \|\mathcal{A}(\bar{\mathbf{x}}_k) - \mathbf{y}\|_2$, *and* $e_{a,k} = \|\bar{\mathbf{x}}_k - \hat{\mathbf{x}}_{0,k}\|_2$. *Then*

$$\|\mathbf{x}_k^{\mathrm{DIP}} - \mathbf{x}_0\|_2 \le e_{M,k} + \frac{r_{M,k} + \sqrt{r_{M,k}^2 + \sigma_y^2 \mu_k e_{a,k}^2 + 2\sigma_y^2 \delta_k^{\mathrm{opt}}}}{\alpha_k}. \tag{25}$$

Proof in Appendix I.2.

The bound separates representation, measurement, anchor, and optimization errors. The anchor contribution scales with $\sqrt{\mu_k}$, while α_k determines how measurement discrepancies translate into image errors. The assumption is conditional: positive singular values on the finitely sampled local directions in Appendix D.5 do not establish the restricted gain over the full image set $\mathcal{M}_k$.

Proposition H.4 (Exact effect of residual gating)**.** *Let* $\mathbf{e}_k = \mathbf{x}_k^{\mathrm{DIP}} - \mathbf{x}_0$ *and* $\mathbf{v}_k = \mathbf{d}_k^{\mathrm{res}}$. *For any* $\lambda \in [0,1]$, *define* $\tilde{\mathbf{x}}_{0,k}(\lambda) = \mathbf{x}_k^{\mathrm{DIP}} - (1-\lambda)\mathbf{v}_k$. *Then*

$$\|\tilde{\mathbf{x}}_{0,k}(\lambda) - \mathbf{x}_0\|_2^2 - \|\mathbf{x}_k^{\mathrm{DIP}} - \mathbf{x}_0\|_2^2 = (1-\lambda)^2\|\mathbf{v}_k\|_2^2 - 2(1-\lambda)\,\mathrm{Re}\langle \mathbf{e}_k, \mathbf{v}_k\rangle. \tag{26}$$

For $\lambda < 1$, *the gate strictly improves the squared reconstruction error if and only if*

$$2\,\mathrm{Re}\langle \mathbf{e}_k, \mathbf{v}_k\rangle > (1-\lambda)\|\mathbf{v}_k\|_2^2. \tag{27}$$

When $\mathbf{v}_k \neq \mathbf{0}$, *the one-step oracle gate for the fixed Stage-B correction over* $\Lambda = [\lambda_{\min}, \lambda_{\max}] \subseteq [0,1]$ *is*

$$\lambda_k^{\mathrm{oracle}} = \mathrm{clip}_\Lambda\left(1 - \frac{\mathrm{Re}\langle \mathbf{e}_k, \mathbf{v}_k\rangle}{\|\mathbf{v}_k\|_2^2}\right). \tag{28}$$

For $\mathbf{v}_k = \mathbf{0}$, *the output is independent of* λ. *The ground-truth-dependent oracle provides a retrospective reference for the adaptive gate.*

Corollary H.5 (Local recovery bound for the gated output)**.** *Under Theorem H.3, let* B_k *denote the right-hand side of Eq.* (25). *Then*

$$\|\tilde{\mathbf{x}}_{0,k} - \mathbf{x}_0\|_2 \le B_k + (1-\lambda_k)\|\mathbf{d}_k^{\mathrm{res}}\|_2. \tag{29}$$

The bound combines Stage-B recovery error with the residual correction withheld by the gate at step k.

Proofs in Appendices I.3 and I.4.

H.3 Diffusion-Operator Mismatch and Gate Attenuation

We compare the denoiser Jacobian with a reference sensitivity operator.

Assumption H.6 (Local sensitivity mismatch)**.** At reverse step k, the denoiser Jacobian and a reference operator $\mathbf{J}_{r,k}$ satisfy

$$\|\mathbf{J}_{\boldsymbol{\theta},k} - \mathbf{J}_{r,k}\|_2 \le \varepsilon_{J,k}. \tag{30}$$

Proposition H.7 (Gate attenuates operator mismatch)**.** *Define* $\mathbf{G}_{r,k}(\lambda) = \lambda\mathbf{I} + (1-\lambda)\mathbf{J}_{r,k}$. *Under Assumption H.6, for a fixed* $\lambda \in [0,1]$ *and vector* $\mathbf{v}$,

$$\left\|\mathbf{G}_{\boldsymbol{\theta},k}(\lambda)\mathbf{v} - \mathbf{G}_{r,k}(\lambda)\mathbf{v}\right\|_2 \le (1-\lambda)\varepsilon_{J,k}\|\mathbf{v}\|_2. \tag{31}$$

For a fixed correction and gate, the direct operator-mismatch contribution is scaled by $1-\lambda$ (Appendix I.5). The next result includes changes in the anchor, fitted correction, and gate.

Proposition H.8 (One-step prior-perturbation bound)**.** *Compare two priors indexed by* $i \in \{\boldsymbol{\theta}, r\}$ *at a fixed reverse step. Let* $\mathbf{c}_i$ *be the respective anchor,* $\mathbf{d}_i$ *the Stage-B correction,* $\mathbf{J}_i$ *the denoiser Jacobian, and* $\lambda_i \in [0,1]$ *the gate. Write* $\tilde{\mathbf{x}}_i = \mathbf{c}_i + \mathbf{G}_i(\lambda_i)\mathbf{d}_i$. *Under Assumption H.6,*

$$\begin{aligned}\|\tilde{\mathbf{x}}_{\boldsymbol{\theta}} - \tilde{\mathbf{x}}_r\|_2 \le{}& \|\mathbf{c}_{\boldsymbol{\theta}} - \mathbf{c}_r\|_2 + \|\mathbf{G}_{\boldsymbol{\theta}}(\lambda_{\boldsymbol{\theta}})\|_2\|\mathbf{d}_{\boldsymbol{\theta}} - \mathbf{d}_r\|_2 \\ &+ (1-\lambda_{\boldsymbol{\theta}})\varepsilon_{J,k}\|\mathbf{d}_r\|_2 + |\lambda_{\boldsymbol{\theta}} - \lambda_r|\|(\mathbf{I} - \mathbf{J}_r)\mathbf{d}_r\|_2.\end{aligned} \tag{32}$$

The bound separates the contributions of anchor, fitted-correction, Jacobian, and gate perturbations to one-step sensitivity (Appendix I.6).

Proposition H.9 (Deviation from the full TD-DIP correction)**.** *The gated operator interpolates between the full TD-DIP correction (*$\lambda = 1$*) and its prior-sensitive part (*$\lambda = 0$*; Appendix I.7):*

$$\|\tilde{\mathbf{x}}_{0,k} - \mathbf{x}_k^{\mathrm{DIP}}\|_2 \;=\; (1-\lambda_k)\,\|\mathbf{d}_k^{\mathrm{res}}\|_2. \tag{33}$$

Effect on measurement consistency. Let $\mathbf{r}_k = \mathcal{A}(\mathbf{x}_k^{\mathrm{DIP}}) - \mathbf{y}$ and $\mathbf{w}_k = \mathcal{A}(\mathbf{d}_k^{\mathrm{res}})$. The exact identity

$$\|\mathcal{A}(\tilde{\mathbf{x}}_{0,k}) - \mathbf{y}\|_2^2 - \|\mathcal{A}(\mathbf{x}_k^{\mathrm{DIP}}) - \mathbf{y}\|_2^2 = (1-\lambda_k)^2\|\mathbf{w}_k\|_2^2 - 2(1-\lambda_k)\,\mathrm{Re}\langle \mathbf{r}_k, \mathbf{w}_k\rangle \tag{34}$$

shows how residual weighting and the alignment of $\mathbf{r}_k$ and $\mathbf{w}_k$ determine the change in measurement residual, complementing the image-error identity in Proposition H.4 (Appendix I.8).

Adaptive residual control. For the raw sigmoid gate $\hat{\lambda}_k$ in Eq. (13),

$$1 - \hat{\lambda}_k \le 1 - \lambda_{\max} + (\lambda_{\max} - \lambda_{\min}) C \exp(-\beta m_k), \tag{35}$$

where $C = \exp(\beta\tau_m + \gamma(1-\tau_a))$ (Appendix I.9). The implemented cap $\lambda_{\max} = 0.5$ retains an operator-mismatch coefficient of at least 0.5; EMA smooths the gate over reverse steps. Proposition H.4 separately characterizes the gate's effect on image error through alignment with the Stage-B error.

H.4 Multi-step error propagation through the sampler

Finally, we propagate per-step deviations through the reverse sampler.

Assumption H.10 (Sampler Lipschitzness)**.** For each k, the sampler update $\mathbf{x}_{k-1} = \mathrm{SamplerStep}(\mathbf{x}_k, \tilde{\mathbf{x}}_{0,k}, \sigma_k, \sigma_{k-1})$ is Lipschitz in $(\mathbf{x}_k, \tilde{\mathbf{x}}_{0,k})$:

$$\|\mathrm{SamplerStep}(\mathbf{x}, \mathbf{u}) - \mathrm{SamplerStep}(\mathbf{x}', \mathbf{u}')\|_2 \le \rho_k\|\mathbf{x} - \mathbf{x}'\|_2 + \tau_k\|\mathbf{u} - \mathbf{u}'\|_2. \tag{36}$$

where we omit (σ_k, σ_{k-1}) for brevity.

Theorem H.11 (Sampler error propagation)**.** *For two reverse trajectories satisfying Assumption H.10, the state discrepancy obeys*

$$\|\mathbf{x}_{k-1} - \mathbf{x}_{k-1}^\dagger\|_2 \le \rho_k\|\mathbf{x}_k - \mathbf{x}_k^\dagger\|_2 + \tau_k\|\tilde{\mathbf{x}}_{0,k} - \tilde{\mathbf{x}}_{0,k}^\dagger\|_2. \tag{37}$$

Unrolling this recursion expresses the final discrepancy as a weighted accumulation of initial and per-step discrepancies (Appendix I.10).

For stochastic samplers, this comparison is pathwise under coupled (identical) random increments. The recursion tracks the accumulation of guided-estimate perturbations along the reverse trajectory.

Consequence for weak diffusion. Local measurement gain controls Stage-B recovery, while the gate regulates dependence on the diffusion operator and deviation from the full TD-DIP correction. The experiments in Sec. 4 evaluate these mechanisms under aggressive undersampling and prior mismatch.

I Proofs for Theoretical Analysis

I.1 Derivation of Proposition H.1

The Gaussian measurement model gives a quadratic measurement term proportional to $\|\mathcal{A}(\mathbf{x}) - \mathbf{y}\|_2^2$. Adding the chosen anchor $\frac{\mu_k}{2}\|\mathbf{x} - \hat{\mathbf{x}}_{0,k}\|_2^2$ yields $\mathcal{L}_k$. Restriction to the TD-DIP image set gives Eq. (23).

I.2 Proof of Theorem H.3

From the objective comparison and the nonnegative anchor term,

$$\|\mathcal{A}(\mathbf{x}_k^{\mathrm{DIP}}) - \mathbf{y}\|_2^2 \le r_{M,k}^2 + \sigma_y^2\mu_k e_{a,k}^2 + 2\sigma_y^2\delta_k^{\mathrm{opt}}.$$

Assumption H.2 and the triangle inequality give

$$\alpha_k\|\mathbf{x}_k^{\mathrm{DIP}} - \bar{\mathbf{x}}_k\|_2 \le \|\mathcal{A}(\mathbf{x}_k^{\mathrm{DIP}} - \bar{\mathbf{x}}_k)\|_2 \le \|\mathcal{A}(\mathbf{x}_k^{\mathrm{DIP}}) - \mathbf{y}\|_2 + r_{M,k}.$$

Combining these inequalities with $\|\mathbf{x}_k^{\mathrm{DIP}} - \mathbf{x}_0\|_2 \le \|\mathbf{x}_k^{\mathrm{DIP}} - \bar{\mathbf{x}}_k\|_2 + e_{M,k}$ proves Eq. (25).

I.3 Proof of Proposition H.4

Set $t = 1 - \lambda \in [0, 1]$, $\mathbf{e} = \mathbf{x}_k^{\mathrm{DIP}} - \mathbf{x}_0$, and $\mathbf{v} = \mathbf{d}_k^{\mathrm{res}}$. Since $\tilde{\mathbf{x}}_{0,k}(\lambda) - \mathbf{x}_0 = \mathbf{e} - t\mathbf{v}$, expansion of the squared norm in the real-valued representation (or, equivalently, of the Hermitian norm for complex images) gives

$$\|\mathbf{e} - t\mathbf{v}\|_2^2 - \|\mathbf{e}\|_2^2 = t^2\|\mathbf{v}\|_2^2 - 2t\,\mathrm{Re}\langle \mathbf{e}, \mathbf{v}\rangle.$$

For $t > 0$, the difference is strictly negative precisely when $2\,\mathrm{Re}\langle \mathbf{e}, \mathbf{v}\rangle > t\|\mathbf{v}\|_2^2$. For $\mathbf{v} \neq \mathbf{0}$, minimizing this convex quadratic over $t \in [1 - \lambda_{\max}, 1 - \lambda_{\min}]$ gives $t^{\mathrm{oracle}} = \mathrm{clip}_{[1-\lambda_{\max}, 1-\lambda_{\min}]}(\mathrm{Re}\langle \mathbf{e}, \mathbf{v}\rangle / \|\mathbf{v}\|_2^2)$, equivalently Eq. (28). If $\mathbf{v} = \mathbf{0}$, the difference vanishes for every λ.

I.4 Proof of Corollary H.5

By the identity preceding Proposition H.4 and the triangle inequality,

$$\|\tilde{\mathbf{x}}_{0,k} - \mathbf{x}_0\|_2 \leq \|\mathbf{x}_k^{\mathrm{DIP}} - \mathbf{x}_0\|_2 + (1 - \lambda_k)\|\mathbf{d}_k^{\mathrm{res}}\|_2.$$

Applying Theorem H.3 to the first term establishes Eq. (29).

I.5 Proof of Proposition H.7

By definition,

$$\mathbf{G}_{\boldsymbol{\theta},k}(\lambda) - \mathbf{G}_{r,k}(\lambda) = (1 - \lambda)(\mathbf{J}_{\boldsymbol{\theta},k} - \mathbf{J}_{r,k}).$$

Thus, for any $\mathbf{v}$,

$$\|\mathbf{G}_{\boldsymbol{\theta},k}(\lambda)\mathbf{v} - \mathbf{G}_{r,k}(\lambda)\mathbf{v}\| \leq (1 - \lambda)\|\mathbf{J}_{\boldsymbol{\theta},k} - \mathbf{J}_{r,k}\|_2\|\mathbf{v}\| \leq (1 - \lambda)\varepsilon_{J,k}\|\mathbf{v}\|.$$

I.6 Proof of Proposition H.8

At the fixed reverse step, add and subtract $\mathbf{G}_{\boldsymbol{\theta}}(\lambda_{\boldsymbol{\theta}})\mathbf{d}_r$ to obtain

$$\tilde{\mathbf{x}}_{\boldsymbol{\theta}} - \tilde{\mathbf{x}}_r = (\mathbf{c}_{\boldsymbol{\theta}} - \mathbf{c}_r) + \mathbf{G}_{\boldsymbol{\theta}}(\lambda_{\boldsymbol{\theta}})(\mathbf{d}_{\boldsymbol{\theta}} - \mathbf{d}_r) + \big[\mathbf{G}_{\boldsymbol{\theta}}(\lambda_{\boldsymbol{\theta}}) - \mathbf{G}_r(\lambda_r)\big]\mathbf{d}_r.$$

By definition of the gate,

$$\mathbf{G}_{\boldsymbol{\theta}}(\lambda_{\boldsymbol{\theta}}) - \mathbf{G}_r(\lambda_r) = (1 - \lambda_{\boldsymbol{\theta}})(\mathbf{J}_{\boldsymbol{\theta}} - \mathbf{J}_r) + (\lambda_{\boldsymbol{\theta}} - \lambda_r)(\mathbf{I} - \mathbf{J}_r).$$

The triangle inequality, submultiplicativity of the operator norm, and Assumption H.6 yield Eq. (32).

I.7 Proof of Proposition H.9

Using Eq. (22) and $\mathbf{x}_k^{\mathrm{DIP}} = \hat{\mathbf{x}}_{0,k} + \mathbf{d}_k$,

$$\tilde{\mathbf{x}}_{0,k} - \mathbf{x}_k^{\mathrm{DIP}} = \hat{\mathbf{x}}_{0,k} + \mathbf{d}_k^{\mathrm{sens}} + \lambda_k\mathbf{d}_k^{\mathrm{res}} - (\hat{\mathbf{x}}_{0,k} + \mathbf{d}_k) = (\lambda_k - 1)\mathbf{d}_k^{\mathrm{res}}.$$

Taking norms gives Eq. (33).

I.8 Derivation of Eq. (34)

The MRI encoding operator $\mathcal{A}$ is linear. By $\tilde{\mathbf{x}}_{0,k} = \mathbf{x}_k^{\mathrm{DIP}} - (1 - \lambda_k)\mathbf{d}_k^{\mathrm{res}}$, we have

$$\mathcal{A}(\tilde{\mathbf{x}}_{0,k}) - \mathbf{y} = \mathbf{r}_k - (1 - \lambda_k)\mathbf{w}_k.$$

Expansion of the Hermitian squared norm gives Eq. (34) without any sign assumption on $\mathrm{Re}\langle \mathbf{r}_k, \mathbf{w}_k\rangle$.

I.9 Derivation of Eq. (35)

For the raw gate $\hat{\lambda}_k$, let $z_k = \beta(m_k - \tau_m) - \gamma(a_k - \tau_a)$. Then

$$1 - \hat{\lambda}_k = 1 - \lambda_{\max} + (\lambda_{\max} - \lambda_{\min})\sigma(-z_k).$$

Using $\sigma(-z) \leq e^{-z}$ and $a_k \leq 1$ yields Eq. (35). The cap $\lambda_{\max} = 0.5$ gives $1 - \hat{\lambda}_k \geq 0.5$. The EMA output averages current and preceding raw gates over reverse steps.

I.10 Proof of Theorem H.11

Assumption H.10 directly gives Eq. (37). Repeated substitution yields

$$\|\mathbf{x}_0 - \mathbf{x}_0^\dagger\| \leq \Big(\prod_{j=1}^{K} \rho_j\Big)\|\mathbf{x}_K - \mathbf{x}_K^\dagger\| + \sum_{k=1}^{K}\Big(\tau_k \prod_{j=1}^{k-1} \rho_j\Big)\|\tilde{\mathbf{x}}_{0,k} - \tilde{\mathbf{x}}_{0,k}^\dagger\|,$$

which expresses the accumulated contribution of initial and per-step discrepancies.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We state the main claims in the abstract and introduction, and we support them with theory and experiments. See Secs. 1–5.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the main limitations of test-time optimization, including latency and scan-adaptive refinement cost. See Appendix Secs. D.2 and G.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: We state the assumptions in Sec. 3 and provide the proofs in Appendix I.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We provide the datasets, protocols, baselines, architectures, optimization details, and evaluation metrics needed to reproduce the main experiments. See Sec. 4 and Appendix Secs. A–E.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We provide a public code repository.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: We provide the data splits, acquisition settings, undersampling patterns, optimizers, and major hyperparameters. See Sec. 4 and Appendix Secs. A and B.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We report mean $\pm$ standard deviation for the ablation study and provide standard deviations for the main quantitative table in Appendix D.3. See Table 3.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We provide the computer resources for reproducing the experiments. See Appendix Secs. C and D.2.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We are not aware of any aspect of this work that conflicts with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We discuss brief positive and negative broader impacts in Appendix F.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: We do not release a high-risk model or dataset that requires special safeguards.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [No]

    Justification: We cite the external datasets and software, but we do not explicitly list all licenses or terms of use in the paper.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: We do not release a new public dataset, code package, or model artifact in this paper.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This paper does not involve crowdsourcing or human annotation tasks.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: We utilized both publicly available MRI datasets and fully de-identified, retrospective internal data strictly for computational algorithm validation. This study does not involve new human subjects research, interventions, or pose any risks to participants.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs are not a core component of the method or the main claims of this paper.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.