Beyond Diffusion: Consistency Models for One-Step, High-Fidelity MRI Reconstruction

Mary-Brenda Akoda[1]   m.akoda24@imperial.ac.uk
Chen Qin[2]   c.qin15@imperial.ac.uk

[1] *Department of Computing & I-X, Imperial College London, UK*

[2] *Department of Electrical and Electronic Engineering & I-X, Imperial College London, UK*

Abstract

Magnetic resonance imaging (MRI) provides excellent soft-tissue contrast, but suffers from long acquisition times, limiting throughput and increasing patient discomfort. Recent diffusion-based generative models have achieved state-of-the-art reconstruction fidelity for undersampled MRI, but typically require hundreds to thousands of neural function evaluations (NFEs), which severely limits their practicality in time-sensitive clinical settings. We introduce C-MORE (Consistency-Model-based One-step REconstruction for MRI), to our knowledge, the first *one-step* consistency-model-based framework for accelerated MRI reconstruction. C-MORE investigates an unconditional one-step consistency model prior and then in one NFE, tackles the inverse problem by leveraging measurement-guided encoding and tunable physics-based refinement updates, thus eliminating multistep diffusion sampling while still providing a controllable trade-off between runtime and reconstruction fidelity. On the CMR×Recon dataset comprising multiple cardiac contrasts and both single- and multi-coil acquisitions, C-MORE achieves higher reconstruction quality than recent state-of-the-art diffusion-based samplers, while reconstructing each image in $0.18-0.52$ seconds on average ($\approx 22-193\times$ speed-up in just 1 NFE). These results establish C-MORE as a practical blueprint for real-time high-fidelity MRI reconstruction across multiple contrasts and acquisition settings.

Keywords: MRI reconstruction, diffusion models, consistency models, inverse problems, cardiac MRI.

1. Introduction

Magnetic resonance imaging (MRI) provides excellent soft-tissue contrast without ionising radiation, but conventional scans remain slow, often between fifteen minutes and over an hour and a half, leading to patient discomfort, motion artefacts, and reduced clinical throughput. Accelerated MRI reduces scan time by undersampling k-space (the frequency domain), but reconstructing an image from sparse measurements is an ill-posed inverse problem that requires strong priors. This challenge is particularly pronounced in quantitative parametric mapping, where each slice is acquired at multiple inversion or echo times to estimate T1/T2 maps: a single examination yields a sequence of contrast images per slice, and training separate reconstruction models for each contrast is inefficient and difficult to scale in practice.

Classical approaches such as parallel imaging (e.g. SENSE, GRAPPA) and compressed sensing (CS) exploit coil sensitivities or sparsity-promoting regularisers to stabilise the inverse problem (Lustig et al., 2007). While effective, these methods rely on hand-crafted priors, require lengthy optimisations, and can be sensitive to noise, parameter tuning, and the choice of sampling pattern. Deep learning–based reconstructions, including unrolled variational networks and image-domain convolutional neural networks (CNNs), learn powerful data-driven priors and can produce high-quality reconstructions in a single forward pass (Sriram et al., 2020; Hammernik et al., 2018). However, such models are typically trained for specific masks, acceleration factors, and acquisition protocols; their performance can degrade when the sampling pattern or contrast distribution shifts, often necessitating retraining or maintaining multiple models for different protocols.

Generative models have recently emerged as strong priors for accelerated MRI. Generative adversarial network (GAN)-based methods encourage realistic image appearance but can be unstable and hallucinate anatomically incorrect details (Mardani et al., 2019; Yang et al., 2018). Diffusion models (DMs) provide a more principled generative formulation, learning a time-dependent score or denoiser that can be combined with the MRI acquisition operator to perform posterior sampling. In score-based diffusion models for accelerated MRI, Chung and Ye (Chung and Ye, 2022) train a continuous-time score network on fully sampled images and alternate between stochastic or deterministic diffusion updates and k-space data consistency (DC) projections during sampling. Related work extends diffusion posterior sampling and plug-and-play guidance to MRI and other inverse problems, typically interleaving denoising steps with explicit DC updates at each iteration (Chung et al., 2023; Peng et al., 2022). More recently, Malyala et al. proposed SPA-MRI, which adapts consistency weights within a diffusion sampler to improve robustness across sampling patterns and contrasts in cardiac MRI (Malyala et al., 2024). Despite their strong reconstruction fidelity, these diffusion-based approaches generally require hundreds to thousands of neural function evaluations (NFEs) per image, with one or more DC steps at each iteration, which limits their practicality for time-sensitive clinical workflows.

Consistency models (CMs) have been proposed as a way to achieve quality comparable to diffusion models while enabling one- or few-step sampling (Song et al., 2023). Instead of integrating many small denoising updates along a diffusion trajectory, a CM learns a direct mapping from a noisy state to a clean state that is consistent with the probability-flow ordinary differential equation (ODE); hence, samples can be generated in a single or small number of NFEs. When trained either by distillation from a pre-trained diffusion model (consistency distillation) or directly from data (consistency training), CMs can approach the performance of diffusion models on natural image benchmarks with one to a few NFEs, offering a promising route towards real-time inverse problems. However, while CM applications to inverse problems are beginning to emerge (Zhao et al., 2024; Garber and Tirer, 2025), prior approaches still rely on few-step sampling or generic inverse problem setups.

In parallel, guided generation methods augment diffusion or consistency model backbones with auxiliary encoders that inject conditioning signals. ControlNet, for example, attaches a parallel encoder with zero-initialised lateral connections to steer the generative process using edge maps, depth maps, or other hints, without retraining the backbone (Zhang et al., 2023). Building on these ideas, Zhao et al. combine a CM prior with a ControlNet-style encoder and an explicit projection step to address generic inverse prob-

lems with one or few NFEs (Zhao et al., 2024). Their formulation, however, is developed for general imaging tasks: it does not exploit MRI-specific acquisition operators, nor is it tailored to multi-contrast parametric mapping scenarios where a single model must handle multiple contrasts per slice under stringent runtime constraints. Moreover, to refine results, their reconstruction pipeline still relies on multistep sampling interleaved with DC updates, and reports no further improvement beyond two NFEs, thereby limiting performance.

In this work, we investigate a one-step CM tailored to accelerated MRI to achieve high-fidelity reconstructions in one NFE while generalising across contrasts and acquisition settings. Specifically, we propose C-MORE (Consistency-Model-based One-step Reconstruction for MRI), a one-step generative reconstruction framework that builds on an unconditional CM prior and leverages an MRI-specific reconstruction strategy based on measurement-guided encoding and tunable conjugate gradient (CG) data consistency. Rather than alternating denoising and DC steps as in prior literature, C-MORE concentrates all learned generative reasoning into one guided CM evaluation and then performs multiple purely physics-based refinement updates. This design collapses the hundreds to thousands of DC corrections used in diffusion samplers into a single tunable K-CG block (where K is the maximum number of refinement updates); this tunable K enables a test-time trade-off between runtime and fidelity while still retaining one NFE and sub-second runtime.

Our contributions can be summarised as follows. 1) We propose, to our knowledge, the first *one-step* consistency-model-based framework for accelerated MRI reconstruction. 2) We introduce a hybrid reconstruction strategy that leverages measurement-guided encoding with tunable K-CG refinement, eliminating the need for multistep diffusion sampling. 3) In just one NFE, a single C-MORE model reconstructs across contrasts and acquisition settings without retraining, achieving higher reconstruction quality than recent state-of-the-art diffusion-based samplers, while running in $0.18 - 0.52\,\mathrm{s}$ on average ($\approx 22 - 193\times$ speed-up compared to state-of-the-art diffusion baselines requiring $100 - 1000$ NFEs).

2. Methods

2.1. Problem Formulation and Framework Overview

Accelerated MRI reconstruction can be formalised as recovering an image $x \in \mathbb{C}^N$ from noisy, undersampled k-space measurements y, thus:

$$y = Ax + \eta, \tag{1}$$

where $A : \mathbb{C}^N \to \mathbb{C}^M$ is the MRI acquisition operator and η models measurement noise; A^{H} denotes its adjoint. For a single-coil acquisition, A typically combines a sampling mask M and a discrete Fourier transform F, giving $y = MFx + \eta$. For multi-coil MRI, each coil c acquires $y_c = MFS_c x + \eta_c$ where S_c encodes coil sensitivities; stacking all coils yields y and the associated operator A.

Our goal is to recover a high-fidelity MR image x from the undersampled acquired k-space data y (Eq. (1)). However, this inverse problem is ill-posed ($M \ll N$), requiring strong priors. Fig. 1 provides an overview of our proposed C-MORE framework. C-MORE comprises four main components: (i) an unconditional diffusion teacher trained to model
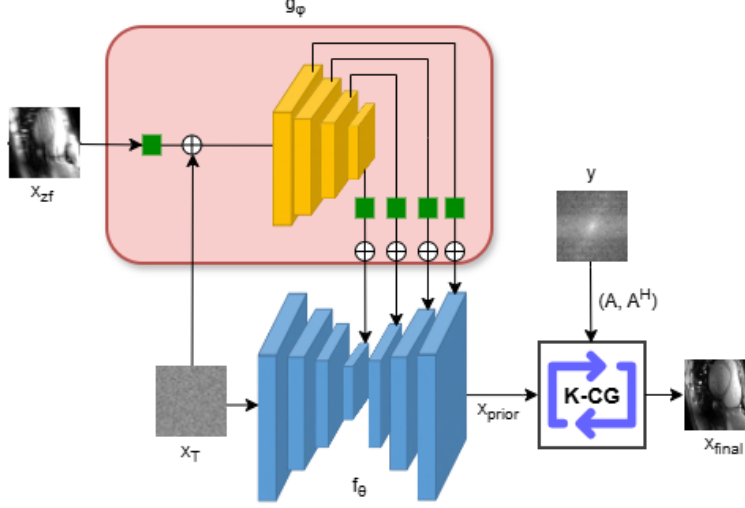
Figure 1: Overview of the proposed C-MORE framework. A pre-trained EDM teacher is distilled into a one-step CM prior. During reconstruction, a measurement-guided encoder g_φ conditions the distilled one-step CM f_θ on x_{zf} via zero-initialised lateral connections (green), and a subsequent CG block performs K physics-based refinement updates using the MRI forward operator, enforcing k-space consistency while keeping NFEs fixed at one.

the distribution of fully sampled MR images; (ii) an unconditional one-step consistency model (CM) prior f_θ obtained by distilling the teacher; (iii) a measurement-guided encoder that processes the zero-filled reconstruction x_{zf}, obtained by applying the inverse Fourier transform, and injects its features into the CM, allowing us to exploit measurement-specific information at inference time without retraining the prior for a specific A; and (iv) a conjugate gradient (CG) refinement module that performs K physics-based updates, thus enforcing k-space consistency with tunable K strength.

2.2. Unconditional Diffusion Teacher Training

To construct a high-fidelity one-step prior, we first train an unconditional Elucidated Diffusion Model (EDM) (Karras et al., 2022) on fully sampled complex-valued MR images. EDM is a continuous-time diffusion formulation that parameterises the corruption process by a noise scale $\sigma \in [\sigma_{\min}, \sigma_{\max}]$ rather than a discrete time index. Compared to earlier discrete-time diffusion models, EDM introduces (i) a tailored noise-level distribution, (ii) a preconditioning scheme that stabilises training and sampling across σ, and (iii) principled choices of parameterisation, loss weighting, and solver discretisation. This design thus yields state-of-the-art image quality with relatively few sampling steps (Karras et al., 2022), making EDM a strong candidate to serve as a teacher in our consistency distillation pipeline.

Let $x_0 \sim p_{\text{data}}$ denote a fully sampled MR image and $z \sim \mathcal{N}(0, I)$, standard Gaussian noise. EDM perturbs x_0 by sampling a noise level σ from a fixed distribution $p(\sigma)$ and forming: $x_\sigma = x_0 + \sigma z$. A U-Net denoiser g_ϕ takes (x_σ, σ) as input and predicts a denoised

sample. Following Karras et al. (2022), we train g_ϕ with a weighted denoising objective:

$$\mathcal{L}_{\text{EDM}}(\phi) = \mathbb{E}_{x_0, \sigma, z}\big[\lambda(\sigma)\,\|g_\phi(x_\sigma, \sigma) - x_0\|_2^2\big], \tag{2}$$

where $\lambda(\sigma)$ compensates for the varying difficulty of denoising at different noise levels. We represent complex-valued MR images as two channels (real and imaginary) and train g_ϕ unconditionally on fully sampled MR images—not dependent on sampling mask, acceleration factor, or coil configuration.

The EDM thus provides a high-quality unconditional generative prior over fully sampled MR images. In the subsection that follows, we treat g_ϕ as a fixed teacher and use its probability-flow ordinary differential equation (ODE) trajectory as the target for consistency distillation. Reconstruction pipelines based on g_ϕ combined with various data-consistency schemes serve as strong diffusion baselines in our experiments (Section 3); here, however, we focus solely on its role as a teacher.

2.3. Perceptual Consistency Distillation

We obtain an unconditional one-step CM prior f_θ via consistency distillation of our high-fidelity EDM teacher g_ϕ into f_θ. Let $\{t_n\}_{n=0}^N$ denote a discretisation of the continuous time interval $[0, 1]$, with $t_0 = 0$ (clean image) and $t_N = 1$ (noisy sample). The EDM teacher defines a reverse trajectory under its probability-flow ODE, which we approximate using a Heun integrator ϕ. For neighbouring times $t_{n+1} > t_n$, the teacher state at t_{n+1} is denoted $x_{t_{n+1}}$, and a single Heun update yields a one-step prediction $\hat{x}_{t_n}^\phi$ of the earlier state x_{t_n}. Consistency distillation then trains f_θ so that, given $x_{t_{n+1}}$, it outputs a state aligned with $\hat{x}_{t_n}^\phi$. Following the online–target scheme of Song et al. (2023), we maintain an exponential moving average (EMA) copy of the parameters θ^- and optimise:

$$\mathcal{L}_{\text{CD}}(\theta, \theta^-; \phi) = \mathbb{E}\big[\lambda(t_n)\,d\big(f_\theta(x_{t_{n+1}}, t_{n+1}), f_{\theta^-}(\hat{x}_{t_n}^\phi, t_n)\big)\big], \tag{3}$$

where $\lambda(t_n)$ is a time-dependent weight and $d(\cdot, \cdot)$ is a distance metric between the online CM output and its EMA counterpart. This objective collapses the multi-step EDM reverse trajectory into a single, globally consistent mapping. For MR images, a purely pixel-wise distance can lead to overly smooth one-step samples since a single U-Net must capture the entire denoising trajectory. Thus, to better preserve local structure, we adopt a perceptual distance based on the Learned Perceptual Image Patch Similarity (LPIPS) metric (Zhang et al., 2018), applied to the magnitude images of the complex-valued outputs. Substituting this choice of d into Eq. (3) yields a perceptual consistency distillation objective that aligns the CM with the EMA target in a feature space that is sensitive to structural similarity. In practice, we find that this design helps maintain sharp boundaries and small anatomical structures in one-step reconstructions, especially at higher acceleration factors.

After training, the CM defines an unconditional one-step prior: given a single noise sample x_T at the largest EDM noise level, we obtain a clean MR image in one forward pass: $x_0 = f_\theta(x_T, T)$. This design allows the same CM to be reused across settings, since acquisition-specific information is injected only at inference time (Section 2.4).

2.4. One-Step Guided Inference with Tunable Data Consistency

The CM prior described in Section 2.3 captures the distribution of fully sampled MR images but, by construction, is independent of the specific measurement y. Consequently, state-of-the-art diffusion-based methods typically solve the inverse problem via posterior sampling, which requires a high number of NFEs and leads to long reconstruction times. To address this limitation, we propose a hybrid reconstruction strategy that: (i) introduces measurement awareness into the generative step via measurement-guided encoding, avoiding the need to retrain the CM or compromise its unconditional, contrast-agnostic nature, and (ii) incorporates tunable physics-based refinement updates, eliminating the need for multistep diffusion sampling.

2.4.1. Measurement-guided encoding

Let x_{zf} denote the zero-filled reconstruction, obtained by applying the inverse Fourier transform and representing the result as a two-channel complex image (real and imaginary parts). Although x_{zf} contains aliasing artefacts, it already encodes coarse anatomy, contrast information, and coil sensitivity effects. Our goal is to use x_{zf} to steer the one-step CM towards solutions that are both likely under the prior and compatible with the measured data, without altering the unconditional CM training.

To this end, we attach a lightweight measurement-guided encoder to the CM in the spirit of ControlNet-style conditioning (Zhang et al., 2023; Zhao et al., 2024). A parallel encoder g_φ with the same multi-scale structure as the CM backbone processes x_{zf} and produces feature maps $\{h_\ell\}$ at multiple resolutions. At each encoder block ℓ of the CM, we fuse these features into the CM activations via a zero-initialised 1×1 convolution:

$$u_\ell = \mathrm{Enc}_\ell^{\mathrm{CM}}(x_T, t), \quad h_\ell = \mathrm{Enc}_\ell^{\mathrm{enc}}(x_{\mathrm{zf}}), \quad u_\ell' = u_\ell + W_\ell h_\ell, \qquad W_\ell\big|_{\mathrm{init}} = 0, \tag{4}$$

where u_ℓ are the original CM features, u_ℓ' the modified features passed to subsequent CM layers and the decoder, and W_ℓ are the learnable 1×1 fusion convolutions.

During training of this measurement-guided encoder, we freeze the CM parameters θ and optimise only φ and $\{W_\ell\}$ using paired fully sampled / undersampled data. Training minimises a reconstruction loss between the encoder-guided CM output and the fully sampled reference image, encouraging the encoder to use x_{zf} to refine the one-step sample towards the correct anatomy and contrast while leaving the unconditional CM backbone unchanged. Because the fusion weights W_ℓ are initialised to zero, the network initially reproduces the unconditional CM output and gradually learns to incorporate measurement information. The resulting measurement-aware one-step sample is:

$$x_{\mathrm{prior}} = f_\theta(x_T, T \mid x_{\mathrm{zf}}), \tag{5}$$

where the conditioning enters only through the additional encoder and lateral connections.

2.4.2. K Physics-based refinement

The guided CM output x_{prior} incorporates information from x_{zf} but is not guaranteed to satisfy the acquisition operator in Eq. (1). To improve data fidelity, we add a final conjugate gradient (CG) refinement stage that solves for the image x in the least-squares problem:

$$A^{\mathrm{H}} A x = A^{\mathrm{H}} y \tag{6}$$

using x_{prior} as the initial iterate and a fixed maximum number of updates, K. Each update requires only applications of A and A^{H}, implemented efficiently via Fast Fourier transforms (FFTs) and sensitivity maps (for multi-coil). We stop after K updates or earlier if the residual norm $\|Ax - y\|_2$ falls below a preset tolerance, and denote the final iterate by x_{final}.

The maximum update count K acts as a test-time knob that trades runtime against the strength of data-fidelity enforcement without retraining: small K values preserve the CM output but apply only mild corrections, whereas larger K values drive the solution closer to the least-squares solution of Eq. (6) at the cost of additional matrix–vector operations. In contrast to posterior-sampling approaches that interleave one or a few data-consistency updates with every network evaluation (Chung and Ye, 2022; Chung et al., 2023; Peng et al., 2022; Chung et al., 2024; Wang et al., 2023; Malyala et al., 2024; Zhao et al., 2024), C-MORE concentrates most of its data-fidelity enforcement into a single CG refinement stage after one measurement-guided CM evaluation. This strategy is more efficient considering diffusion-based posterior-sampling methods accumulate hundreds to thousands of DC steps across many NFEs, whereas our design uses a cumulatively smaller number of maximum updates ($K \in [100, 200]$ in our experiments), without increasing the number of NFEs and while keeping the refinement purely physics-based and independent of the generative prior.

2.5. C-MORE (Consistency-Model-based One-step REconstruction)

Combining the components above, C-MORE performs one-step reconstruction as in Algorithm 1: Given a measurement y, we first form the zero-filled reconstruction, then sample a single noise realisation x_T at the largest EDM noise level and run the measurement-guided CM to obtain the prior estimate x_{prior}. Finally, we run K updates of CG on Eq. (6), initialised at x_{prior}, to obtain the refined reconstruction x_{final}. In this way, all learned generative reasoning is concentrated into a single guided CM evaluation, and subsequent updates are governed solely by the MRI acquisition operator, thus achieving one–to–two orders of magnitude speed-ups over diffusion samplers that require 100–1000 NFEs.

Algorithm 1: C-MORE (Consistency-Model-based One-step REconstruction)

Input: Measurement y, sampling mask M, Fourier operator F, acquisition operator A
and its adjoint A^{H}, CM prior f_θ, measurement encoder g_φ, maximum number
of CG updates K, maximum noise level $\sigma_{\max}$
Output: x_{final}, the reconstructed MR image
$x_{\text{zf}} \leftarrow F^{-1}(M^\top y)$ // Zero-filled reconstruction
Sample $x_T \sim \mathcal{N}(0, \sigma_{\max}^2 I)$
$x_{\text{prior}} \leftarrow f_\theta(x_T, T \mid x_{\text{zf}})$ // Measurement-guided CM
$x_{\text{final}} \leftarrow \text{CG}(A^{\text{H}}A, A^{\text{H}}y, x_{\text{prior}}, K)$
return x_{final}

---

3. Experiments and Results

Dataset We evaluate on the CMR×Recon 2023 challenge dataset (Wang et al., 2024), which contains single- and multi-coil cardiac MR images for T1 and T2 mapping. Fully

sampled complex-valued images are retrospectively undersampled to simulate accelerated acquisitions. We use 15,576 complex 2D images (real and imaginary channels) for training and 7,758 undersampled images for testing, covering 9 T1-weighted and 3 T2-weighted contrasts per slice. A single C-MORE model is trained across all T1/T2 mapping contrasts, and results are reported on a multi-coil test set. We consider acceleration factors $R \in \{4, 8, 10\}$ with uniform Cartesian undersampling masks.

Baselines and Metrics To evaluate the effectiveness of C-MORE, we compare against four state-of-the-art diffusion-based samplers, each applied using the same pre-trained EDM prior alongside their respective posterior sampling schemes: DDNM (Wang et al., 2023); SPA-MRI (Malyala et al., 2024); DPS (Chung et al., 2023); and a DDS variant (Chung et al., 2024), i.e., we adapt DDS using EDM as its prior with CG at K=5 updates per NFE, also referring to this enhanced configuration as the *Teacher*. All baselines are configured for high-quality reconstructions, with NFEs between 100 and 1000 and data consistency (DC) enforcement at each NFE, reflecting typical settings in the literature. We evaluate C-MORE ("Ours") at K=100 and K=200. peak signal-to-noise ratio (PSNR, dB), structural similarity index measure (SSIM), NFE, and runtime in seconds (measured on a single NVIDIA A800 GPU) are used as evaluation metrics for comparisons.

Reconstruction Performance Table 1 summarises contrast-wise reconstruction quality and runtime for multi-coil setting across T1 and T2 at R=4, 8, 10. C-MORE with K=200 consistently achieves the highest PSNR and SSIM across all contrasts and acceleration factors. Even at K=100, C-MORE surpasses the teacher, which required 500 cumulative CG updates in comparison (K=5/NFE), in all cases except T1 at 10×, while maintaining a single NFE. In terms of efficiency, C-MORE runs in 0.18–0.52 s per image (depending on K and contrast) vs. 11.48–34.72 s for other methods, yielding speed-ups of $\approx$ 22–193× relative to diffusion-based baselines. Crucially, these gains are achieved without increasing NFE beyond 1. Note that increasing K from 100 to 200 adds negligible cost (at most $\approx$ 0.1 s per image), and with residual-based early stopping at a tolerance of 10^{-8}, some K=200 runs terminate in about the same time as K=100.

Fig. 2 shows qualitative T1 reconstructions at R=4, 8, 10. The zero-filled images exhibit severe aliasing at higher accelerations. Among the baselines, the teacher and a DDNM variant using CG for further enhancement yield visually high-quality reconstructions but require 100 NFEs and 500 cumulative CG updates (K=5/NFE). In contrast, C-MORE (K=100 and K=200) shows no visible artefacts, while running in one NFE and sub-second time. Full T1 and T2 parametric-sequence reconstructions are provided in Appendix A.

Ablation Studies To analyse the choice of refinement operator and the effect of varying K, we fix the CM (without measurement-guided encoding) and vary K for different posterior sampling schemes at R=10 while keeping NFE= 1. For K from 1 to 100, CG yields the strongest and most monotonic improvements in PSNR/SSIM, outperforming other methods (Fig. 3). Increasing K from 100 to 200 further improves PSNR/SSIM at R=10, while increasing runtime by only $\approx$ 0.1 s per image. In practice, we report results for K=100 and K=200 balancing efficiency and fidelity, while maintaining sub-second runtime. We also compare three one-step variants at NFE= 1: (a) a CM with measurement-guided encoding only, (b) a CM with CG at K=100 (no guided encoding), and (c) the full C-MORE hybrid (guided encoding + CG at K=100). Fig. 3 shows that the hybrid approach
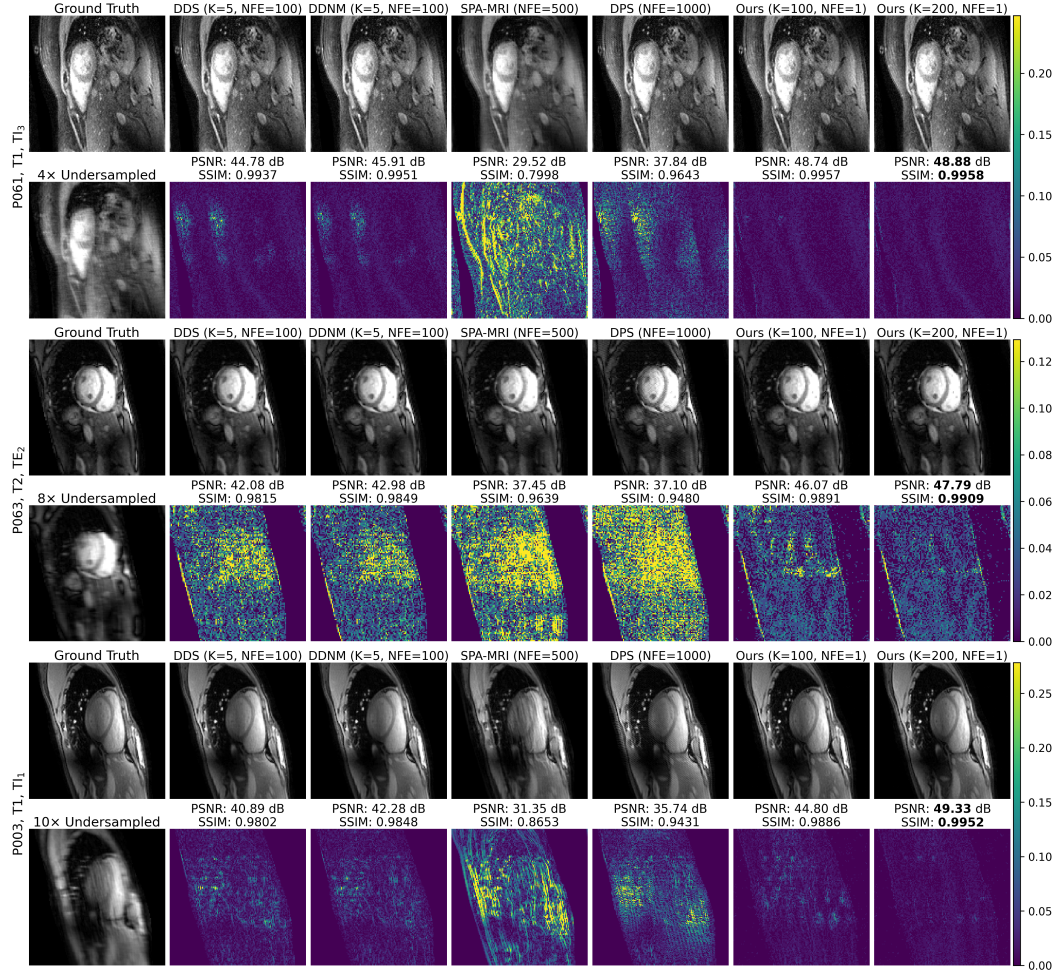
Figure 2: Cardiac MR reconstructions at $4\times$, $8\times$, $10\times$ across different patients and tissue contrasts, acquired at varying inversion times (TI) and echo times (TE). We compare ground truth, zero-filled, teacher ("DDS") and a DDNM variant (both using CG at K=5/NFE for further enhancement), SPA-MRI, DPS, C-MORE ("Ours") at $K \in [100, 200]$, plus error maps. Best results are in **bold**.
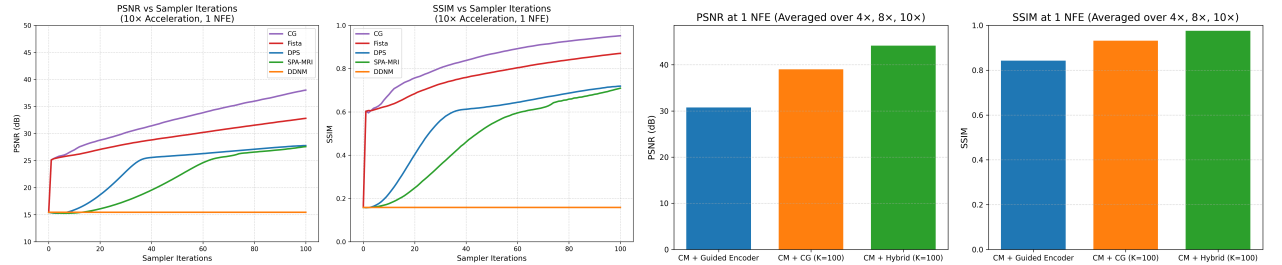

Figure 3: Ablation study across T1/T2. Left: Results of increasing K for different DC schemes at $10\times$ (1 NFE, no guided encoding). CG (purple) shows largest first K gain and remains best at K=100. Right: Effect of C-MORE components; hybrid (green): guided encoding + K-CG shows highest PSNR/SSIM.

Table 1: Contrast-wise PSNR, SSIM, NFEs, and per-contrast runtime for T1 and T2 at $4\times$, $8\times$, and $10\times$ undersampling. Values are reported as mean $\pm$ standard deviation across the test set. Best results are in **bold**; second-best are underlined.

Acc	Method	NFE	T1			T2		
			PSNR ↑	SSIM ↑	Runtime (s) ↓	PSNR ↑	SSIM ↑	Runtime (s) ↓
$4\times$	DDS (Teacher, K=5)	100	48.44 ± 0.97	0.9920 ± 0.0057	17.14 ± 2.40	46.56 ± 1.52	0.9941 ± 0.0047	11.38 ± 0.23
	DDNM	100	38.03 ± 2.29	0.9240 ± 0.0279	11.48 ± 0.30	36.47 ± 1.65	0.9274 ± 0.0367	11.48 ± 0.27
	SPA-MRI	500	40.18 ± 6.15	0.9441 ± 0.0514	34.70 ± 15.11	39.09 ± 3.59	0.9629 ± 0.0344	22.22 ± 0.24
	DPS	500	44.86 ± 2.87	0.9827 ± 0.0102	23.53 ± 0.22	41.77 ± 2.19	0.9804 ± 0.0126	23.50 ± 0.21
	C-MORE (K=100)	1	$\underline{49.05} \pm 0.45$	$\mathbf{0.9925} \pm 0.0051$	$\underline{0.37} \pm 0.12$	$\underline{48.81} \pm 0.57$	$\mathbf{0.9965} \pm 0.0030$	$\mathbf{0.18} \pm 0.05$
	C-MORE (K=200)	1	$\mathbf{49.06} \pm 0.45$	$\mathbf{0.9925} \pm 0.0051$	$\mathbf{0.31} \pm 0.11$	$\mathbf{48.92} \pm 0.48$	$\mathbf{0.9965} \pm 0.0030$	$\underline{0.24} \pm 0.05$
$8\times$	DDS (Teacher, K=5)	100	43.03 ± 2.91	0.9754 ± 0.0100	17.12 ± 2.40	40.15 ± 2.04	0.9750 ± 0.0090	11.38 ± 0.24
	DDNM	100	36.14 ± 2.61	0.9050 ± 0.0334	11.49 ± 0.29	34.79 ± 1.84	0.9128 ± 0.0397	11.47 ± 0.27
	SPA-MRI	500	34.59 ± 4.58	0.8840 ± 0.0605	34.71 ± 15.12	33.85 ± 3.07	0.9221 ± 0.0367	22.24 ± 0.25
	DPS	500	38.95 ± 4.20	0.9437 ± 0.0313	23.54 ± 0.22	36.59 ± 2.48	0.9490 ± 0.0213	23.52 ± 0.20
	C-MORE (K=100)	1	$\underline{43.85} \pm 2.31$	$\underline{0.9761} \pm 0.0109$	$\mathbf{0.42} \pm 0.13$	$\underline{42.49} \pm 1.96$	$\underline{0.9842} \pm 0.0080$	$\mathbf{0.18} \pm 0.05$
	C-MORE (K=200)	1	$\mathbf{47.20} \pm 1.59$	$\mathbf{0.9885} \pm 0.0064$	$\underline{0.52} \pm 0.15$	$\mathbf{46.36} \pm 1.58$	$\mathbf{0.9933} \pm 0.0045$	$\underline{0.35} \pm 0.05$
$10\times$	DDS (Teacher, K=5)	100	$\underline{41.21} \pm 3.09$	$\underline{0.9678} \pm 0.0133$	17.12 ± 2.40	38.89 ± 2.10	0.9696 ± 0.0108	11.39 ± 0.23
	DDNM	100	35.32 ± 2.71	0.8985 ± 0.0348	11.49 ± 0.30	34.44 ± 1.90	0.9131 ± 0.0395	11.47 ± 0.29
	SPA-MRI	500	32.55 ± 3.97	0.8574 ± 0.0635	34.72 ± 15.12	32.98 ± 3.05	0.9167 ± 0.0332	22.23 ± 0.26
	DPS	500	36.22 ± 4.22	0.9243 ± 0.0407	23.52 ± 0.22	35.38 ± 2.51	0.9401 ± 0.0231	23.51 ± 0.21
	C-MORE (K=100)	1	40.07 ± 2.52	0.9505 ± 0.0166	$\mathbf{0.41} \pm 0.12$	$\underline{39.29} \pm 1.88$	$\underline{0.9700} \pm 0.0104$	$\mathbf{0.18} \pm 0.05$
	C-MORE (K=200)	1	$\mathbf{43.25} \pm 2.42$	$\mathbf{0.9722} \pm 0.0124$	$\underline{0.52} \pm 0.15$	$\mathbf{42.54} \pm 1.85$	$\mathbf{0.9838} \pm 0.0077$	$\underline{0.36} \pm 0.04$

consistently achieves the highest average PSNR/SSIM across $R=4, 8, 10$, and improves with higher K. Training the encoder for longer (200,000 vs. 100,000 epochs) also provides modest additional gains, and we use the 200,000-epoch guided encoder for qualitative visualisations. This indicates that the measurement-guided encoding and physics-based refinement play complementary roles.

4. Conclusion

We introduced the first one-step consistency-model-based framework for MRI reconstruction, uniting the power of diffusion generative priors with the efficiency of a single-pass network. By distilling a diffusion teacher into a one-step CM prior and augmenting it with measurement-guided encoding and K-CG refinement updates, our method achieves high-fidelity reconstructions in a single NFE while rigorously enforcing measurement constraints. Across diverse settings, we demonstrated that C-MORE consistently outperforms recent state-of-the-art diffusion-based samplers, while achieving $\approx 22 - 193\times$ speed-up in just 1 NFE. This work therefore provides a flexible, real-time reconstruction pipeline that transforms a previously slow capability into a clinically plausible, on-the-fly solution. Future work will extend C-MORE to multi-centre cohorts with diverse pathology. Ultimately, our proposed one-step framework offers a promising blueprint for achieving state-of-the-art reconstruction quality with real-time performance.

References

Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated mri. *Medical image analysis*, 80:102479, 2022.

Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023.

Hyungjin Chung, Suhyeon Lee, and Jong Chul Ye. Decomposed diffusion sampler for accelerating large-scale inverse problems. In *12th International Conference on Learning Representations, ICLR 2024*, 2024.

Tomer Garber and Tom Tirer. Zero-shot image restoration via few-step guidance of consistency models. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2025.

K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, and T. et al. Pock. Learning a variational network for reconstruction of accelerated mri data. *Magnetic Resonance in Medicine*, 79:3055–3071, 6 2018.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, 35:26565–26577, 2022.

Michael Lustig, David Donoho, and John M Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

Anurag Malyala, Zhenlin Zhang, Chengyan Wang, and Chen Qin. Sampling-pattern-agnostic mri reconstruction through adaptive consistency enforcement with diffusion model. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 140–151. Springer, 2024.

M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, and L. et al. Xing. Deep generative adversarial neural networks for compressive sensing mri. *IEEE Transactions on Medical Imaging*, 38:167–179, 1 2019.

Cheng Peng, Pengfei Guo, S Kevin Zhou, Vishal M Patel, and Rama Chellappa. Towards performant and reliable undersampled mr reconstruction via diffusion model sampling. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 623–633. Springer, 2022.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *International Conference on Machine Learning*, pages 32211–32252. PMLR, 2023.

Anuroop Sriram, Jure Zbontar, Tullie Murrell, Aaron Defazio, C Lawrence Zitnick, Nafissa Yakubova, Florian Knoll, and Patricia Johnson. End-to-end variational networks for

accelerated mri reconstruction. In *International conference on medical image computing and computer-assisted intervention*, pages 64–73. Springer, 2020.

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.

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.

Guang Yang, Simiao Yu, Hao Dong, Greg Slabaugh, Pier Luigi Dragotti, Xujiong Ye, Fangde Liu, Simon Arridge, Jennifer Keegan, Yike Guo, and David Firmin. Dagan: Deep dealiasing generative adversarial networks for fast compressed sensing mri reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1310–1321, 2018. doi: 10.1109/TMI.2017. 2785879.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3836–3847, 2023.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018.

Jiankun Zhao, Bowen Song, and Liyue Shen. Cosign: Few-step guidance of consistency model to solve general inverse problems. In *European Conference on Computer Vision*, pages 108–126. Springer, 2024.

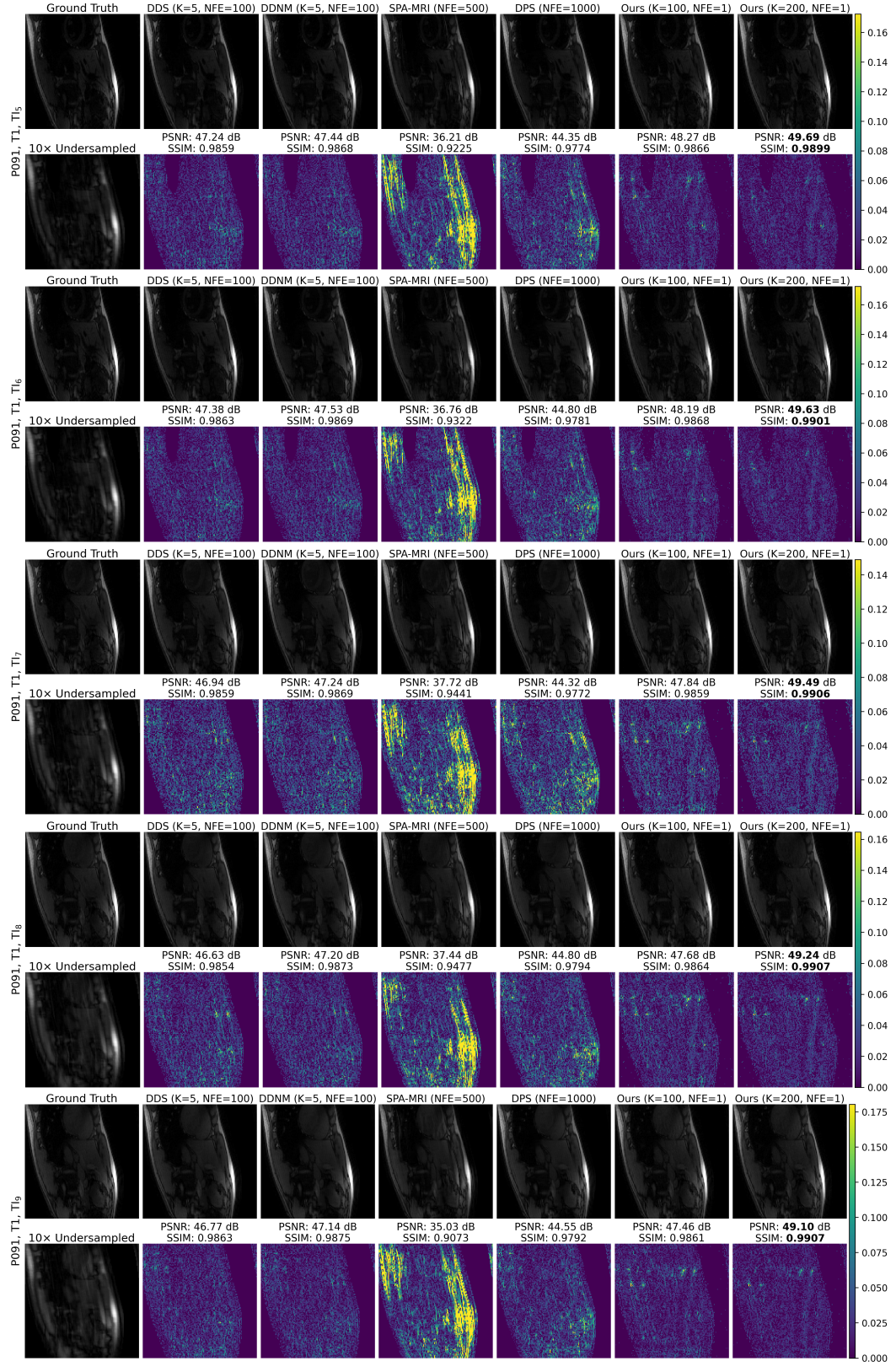## Appendix A. Reconstructions across a full T1 and T2 parametric mapping sequence
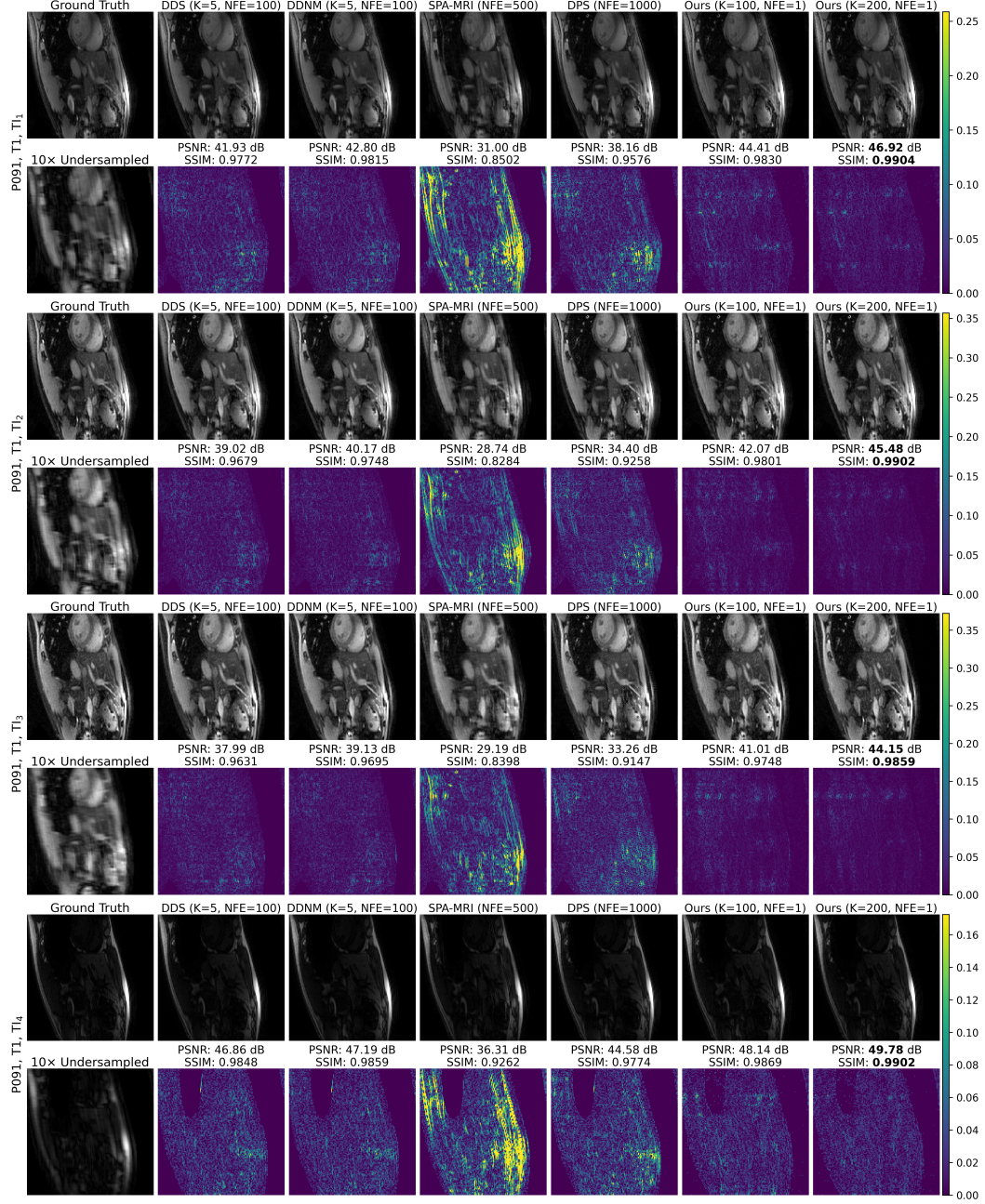
Figure 3: Reconstructions across the full T1 parametric mapping sequence (9 inversion times) for patient 91 at 10× acceleration.
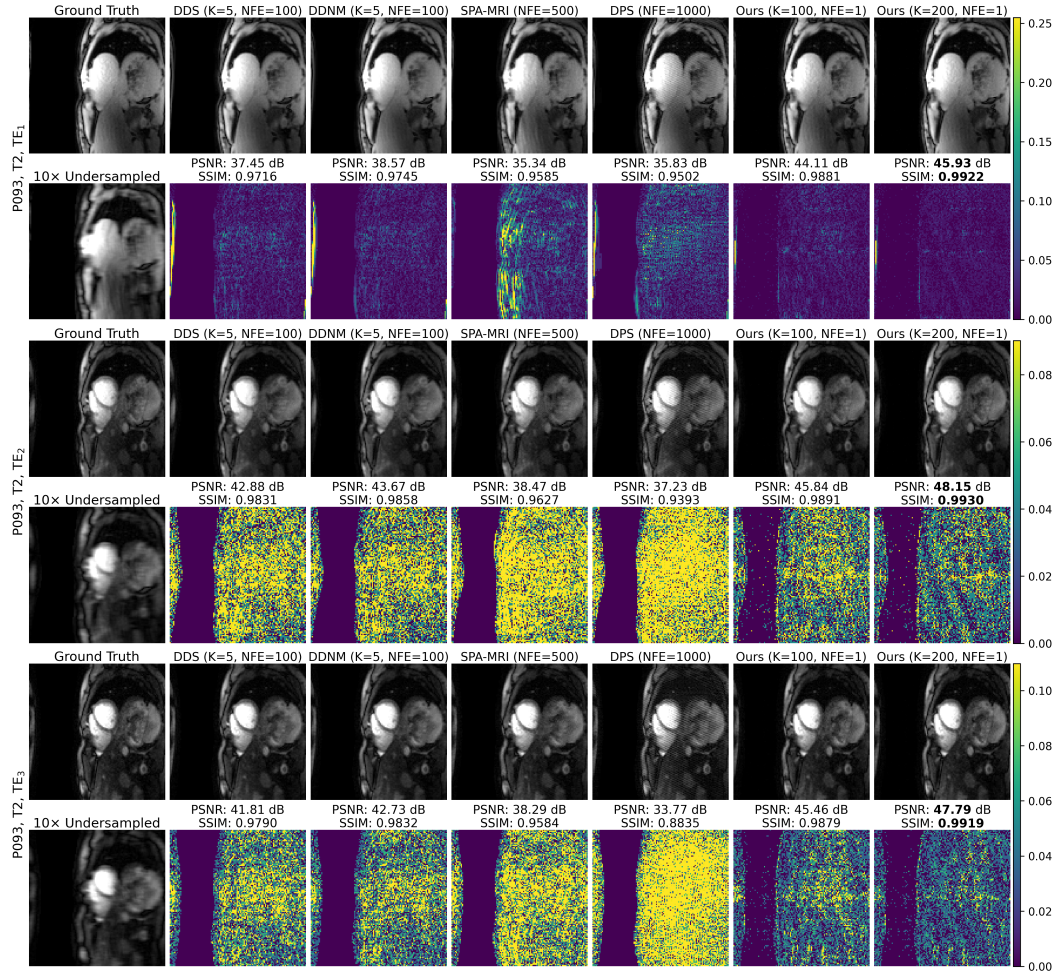
Figure 4: Reconstructions across the full T2 parametric mapping sequence (3 echo times) for patient 93 at $10\times$ acceleration.