Low-Rank Velocity Fields as a Structural Prior for Unsupervised 4D Medical Image Interpolation

Haojin Li[1], Hengzhuo Wang[2,3], Chang Liu[2], Zhiheng Ma[4], Heng Li[2,5]✉, and Jiang Liu[1,5]✉

[1] Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China
[2] Faculty of Biomedical Engineering, Shenzhen University of Advanced Technology, Shenzhen, China
[3] School of Biomedical Engineering, Shenzhen University Medical School, Shenzhen University, Shenzhen, China
[4] Faculty of Computility Microelectronics, Shenzhen University of Advanced Technology, Shenzhen, China
[5] Research Institute of Trustworthy Autonomous Systems, Southern University of Science and Technology, Shenzhen, China
✉ Corresponding authors: liheng@suat-sz.edu.cn, liuj@sustech.edu.cn

Abstract. Endpoint-only unsupervised 4D medical image interpolation synthesizes intermediate volumes from sparsely sampled sequences with only the start and end volumes available for training; however, this weakly constrained setting often yields intermediates with unstable boundaries and non-physiological motion, limiting interpretability and downstream analysis. We propose low-rank velocity fields as a structural prior, constraining motion to a structured Tucker low-rank velocity field space that decomposes motion into globally shared spatial bases and a compact sample-specific core, thereby encouraging spatially correlated, anatomy-consistent deformation while suppressing voxel-wise high-frequency artifacts. To capture global coordination and local non-rigid details, we model motion in a coarse-to-fine multi-scale scheme and compose scale-wise deformations at inference to synthesize volumes at arbitrary times. We further provide a theoretical analysis showing that, under Tucker parameterization, bounded low-rank parameters control the smoothness energy of the velocity field, offering an energy-based interpretation of smoother motion. Experiments on ACDC and 4D-Lung demonstrate state-of-the-art performance, remaining competitive with methods trained with intermediate-frame supervision, and producing intermediates with improved structural coherence and more stable anatomical contours. The code is available at https://github.com/JingHuaMan/LRVF.

Keywords: 4D Medical Imaging · 4D Image Interpolation · Low-Rank Velocity Fields · Motion Modeling

1 Introduction

4D medical imaging captures anatomical dynamics and supports functional assessment, therapy monitoring, and longitudinal disease tracking [22]. However, practical acquisition constraints often lead to sparse and irregular temporal sampling, motivating 4D interpolation to synthesize intermediate volumes and increase temporal resolution for downstream analysis such as segmentation [17] and reconstruction [15]. Here, we study endpoint-only unsupervised 4D medical interpolation, where training provides only two endpoint volumes.

Prior volumetric interpolation methods are broadly motion-driven, warping endpoints via estimated deformations (or flows) [7, 21, 12], or synthesis-driven, directly generating intermediate volumes [23, 24], and are sometimes combined [11, 5]. For medical imaging, plausible intermediates must preserve anatomical continuity and topology while evolving smoothly [25]. Under endpoint-only supervision, fully voxel-wise, high-dimensional motion parameterizations can overfit locally inconsistent high-frequency motion, leading to boundary instability, local misalignment, and non-physiological warping that degrades interpretability and downstream usability.

Anatomical motion is strongly spatially correlated and partially compressible: coordinated global patterns dominate, while local non-rigid details remain residual variation [18, 9]. This motivates an anatomy-aligned, compact motion representation rather than increased voxel-level freedom; thus we adopt structured low-rank motion modeling to suppress spurious high-frequency motion and improve anatomical consistency and temporal coherence of synthesized intermediates.

Two challenges arise when introducing structured low-rank motion into endpoint-only unsupervised 4D interpolation: (1) how to impose a compact, low-dimensional motion parameterization that captures shared spatial coordination yet adapts to each sample under weak supervision; (2) how to model motion across scales, as coarse global trends and fine non-rigid details call for a multi-scale representation and refinement [19].

To tackle these challenges, we propose a framework for endpoint-only unsupervised 4D medical image interpolation that introduces low-rank velocity fields as a structural prior. The framework separates shared coordinated motion from sample-specific variation and organizes motion modeling in a multi-scale hierarchy for coarse-to-fine refinement, thereby guiding endpoint-only learning toward anatomically plausible dynamics. Our contributions are summarized as follows:

- We present low-rank velocity fields as a structural prior for endpoint-only unsupervised 4D medical interpolation, improving anatomical consistency and temporal coherence, with analysis linking bounded low-rank parameters to smoothness-energy control and DCT bases to high-frequency motion suppression.
- We introduce a Tucker-parameterized low-rank velocity representation with shared spatial bases and a compact, sample-dependent core to capture coordinated motion while retaining individual variation.
- We propose a multi-scale low-rank motion modeling strategy that allocates capacity across resolutions for coarse-to-fine refinement from global coordination to local non-rigid deformation.

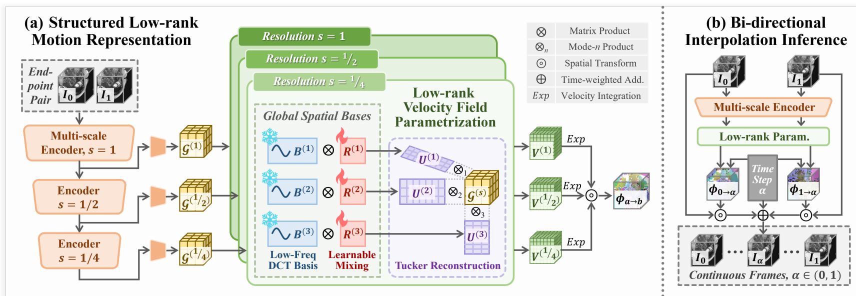

Fig. 1. Overview of the proposed framework. (a) Multi-scale endpoint features are projected into structured Tucker low-rank velocity representations across resolutions. (b) Inference performs bi-directional continuous-time interpolation to produce structurally consistent intermediate volumes.

– Experiments on ACDC and 4D-Lung show strong results, achieving the best structure-oriented scores with competitive reconstruction accuracy.

2 Method

2.1 Problem Definition

Unsupervised 4D interpolation models an image sequence on $\Omega \subset \mathbb{R}^3$ as a continuous-time function $I(t) : \Omega \to \mathbb{R}$. We adopt endpoint-only training, observing only $I_0 := I(0)$ and $I_1 := I(1)$, while intermediate volumes I_α with $\alpha \in (0, 1)$ are unavailable.

Given (I_0, I_1), a model with parameters θ predicts an inter-endpoint velocity field for either ordered pair (I_a, I_b) with $(a, b) \in \{(0, 1), (1, 0)\}$, represented on the voxel grid as a tensor $\mathcal{V}_{a\to b}$. It is converted to a deformation $\phi_{a\to b} = \mathrm{Exp}(\mathcal{V}_{a\to b})$, where $\mathrm{Exp}(\cdot)$ is the exponential map of a stationary velocity field implemented via scaling-and-squaring, and the target endpoint is reconstructed by warping $\hat{I}_b = I_a \circ \phi_{a\to b}$, where $\circ$ applies a deformation field to an image. Learning is driven by an endpoint reconstruction objective:

$$\theta^* \; = \; \arg\min_{\theta} \; \mathbb{E}_{(I_a, I_b)} \Big[\mathcal{L}\big(I_b, \hat{I}_b\big) \; + \; \lambda \, \mathcal{R}\big(\mathcal{V}_{a\to b}\big) \Big], \tag{1}$$

where $\mathcal{L}(\cdot, \cdot)$ is an endpoint similarity measure (e.g., normalized cross-correlation; NCC), and $\mathcal{R}(\cdot)$ regularizes the predicted velocity to promote plausible motion.

2.2 Structured Low-rank Motion Representation

Tucker Parameterization of Velocity Fields: To model high-dimensional non-rigid motion stably under endpoint-only training (Fig. 1(a)), we constrain the predicted velocity to a structured low-rank subspace instead of a fully voxel-wise free form. A 3D velocity field on $\Omega \subset \mathbb{R}^3$, denoted $v(x) = (v_x(x), v_y(x), v_z(x))$

for $x \in \Omega$, is represented on the voxel grid as a fourth-order tensor $\mathcal{V} \in \mathbb{R}^{H\times W\times D\times 3}$. Below, $\mathcal{V}$ denotes this voxel-grid tensor form for low-rank parameterization and regularization.

We decompose $\mathcal{V}$ into globally shared spatial bases and a compact core of combination coefficients [13, 6]:

$$\mathcal{V} \;=\; \mathcal{G} \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^{(3)}, \tag{2}$$

where $\mathcal{G} \in \mathbb{R}^{r\times r\times r\times 3}$ is the low-dimensional core tensor, $U^{(1)} \in \mathbb{R}^{H\times r}$, $U^{(2)} \in \mathbb{R}^{W\times r}$, and $U^{(3)} \in \mathbb{R}^{D\times r}$ are spatial bases along the three axes, and $\times_n$ denotes the mode-n tensor–matrix product. The bases capture domain-consistent, globally shared deformation structure, while $\mathcal{G}$ provides sample-dependent coefficients to combine them; thus, spatial correlations are absorbed by the shared bases and instance variation is governed by the compact core.

To stably construct the low-rank subspace, each spatial basis uses a fixed analytic DCT basis with learnable mixing within the same low-frequency span:

$$U^{(i)} = B^{(i)} R^{(i)}_{:,1:r}, \qquad B^{(i)} \in \mathbb{R}^{N_i\times K},\; R^{(i)} \in \mathbb{R}^{K\times K},\; U^{(i)} \in \mathbb{R}^{N_i\times r}, \tag{3}$$

where $B^{(i)}$ is a fixed truncated (low-frequency) DCT basis, $R^{(i)}$ is identity-initialized and optimized to mix DCT directions within this span, and $U^{(i)}$ is the resulting rank-r spatial basis. Restricting $U^{(i)}$ to a low-frequency span serves as a band-limited structural prior that discourages voxel-wise high-frequency oscillations in the induced motion [10, 14]. We treat K as a frequency budget; if $K > N_i$, the effective span is naturally limited by N_i.

Multi-scale Motion Modeling: Building on the Tucker parameterization, we introduce multi-scale modeling to capture both global and local motion by predicting the velocity field coarse-to-fine at three resolutions, 1/4, 1/2, and 1. Each scale $s \in \{1/4, 1/2, 1\}$ outputs a scale-specific low-rank velocity tensor $\mathcal{V}^{(s)}$ with its own spatial bases and core $\mathcal{G}^{(s)}$; the Tucker rank at scale s is denoted by $r^{(s)}$, so that $\mathcal{G}^{(s)} \in \mathbb{R}^{r^{(s)}\times r^{(s)}\times r^{(s)}\times 3}$. The core $\mathcal{G}^{(s)}$ is regressed from scale-specific endpoint features, and all scale-wise branches are trained jointly under a unified multi-scale objective, with scales coupled via coarse-to-fine deformation composition for progressive refinement and without explicit cross-scale feature fusion.

During training, we compute and aggregate scale-wise endpoint reconstruction and smoothness losses. Let $I^{(s)}$ denote endpoints downsampled to scale s. Using $\mathrm{Exp}(\cdot)$ and the warping operator defined above, we integrate the scale-wise velocity $\mathcal{V}^{(s)}_{a\to b}$ into a deformation $\phi^{(s)}_{a\to b}$ (progressively refined from coarse to fine) and warp $I^{(s)}_a$ to reconstruct $\hat{I}^{(s)}_b$:

$$\mathcal{L}_{\text{total}} = \sum_s \omega^{(s)} \Big(-\lambda_1\, \mathrm{NCC}(I^{(s)}_b, \hat{I}^{(s)}_b) + \lambda_2 \left\| I^{(s)}_b - \hat{I}^{(s)}_b \right\|_{\mathrm{C}} + \lambda_3\, \mathcal{R}_{H^1}(\mathcal{V}^{(s)}_{a\to b}) \Big), \tag{4}$$

where $\sum_s \omega^{(s)} = 1$, $\mathrm{NCC}(\cdot,\cdot)$ denotes local normalized cross-correlation, $\|\cdot\|_{\mathrm{C}}$ is the Charbonnier norm, and $\mathcal{R}_{H^1}$ penalizes velocity magnitude and spatial gradients as an additional safeguard for smooth, integrable motion. All reconstruction terms at scale s are evaluated between $I^{(s)}_b$ and $\hat{I}^{(s)}_b$ at the same resolution.

2.3 Inference for Continuous-Time Interpolation

At inference (Fig. 1(b)), we evaluate the learned multi-scale low-rank representation to synthesize intermediate volumes at arbitrary $\alpha \in (0,1)$. Given endpoints (I_0, I_1), the predicted velocity tensors at three spatial scales, $\mathcal{V}^{(1/4)}, \mathcal{V}^{(1/2)}, \mathcal{V}^{(1)}$, are resampled to a common voxel grid, with magnitudes scaled according to resolution. For each direction $a \to b$, we integrate the scale-specific velocity via $\mathrm{Exp}(\cdot)$ to obtain a deformation $\phi^{(s)}_{a\to b}$, and form the endpoint deformation by coarse-to-fine composition:

$$\phi_{a\to b} = \phi^{(1)}_{a\to b} \circ \phi^{(1/2)}_{a\to b} \circ \phi^{(1/4)}_{a\to b}. \tag{5}$$

Composition follows the right-to-left coarse-to-fine order, with each scale-wise deformation integrated and resampled to the common grid beforehand.

To obtain an intermediate time α, we use both predicted directions ($0 \to 1$ and $1 \to 0$) and scale each stationary velocity by the corresponding travel time (proportional to α or $1-\alpha$) prior to exponential-map integration [20], yielding $\phi_{0\to\alpha}$ and $\phi_{1\to\alpha}$. The intermediate prediction is then computed by warping both endpoints and fusing them with time-based weights:

$$\hat{I}_\alpha = (1-\alpha)\big(I_0 \circ \phi_{0\to\alpha}\big) + \alpha\big(I_1 \circ \phi_{1\to\alpha}\big). \tag{6}$$

2.4 Theoretical Analysis of Energy Bounds

Proposition: Energy control by low-rank parameters. In Tucker form, the smoothness energy of the full velocity field can be bounded by the core magnitude, smoothing operator, and spatial bases; consequently, bounding the factor norms and core norm controls the regularity of the induced deformation [13]. Define the quadratic smoothness energy:

$$E(\mathcal{V}) \;:=\; \big\|L\,\mathrm{vec}(\mathcal{V})\big\|_2^2, \tag{7}$$

where L is a fixed linear smoothing/differential operator and $\mathrm{vec}(\cdot)$ stacks tensor entries into a vector. Let $\lambda_{\max}$ be the largest eigenvalue of $L^\top L$, i.e., $\lambda_{\max} = \|L\|_2^2$. For a Tucker-parameterized velocity tensor $\mathcal{V} = \mathcal{G} \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^{(3)}$, the following upper bound holds:

$$E(\mathcal{V}) \;\le\; \lambda_{\max}\Big(\prod_{i=1}^{3} \|U^{(i)}\|_2\Big)^2 \|\mathcal{G}\|_F^2, \tag{8}$$

where $\|\cdot\|_2$ is the matrix spectral norm and $\|\cdot\|_F$ is the Frobenius norm.

Proof. By the operator norm inequality, $\|L\,\mathrm{vec}(\mathcal{V})\|_2 \le \|L\|_2\,\|\mathrm{vec}(\mathcal{V})\|_2$, we have $E(\mathcal{V}) \le \lambda_{\max}\|\mathcal{V}\|_F^2$. Moreover, for any mode-n product, $\|\mathcal{X} \times_n U\|_F \le \|U\|_2\|\mathcal{X}\|_F$ [2]; applying this sequentially to the Tucker reconstruction yields $\|\mathcal{V}\|_F \le (\prod_{i=1}^{3}\|U^{(i)}\|_2)\|\mathcal{G}\|_F$. Combining the two bounds proves (8).

Corollary (Rank scaling). Under bounded factors $\|U^{(i)}\|_2 \leq c$ and bounded core amplitude $|\mathcal{G}_{pqr\ell}| \leq \beta$ for constants c and β, the worst-case energy bound scales at most cubically with the rank: for $\mathcal{G} \in \mathbb{R}^{r\times r\times r\times 3}$, $\|\mathcal{G}\|_F^2 \leq 3r^3\beta^2$, hence

$$E(\mathcal{V}) \;\leq\; 3\,\lambda_{\max}\, c^6\, r^3\, \beta^2. \tag{9}$$

The same bound applies to each scale s by replacing r with $r^{(s)}$.

3 Experiment

3.1 Experimental Settings

Datasets: We evaluate on ACDC [3] and 4D-Lung [8]. ACDC contains 150 cardiac MRI cases with an 80/20/50 (train/val/test) split, resized to $160\times160\times16$; end-diastolic and end-systolic phases are used as endpoints. 4D-Lung contains 500 cone-beam CT volumes with a patient-level 306/84/110 split, resized to $128\times128\times32$; 0% and 50% respiratory phases are used as start/end frames. All volumes are histogram-equalized and linearly normalized to $[0,1]$.
Metrics: We report Normalized Mutual Information (NMI), Normalized Mean Squared Error (NMSE), and Structural Similarity (SSIM) to measure statistical dependence, reconstruction error, and structural similarity of results.
Implementation Settings: All methods are implemented in PyTorch and trained on NVIDIA RTX A6000 GPUs under a unified protocol. We use AdamW with an initial learning rate of 2×10^{-4} and a cosine annealing scheduler. Models are trained for 500/200 epochs on ACDC/4D-Lung with early stopping; batch size is 1 with gradient accumulation of 8. For our method, endpoint features are extracted at three resolutions (1/4, 1/2, 1) using lightweight scale-specific 3D convolutional encoder branches trained jointly; scale-wise motion heads refine from coarse to fine, and low-frequency bases are truncated to $K = 128$. Scale weights are set to $\omega^{(1/4)} = 0.2$, $\omega^{(1/2)} = 0.3$, and $\omega^{(1)} = 0.5$; Tucker ranks are set to $r^{(1/4)} = 64$, $r^{(1/2)} = 32$, and $r^{(1)} = 8$. We set $\lambda_1 = \lambda_2 = 1$ and $\lambda_3 = 0.05$.

3.2 Comparison Experiment

We compare with unsupervised deformation models (Voxelmorph (VM) [1], Transmorph (TM) [4]), unsupervised interpolation networks (SVIN [7], UVINet [12]), and intermediate-supervised methods (DDM [11], LDDM [5], FB-Diff [23], TMSDF [24]).

The quantitative results in Tab. 1 reveal distinct performance patterns across the evaluated methods. While recent unsupervised interpolation models perform strongly on ACDC, they exhibit a notable gap on 4D-Lung due to the challenge of preserving globally coordinated respiratory motion. Conversely, methods utilizing intermediate-frame supervision often reduce reconstruction error, yet their NMI/SSIM scores do not consistently improve, confirming that appearance gains do not necessarily translate into stronger anatomical correspondence. Our approach achieves leading NMI/SSIM on both datasets, particularly on 4D-Lung,

Table 1. Quantitative comparison on ACDC and 4D-Lung datasets.

Method	Year	ACDC			4D-Lung		
		NMI$^{\uparrow}_{\times 10^{-2}}$	NMSE$^{\downarrow}_{\times 10^{-3}}$	SSIM$^{\uparrow}_{\times 10^{-2}}$	NMI$^{\uparrow}_{\times 10^{-2}}$	NMSE$^{\downarrow}_{\times 10^{-3}}$	SSIM$^{\uparrow}_{\times 10^{-2}}$
VM	2019	53.54$_{\pm 0.37}$	19.18$_{\pm 0.34}$	94.68$_{\pm 0.10}$	43.98$_{\pm 0.36}$	39.02$_{\pm 0.60}$	86.06$_{\pm 0.29}$
TM	2022	64.63$_{\pm 0.67}$	14.43$_{\pm 0.62}$	96.64$_{\pm 0.12}$	48.95$_{\pm 0.10}$	34.96$_{\pm 0.40}$	84.78$_{\pm 0.12}$
SVIN	2020	60.18$_{\pm 0.33}$	9.85$_{\pm 0.89}$	97.35$_{\pm 0.15}$	43.17$_{\pm 0.39}$	30.28$_{\pm 2.03}$	85.44$_{\pm 0.29}$
UVINet	2024	72.25$_{\pm 0.17}$	4.81$_{\pm 0.12}$	98.62$_{\pm 0.03}$	55.16$_{\pm 0.84}$	20.93$_{\pm 1.27}$	88.75$_{\pm 0.67}$
DDM	2022	64.14$_{\pm 1.35}$	10.36$_{\pm 0.30}$	97.24$_{\pm 0.87}$	42.54$_{\pm 0.89}$	33.19$_{\pm 0.73}$	85.23$_{\pm 0.27}$
LDDM	2024	72.46$_{\pm 0.30}$	4.88$_{\pm 0.04}$	98.56$_{\pm 0.01}$	56.58$_{\pm 8.08}$	28.62$_{\pm 0.80}$	83.50$_{\pm 6.09}$
FB-Diff	2025	61.33$_{\pm 0.21}$	8.64$_{\pm 0.21}$	97.29$_{\pm 0.09}$	49.28$_{\pm 1.34}$	27.39$_{\pm 6.26}$	80.91$_{\pm 3.71}$
TMSDF	2025	66.40$_{\pm 0.83}$	**4.72**$_{\pm 0.23}$	98.56$_{\pm 0.07}$	48.36$_{\pm 0.11}$	**16.85**$_{\pm 0.09}$	90.26$_{\pm 1.09}$
Ours	-	**72.61**$_{\pm 0.13}$	4.74$_{\pm 0.05}$	**98.62**$_{\pm 0.02}$	**62.34**$_{\pm 0.20}$	17.00$_{\pm 0.43}$	**91.29**$_{\pm 0.13}$

Gray background : methods trained with intermediate-frame supervision.

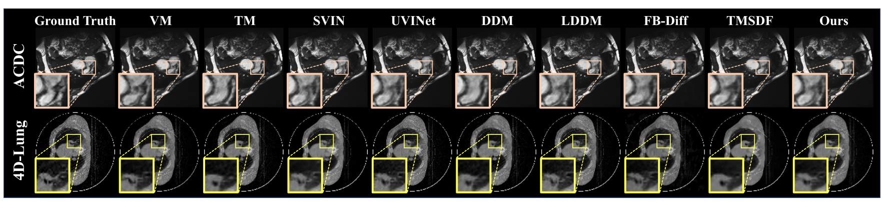

Fig. 2. Qualitative comparison on ACDC and 4D-Lung datasets.

while maintaining competitive reconstruction error. Despite relying only on endpoint supervision, it reaches supervised-level overall quality and yields more anatomically consistent, temporally coherent intermediates.

Fig. 2 shows qualitative comparisons. On ACDC, our method produces more regular contours and stable anatomy, especially in complex regions; on 4D-Lung, it better preserves coherent global organization. Baselines often exhibit local inconsistencies, while diffusion-style methods may improve appearance but compromise structural stability, yielding more anatomically consistent and visually stable intermediates.

3.3 Ablation Study

Effect of Multi-scale Rank: As shown in Fig. 3(a), we fix the coarse Tucker rank (given its low sensitivity) and vary the mid/fine ranks, reporting NMI. Increasing Tucker rank improves interpolation quality in low-capacity regimes, while the gain saturates once the fine-scale rank becomes large. Fig. 3(b) reports the Effective Rank of the resulting velocity fields (computed following [16]) as a proxy for motion complexity: lower Effective Rank corresponds to simpler, more compressible motion, and it increases monotonically with Tucker rank in our setting. Performance is best at low-to-moderate Effective Ranks, whereas

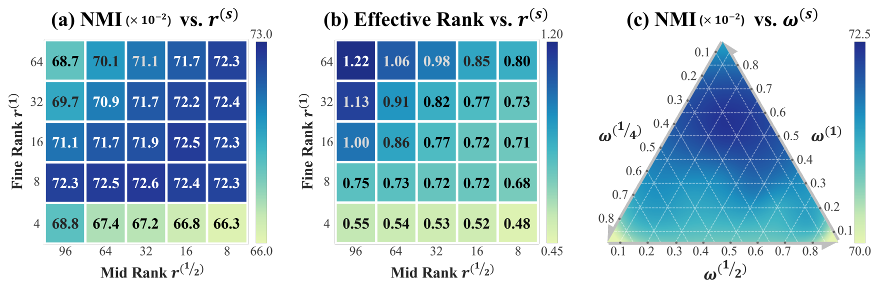

Fig. 3. Ablation on ACDC. (a) NMI under different Tucker rank settings. (b) Velocity field Effective Rank [16] ($\log_{10}$-scaled). (c) NMI vs. scale weights (ternary diagram).

Table 2. Loss term ablation on ACDC.

NCC	Charb	Reg	$\text{NMI}^{\uparrow}_{\times 10^{-2}}$	$\text{NMSE}^{\downarrow}_{\times 10^{-3}}$	$\text{SSIM}^{\uparrow}_{\times 10^{-2}}$
✓		✓	72.22	**4.69**	**98.63**
	✓	✓	72.35	4.81	98.59
✓	✓		71.45	5.18	98.49
✓	✓	✓	**72.61**	4.74	98.62

overly small ranks underfit and degrade accuracy. Overall, Tucker rank offers a direct knob to tune motion complexity via the induced Effective Rank, and both under-constraint and excessive capacity can lead to suboptimal results.

Effect of Scale Weights: We ablate the scale-wise loss weights in Fig. 3(c) by enforcing $\sum_s \omega^{(s)} = 1$ and visualizing NMI on a ternary plot. Overall, allocating more weight to the fine scale tends to improve performance, but overly extreme fine-dominant settings degrade NMI; the best region favors a fine-leaning yet balanced combination that retains non-negligible mid/coarse contributions. This indicates that fine-scale supervision is most effective for refinement, while mid/coarse guidance helps preserve global coherence and prevent locally driven inconsistencies, yielding the most reliable training signal across scales.

Loss-Term Ablation: Tab. 2 ablates the loss by enabling/disabling NCC, the Charbonnier term, and the regularizer. Dropping Charbonnier can slightly improve NMSE/SSIM but may destabilize optimization, suggesting it complements NCC by tolerating local intensity deviations under unsupervised reconstruction. Regularization is also beneficial: without it, the model tends to produce larger and noisier motions, whereas enabling it stabilizes the scale of core tensors and curbs excessive growth of low-rank parameters, yielding smoother fields and more robust training.

4 Conclusion

We proposed a structured low-rank framework for endpoint-only 4D medical image interpolation. By parameterizing velocity fields with a Tucker low-rank form and modeling motion coarse-to-fine, the method favors spatially correlated, anatomy-consistent deformation. Our analysis links velocity-field smoothness energy to low-rank parameters, supporting the structural prior. Experiments on ACDC and 4D-Lung show best performance on structure-oriented metrics with competitive interpolation error.

Acknowledgement This work was supported in part by the National Natural Science Foundation of China under Grant No. 62401246, and the Shenzhen Science and Technology Program under Grant Nos. JCYJ20250604185805008 and JCYJ20240813095112017. The authors gratefully acknowledge Beijing Tiromu Medical Technology Co., Ltd. and Shenzhen DE Sci&Tech Co., Ltd. for their support in technical validation and for providing a practical platform for this work.

Disclosure of Interests The authors declare that they have no competing interests relevant to this work.

References

1. 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)
2. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International journal of computer vision **61**(2), 139–157 (2005)
3. Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Ballester, M.A.G., et al.: Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE transactions on medical imaging **37**(11), 2514–2525 (2018)
4. Chen, J., Frey, E.C., He, Y., Segars, W.P., Li, Y., Du, Y.: Transmorph: Transformer for unsupervised medical image registration. Medical image analysis **82**, 102615 (2022)
5. Chen, T., Shi, Y., Zheng, Z., Yan, B., Hu, J., Zhu, X.X., Mou, L.: Ultrasound image-to-video synthesis via latent dynamic diffusion models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 764–774. Springer (2024)
6. De Lathauwer, L., De Moor, B., Vandewalle, J.: A multilinear singular value decomposition. SIAM journal on Matrix Analysis and Applications **21**(4), 1253–1278 (2000)
7. Guo, Y., Bi, L., Ahn, E., Feng, D., Wang, Q., Kim, J.: A spatiotemporal volumetric interpolation network for 4d dynamic medical image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4726–4735 (2020)

8. Hugo, G.D., Weiss, E., Sleeman, W.C., Balik, S., Keall, P.J., Lu, J., Williamson, J.F.: Data from 4d lung imaging of nsclc patients. (No Title) (2016)
9. Huttinga, N.R., Bruijnen, T., Van Den Berg, C.A., Sbrizzi, A.: Real-time non-rigid 3d respiratory motion estimation for mr-guided radiotherapy using mr-motus. IEEE Transactions on Medical Imaging **41**(2), 332–346 (2021)
10. Jia, X., Bartlett, J., Chen, W., Song, S., Zhang, T., Cheng, X., Lu, W., Qiu, Z., Duan, J.: Fourier-net: Fast image registration with band-limited deformation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 1015–1023 (2023)
11. Kim, B., Ye, J.C.: Diffusion deformable model for 4d temporal medical image generation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 539–548. Springer (2022)
12. Kim, J., Yoon, H., Park, G., Kim, K., Yang, E.: Data-efficient unsupervised interpolation without any intermediate frame for 4d medical images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11353–11364 (2024)
13. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM review **51**(3), 455–500 (2009)
14. Li, H., Li, H., Chen, J., Zhong, R., Niu, K., Fu, H., Liu, J.: Aif-sfda: Autonomous information filter driven source-free domain adaptation for medical image segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 4716–4724 (2025)
15. Li, Z., Sun, A., Wei, H., Chen, W., Liu, C., Sun, H., Du, C., Li, R.: Unsupervised 4d-flow mri reconstruction based on partially-independent generative modeling and complex-difference sparsity constraint. Medical Image Analysis p. 103769 (2025)
16. Park, G.Y., Jung, C., Lee, S., Ye, J.C., Lee, S.W.: Self-supervised debiasing using low rank regularization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12395–12405 (2024)
17. Perrin, S., Levilly, S., Mouchère, H., Serfaty, J.M.: Super-resolution and segmentation of 4d flow mri using deep learning and weighted mean frequencies. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 552–561. Springer (2025)
18. Pramanik, A., Aggarwal, H.K., Jacob, M.: Deep generalization of structured low-rank algorithms (deep-slr). IEEE transactions on medical imaging **39**(12), 4186–4197 (2020)
19. Tian, M., Yang, Q., Gao, Y.: Multi-scale multi-task distillation for incremental 3d medical image segmentation. In: European Conference on Computer Vision. pp. 369–384. Springer (2022)
20. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage **45**(1), S61–S72 (2009)
21. Wang, M., Li, C., Vaxman, A.: Canfields: Consolidating diffeomorphic flows for non-rigid 4d interpolation from arbitrary-length sequences. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 28587–28598 (2025)
22. Yan, J., Tan, Q., Kawara, S., Zhu, J., Wang, B., Toulemonde, M., Liu, H., Tan, Y., Tang, M.X.: Online 4d ultrasound-guided robotic tracking enables 3d ultrasound localisation microscopy with large tissue displacements. IEEE transactions on medical imaging (2025)
23. You, X., Yang, R., Zhang, C., Jiang, Z., Yang, J., Navab, N.: Fb-diff: Fourier basis-guided diffusion for temporal interpolation of 4d medical imaging. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 28010–28020 (2025)

24. Zhang, J., Ai, D., Gan, Z., Fu, T., Fan, J., Song, H., Xiao, D., Yang, J.: Temporal modulated multi-scale deformation fusion via knowledge distillation for 4d medical image interpolation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 551–561. Springer (2025)
25. 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)