This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Dual-Registration Reciprocal Super-Resolution for Prior-Guided Real-Time 4D-MRI

Yinghui Wang[1†], Mingyuan Meng[2†], Lu Wang[3], and Jing Cai[1] ✉

[1] The Hong Kong Polytechnic University, Hong Kong SAR, China
[2] Zhongguancun Academy, Beijing, China
[3] Northeastern University, Liaoning, China
Jing.Cai@polyu.edu.hk

Abstract. Real-time 4D-MRI promises to enhance adaptive radiotherapy through superior soft-tissue contrast and instantaneous motion visualization. Unfortunately, its clinical implementation is confronted by aggressive spatial undersampling to achieve real-time acquisition, which incurs severe image degradation. Super-resolution (SR) offers a viable solution to recover image quality from degraded inputs; however, applying it to real-time 4D-MRI faces two key challenges: how to precisely reconstruct intricate human anatomy (i) from highly corrupted inputs without essential structural details, and (ii) in the absence of ground-truth for supervised training. Regarding these challenges, we propose DuReg-SR, a dual-registration reciprocal SR framework. Through dual reciprocal registration, DuReg-SR effectively (i) incorporates routinely acquired diagnostic MRI scans as anatomical priors for recovering authentic anatomy, and (ii) leverages static breath-hold high-resolution (HR) scans as pseudo ground-truth for training. For reciprocal SR and registration, DuReg-SR introduces a Hierarchically Recursive Mutual Network (HRMNet), which performs both tasks in a coarse-to-fine hierarchy to progressively infuse anatomical priors for anatomy reconstruction. Further, an Anatomically-Constrained Mutual Training (ACMT) scheme is proposed, which spatially warps a single static HR scan to supervise dynamic 4D-MRI frames while preserving anatomical consistency across temporal sequences. Extensive experiments show that DuReg-SR achieves state-of-the-art performance in both visual quality and quantitative metrics.

Keywords: Real-time 4D-MRI · Image Registration · Super-Resolution.

1 Introduction

Real-time four-dimensional magnetic resonance imaging (4D-MRI) holds considerable promise in MR-guided radiotherapy of abdominal cancers, offering continuous high-contrast volumetric visualization of moving anatomies [19, 21]. This is critical for real-time adaptive beam delivery and gating, facilitating safe dose

†Yinghui Wang and Mingyuan Meng contributed equally to this work.

escalation precisely to the target while sparing surrounding normal tissues [15, 24]. However, acquiring high-quality real-time 4D-MRI remains an unmet problem. Conventional MRI acquisition requires tens of seconds to minutes per scan [17], far exceeding the sub-300 ms latency demanded by real-time imaging protocols [25]. While aggressive k-space undersampling can accelerate acquisition, it inevitably degrades spatial resolution and introduces severe aliasing artifacts.

Super-resolution (SR) is widely recognized for reconstructing high-resolution (HR) images from low-resolution (LR) inputs [12]. However, its application to real-time 4D-MRI faces two key challenges. First, the extreme undersampling rates required for real-time acquisition severely corrupt 4D-MRI frames, stripping away critical anatomical details. Reconstructing complex human anatomy from such degraded inputs is highly intractable and often yields over-smoothed and implausible outputs [20]. Second, SR typically relies on spatially aligned HR-LR image pairs for supervised training [6]. Unfortunately, perfectly aligned ground-truth is technically unattainable in 4D-MRI due to hardware constraints.

For the first challenge, we identify a valuable yet underutilized resource within standard clinical workflows: diagnostic scans (e.g., T1/T2-weighted MRI). Acquired prior to radiotherapy, these scans capture detailed patient-specific anatomy that can serve as the reference for reconstructing authentic structures from heavily corrupted 4D-MRI frames. For the second challenge, high-quality breath-hold MRI scans (acquired over 10-20 seconds) present a potential alternative as pseudo ground-truth for model training [23, 27]. It should be noted that, unlike routinely available diagnostic scans, breath-hold HR scans require both identical acquisition protocols to 4D-MRI and a prolonged, stable breath-hold, which is often infeasible for vulnerable patient populations [1]. Therefore, diagnostic imaging offers a more clinically accessible prior for deployment, while breath-hold HR MRI is only used during training.

Effective utilization of the reference's anatomical prior and breath-hold HR supervision necessitates addressing spatial misalignment between these static scans and each dynamic 4D-MRI frame. This misalignment, caused by physiological motion, leads to anatomical distortions and limits reconstruction accuracy. Image registration thus becomes a critical enabling technique. Recent multi-contrast SR methods have explored integrating registration to erase spatial misalignment between the LR and reference [5]. For instance, Chen et al. [3] leveraged deformable attention mechanisms to estimate spatial misalignment, while Lei et al. [11] incorporated a spatial alignment module to compensate for misaligned structures across images. Nevertheless, these methods are ill-adapted to abdominal 4D-MRI due to two technical barriers. First, designed primarily for brain imaging with relatively rigid deformations, these methods employed one-pass registration without considering advanced techniques such as progressive or recursive registration [22, 16], which can be difficult to handle complex non-linear deformations stemming from respiratory and peristaltic motion. Second, their training relied on synthetically aligned LR-HR pairs generated via retrospective downsampling, thereby artificially sidestepping the real-world issue

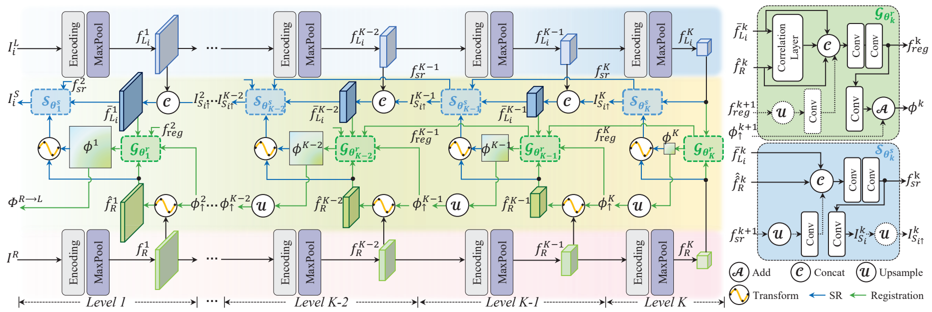

Fig. 1: Detailed Architecture of Hierarchically Recursive Mutual Network (HRMNet).

of inter-fraction misalignment [13, 14]. Thus, applying breath-hold HR scans as supervised labels necessitates additional registration-guided training strategies.

In this study, we propose a Dual-Registration Reciprocal Super-resolution framework (dubbed **DuReg-SR**) for real-time 4D-MRI, wherein SR and registration interact through two mutually-reinforcing cycles to exploit anatomical priors and supervisions. To our knowledge, this is the first work to incorporate anatomical priors for real-time 4D-MRI, marking a paradigm shift from conventional single-image SR to prior-guided reconstruction. Here, we adopt T1-weighted (T1w) MRI as the reference, given its protocol similarity to our 4D-MRI. Two technical contributions are introduced by our DuReg-SR: (i) A Hierarchically Recursive Mutual Network (**HRMNet**) for T1w-guided prior infusion, where SR and registration are progressively integrated in a coarse-to-fine manner, enabling reciprocal feature interaction and prior knowledge infusion at each level. (ii) An Anatomically-Constrained Mutual Training (**ACMT**) scheme for HR-constrained anatomical supervision, where our HRMNet and an add-on registration network are reciprocally trained with a contrastive-based artifact-resilient loss. Extensive experiments on in-house 4D-MRI and public CHAOS [8] demonstrate the state-of-the-art performance of DuReg-SR in both visual quality and quantitative metrics.

2 Methodology

Let $\{I_i^L\}_{i=1}^N$ (with $I_i^L \in \mathbb{R}^{H\times W\times D}$) denote the LR 4D-MRI sequence of N frames, $I^R \in \mathbb{R}^{H\times W\times D}$ the patient-specific reference, and $I^H \in \mathbb{R}^{H\times W\times D}$ the static HR scan. Our objective is to learn an SR network, HRMNet ($\mathcal{H}_{\theta_1}$), that reconstructs each high-quality frame as $I_i^S = \mathcal{H}_{\theta_1}(I_i^L, I^R)$ by fusing anatomical priors from I^R. Additionally, the ACMT scheme supervises the entire output sequence $\{I_i^S\}_{i=1}^N$ using the static scan I^H, achieving visual fidelity to I^H while preserving the spatial geometry of the original LR input.

2.1 Hierarchically Recursive Mutual Network

As shown in Fig. 1, HRMNet adopts a dual-stream U-shaped architecture with two encoders and an inverted-pyramid decoder. The decoder interweaves SR and registration pathways that cooperate across K hierarchical scales. At each level, the SR branch uses aligned reference features as semantic guidance, while the registration branch leverages the structurally enhanced SR output to refine spatial correspondences. Through this recursive mutual machanism, I^R is progressively aligned to each frame I_i^L, thereby enabling high-fidelity SR reconstruction.

Two separate encoders (without weight sharing) are employed to extract multi-scale feature pyramids $\{f_{L_i}^k\}_{k=1}^K$ and $\{f_R^k\}_{k=1}^K$ from I_i^L and I^R, respectively. Each encoder comprises K encoding blocks. Every block consists of two consecutive convolutions, each followed by a PReLU activation. After the first block, a max-pooling layer precedes each block to halve the spatial resolution.

The decoder mirrors the encoder's hierarchical structure, performing recursive deformation field estimation and SR reconstruction from the coarsest level $k = K$ to the finest level $k = 1$. At each level k ($k \neq K$), it takes as input not only the current-level encoder features $f_{L_i}^k$ and f_R^k, but also the deformation field ϕ^{k+1} and SR result $I_{S_i}^{k+1}$ propagated from the previous level. Consequently, only residual components at the current scale need to be estimated for both registration and super-resolution, enabling progressive refinement along a coarse-to-fine trajectory and facilitating detailed recovery in a cascaded manner.

Specifically, the deformation field ϕ^{k+1} from the previous level is upsampled via trilinear interpolation to spatially pre-align the prior feature f_R^k, producing $\hat{f}_R^k = f_R^k \circ \mathrm{Up}(\phi^{k+1})$ as the moving feature for the registration module. Within each registration module, a correlation layer computes feature matching responses between the fixed and moving features within a local window. These responses are concatenated with the original features and upsampled registration features from the previous level, then refined through several convolutional layers with kernel size of 3. A final convolutional layer with kernel size of 3 initialized from $\mathcal{N}(1, 10^{-5})$ predicts the residual field $\Delta\phi^k$, which is added to the upsampled field to produce $\phi^k = \mathrm{Up}(\phi^{k+1}) + \Delta\phi^k$. On the other hand, the SR result $I_{S_i}^{k+1}$ from the previous level is upsampled and encoded into enhanced features, then concatenated with the current-level LR features $f_{L_i}^k$ to form the fused feature $\bar{f}_{L_i}^k$. This fused feature is refined through convolutional layers with kernel size of 3, followed by a final convolution with kernel size of 1 to produce the SR result $I_{S_i}^k$ for the current level. Except for the final level, each SR level concludes with a trilinear upsampling layer that doubles the spatial resolution for the next level.

Notably, the fused feature $\bar{f}_{L_i}^k$ also serves as the fixed feature input to the registration module at the same level, leveraging the reconstruction priors from the previous level to enhance feature robustness and thereby improve the accuracy of deformation field estimation. The refined deformation field estimated at the current level can, in turn, warp the reference feature more precisely to achieve structural alignment with the dynamic features, providing high-quality

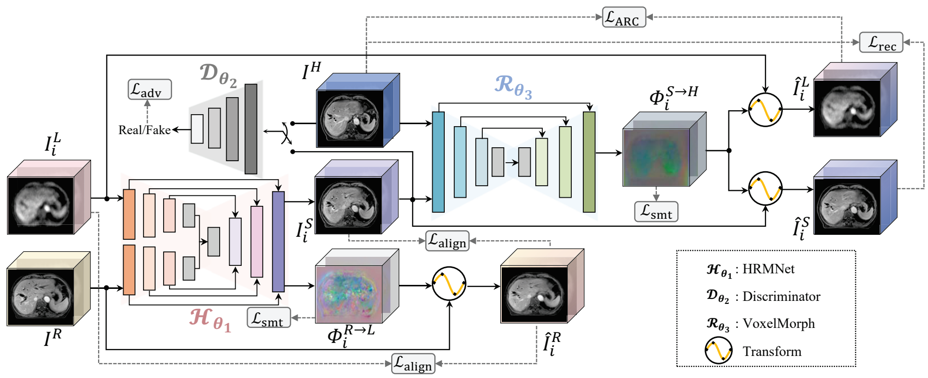

Fig. 2: Illustration of Anatomically-Constrained Mutual Training (ACMT) scheme.

spatial guidance for SR reconstruction: $\hat{\hat{f}}_R^k = \hat{f}_R^k \circ \phi^k$. In this manner, registration and SR are deeply coupled and mutually reinforcing at each level, forming a synergistic optimization mechanism. The overall process is formulated as:

$$\phi^k = \mathcal{G}_{\Theta_k^r}\left(\bar{f}_{L_i}^k,\, \hat{f}_R^k,\, f_{\mathrm{reg}}^{k+1}\right) + \mathrm{Up}(\phi^{k+1}), \tag{1a}$$

$$I_{S_i}^k = \mathcal{S}_{\Theta_k^s}\left(\bar{f}_{L_i}^k \oplus \hat{\hat{f}}_R^k,\, f_{\mathrm{sr}}^{k+1}\right), \tag{1b}$$

where $\mathcal{G}_{\Theta_k^r}$ and $\mathcal{S}_{\Theta_k^s}$ are the registration and SR modules at level k; $\oplus$ denotes channel-wise concatenation; and f_{reg}^{k+1} and f_{sr}^{k+1} are propagated features (available only for $k < K$). After recursive refinement across all levels, the decoder outputs the full-scale deformation field $\Phi_i^{R\to L}$ and the final SR result I_{S_i}.

2.2 Anatomically-Constrained Mutual Training

Our ACMT scheme is illustrated in Fig. 3. A PatchGAN [10] discriminator imposes an adversarial loss $\mathcal{L}_{\mathrm{adv}}$, encouraging I_i^S to visually approximate I^H:

$$\mathcal{L}_{\mathrm{adv}} = \mathbb{E}_{I^H \sim p_{\mathrm{data}}(I^H)}\left[\log \mathcal{D}_{\theta_2}(I^H)\right] + \mathbb{E}_{I_i^S \sim p_{\mathcal{H}_{\theta_1}}(I_i^S)}\left[\log\left(1 - \mathcal{D}_{\theta_2}(I_i^S)\right)\right]. \tag{2}$$

Following [11], we also introduce a mutual information-based alignment loss that steers the reference image I^R toward each LR frame I_i^L, allowing the SR network to exploit structural cues from anatomically corresponding regions. Given the severe degradation of LR inputs, we further leverage the enhanced image I_i^S for more robust alignment assessment:

$$\mathcal{L}_{\mathrm{align}} = \mathcal{L}_{\mathrm{MI}}(I^R \circ \Phi_i^{R\to L}, I_i^L) + \mathcal{L}_{\mathrm{MI}}(I^R \circ \Phi_i^{R\to L}, I_i^S), \tag{3}$$

where $\Phi_i^{R\to L}$ denotes the deformation fields from I^R to I_i^L.

However, GAN-based reconstruction risks incurring geometric distortions, which can be exacerbated by misalignment between I^R and I_i^L. This is particularly detrimental in radiotherapy, where spatial fidelity is paramount. Prior works enforce geometric consistency via patch-wise contrastive losses (e.g., patch-NCE) applied between input and output images (i.e., I_i^L and I_i^S) [4, 18]. In our context, such constraints tend to misinterpret artifacts as valid anatomical structures, impeding both artifact suppression and fine-detail recovery.

To tackle this issue, we introduce a lightweight VoxelMorph registration module $\mathcal{R}_{\theta_3}$ jointly optimized with the SR network $\mathcal{H}_{\theta_1}$ [2]. Rather than aligning the high-quality static scan I^H to each artifact-corrupted LR input I_i^L, $\mathcal{R}_{\theta_3}$ estimates the deformation from the reconstructed SR image I_i^S to I^H. The warped SR image is then supervised against I^H via pixel-wise ℓ_1 loss:

$$\mathcal{L}_{\text{recon}} = \|I_i^S \circ \Phi_i^{S\to H} - I^H\|_1, \tag{4}$$

where $\Phi_i^{S\to H}$ is the deformation fields from I_i^S to I^H. This creates the second mutually reinforcing cycle: enhanced structural clarity in I_i^S improves correspondence estimation, while accurate alignment provides faithful supervision for SR. This reciprocal design ensures both tasks benefit progressively during training.

Beyond this reciprocal cycle, we further enforce geometric consistency through a novel Artifact-Resilient Contrastive (ARC) loss $\mathcal{L}_{\text{ARC}}$. Instead of applying contrastive constraints directly between I_i^L and I_i^S, we leverage the deformation estimated by $\mathcal{R}_{\theta_3}$ to warp the original LR image, and enforce alignment between the warped LR and fixed HR. The underlying rationale is that if I_i^S and I_i^L are anatomically consistent, they should share the same deformation relative to I^H:

$$\mathcal{L}_{\text{ARC}} = -\frac{1}{BN_p}\sum_{b,p}\log\frac{e^{\text{sim}(u_{b,p},v_{b,p})/\tau}}{e^{\text{sim}(u_{b,p},v_{b,p})/\tau} + \sum_{q\neq p} e^{\text{sim}(u_{b,p},v_{b,q})/\tau}}, \tag{5}$$

where B is batch size, N_p patches per image, τ temperature, and $u_{b,p}, v_{b,p}$ are features from corresponding patches of warped LR $\hat{I}_i^L = I_i^L \circ \mathcal{R}_{\theta_3}(I_i^S, I^H)$ and fixed HR I^H, with $\text{sim}(\cdot,\cdot)$ as cosine similarity. By operating on the warped LR rather than the SR output directly, our formulation preserves faithful anatomy while maintaining the SR network's capacity.

The overall objective is a weighted combination of five loss terms:

$$\mathcal{L}_{\text{total}} = \lambda_1\mathcal{L}_{\text{adv}} + \lambda_2\mathcal{L}_{\text{align}} + \lambda_3\mathcal{L}_{\text{recon}} + \lambda_4\mathcal{L}_{\text{ARC}} + \lambda_5\mathcal{L}_{\text{smt}}. \tag{6}$$

We empirically set $\lambda_1 = 1$, $\lambda_2 = 1$, $\lambda_3 = 10$, $\lambda_4 = 0.1$, and $\lambda_5 = 5$. $\mathcal{L}_{\text{smt}}$ applies an ℓ_2 gradient penalty to regularize $\Phi_i^{R\to L}$ and $\Phi_i^{S\to H}$.

3 Experiments

Experiment Setup. We evaluated our method on two datasets: an in-house 4D-MRI dataset of 26 liver cancer patients and the public CHAOS dataset [7]. For in-house dataset, free-breathing 4D-MRI was acquired using a TWIST-VIBE sequence (690ms/frame). To emulate real-time acquisition, we further undersampled the acquired 4D-MRI k-space by retaining only 50% and 25% of the

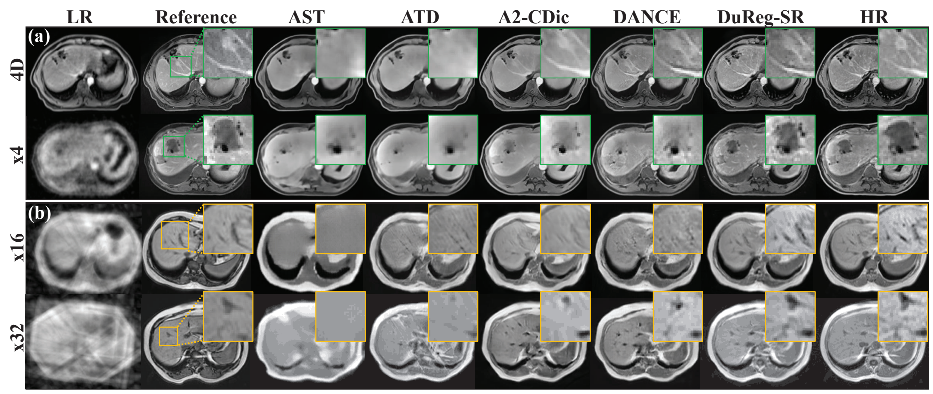

Fig. 3: Visual comparison of SR results on (a) 4D-MRI and (b) CHAOS datasets.

radial spokes, corresponding to frame rates of 345 ms and 173 ms, respectively. Breath-hold scans (11s/frame) served as HR ground-truth, and T1w STAR-VIBE images were employed as the reference. For quantitative evaluation, each static HR scan was registered to its matched motion phase using Elastix [9], with all reported metrics computed exclusively on the aligned phase.

For CHAOS, we retrospectively applied radial undersampling in-phase T1w images at acceleration factors of $\times16$, $\times24$, and $\times32$ to generate highly corrupted LR inputs, while out-of-phase T1w images were retained as the reference. During training, synthetic Gaussian deformations were applied between the LR, reference, and HR to simulate respiratory-induced misalignment [10].

Our framework was implemented in PyTorch 1.10.2 and trained on a single RTX A6000 GPU. The model was optimized with Adam over 100 epochs, using an initial learning rate of 1e$-$4 that decayed cosine-annealing to 1e$-$6. We applied a batch size of 1 for the 4D-MRI dataset and 8 for the CHAOS dataset.

Comparison Study. We compared our DuReg-SR with state-of-the-art SR approaches, including single-image SR (AST [28], ATD [26]) and multi-contrast SR (A2-CDic [11], DANCE [3]) that can utilize misaligned references. For the lack of paired HR-LR ground-truth, all competitors were trained using HR–LR image pairs pre-aligned with Elastix, following [27].

As observed in Fig. 3, our method effectively restores fine anatomical details, such as vascular structures and well-defined tumor boundaries. AST and ATD, lacking patient-specific anatomical priors, produce over-smoothed outputs that miss subtle structures. While A2-CDic and DANCE partially recover missing details, they often introduce structural inconsistencies that resemble the reference rather than the input anatomy. Quantitative results in Table 1 showcase that DuReg-SR consistently outperforms all baselines across PSNR, SSIM, and LPIPS (with AlexNet as the backbone), with gains especially pronounced at higher undersampling factors. These findings establish DuReg-SR as an effective solution for high-quality real-time 4D-MRI under extreme acceleration.

Table 1: Quantitative results on 4D-MRI and CHAOS. Best and second-best results are highlighted in red and blue, respectively. † indicates single-image super-resolution methods; ∗ denotes statistically significant differences compared to our DuReg-SR.

4D-MRI	Original 4D			x2			x4		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
AST†	21.30*	89.67*	13.57*	20.77*	87.60*	16.16*	20.13*	85.78*	17.58*
ATD†	22.30*	91.81*	12.22*	21.69*	89.85*	15.13*	20.71*	87.77*	17.38*
A2-CDic	23.01	93.01*	11.79*	21.82*	92.04*	12.97*	21.60*	90.00*	15.04*
DANCE	22.47*	91.59*	11.56*	21.89*	90.45*	13.12*	21.40*	89.38*	13.73*
DuReg-SR	22.84	93.55	5.11	22.62	93.00	5.77	22.04	92.34	5.94
CHAOS	x16			x24			x32		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
AST†	21.39*	87.20*	18.71*	19.40*	83.74*	22.39*	18.04*	79.56*	26.79*
ATD†	21.46*	87.82*	12.83*	19.11*	83.30*	15.24*	17.62*	79.18*	21.40*
A2-CDic	21.79*	88.43*	17.46*	19.95*	86.26*	18.75*	18.77*	82.98*	21.13*
DANCE	21.54*	88.10*	10.23*	20.59*	86.20*	11.75*	18.61*	82.61*	13.84*
DuReg-SR	24.48	91.35	8.40	22.79	88.88	9.74	21.58	88.35	10.43

Table 2: Ablation Analysis on HRMNet at ×32 (VM: VoxelMorph; Casc.: Cascaded).

Settings	SR		Registration	
	SSIM	LPIPS	NCC	DSC
SISR	77.93	25.38	–	–
Casc. VM & SR	81.84	15.84	85.68	86.53
Joint VM & SR	84.74	13.00	88.71	88.53
Ours ($K = 1$)	84.87	13.64	90.12	89.16
Ours ($K = 2$)	86.20	11.69	90.51	89.30
Ours ($K = 3$)	87.07	11.15	90.84	89.67
Ours ($K = 4$)	88.35	10.43	90.91	90.09

Table 3: Ablation Analysis on ACMT.

Settings	×24		×32	
	SSIM	LPIPS	SSIM	LPIPS
Baseline	85.86	10.60	82.68	13.33
+ $\mathcal{R}_{\theta_3}$	85.63	10.67	83.09	11.71
+ PatchNCE	82.69	11.76	81.71	12.04
w ACMT	88.88	9.74	88.35	10.43
ATD	83.30	15.24	79.18	21.40
w ACMT	85.80	13.81	81.27	20.46
DANCE	86.20	11.75	82.61	13.84
w ACMT	87.68	10.39	84.25	12.45

Ablation Study. We conducted two ablation studies to validate the proposed HRMNet and ACMT scheme. As shown in Table 2, replacing SISR with a pre-trained VoxelMorph and SR network improves performance, confirming the benefit of anatomical priors from the reference. Joint SR-registration training further enhances both tasks, as refined SR outputs improve registration via $\mathcal{L}_{\text{align}}$. Our HRMNet achieves progressive gains in both SR and registration metrics as recursion depth K increases, demonstrating that its coarse-to-fine interaction enables mutual reinforcement between the two tasks.

Table 3 compares different training strategies. The baseline, trained on pre-aligned HR images, yields suboptimal results. Incorporating $\mathcal{R}_{\theta_3}$ provides more faithful supervision for SR, as improved SR outputs in turn benefit registration through joint training. However, introducing PatchNCE loss between SR input and output degrades all metrics, as enforcing strong consistency between SR and LR conflicts with artifact removal and structural recovery. Our ACMT framework, which substitutes PatchNCE with $\mathcal{L}_{\text{ARC}}$, outperforms all variants, underscoring its effectiveness in maintaining anatomical consistency while restoring delicate details. Moreover, applying ACMT to ATD and DANCE also yields consistent improvements, confirming its general applicability.

4 Conclusion

In this study, we have outlined DuReg-SR, a novel framework for real-time 4D-MRI super-resolution that integrates diagnostic T1w MRI as anatomical priors and breath-hold HR as pseudo-labels. It incorporates HRMNet for prior-infused reconstruction and ACMT for anatomically-constrained supervision. Extensive evaluations on two datasets validate that our method accurately reconstructs anatomical structures even under extremely high downsampling factors required for real-time acquisition, consistently outperforming existing approaches by a significant margin.

Acknowledgments. This work was supported by General Research Fund (GRF 15104323), Theme-based Scheme (T45-401/22-N), Seed Funding with Peking University Third Hospital (P0063898), Zhongguancun Academy (XTS0071), and Liaoning Doctoral Startup Fund (2025-BS-13).

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Ahmed, A.H., Zhou, R., Yang, Y., Nagpal, P., Salerno, M., Jacob, M.: Free-breathing and ungated dynamic mri using navigator-less spiral storm. IEEE Transactions on Medical Imaging **39**(12), 3933–3943 (2020)
2. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. IEEE Transactions on Medical Imaging **38**(8), 1788–1800 (2019)
3. Chen, W., Wu, S., Wang, S., Li, Z., Yang, J., Yao, H., Tian, Q., Song, X.: Multi-contrast image super-resolution with deformable attention and neighborhood-based feature aggregation (dance): Applications in anatomic and metabolic mri. Medical Image Analysis **99**, 103359 (2025)
4. Chen, Z., Wei, J., Li, R.: Unsupervised multi-modal medical image registration via discriminator-free image-to-image translation. arXiv preprint arXiv:2204.13656 (2022)
5. Feng, Y., Deng, S., Lyu, J., Cai, J., Wei, M., Qin, J.: Bridging mri cross-modality synthesis and multi-contrast super-resolution by fine-grained difference learning. IEEE Transactions on Medical Imaging **44**(1), 373–383 (2024)
6. Hassan, E., El-Rashidy, N., Elbedwehy, S., Abd El-Hafeez, T., Saber, A., Shams, M.Y.: Exploring the frontiers of image super-resolution: a review of modern techniques and emerging applications. Neural Computing and Applications **37**(22), 17913–17961 (2025)

7. Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., Conze, P.H., Groza, V., Pham, D.D., Chatterjee, S., Ernst, P., Özkan, S., et al.: Chaos challenge-combined (ct-mr) healthy abdominal organ segmentation. Medical Image Analysis **69**, 101950 (2021)
8. Kavur, A.E., Gezer, N.S., Barış, M., Şahin, Y., Özkan, S., Baydar, B., Yüksel, U., Kılıkçıer, , Olut, , Bozdağı Akar, G., Ünal, G., Dicle, O., Selver, M.A.: Comparison of semi-automatic and deep learning based automatic methods for liver segmentation in living liver transplant donors. Diagnostic and Interventional Radiology **26**, 11–21 (Jan 2020). https://doi.org/10.5152/dir.2019.19
9. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: Elastix: a toolbox for intensity-based medical image registration. IEEE Transactions on Medical Imaging **29**(1), 196–205 (2009)
10. Kong, L., Lian, C., Huang, D., Hu, Y., Zhou, Q., et al.: Breaking the dilemma of medical image-to-image translation. Advances in Neural Information Processing Systems **34**, 1964–1978 (2021)
11. Lei, P., Zhang, M., Fang, F., Zhang, G.: Robust deep convolutional dictionary model with alignment assistance for multi-contrast mri super-resolution. IEEE Transactions on Medical Imaging (2025)
12. Lepcha, D.C., Goyal, B., Dogra, A., Goyal, V.: Image super-resolution: A comprehensive review, recent trends, challenges and applications. Information Fusion **91**, 230–260 (2023)
13. Li, G., Lv, J., Tian, Y., Dou, Q., Wang, C., Xu, C., Qin, J.: Transformer-empowered multi-scale contextual matching and aggregation for multi-contrast mri super-resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 20636–20645 (2022)
14. Li, J., Yang, H., Yi, Q., Lu, M., Shi, J., Zeng, T.: High-frequency modulated transformer for multi-contrast mri super-resolution. IEEE Transactions on Medical Imaging (2025)
15. Liao, Y.P., Xiao, H., Wang, P., Li, T., Aguilera, T.A., Visak, J.D., Godley, A.R., Zhang, Y., Cai, J., Deng, J.: Internal target volume estimation for liver cancer radiation therapy using an ultra quality 4-dimensional magnetic resonance imaging. Advances in radiation oncology **10**(6), 101774 (2025)
16. Meng, M., Feng, D., Bi, L., Kim, J.: Correlation-aware coarse-to-fine mlps for deformable medical image registration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9645–9654 (2024)
17. Munoz, C., Fotaki, A., Botnar, R.M., Prieto, C.: Latest advances in image acceleration: all dimensions are fair game. Journal of Magnetic Resonance Imaging **57**(2), 387–402 (2023)
18. Park, T., Efros, A.A., Zhang, R., Zhu, J.Y.: Contrastive learning for unpaired image-to-image translation. In: European Conference on Computer Vision. pp. 319–345. Springer (2020)
19. Stemkens, B., Paulson, E.S., Tijssen, R.H.: Nuts and bolts of 4d-mri for radiotherapy. Physics in Medicine Biology **63**(21), 21TR01 (2018)
20. Terpstra, M.L., Maspero, M., d'Agata, F., Stemkens, B., Intven, M.P., Lagendijk, J.J., Van den Berg, C.A., Tijssen, R.H.: Deep learning-based image reconstruction and motion estimation from undersampled radial k-space for real-time mri-guided radiotherapy. Physics in Medicine Biology **65**(15), 155015 (2020)
21. Wang, C., Yin, F.F.: 4d-mri in radiotherapy. Magnetic Resonance Imaging (2019)
22. Wang, H., Ni, D., Wang, Y.: Recursive deformable pyramid network for unsupervised medical image registration. IEEE Transactions on Medical Imaging **43**(6), 2229–2240 (2024)

23. Wang, Y., Wang, L., Feng, Y., Chen, Z., Qin, J., Li, T., Cai, J.: High spatiotemporal-resolution abdominal 4d-mri through respiratory-synchronized frame collaborative reconstruction. Medical Physics **52**(9), e18101 (2025)
24. Wang, Y., Xiao, H., Wang, J., Wang, L., Li, W., Jiang, Z., Ren, G., Zhi, S., Qian, J., Dai, J., et al.: Artificial intelligence in four-dimensional imaging for motion management in radiation therapy. Artificial Intelligence Review **58**(4), 103 (2025)
25. Xiao, H., Han, X., Zhi, S., Wong, Y.L., Liu, C., Li, W., Liu, W., Wang, W., Zhang, Y., Wu, H., et al.: Ultra-fast multi-parametric 4d-mri image reconstruction for real-time applications using a downsampling-invariant deformable registration (d2r) model. Radiotherapy and Oncology **189**, 109948 (2023)
26. Zhang, L., Li, Y., Zhou, X., Zhao, X., Gu, S.: Transcending the limit of local window: Advanced super-resolution transformer with adaptive token dictionary. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2856–2865 (2024)
27. Zhi, S., Wang, Y., Xiao, H., Bai, T., Li, B., Tang, Y., Liu, C., Li, W., Li, T., Ge, H., et al.: Coarse–super-resolution–fine network (cosf-net): a unified end-to-end neural network for 4d-mri with simultaneous motion estimation and super-resolution. IEEE Transactions on Medical Imaging **43**(1), 162–174 (2023)
28. Zhou, S., Chen, D., Pan, J., Shi, J., Yang, J.: Adapt or perish: Adaptive sparse transformer with attentive feature refinement for image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2952–2963 (2024)