SDUM: A Scalable Deep Unrolled Model for Universal Cardiac MRI Reconstruction

Puyang Wang[1⋆], Pengfei Guo[2*], Keyi Chai[1], Jinyuan Zhou[1], Daguang Xu[2], and Shanshan Jiang[1]

[1] Johns Hopkins University
[2] NVIDIA
pwang47@jhu.edu, pengfeig@nvidia.com, sjiang21@jhmi.edu

Abstract. Clinical cardiac MRI spans diverse contrasts, sampling trajectories, accelerations, scanners, and patient populations, yet many deterministic deep-learning reconstruction methods remain protocol specific. We present the Scalable Deep Unrolled Model (SDUM), which integrates a Restormer-based unrolled reconstructor, per-cascade coil sensitivity estimation, sampling aware weighted data consistency, and universal conditioning on cascade index and acquisition metadata. A single SDUM model achieves state-of-the-art performance across all CMRxRecon2025 tracks without task-specific fine-tuning and outperforms PromptMR+ on CMRxRecon2024 by +0.55 dB. Scaling experiments show near-logarithmic gains with depth up to 18 cascades (r=0.986, R^2=0.973) and continued but diminishing gains from data scaling (32.72 dB at 40% to 33.18 dB at 100%). SDUM also generalizes in a zero-shot setting to unseen in-house chemical exchange saturation transfer (CEST) MRI (43.57 dB PSNR, 0.9769 SSIM). When trained separately on fastMRI brain, SDUM surpasses PC-RNN by +1.8 dB. These results support SDUM as a scalable framework for robust MRI reconstruction across heterogeneous acquisition settings beyond cardiac MRI.

Code: https://github.com/NVIDIA-Medtech/NV-Raw2insights-MRI
Model: https://huggingface.co/nvidia/NV-Raw2insights-MRI

Keywords: MRI reconstruction · Deep unrolling · Cardiac MRI · Scaling behavior · Foundation model

1 Introduction

Cardiac MRI (CMR) is the gold standard for non-invasive assessment of cardiac morphology, function, and tissue characterization. A single clinical CMR exam routinely acquires a dozen or more sequences, cine for wall motion, T1/T2 mapping for tissue characterization, late gadolinium enhancement (LGE) for fibrosis, perfusion for ischemia, flow for valvular assessment, tagging for strain, using Cartesian, radial, spiral, or kt-space sampling at acceleration factors ranging from 4× to 24×. Scanner field strengths (1.5T, 3T, emerging 5T), coil arrays, vendor software, and patient populations (adult, pediatric, multi-disease)

⋆ Equal contribution.

add further heterogeneity across clinical sites. While recent deterministic supervised deep learning methods [1,7,13,26,28] have delivered strong performance in controlled settings, they often remain sensitive to protocol boundaries: a model trained for one sampling mask or acceleration often degrades when the acquisition changes. This brittleness hinders deployment in multi-site clinical environments and motivates a *single, universally applicable* cardiac MRI reconstruction model that can robustly handle heterogeneous inputs without per-protocol retraining.

Hypothesis. We hypothesize that despite the exceptional diversity of cardiac MRI protocols, the underlying cardiac anatomy and physiology impose a *shared image prior* that can be learned by a single model. Specifically, a deep unrolled reconstruction network, when conditioned on the acquisition physics (sampling pattern, trajectory type, acceleration factor) via universal conditioning and trained with a progressive curriculum, can disentangle protocol-specific acquisition artifacts from anatomy-specific content. The shared prior across protocols acts as implicit regularization, enabling robustness to distribution shifts across centers, diseases, field strengths, and patient populations.

The scaling gap. In adjacent domains, empirical scaling analyses [14,16,42], quantifying how error decays with parameters, data, or compute, have become a practical compass for model design and resource allocation. MRI reconstruction lacks such guidance. Practitioners must rely on trial-and-error to decide whether to invest in wider backbones, deeper cascade unrolling, or larger training sets, with no principled understanding of returns or saturation.

Limitations of existing approaches. Current attempts at universal MRI reconstruction remain incomplete. Many deep unrolled models [1, 13] employ a scalar data-consistency (DC) weight that ignores sampling-density and noise characteristics specific to each trajectory. Conditioning mechanisms, when present, are often ad hoc or task-specific [34]. Coil sensitivity maps (CSMs) are typically precomputed and fixed, limiting robustness to motion and field inhomogeneities. Score- and diffusion-prior approaches offer a complementary route by decoupling the learned prior from the forward model, but their iterative sampling/projection procedures target a different operating regime from the deterministic raw multicoil CMR challenge setting studied here. No existing approach has been validated as a single model across the full spectrum of cardiac MRI heterogeneity: multi-contrast, multi-trajectory, multi-center, multi-field-strength, and multi-population, simultaneously.

Our approach: SDUM. We introduce the Scalable Deep Unrolled Model (SDUM), a framework designed for universal cardiac MRI reconstruction through five synergistic components:

1. **Restormer-based reconstruction backbone**: Multi-Dconv Head Transposed Attention (MDTA) and Gated-Dconv Feedforward Network (GDFN) [40] jointly capture long-range dependencies (to unfold aliasing) and local structures (to preserve edges) with favorable memory efficiency. A shallow two-stage pyramid preserves high-resolution detail while enabling global context aggregation.

2. **Learned per-cascade coil sensitivity map estimation (CSME)**: A U-Net-based estimator refines CSMs at each cascade, mitigating errors from motion, noise, and field inhomogeneities without relying on autocalibration regions [23, 30].
3. **Sampling-aware weighted data consistency (SWDC)**: Instead of a scalar DC weight, we learn spatially varying k-space weight maps conditioned on the sampling pattern, enabling mask-specific and density-aware fidelity enforcement that unifies Cartesian and non-Cartesian trajectories in a single learnable module.
4. **Universal conditioning (UC)**: Sinusoidal embeddings of cascade index t and protocol metadata (mask type, acceleration, modality) are mapped through MLPs and injected into every Restormer block [21, 22], allowing a single model to adapt its behavior across diverse acquisitions.
5. **Progressive cascade expansion**: Training proceeds by doubling interior cascades while fixing endpoints, stabilizing optimization and reusing learned weights as depth increases [3, 17].

We train SDUM on heterogeneous CMRxRecon data covering cardiac protocols (BlackBlood, Cine, Flow2d, LGE, Mapping, Perfusion, T1rho, T1w, T2w, Aorta, Tagging) with Cartesian and non-Cartesian sampling. On the CMRxRecon2025 challenge [37], which evaluates generalization to unseen centers, diseases, 5T field strength, and pediatric populations, a *single* SDUM achieves state-of-the-art performance across all four tracks without task-specific fine-tuning, exceeding specialized baselines by up to +1.0 dB. On CMRxRecon2024 [32], SDUM surpasses the winning method PromptMR+ [34] by +0.55 dB and wins in over 90% of paired validation cases. We further evaluate strict zero-shot transfer on in-house CEST MRI from an unseen scanner vendor/protocol, where SDUM reaches 43.57 dB PSNR and 0.9769 SSIM without adaptation. To examine whether SDUM's design principles transfer beyond cardiac MRI, we train a separate model on fastMRI brain data [41], where it exceeds PC-RNN [4] by +1.8 dB. We also conduct (to our knowledge) the first **empirical scaling analysis for CMR reconstruction**, sweeping cascade depth T from 1 to 18 and training data volume from 40% to 100%. Reconstruction quality follows PSNR $\sim$ log(parameters) with correlation r=0.986. Beyond reporting numbers, we distill practical scaling lessons for building MRI reconstruction foundation models, including the need for diversity-aware data scaling and compute-optimal training.

2 Related Work

Classical and deep learning MRI reconstruction. Classical accelerated MRI reconstruction employs compressed sensing [19], parallel imaging (SENSE [23], GRAPPA [10], ESPIRiT [30]), and dynamic imaging methods [9, 20]. Deep unrolled models (DUMs) [1, 7, 13, 26, 28] integrate iterative optimization into neural networks, alternating data-consistency (DC) and learned regularization steps. Related approaches include KIKI-Net [8], DeepComplexMRI [33], PC-RNN [4],

and Plug-and-Play/RED frameworks [24,31,43]. Most unrolled models use scalar DC weights; our SWDC learns spatially varying k-space weights conditioned on sampling patterns. E2E-VarNet [28] jointly learns reconstruction and sensitivity estimation, inspiring our learned CSME approach.

Score and diffusion priors for MRI. Deep generative priors and score-/diffusion-based reconstruction methods can separate a learned image prior from the acquisition forward model, making them naturally flexible with respect to masks and acceleration factors [5, 6, 15]. This flexibility is complementary to SDUM. In contrast, SDUM is a deterministic supervised reconstructor trained on raw multicoil data, where a single feed-forward unrolled model must handle heterogeneous cardiac contrasts, sampling trajectories, centers, field strengths, and patient populations with near-real-time inference.

Cardiac MRI reconstruction. Cardiac MRI poses unique challenges due to the diversity of contrasts (cine, mapping, LGE, perfusion, flow), dynamic acquisitions requiring temporal modeling, and the need for robustness across field strengths and patient populations. Recent cardiac-focused methods include PromptMR [35] and PromptMR+ [34], which use prompt-based conditioning for multi-contrast cardiac reconstruction, and HierAdaptMR [38], which employs hierarchical adaptation across centers and protocols. The CMRxRecon challenge series [32, 37] has catalyzed progress by providing large-scale multi-contrast, multi-trajectory cardiac k-space datasets and standardized evaluation across clinically relevant distribution shifts. Despite this progress, no prior method has demonstrated a single model that generalizes across the full spectrum of cardiac MRI heterogeneity—multi-contrast, multi-trajectory, multi-center, multi-field-strength, and multi-population—without task-specific fine-tuning.

Architectures and conditioning. U-Net [25] and restoration transformers (SwinIR [18], Restormer [40]) are widely used backbones; we adopt Restormer for its favorable memory/compute trade-off. Conditional normalization (FiLM [22], AdaLN [21]) enables task adaptation; in MRI, PromptMR+ [34] and HierAdaptMR [38] use conditioning for multi-contrast reconstruction. We extend this paradigm with physics-informed universal conditioning.

Scaling behavior and training. Empirical scaling analyses [14,16,42] quantify performance vs. parameters/data/compute in NLP and vision, guiding resource allocation. Prior work has shown that increasing unrolled depth (number of cascades) is important for reconstruction performance: ReconFormer [11] demonstrated this with recurrent transformers, and similar findings were reported for deep unrolled models [36]. However, no prior study has systematically characterized the scaling behavior of unrolled MRI models across both depth and data axes. For training, our cascade expansion applies progressive training [3, 17] to stabilize optimization during deep unrolling.

3 Methods

3.1 Problem Formulation

MRI reconstruction aims to recover an image $x \in \mathbb{C}^{H\times W}$ from undersampled multi-coil k-space measurements $y \in \mathbb{C}^{C\times H\times W}$ with C receiver coils. The forward model is

$$y = M \cdot \mathcal{F}(S \cdot x) + n, \tag{1}$$

where S denotes coil sensitivity maps, $\mathcal{F}$ is the Fourier transform, M is a binary sampling mask, and n is acquisition noise. Real-world cardiac MRI protocols vary in contrast (cine, mapping, LGE, perfusion, flow), sampling pattern (Cartesian, radial, spiral, kt), and acceleration, so the model must handle heterogeneous acquisition conditions.

Reconstruction is posed as $\hat{x} = \arg\min_x \big(\frac{1}{2}\|M\mathcal{F}(Sx) - y\|_2^2 + \lambda R(x)\big)$, balancing a data-consistency (DC) term with a regularizer $R(x)$.

Deep Unrolled Reconstruction. Let $A = M \cdot \mathcal{F}$. Starting from the zero-filled image $x^{(0)}$, cascade $t \in \{0, \ldots, T-1\}$ performs proximal-gradient unrolling:

$$z^{(t)} = x^{(t)} - \tau_t S^H A^H \big(AS \cdot x^{(t)} - y\big), \tag{2}$$

$$x^{(t+1)} = D_\theta^{(t)}\big(z^{(t)}\big), \tag{3}$$

where τ_t is a learnable step size and $D_{\theta^{(t)}}$ is a learnable proximal operator. The full model $\hat{x} = x^{(T)}$ is trained end-to-end with conditioning on cascade index and protocol metadata.

3.2 Scalable Deep Unrolled Models

Deep unrolled MRI reconstruction methods often rely on precomputed coil sensitivity maps (CSMs), which are inevitably imperfect due to noise, motion, and field inhomogeneities.

We adopt a deep-unrolled architecture with T cascades. Each cascade comprises (i) a Restormer-based image reconstructor, (ii) a U-Net-based coil-sensitivity map estimator (CSME) that refines CSMs during reconstruction and does not rely on ACS regions, and (iii) a sampling-aware data-consistency operator that enforces the forward model with mask-dependent weighting in k-space. We introduce two key enhancements:

Adaptive unrolling with cascade skip connections. Each cascade receives not only the output of the previous stage, but also skip connections of decoder feature at downsampling level f_l from previous cascade. This mitigates vanishing gradients, facilitates training deeper models, and supports effective reuse of intermediate features.

Adjacent slice/frame input. For dynamic (eg. cinematic cardiac MRI) and 3D volumetric MRI, each cascade can access temporally or spatially adjacent frames. These inputs are fed into the reconstructor alongside the current slice, enabling the model to leverage temporal redundancy or 3D anatomical context.

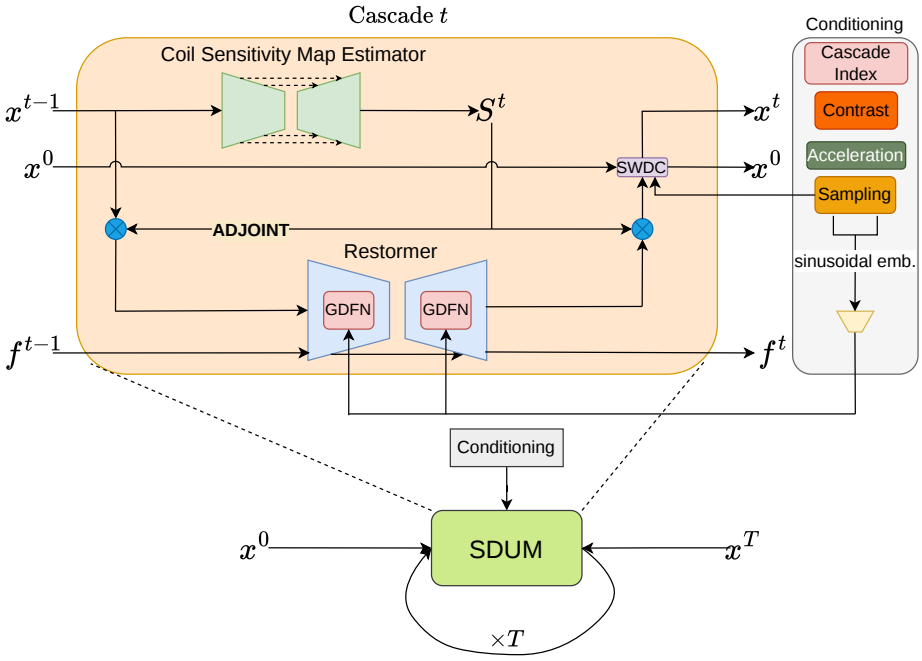

Fig. 1: SDUM overview. Each cascade combines a Restormer-based reconstructor, a learned CSME, sampling-aware weighted data consistency (SWDC), and universal conditioning (UC) on cascade index and protocol metadata.

3.3 Restormer Backbone

We adopt Restormer [40] as the per-cascade reconstructor because its Multi-Dconv Head Transposed Attention (MDTA) and Gated-Dconv Feedforward Network (GDFN) jointly capture non-local and local dependencies at $O(HW \cdot C^2)$ complexity. We use a shallow two-stage pyramid (one down/up) with MDTA+GDFN blocks, which preserves high-resolution detail while providing global context. Empirically, this shallow design outperforms deeper pyramids in our setting: additional downsampling increases receptive field but tends to oversmooth fine structures and weaken small-lesion contrast. A wider-but-shallow per-cascade design also stabilizes deep unrolling by improving gradient flow across cascades. Additional implementation details are provided in the supplementary material.

At cascade $t \in \{0, \ldots, T-1\}$, we estimate coil sensitivities with a U-Net CSME and then apply DC and a Restormer prior:

$$S^{(t)} = \mathrm{Norm}\Big(\mathrm{CSME}_{\phi_t}\big(x^{(t)}\big)\Big), \tag{4}$$

$$z^{(t)} = x^{(t)} - \tau_t\, {S^{(t)}}^H A^H \Big(A\big(S^{(t)} \cdot x^{(t)}\big) - y\Big), \tag{5}$$

$$x^{(t+1)} = \mathrm{Restormer}_{\theta_t}\big(z^{(t)}\big), \tag{6}$$

where $A = M\mathcal{F}$, τ_t is a learnable DC step size, and $\mathrm{Norm}(\cdot)$ enforces per-pixel coil normalization ($\sum_c |S_c|^2 = 1$).

Each cascade has its own CSME and Restormer parameters. The CSME is a compact complex U-Net with channel progression $\{12, 24, 48, 96, 192\}$, instance normalization, and deconvolution-based upsampling; its output is normalized by $S \leftarrow S/\sqrt{\sum_c |S_c|^2}$ to keep physically meaningful coil-energy scaling. For the image prior, we use a two-level Restormer with channel widths $\{256, 512\}$, attention heads $\{1, 2\}$, and block depths $\{3, 6\}$, followed by two refinement blocks. This shallow-but-wide design preserves high-resolution structure while still enabling long-range interaction through MDTA. Complex-valued MRI data are represented as stacked real/imaginary channels and processed with real-valued operators, and inputs are padded to sizes divisible by the down/up-sampling hierarchy for numerically stable unrolling.

3.4 Sampling-aware Weighted Data Consistency

We replace the standard scalar DC weight with a learned, spatially varying k-space weight map conditioned on the sampling pattern. The standard DC gradient in cascade t is

$$g_{\mathrm{DC}}^{(t)} \;=\; S^{(t)^H} A^H \Big(A\big(S^{(t)} \cdot x^{(t)}\big) - y \Big), \tag{7}$$

leading to the update $z^{(t)} = x^{(t)} - \tau_t \, g_{\mathrm{DC}}^{(t)}$.

Weighted residual in k-space. Let $w^{(t)} \in \mathbb{R}^{H \times W}$ be a learned weight map defined in k-space (broadcast across coils), with $w^{(t)} = 0$ on unsampled locations. We form a weighted residual

$$r_w^{(t)} \;=\; w^{(t)} \odot \Big(A\big(S^{(t)} \cdot x^{(t)}\big) - y \Big), \tag{8}$$

and replace (7) by

$$g_{\mathrm{SWDC}}^{(t)} \;=\; S^{(t)^H} A^H\big(r_w^{(t)}\big) \tag{9}$$

yielding the SWDC update $z^{(t)} \;=\; x^{(t)} - \tau_t \, g_{\mathrm{SWDC}}^{(t)}$. When $w^{(t)} = M$, (9) recovers the standard DC. As shown in Fig. 2, the SWDC module learns distinct weight maps tailored to uniform, Gaussian, and radial sampling patterns. The learnable weight $w^{(t)}$ is initialized at maximum resolution $H \times W$ and center-cropped to match y; we maintain distinct $w_s^{(t)}$ for each sampling pattern s to enable pattern-specific DC. Although these maps are not explicitly constrained to remain positive, they are initialized to the standard sampled-mask DC solution and are learned independently for each cascade and sampling pattern. Empirically, we did not observe collapse toward zero-valued maps; instead, the learned weights remain spatially structured and pattern-specific, consistent with the ablation gain over scalar learnable DC. A detailed justification of SWDC over classical density compensation is provided in the supplementary material.

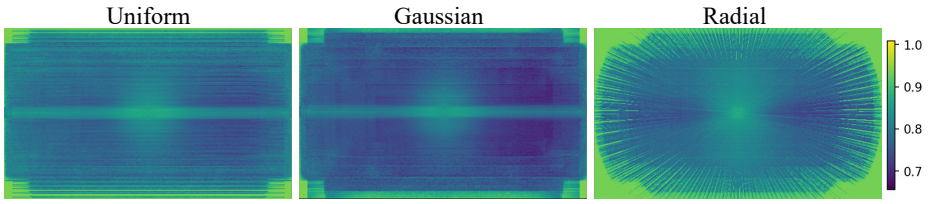

Fig. 2: Example learned sampling-aware data-consistency weight maps showing spatial adaptation to sampling patterns.

3.5 Universal Conditioning Mechanism

We condition every cascade on (i) its cascade index t and (ii) a discrete protocol label y encoding mask type, acquisition type, and acceleration factor. Both are mapped by sinusoidal embeddings followed by two-layer MLPs and *summed* to form a single conditioning vector injected into all transformer blocks, enabling a single model to adapt across cascade depth and protocol.

Let $\phi(\cdot) : \mathbb{R} \to \mathbb{R}^{d_0}$ denote a sinusoidal encoder and let $h_t, h_y : \mathbb{R}^{d_0} \to \mathbb{R}^d$ be two-layer MLPs,

$$c = h_t\big(\phi(t)\big) + h_y\big(\phi(y)\big) \in \mathbb{R}^d. \tag{10}$$

Injection into Transformer blocks. Let $f_\ell^{(t)} \in \mathbb{R}^{H_\ell \times W_\ell \times C_\ell}$ be features at layer ℓ in cascade t. Each transformer block receives c and applies attention and a gated depthwise FFN (GDFN) with an *additive, spatially broadcast* bias derived from c:

$$\tilde{f}_\ell^{(t)} = \mathrm{MDTA}\big(f_\ell^{(t)}\big) \;+\; f_\ell^{(t)}, \tag{11}$$

$$f_{\ell+1}^{(t)} = \mathrm{GDFN}\big(\tilde{f}_\ell^{(t)}\big) \;+\; B_\ell(c), \tag{12}$$

where $B_\ell(c)$ produces a C_ℓ-dimensional vector that is broadcast across (H_ℓ, W_ℓ) using single linear layer. The same c is passed (via a timestep-aware sequential wrapper) to *all* encoder, decoder, and refinement blocks of each cascade in the Restormer.

3.6 Progressive Cascade Expansion

We train SDUM with a curriculum that progressively grows the cascade depth [3, 17]. From a stage with T_{k-1} cascades, the next stage expands to $T_k = 2\,(T_{k-1} - 1)$ by keeping the first and last cascades fixed and duplicating only the interior cascades. All parameters (Restormer weights, CSME weights, DC step sizes, SWDC weights) are warm-started from the previous stage. This "end-fixed, middle-doubling" schedule refines the interior discretization, where residual correction is largest, without disturbing already-converged endpoint parameters, improving gradient flow and convergence at large T.

Concretely, we define a stage-transition index map

$$\pi_k(t) = \begin{cases} 0, & t = 0, \\ T_{k-1} - 1, & t = T_k - 1, \\ 1 + \lfloor (t-1)/2 \rfloor, & \text{otherwise}, \end{cases}$$

so each new cascade $t \in \{0, \ldots, T_k - 1\}$ copies parameters from cascade $\pi_k(t)$ at the previous stage. In practice, this means only interior cascades are duplicated, while endpoint cascades are inherited unchanged:

$$\begin{aligned} \theta^{(t)} &\leftarrow \theta_{\text{prev}}^{(\pi_k(t))}, & \phi^{(t)} &\leftarrow \phi_{\text{prev}}^{(\pi_k(t))}, \\ \tau_t &\leftarrow \tau_{\text{prev}}^{(\pi_k(t))}, & w^{(t)} &\leftarrow w_{\text{prev}}^{(\pi_k(t))}. \end{aligned} \tag{13}$$

Starting from a base depth T_0=6, the schedule used in this work yields T=6 → 10 → 18, which empirically provides stable optimization and consistent gains over direct one-shot training at T=18.

4 Experiments

We evaluate SDUM on **CMRxRecon2024** [32], **CMRxRecon2025** [37], and **fastMRI multi-coil brain** [41]. CMRxRecon2025 extends CMRxRecon2024 by evaluating reconstruction performance under distribution shifts arising from unseen centers, disease categories, field strengths, and pediatric cohorts. Unless stated otherwise, results are reported on official validation splits or challenge leaderboards and measured by SSIM ↑, PSNR ↑, and NMSE ↓.

4.1 Baselines

On fastMRI brain, we compare against classical CS [19], U-Net [25], D5C5 [26], KIKI-Net [8], VN [13], VS-Net [7], ComplexMRI [33], and PC-RNN [4] (fastMRI winner). For CMRxRecon2024/2025, we follow official protocols and compare against top leaderboard methods, including PromptMR+ [34], Hamedani *et al.* [2], vSHARP+ARN [39], IMR [32], Shen *et al.* [27], HierAdaptMR [38], and GENRE-CMR [12].

4.2 Comparisons with State-of-the-Art

CMRxRecon2025: Universal cardiac MRI reconstruction. Tab. 1 reports results on the CMRxRecon2025 leaderboard, which evaluates model generalization across unseen centers, diseases, field strengths, and age groups. A ***single SDUM model*** (T=18) is used for all four subtasks without any task-specific fine-tuning, and it achieves the highest SSIM and PSNR throughout, surpassing strong baselines including PromptMR+ (*32 cas.*) [34], HierAdaptMR [38], Shen *et al.* [27], and GENRE-CMR [12]. On Regular Task 1 (**multi-center generalization**), where both validation and test data come from entirely unseen

Table 1: Quantitative comparison on the CMRxRecon2025 challenge. Results are reported on the official leaderboard across four subtasks. A single SDUM (T=18) is used for all tracks without task-specific fine-tuning.

Regular Task 1: Multi-center				**Regular Task 2: Multiple diseases**			
Method	SSIM	PSNR	NMSE	Method	SSIM	PSNR	NMSE
HierAdaptMR [38]	0.870	31.702	0.019	HierAdaptMR [38]	0.864	32.520	0.021
Shen *et al.* [27]	0.878	32.19	0.017	Shen *et al.* [27]	0.872	32.998	0.017
PromptMR+ [34]	0.891	32.919	0.014	PromptMR+ [34]	0.879	33.422	0.014
SDUM (Ours)	**0.895**	**33.179**	**0.014**	**SDUM (Ours)**	**0.880**	**33.538**	**0.014**
Special Task 1: 5T				**Special Task 2: Pediatric imaging**			
Method	SSIM	PSNR	NMSE	Method	SSIM	PSNR	NMSE
HierAdaptMR [38]	0.888	33.290	0.018	GENRE-CMR [12]	0.872	31.656	0.033
Shen *et al.* [27]	0.893	33.605	0.017	HierAdaptMR [38]	0.880	32.014	0.030
PromptMR+ [34]	0.895	33.824	0.016	Shen *et al.* [27]	0.887	32.436	0.028
SDUM (Ours)	**0.901**	**34.225**	**0.015**	**SDUM (Ours)**	**0.905**	**33.478**	**0.020**

Table 2: Quantitative comparison on the CMRxRecon2024 challenge. Results are reported on the official validation leaderboard across Task 1 and Task 2. SDUM is configured to T=18.

Task 1: Uniform (4×, 8×, 10×)				**Task 2: k-t Gauss./Rad./Unif.** (4×–24×)			
Method	SSIM	PSNR	NMSE	Method	SSIM	PSNR	NMSE
vSHARP w/ ARN [39]	0.910	34.449	0.022	IMR [44]	0.885	32.479	0.034
Hamedani *et al.* [2]	0.912	34.522	0.022	vSHARP w/ ARN [39]	0.888	32.712	0.033
PromptMR [35]	0.907	34.053	0.023	PromptMR [35]	0.886	32.492	0.033
PromptMR+ (*12 cas.*) [36]	0.915	34.677	0.020	PromptMR+ (*12 cas.*) [36]	0.896	33.078	0.030
PromptMR+ (*32 cas.*) [34]	0.921	35.152	0.018	PromptMR+ (*32 cas.*) [34]	0.907	33.812	0.026
SDUM (Ours)	**0.931**	**35.700**	**0.016**	**SDUM (Ours)**	**0.911**	**33.948**	**0.025**

medical centers, SDUM attains an absolute gain of +0.26 dB over the second-best method. On Regular Task 2 (**multi-disease evaluation**), which assesses robustness to unseen cardiovascular conditions with undisclosed disease labels, SDUM maintains superior performance. For the two special tracks (**5T** and **Pediatric**), SDUM generalizes to field strengths and age distributions absent from training, exceeding the next-best methods by up to +1.0 dB PSNR. These results validate our hypothesis: a single model conditioned on protocol physics can generalize across the full spectrum of cardiac MRI heterogeneity.

CMRxRecon2024. Tab. 2 reports results on the official CMRxRecon2024 leaderboard across two subtasks. SDUM outperforms the 2024 winning method PromptMR+ [34] by 0.55 dB in Task 1 and provides consistently higher fidelity under both spatially uniform and temporally varying sampling conditions. A paired-case analysis over the full validation set further shows that SDUM beats PromptMR+ in 90.3% of Task 1 cases and 93.9% of Task 2 cases, with mean gains of +1.091 dB and +1.060 dB, respectively. Qualitative examples (Fig. 3) show clearer myocardial boundaries and fewer residual aliasing artifacts.

4.3 Ablation Study

We conduct ablations on the CMRxRecon2025 challenge validation leaderboard. Results are summarized in Tab. 3.

Backbone variants. We evaluate different reconstruction backbones, U-Net [25], DiT [21], TAU [29], and Restormer [40] using T=6 cascades with approximately the same number of trainable parameters. Restormer achieves the best overall

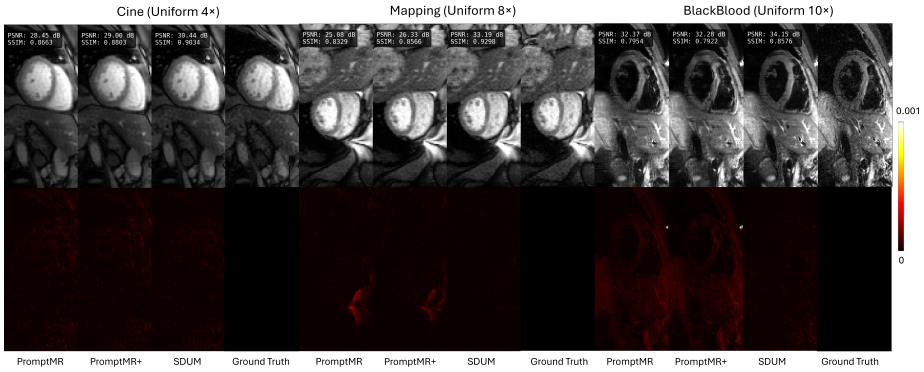

Fig. 3: Qualitative comparison across cardiac imaging modalities. Three representative cases from CMRxRecon2024 Task 1 validation set show the superior performance of SDUM over PromptMR and PromptMR+. For each modality, the top row shows reconstructions and the bottom row shows absolute-error maps, where brighter red/yellow regions indicate larger residual artifacts relative to the fully sampled reference.

performance (backbone block in Tab. 3(a)) by coupling efficient global attention with strong local modeling.

Down-sampling depth. We vary the number of down-sampling layers within each SDUM cascade with T=6. As shown in the DS block of Tab. 3(b), two layers provide the best balance between contextual coverage and spatial fidelity (32.09 dB). Fewer layers hinder abstraction (training fails at one layer), while deeper hierarchies lose fine texture fidelity.

CSME variants. We compare using a single shared CSME versus per-cascade estimators. The CSME block in Tab. 3(d) shows that the multi-CSME configuration improves PSNR by +0.51 dB, confirming the value of progressive sensitivity refinement.

Data consistency formulations. We ablate different DC layers: simple learnable DC, weighted DC (WDC), and our SWDC. The DC block in Tab. 3(e) shows that SWDC achieves the best result, outperforming learnable DC by +0.43 dB.

UC and iterative behavior. The UC/iteration block in Tab. 3(f) shows that enabling UC consistently improves quality (+0.38 dB). Three cascades without iterative refinement outperform a single cascade with three iterations at similar compute, indicating that unrolled depth is more beneficial than additional iterations under limited compute.

4.4 Scaling Analysis

Width scaling. We adjust network width from 7.2M to 78.7M parameters while fixing T=1. The single-cascade capacity block in Tab. 3(c) shows that performance improves steadily up to ∼42M parameters but saturates thereafter, suggesting that depth T yields more efficient capacity scaling than width.

Table 3: Ablation study on CMRxRecon2025 challenge. Results are reported on the official leaderboard of Regular Task 1 (multi-center evaluation). Default configurations are marked in gray.

Study	Setting	SSIM	PSNR	NMSE
(a) Backbone variants (T=6)				
	U-Net [25]	0.811	28.734	0.032
	DiT [21]	0.825	29.542	0.028
	TAU [29]	0.873	31.597	0.018
	Restormer [40]	**0.881**	**32.090**	**0.017**
(b) Down-sampling (DS) layers				
	4	0.872	31.544	0.019
	3	0.877	31.787	0.018
	2	**0.881**	**32.090**	**0.017**
	1	*Training failed*		
(c) Single-cascade capacity scaling (different widths, T=1)				
	7.2 M params.	0.784	27.605	0.042
	13.4 M params.	0.803	28.271	0.036
	42.2 M params.	0.814	28.728	0.033
	78.7 M params.	**0.821**	**28.978**	**0.031**
(d) CSME variants (*Multiple*: one CSME per cascade)				
	Single	0.840	29.694	0.026
	Multiple	**0.845**	**30.205**	**0.025**
(e) Data consistency (DC) layer variants (WDC: Weighted DC)				
	Simple Learnable	0.872	31.661	0.018
	WDC	0.877	31.787	0.017
	SWDC	**0.881**	**32.090**	**0.017**
(f) Effect of UC and iterative reconstruction				
	T=1, Iter.=3, w/o UC	0.835	29.693	0.027
	T=1, Iter.=3, w/ UC	0.843	30.071	0.025
	T=3, Iter.=1, w/ UC	**0.845**	**30.205**	**0.025**

Depth scaling. Fig. 4 shows the scaling trend with respect to unrolled depth T. As the number of cascades increases from 1 to 18 (42 M to 759 M parameters), PSNR improves monotonically, following an approximately linear relationship with log(# parameters) (r=0.986, R^2=0.973). This predictable scaling provides practical guidance for trading compute against reconstruction fidelity.

Data scaling. Tab. 4 reports SDUM (T=18) trained on 40%, 80%, and 100% of the training data and evaluated on CMRxRecon2025 Regular Task 1. Performance improves consistently from 32.72 to 33.18 dB PSNR, from 0.882 to 0.895 SSIM, and from 0.016 to 0.014 NMSE.

The gains are monotonic but not uniform: 40%→80% yields +0.33 dB, while 80%→100% yields +0.13 dB. This indicates diminishing marginal returns with scale, yet no saturation at the current regime. For the community, this suggests that simply adding more samples helps, but scaling should increasingly empha-

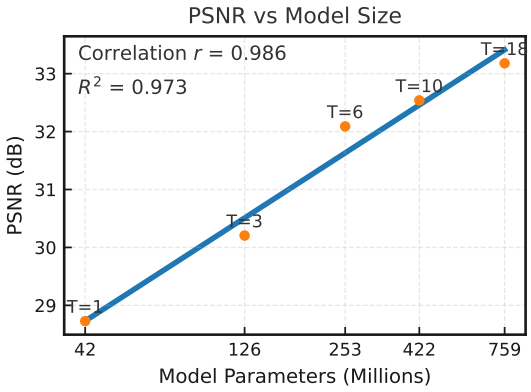

Fig. 4: Scaling behavior of unrolled depth T. Reconstruction PSNR follows an approximately linear relationship with log(# parameters), with r=0.986 and R^2=0.973 up to T=18.

Table 4: Data scaling behavior. SDUM (T=18) trained on different fractions of the training data, evaluated on CMRxRecon2025 Regular Task 1.

Data Fraction	SSIM	PSNR (dB)	NMSE
40%	0.882	32.72	0.016
80%	0.890	33.05	0.015
100%	**0.895**	**33.18**	**0.014**

size *data diversity* (vendors, trajectories, pathologies, field strengths, and age groups) rather than only increasing volume within already-dominant distributions.

Inference efficiency. On a single NVIDIA H100 GPU, the full T=18 model completes inference in ∼1.0 s using ∼6 GB memory at 256×256, and under 4 s with ∼11 GB peak memory at 328×768, making it practical for near–real-time clinical deployment (detailed hardware settings are in the supplementary material).

4.5 Cross-Anatomy Generalization: fastMRI Brain

To verify that SDUM's design principles transfer beyond cardiac MRI, we train a *separate* model on fastMRI brain data [41] with T=6 cascades. Tab. 5 reports results on the fastMRI multi-coil brain benchmark at 4× and 6× acceleration. SDUM consistently outperforms prior methods, exceeding the strong recurrent baseline PC-RNN [4] by +1.8 dB PSNR, validating the scalable unrolled Restormer backbone and SWDC. These results demonstrate that the architectural design, while developed and validated primarily for cardiac MRI, generalizes to a different anatomy and dataset, suggesting the approach is not overfit to cardiac-specific characteristics.

Table 5: Quantitative comparison on fastMRI multi-coil brain. We follow the setting of PC-RNN [4] at 4× and 6× acceleration.

Method	4×		6×	
	SSIM	PSNR	SSIM	PSNR
CS [19]	0.904	35.2	0.851	31.2
U-Net [25]	0.944	36.3	0.925	34.3
VN [13]	0.938	36.2	0.900	32.6
VS-Net [7]	0.950	37.6	0.921	34.3
KIKI-Net [8]	0.956	37.9	0.935	35.3
D5C5 [26]	0.955	38.0	0.931	34.9
ComplexMRI [33]	0.953	37.9	0.933	35.2
PC-RNN [4]	0.971	40.8	0.962	38.9
SDUM (Ours)	**0.979**	**42.6**	**0.973**	**40.8**

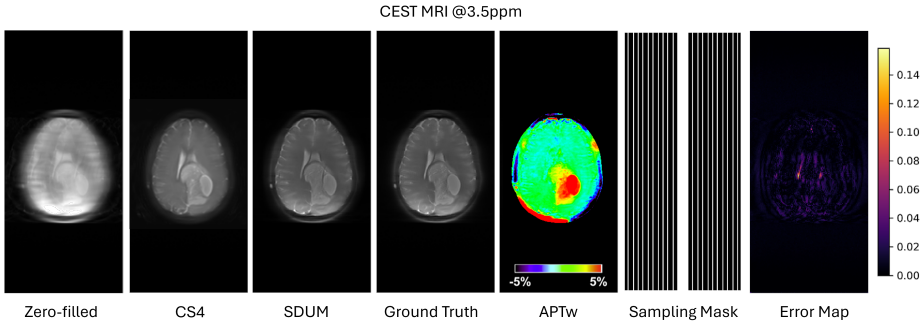

Fig. 5: Zero-shot CEST MRI reconstruction on in-house data. Left to right: zero-filled, scanner CS4, SDUM, fully sampled reference, amide proton transfer-weighted (APTw) map (± 3.5 ppm), sampling mask, and absolute error. SDUM reaches 43.57 dB PSNR and 0.9769 SSIM without fine-tuning.

4.6 Zero-Shot Reconstruction of Chemical Exchange Saturation Transfer (CEST) MRI

For out-of-distribution evaluation, we apply SDUM trained on fastMRI brain directly to in-house chemical exchange saturation transfer (CEST) MRI from brain-tumor patients acquired on a 3T Philips scanner, with retrospectively 4× undersampled Cartesian multi-coil k-space. Because this vendor, pulse sequence, and image contrast are absent from training, this is a strict zero-shot test. At test time, the conditioning inputs use the cascade index, Cartesian sampling label, 4× acceleration factor, and an "unknown contrast" label for the unavailable CEST-specific protocol category; no CEST data are used for fine-tuning. As shown in Fig. 5, SDUM suppresses strong aliasing and recovers finer structures and lesion contrast more faithfully than zero-filled and scanner CS4 reconstructions. Quantitatively, SDUM reaches 43.57 dB PSNR and 0.9769 SSIM against the fully sampled reference while preserving downstream APTw-map fidelity, indicating robust cross-vendor and cross-protocol generalization.

5 Discussion

SDUM integrates a Restormer backbone, per-cascade CSME, sampling-aware weighted data consistency, universal conditioning, and progressive cascade expansion within a single cardiac MRI reconstruction framework. Across CMRxRecon 2024/2025 tracks, the model maintains strong performance without task-specific fine-tuning, outperforms PromptMR+ on more than 90% of paired CMRxRecon2024 validation cases, and shows zero-shot transfer to unseen-protocol CEST MRI (43.57 dB). Results on fastMRI (+1.8 dB over PC-RNN) further suggest potential generalization beyond the primary training domain.

To connect benchmark improvements with perceived image quality, the supplementary material reports a preliminary blinded radiologist scoring study on CMRxRecon2025. The results suggest that SDUM is often indistinguishable from fully sampled references and is preferred for artifact suppression, but they should be interpreted as early reader evidence rather than clinical deployment validation.

The scaling experiments indicate that depth and data contribute differently to performance. PSNR follows an approximately logarithmic trend with parameter count up to T=18 (r=0.986), while gains from additional data diminish at higher fractions (40%→80%: +0.33 dB; 80%→100%: +0.13 dB). These observations motivate future work on compute-efficient scaling, diversity-aware data selection, and standardized conditioning metadata, alongside broader evaluation protocols that include downstream clinical consistency, robustness, and uncertainty calibration.

Several limitations remain. SWDC weight maps are not yet resolution-adaptive at inference, and training deeper variants is computationally expensive. In addition, the present scaling analysis does not fully characterize compute-optimal trade-offs. Although preliminary cross-anatomy results are encouraging, establishing a single anatomy-agnostic MRI reconstruction model requires further validation. Finally, prospective clinical studies and task-specific diagnostic evaluations are still needed before SDUM can be considered deployment-ready.

6 Conclusion

In summary, SDUM provides a single conditional reconstruction framework that achieves strong performance across multiple cardiac MRI reconstruction settings without task-specific fine-tuning. The observed depth and data scaling behavior indicates that further progress is likely to come from jointly optimizing model capacity, data diversity, and training compute. These findings support continued investigation of unified, scalable reconstruction models as a practical path toward more general MRI reconstruction systems.

Acknowledgments. This study was supported partially by grants from the NIH (R01AG069179, R01CA276221, and R37CA248077).

References

1. Aggarwal, H.K., Mani, M.P., Jacob, M.: Modl: Model-based deep learning architecture for inverse problems. IEEE Transactions on Medical Imaging **38**(2), 394–405 (2019)
2. Anvari Hamedani, K., Razizadeh, N., Nabavi, S., Ebrahimi Moghaddam, M.: An all-in-one approach for accelerated cardiac mri reconstruction. In: International Workshop on Statistical Atlases and Computational Models of the Heart. pp. 464–475. Springer (2024)
3. Bengio, Y., Louradour, J., Collobert, R., Weston, J.: Curriculum learning. In: Proceedings of the 26th Annual International Conference on Machine Learning. pp. 41–48 (2009)
4. Chen, E.Z., Wang, P., Chen, X., Chen, T., Sun, S.: Pyramid convolutional rnn for mri image reconstruction. IEEE Transactions on Medical Imaging **41**(8), 2033–2047 (2022)
5. Chung, H., Ye, J.C.: Score-based diffusion models for accelerated mri. Medical Image Analysis **80**, 102479 (2022)
6. Chung, H., Ye, J.C.: Deep diffusion image prior for efficient OOD adaptation in 3D inverse problems. In: Computer Vision – ECCV 2024. Lecture Notes in Computer Science, vol. 15133, pp. 432–455. Springer (2025). https://doi.org/10.1007/978-3-031-73226-3_25
7. Duan, J., Schlemper, J., Qin, C., Ouyang, C., Bai, W., Biffi, C., Bello, G., Statton, B., O'regan, D.P., Rueckert, D.: Vs-net: Variable splitting network for accelerated parallel mri reconstruction. In: International conference on medical image computing and computer-assisted intervention. pp. 713–722. Springer (2019)
8. Eo, T., Jun, Y., Kim, T., Jang, J., Lee, H.J., Hwang, D.: Kiki-net: cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. Magnetic resonance in medicine **80**(5), 2188–2201 (2018)
9. Feng, L., Grimm, R., Block, K.T., Chandarana, H., Kim, S., Xu, J., Axel, L., Sodickson, D.K., Otazo, R.: 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)
10. Griswold, M.A., Jakob, P.M., Heidemann, R.M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., Haase, A.: Generalized autocalibrating partially parallel acquisitions (grappa). Magnetic Resonance in Medicine **47**(6), 1202–1210 (2002)
11. Guo, P., Mei, Y., Zhou, J., Jiang, S., Patel, V.M.: Reconformer: Accelerated mri reconstruction using recurrent transformer. IEEE transactions on medical imaging **43**(1), 582–593 (2023)
12. Hamedani, K.A., Razizadeh, N., Nabavi, S., Moghaddam, M.E.: Genre-cmr: Generalizable deep learning for diverse multi-domain cardiac mri reconstruction. arXiv preprint arXiv:2508.20600 (2025)
13. Hammernik, K., Klatzer, T., Kobler, E., Recht, M.P., Sodickson, D.K., Pock, T., Knoll, F.: Learning a variational network for reconstruction of accelerated mri data. Magnetic resonance in medicine **79**(6), 3055–3071 (2018)
14. Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D.d.L., Hendricks, L.A., Welbl, J., Clark, A., et al.: Training compute-optimal large language models. arXiv preprint arXiv:2203.15556 (2022)
15. Jalal, A., Arvinte, M., Daras, G., Price, E., Dimakis, A.G., Tamir, J.I.: Robust compressed sensing mri with deep generative priors. In: Advances in Neural Information Processing Systems. vol. 34, pp. 14938–14954 (2021)

16. Kaplan, J., McCandlish, S., Henighan, T., Brown, T.B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., Amodei, D.: Scaling laws for neural language models. arXiv preprint arXiv:2001.08361 (2020)
17. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of gans for improved quality, stability, and variation. In: International Conference on Learning Representations (2018)
18. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. pp. 1833–1844 (2021)
19. Lustig, M., Donoho, D.L., Santos, J.M., Pauly, J.M.: Compressed sensing mri. IEEE signal processing magazine **25**(2), 72–82 (2008)
20. Otazo, R., Candès, E., Sodickson, D.K.: 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)
21. Peebles, W., Xie, S.: Scalable diffusion models with transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4195–4205 (2023)
22. Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: Film: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 32 (2018)
23. Pruessmann, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P.: Sense: sensitivity encoding for fast mri. Magnetic Resonance in Medicine **42**(5), 952–962 (1999)
24. Romano, Y., Elad, M., Milanfar, P.: The little engine that could: Regularization by denoising (red). SIAM Journal on Imaging Sciences **10**(4), 1804–1844 (2017)
25. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
26. Schlemper, J., Caballero, J., Hajnal, J.V., Price, A., Rueckert, D.: A deep cascade of convolutional neural networks for mr image reconstruction. In: International conference on information processing in medical imaging. pp. 647–658. Springer (2017)
27. Shen, C.: "chushu shen (chushu) — synapse profile". https://www.synapse.org/Profile:3534520 (2025), accessed: 2025-11-12
28. Sriram, A., Zbontar, J., Murrell, T., Defazio, A., Zitnick, C.L., Yakubova, N., Knoll, F., Johnson, P.: End-to-end variational networks for accelerated mri reconstruction. arXiv preprint arXiv:2004.06688 (2020)
29. Tan, C., Gao, Z., Wu, L., Xu, Y., Xia, J., Li, S., Li, S.Z.: Temporal attention unit: Towards efficient spatiotemporal predictive learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 18770–18782 (2023)
30. Uecker, M., Lai, P., Murphy, M.J., Virtue, P., Elad, M., Pauly, J.M., Vasanawala, S.S., Lustig, M.: Espirit—an eigenvalue approach to autocalibrating parallel mri: where sense meets grappa. Magnetic Resonance in Medicine **71**(3), 990–1001 (2014)
31. Venkatakrishnan, S.V., Bouman, C.A., Wohlberg, B.: Plug-and-play priors for model based reconstruction. In: 2013 IEEE Global Conference on Signal and Information Processing. pp. 945–948. IEEE (2013)
32. Wang, F., Wang, Z., Li, Y., Lyu, J., Qin, C., Wang, S., Guo, K., Sun, M., Huang, M., Zhang, H., et al.: Towards universal learning-based model for cardiac image reconstruction: Summary of the cmrxrecon2024 challenge. arXiv preprint arXiv:2503.03971 (2025)

33. Wang, S., Cheng, H., Ying, L., Xiao, T., Ke, Z., Zheng, H., Liang, D.: Deepcomplexmri: Exploiting deep residual network for fast parallel mr imaging with complex convolution. Magnetic resonance imaging **68**, 136–147 (2020)
34. Xin, B., Ye, M., Axel, L., Metaxas, D.N.: Enhanced deep unrolled models applied to the cmrxrecon2024 challenge. In: International Workshop on Statistical Atlases and Computational Models of the Heart. pp. 289–300. Springer (2024)
35. Xin, B., Ye, M., Axel, L., Metaxas, D.N.: Fill the k-space and refine the image: Prompting for dynamic and multi-contrast mri reconstruction. In: Camara, O., Puyol-Antón, E., Sermesant, M., Suinesiaputra, A., Tao, Q., Wang, C., Young, A. (eds.) Statistical Atlases and Computational Models of the Heart. Regular and CMRxRecon Challenge Papers. pp. 261–273. Springer Nature Switzerland, Cham (2024)
36. Xin, B., Ye, M., Axel, L., Metaxas, D.N.: Rethinking deep unrolled model for accelerated mri reconstruction. In: European Conference on Computer Vision. pp. 164–181. Springer (2024)
37. Xu, D.: Cmrxrecon2025. Web page, https://www.synapse.org/Synapse:syn59814210/wiki/628454 (2025), accessed: 2025-11-05
38. Xu, R., Oksuz, I.: Hieradaptmr: Cross-center cardiac mri reconstruction with hierarchical feature adapters. arXiv preprint arXiv:2508.13026 (2025)
39. Yiasemis, G., Moriakov, N., Sonke, J.J., Teuwen, J.: Deep multi-contrast cardiac mri reconstruction via vsharp with auxiliary refinement network. In: International Workshop on Statistical Atlases and Computational Models of the Heart. pp. 183–192. Springer (2024)
40. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5728–5739 (2022)
41. Zbontar, J., Knoll, F., Sriram, A., Murrell, T., Huang, Z., Muckley, M.J., Defazio, A., Stern, R., Johnson, P., Bruno, M., et al.: fastmri: An open dataset and benchmarks for accelerated mri. arXiv preprint arXiv:1811.08839 (2018)
42. Zhai, X., Kolesnikov, A., Houlsby, N., Beyer, L.: Scaling vision transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12104–12113 (2022)
43. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Transactions on Image Processing **26**(7), 3142–3155 (2017)
44. Zhang, L.: "synapse profile 3472363". https://www.synapse.org/Profile:3472363 (2025), accessed: 2025-11-12