Temporal Modulated Multi-Scale Deformation Fusion via Knowledge Distillation for 4D Medical Image Interpolation

Jiaju Zhang[1], Danni Ai[1*], Zhikun Gan[1], Tianyu Fu[2], Jingfan Fan[1], Hong Song[3], Deqiang Xiao[1], and Jian Yang[1]

[1] Beijing Engineering Research Center of Mixed Reality and Advanced Display, School of Optics and Photonics, Bejing Institute of Technology, No 5 Zhongguancun South Street, Haidian District, 100081, Beijing, China.
danni@bit.edu.cn
https://www.inavilab.com/

[2] School of Medical Technology, Bejing Institute of Technology, No 5 Zhongguancun South Street, Haidian District, 100081, Beijing, China.

[3] School of Computer Science and Technology, Bejing Institute of Technology, No 5 Zhongguancun South Street, Haidian District, 100081, Beijing, China.

Abstract. The acquisition of 4D medical images, which are crucial for monitoring disease progression, poses significant challenges due to the expensive cost and the imaging mechanism constraints. Existing solutions attempt to interpolate the volumes between the acquired volumes with linearly scaling the initial bidirectional deformation between two distant phases like end-systole and end-diastole, to generate detailed 4D image. However, the simple linear motion assumption fails to accurately model the anisotropic deformation induced by respiration and heartbeat. In this paper, we propose a temporal modulated multi-scale deformation fusion framework for 4D medical image interpolation via knowledge distillation, to directly generate the bidirectional deformation and volume at any intermediate time without the sub-optimal linear motion assumption. Guided by the teacher model with extensive priors, the student model, modulated by surrogate timestamps, learns to approximate the deformation modeling ability of teacher without any need for intermediate volumes. Particularly, a multi-scale deformation fusion decoder is proposed including the temporal modulated deformation feature generator and the deformation fusion module. The former generates modulation parameters with timestamps for temporal-aware transformation and then models the bidirectional deformation in a coarse-to-fine manner. While the latter adaptively fuses deformation features at different scales to improve the accuracy of predicted deformation. Compared with nine state-of-the-art methods, the proposed method achieves superior performance on two public datasets, fully demonstrating its effectiveness and generalization.

Keywords: Deep Learning · Volume Interpolation · Knowledge Distillation.

1 Introduction

Four-dimensional (4D) medical image, crucial for disease monitoring and clinical diagnostics [18,14,6], contains both spatial anatomical and temporal dynamic information. However, the acquisition is limited by the extensive costs and long examination time. The scanning time of 4D MRI lasts up to 60 minutes [23], while the exposure risk of radiation significantly increases over time during CT imaging [24]. Moreover, 4D image may occur quality degradation due to the movement and respiration of patients during scanning [3]. Achieving the optimal balance between image quality and acquisition efficiency poses significant challenges.

As a post-processing technique, volume interpolation enables the synthesis of intermediate volumes from acquired 4D images, thereby enhancing temporal resolution and reducing acquisition constraints. The similar video frame interpolation (VFI) has been adopted in many scenarios like video compression [27,12], which requires massive training data and focuses on 2D+t scenes [7]. However, 4D medical images exhibit significantly lower temporal sampling rates and more complex deformation patterns compared to videos. Besides, the transition from 2D+t to 3D+t substantially elevates computational burdens [21], hindering the direct application of VFI in 4D medical images.

Existing volume interpolation methods establish the deformation between adjacent volumes based on time-consuming iterative-optimized registration, and approximate the intermediate volume by linear scaling [28,15,20]. Kim et al. [16] proposed a diffusion deformable model to synthesize the intermediate deformation by linearly scaling the latent code generated by the diffusion module. Guo et al. [8] developed a dual-network SVIN that predicted the bidirectional deformation of the intermediate volume by linearly scaling the initial deformation between end-systole and end-diastole. And Kim et al. [17] introduced an unsupervised strategy based on cycle consistency with the similar framework. However, these methods fail to consider the impact of large difference between two endpoint phases. Moreover, the linear motion assumption hinders accurate capture of complex physiological motion, ultimately leading to spatial distortion [8]. Although Li et al. [19] used implicit neural representation to model patient anatomic motion without linear scaling, the performance and generalization was constrained by the case-specific optimization and complex hyper-parameters tuning.

We propose a temporal modulated multi-scale deformation fusion volume interpolation framework via knowledge distillation to synthesize intermediate volumes. The teacher model, comprising a feature extraction encoder, multi-scale deformation fusion decoder (MSDF), and refinement network, accurately predicts bidirectional deformation using real volumes at intermediate and two endpoints as priors. The MSDF integrates deformation feature generator (DFGen) and deformation fusion module (DFM) across multiple scales. DFGen captures both global and local deformation in a coarse-to-fine manner, while DFM adaptively fuses multi-scale features to directly predict the intermediate deformation, eliminating the sub-optimal linear scaling operation. The DFGen in student is time-modulated to use surrogate timestamps to emulate the teacher's deforma-

tion features, directly predicting the intermediate deformation without the need for real intermediate volumes. To accurately model the latent deformation features, we developed a distillation loss that facilitates knowledge transfer from teacher with richer priors to student. The contributions are as follows:

1. We propose a novel temporal modulated multi-scale deformation fusion volume interpolation framework to directly formulate the deformation of intermediate volumes at arbitrary time in the student model via knowledge distillation with using surrogate timestamps and the guidance of the teacher model.
2. We propose a multi-scale deformation fusion decoder to simultaneously capture global and local deformation, mitigating the impact of large difference between two endpoint phases.
3. Extensive experiments on two public datasets compared with nine state-of-the-art methods verify the effectiveness of the proposed method.

2 Method

2.1 Overview

The volumes at two endpoint phases and intermediate time are represented as I_0, I_1, and I_t, respectively. Existing volume interpolation solutions synthesize I_t based on its bidirectional deformation with linearly scaling the initial deformation between I_0 and I_1. It has been demonstrated that explicitly modeling the high-order complex motion solely based on two endpoint phases is challenging [9]. The proposed temporal modulated multi-scale deformation fusion volume interpolation framework is illustrated in Fig.1(a). The teacher model takes I_0, I_t, and I_1 as input priors to explore the latent deformation features and accurately predict the bidirectional deformation of I_t at different resolutions and the synthesized volume $\tilde{I}_t^{Tch}$. The architecture of student is similar to that of teacher. While the student model only takes I_0, I_1 as input volumes and introduces the relevant timestamp t as surrogates to replace I_t. The student performs temporal-aware transformation on the deformation features at multi resolutions to emulate the intermediate features of teacher with real I_t as input, eventually generating the bidirectional deformation and synthesizing the intermediate volume $\tilde{I}_t^{Stu}$. The teacher effectively transfers deformation clues learned from rich priors to student via knowledge distillation, enhancing the deformation modeling ability of student with limited inputs under the guidance of surrogate timestamps.

2.2 Teacher Model

The teacher model includes feature extraction encoder, multi-scale deformation fusion decoder (MSDF), and refinement network. The feature extraction encoder takes 4 convolution units to extract multi-scale features $\{\mathcal{F}_0^{Tch_i}\}$, $\{\mathcal{F}_t^{Tch_i}\}$, and $\{\mathcal{F}_1^{Tch_i}\}, i = 1, \ldots, 4$ from I_0, I_t, and I_1. Each unit has 2 3D Conv Blocks, comprising a 3D Conv layer, an InstanceNorm layer, and LeakyReLU function.

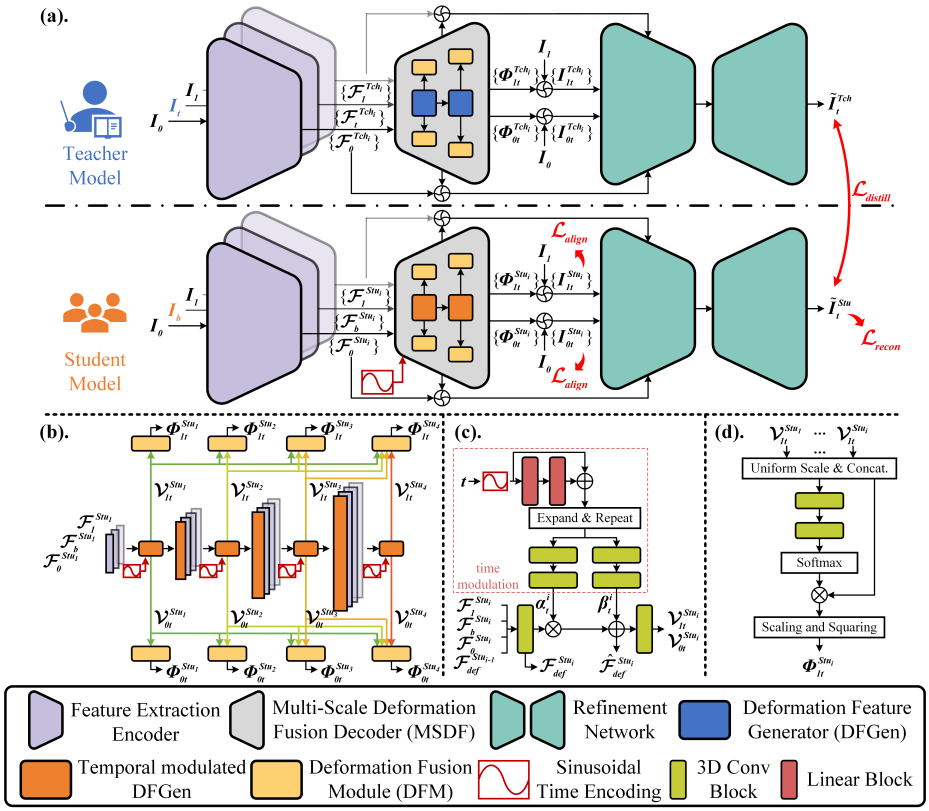

Fig. 1. (a). The proposed knowledge distillation-based volume interpolation framework; (b). The architecture of MSDF in student. The MSDF in teacher is similar without the surrogate timestamps; (c). The architecture of DFGen. The time modulation part only exists in student; (d). The architecture of DFM; Intermediate variables of student are taken as examples for easily understanding in (b-d).

The MSDF includes the deformation feature generator (DFGen) and deformation fusion module (DFM) at different scales (Fig.1(b)). When i=1, $\mathcal{F}_0^{Tch_1}$, $\mathcal{F}_t^{Tch_1}$, and $\mathcal{F}_1^{Tch_1}$ are concatenated and sent into DFGen to model the deformation feature $\mathcal{F}_{def}^{Tch_1}$ and bidirectional velocity field $\mathcal{V}_{0t}^{Tch_1}$ and $\mathcal{V}_{01t}^{Tch_1}$. When i>1, $\mathcal{F}_{def}^{Tch_{(i-1)}}$ output by preceding DFGen are upsampled and fused with other input of current DFGen to generate $\mathcal{F}_{def}^{Tch_i}$, achieving progressive coarse-to-fine deformation (Fig.1(c)). In the teacher model without time modulation:$\hat{\mathcal{F}}_{def}^{Tch_i} = \mathcal{F}_{def}^{Tch_i}$. The velocity fields from all i current scales are unified and stacked in DFM to be convolved into i weight maps to re-weight the contribution of each velocity vector in the corresponding velocity field. The weighted summation is

integrated to generate the deformation field $\Phi_{0t}^{Tch_i}$ and $\Phi_{1t}^{Tch_i}$ via scaling and squaring [5]. Similarly, $\{\Phi_{0t}^{Tch_i}\}$ and $\{\Phi_{1t}^{Tch_i}\}$ at different scales are obtained.

I_0, $\{\mathcal{F}_0^{Tch_i}\}$, and I_1, $\{\mathcal{F}_1^{Tch_i}\}$ are respectively warped by the corresponding fields to obtain the initial deformed volumes $\{I_{0t}^{Tch_i}\}$, $\{I_{1t}^{Tch_i}\}$ and features $\{\tilde{\mathcal{F}}_{0t}^{Tch_i}\}$, $\{\tilde{\mathcal{F}}_{1t}^{Tch_i}\}$, which are taken as the input of refinement network to eventually generate the synthesized volume $\tilde{I}_t^{Tch}$. An encoder-decoder architecture is adopted in the refinement network with 8 3D convolution units. Each deformed feature is fed into individual encoder layers and concatenated channel-wise.

2.3 Student Model

The architecture of student is similar to that of teacher, differing in: (1). The input I_t in teacher is replaced with the weighted summation of I_0 and I_1: $I_b = t \times I_0 + (1-t) \times I_1$. (2). The surrogate timestamp and time modulation are added in DFGen. The high-dimensional embedding of timestamp is obtained with a sinusoidal time encoding function, which is expanded to the size of current feature to generate the modulation parameters α_t^i and β_t^i after 2 Linear Blocks and 2 3D Conv Blocks. And $\mathcal{F}_{def}^{Tch_i}$ is modulated as: $\hat{\mathcal{F}}_{def}^{Stu_i} = \alpha_t \times \mathcal{F}_{def}^{Stu_i} + \beta_t$. Each Linear Block contains a linear layer and LeakyReLU activation function.

2.4 Model Objectives

The overall objective function is as follows:

$$\mathcal{L} = \mathcal{L}_{align} + \mathcal{L}_{recon} + \lambda_d \mathcal{L}_{distill} \tag{1}$$

The mean square error (MSE) between the synthesized volume $\tilde{I}_t^{Stu}$ and I_t is calculated as the reconstruction loss $\mathcal{L}_{recon}$. The alignment loss $\mathcal{L}_{align}$ includes the similarity term $\mathcal{L}_{sim}$ and regularization term $\mathcal{L}_{reg}$. $\mathcal{L}_{sim}$ calculates MSE between the deformed volume after MSDF and target, while $\mathcal{L}_{reg}$ use a diffusion regularizer [1] to keep the deformation smooth and continuous:

$$\mathcal{L}_{align} = \sum_{k=0,1} \sum_{i=1}^{4} 2^{i-4} \times \left(\mathcal{L}_{sim}\left(I_t, I_{kt}^{Stu_i}\right) + \lambda_r \mathcal{L}_{reg}(\Phi_{kt}^{Stu_i}) \right) \tag{2}$$

The teacher model is initially optimized to learn precise nonlinear deformation with rich priors. Then the distillation loss $\mathcal{L}_{distill}$ is employed to guide student in extracting potentially valuable information with limited prior knowledge:

$$\mathcal{L}_{distill} = \mathcal{L}_{sim}\left(\tilde{I}_t^{Stu}, \tilde{I}_t^{Tch}\right) + \sum_{i=1}^{4} \left(\mathcal{L}_{sim}\left(I_{0t}^{Stu_i}, I_{0t}^{Tch_i}\right) + \mathcal{L}_{sim}\left(I_{1t}^{Stu_i}, I_{1t}^{Tch_i}\right) \right) \tag{3}$$

3 Experiments

3.1 Experiments Settings

Datasets. Two public 4D image datasets ACDC [2] and 4D_Lung [10] are used to verify the performance of our method. We follow the data partitioning and prepossessing in [17]. The ACDC dataset consists of 100 4D cardiac MRI images where 90, 10 images are selected for training and test, respectively. While 68, 14 images are for training and test in 4D_Lung which contains 82 4D CT images.
Metrics. To quantitatively evaluate the performance, we introduce various metrics: peak signal-to-noise ratio (PSNR), normalized cross-correlation coefficient (NCC), structural similarity (SSIM), and normalized mean square error (NMSE).
Compared Methods. We conduct extensive comparisons with 9 state-of-the-art methods: SVIN [8], MPVF [25], VoxelMorph (VM) [1], TransMorph (TM) [4], Fourier-Net+ [11], R2Net [13], DDM [16], IDIR [26], and UVI-Net [17].
Implementation Details. The proposed method is implemented with PyTorch [22] using an NVIDIA RTX 4090 GPU. We employ the Adam optimizer with a learning rate 10^{-4} for 200 epochs, and a batch size of 1. λ_r and λ_d are set to 0.1. The training has two stages: we first train the teacher model via the combination of $\mathcal{L}_{recon}$ and $\mathcal{L}_{align}$; and then we train the student model via Eq. 1.

3.2 Comparisons

The interpolation performance of each method on two datasets is presented in Table 1. The proposed method significantly outperforms others across all metrics. From multiple quantitative perspectives including voxel-wise similarity, correlation, structural similarity, and reconstruction error, the synthesized volume of the proposed method is the closest to the ground truth volume. The visual-

Table 1. Quantitative comparison of interpolation results. The best results and the second best results in each column are bolded and underlined respectively.

Method	ACDC				4D_Lung			
	PSNR	NCC	SSIM	NMSE	PSNR	NCC	SSIM	NMSE
SVIN	32.51	0.559	0.972	2.930	30.99	0.312	0.973	0.852
MPVF	33.15	0.561	0.971	2.435	31.18	0.310	0.972	0.761
VM	31.02	0.555	0.966	4.254	32.29	0.316	0.977	0.641
TM	30.45	0.547	0.958	4.826	30.92	0.313	0.973	0.786
Fourier-Net+	29.98	0.544	0.957	5.503	30.26	0.308	0.971	0.959
R2Net	28.59	0.509	0.930	7.281	29.34	0.294	0.962	1.061
DDM	29.71	0.541	0.956	5.007	30.37	0.308	0.971	0.905
IDIR	31.56	0.557	0.968	3.806	32.91	<u>0.321</u>	0.980	0.586
UVI-Net	<u>33.59</u>	<u>0.565</u>	<u>0.978</u>	<u>2.384</u>	<u>34.00</u>	0.320	<u>0.980</u>	<u>0.552</u>
ours	**35.21**	**0.579**	**0.982**	**1.869**	**35.47**	**0.325**	**0.984**	**0.386**

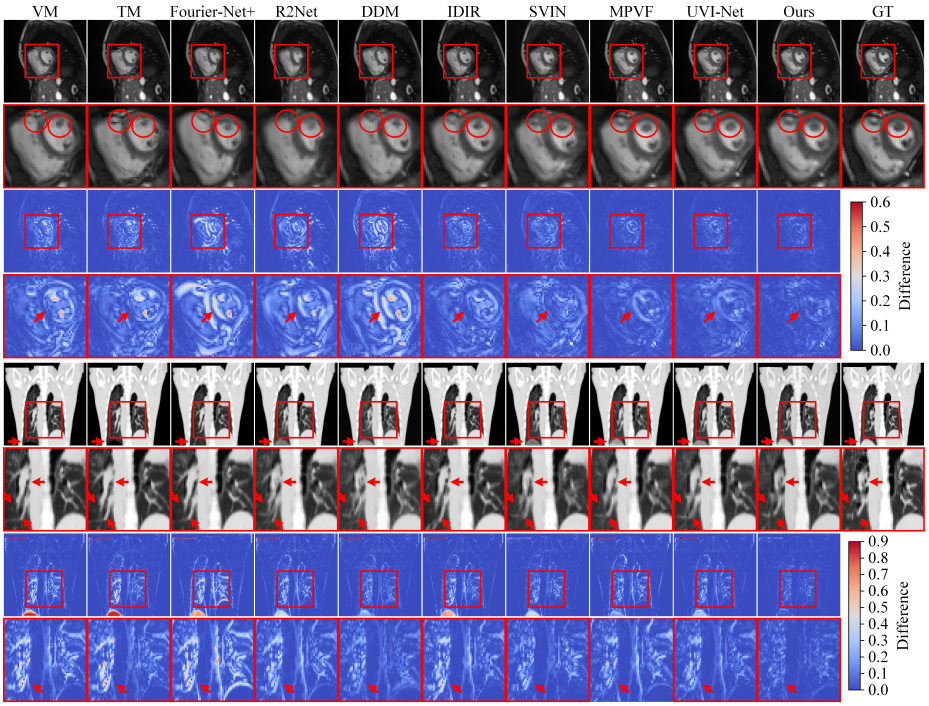

Fig. 2. Visualization of synthesized volumes and error maps on ACDC and 4D_Lung.

ization of synthesized volumes and the relevant difference maps on ACDC and 4D_Lung are shown in Fig.2. The volumes generated by our method are visually more realistic and reliable, with more accurate edges of anatomical regions such as the heart and pulmonary vessels. Fig.3 presents the prediction of 4D image series over time. Compared with the linear scaling based VM, the proposed method more effectively captures fine-grained details.

3.3 Ablations

We conduct ablations on ACDC. Table 2 presents the influence on interpolation of each modules, where *TM_DFGen* represents the temporal modulated DFGen in student, *Ref.* represents the refinement network and *MsDFM* represents the usage of DFM at multi scales. Without any modules incorporated, the student model generates intermediate deformation through linear scaling with solely taking I_0 and I_1 as inputs, whose performance is limited. The utilize of DFM significantly enhances performance, benefiting from the multi-scale deformation fusion. Incorporating the time modulation in DFGen of student effectively alleviates the input prior deficiency, enabling direct prediction of bidirectional deformation with surrogate timestamps instead of linear scaling. And the refinement network further improves detail features in synthesized volumes, increasing

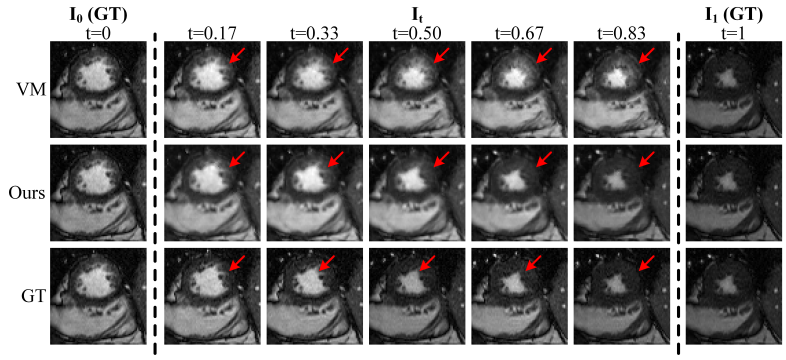

Fig. 3. Qualitative results on the prediction of 4D image series over time on ACDC.

the similarity to ground truth. Expanding the DFM from only used at final scale to all scales simultaneously improve the accuracy of bidirectional deformation at different scales, further enhancing the authenticity of the synthesized volumes.

Table 2. The influence of each modules on interpolation performance.

DFM	*TM_DFGen*	*Ref.*	*MsDFM*	PSNR	NCC	SSIM	NMSE
				30.85	0.553	0.964	4.451
✓				31.73	0.562	0.967	3.658
✓	✓			32.52	0.569	0.972	2.728
✓	✓	✓		33.92	0.576	0.978	2.197
✓	✓	✓	✓	**35.21**	**0.579**	**0.982**	**1.869**

Table 3. The influence of each loss term on interpolation performance.

$\mathcal{L}_{recon}$	$\mathcal{L}_{align}$	$\mathcal{L}_{distill}$	PSNR	NCC	SSIM	NMSE
✓			32.08	0.565	0.963	2.735
✓	✓		33.14	0.570	0.973	2.298
✓	✓	✓	**35.21**	**0.579**	**0.982**	**1.869**

Table 3 shows the impact of each component in the overall objective function on interpolation. Compared to directly training student, $\mathcal{L}_{distill}$ provides additional constraints from teacher, reducing the difficulty of directly modeling the relationship between surrogate timestamp and the bidirectional deformation.

4 Conclusion

We propose a knowledge distillation-based volume interpolation framework with temporal modulated multi-scale deformation fusion to directly predict bidirectional deformation and corresponding volumes at arbitrary time. The teacher model accurately models multi-scale bidirectional deformation features with taking two endpoint phase and intermediate volumes as priors. The student model innovatively introduces time modulation and surrogate timestamps during multi-scale deformation fusion to achieve time-aware transformation of deformation features, enabling deformation modeling and volume synthesis without the need for real intermediate volume. A distillation loss is developed to facilitate knowledge transfer from teacher with richer prior knowledge to student. Experimental results demonstrate that our method outperforms existing volume interpolation approaches, showing promising potential for clinical applications.

Acknowledgments. This study was supported by Beijing Natural Science Foundation(7252288).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

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. 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)
3. Caines, R., Sisson, N.K., Rowbottom, C.G.: 4dct and vmat for lung patients with irregular breathing. Journal of Applied Clinical Medical Physics **23**(1), e13453 (2022)
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. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. Medical image analysis **57**, 226–236 (2019)
6. Dey, S., Konar, D., De, S., Bhattacharyya, S.: An introductory illustration of medical image analysis. In: Advanced Machine Vision Paradigms for Medical Image Analysis, pp. 1–9. Elsevier (2021)
7. Ferdian, E., Suinesiaputra, A., Dubowitz, D.J., Zhao, D., Wang, A., Cowan, B., Young, A.A.: 4dflownet: super-resolution 4d flow mri using deep learning and computational fluid dynamics. Frontiers in Physics **8**, 138 (2020)
8. 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)

9. Hu, M., Jiang, K., Zhong, Z., Wang, Z., Zheng, Y.: Iq-vfi: Implicit quadratic motion estimation for video frame interpolation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6410–6419 (2024)
10. 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)
11. Jia, X., Thorley, A., Gomez, A., Lu, W., Kotecha, D., Duan, J.: Fourier-net+: Leveraging band-limited representation for efficient 3d medical image registration. arXiv preprint arXiv:2307.02997 (2023)
12. Jia, Z., Lu, Y., Li, H.: Neighbor correspondence matching for flow-based video frame synthesis. In: Proceedings of the 30th ACM International Conference on Multimedia. pp. 5389–5397 (2022)
13. Joshi, A., Hong, Y.: R2net: Efficient and flexible diffeomorphic image registration using lipschitz continuous residual networks. Medical Image Analysis **89**, 102917 (2023)
14. Kanaga, E.G.M., Anitha, J., Juliet, D.S.: 4d medical image analysis: a systematic study on applications, challenges, and future research directions. In: Advanced Machine Vision Paradigms for Medical Image Analysis, pp. 97–130. Elsevier (2021)
15. Karani, N., Tanner, C., Kozerke, S., Konukoglu, E.: Temporal interpolation of abdominal mris acquired during free-breathing. In: Medical Image Computing and Computer-Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part II 20. pp. 359–367. Springer (2017)
16. 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)
17. 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)
18. Li, G., Citrin, D., Camphausen, K., Mueller, B., Burman, C., Mychalczak, B., Miller, R.W., Song, Y.: Advances in 4d medical imaging and 4d radiation therapy. Technology in cancer research & treatment **7**(1), 67–81 (2008)
19. Li, X., Yang, R., Li, X., Lomax, A., Zhang, Y., Buhmann, J.: Cpt-interp: Continuous spatial and temporal motion modeling for 4d medical image interpolation. arXiv preprint arXiv:2405.15385 (2024)
20. Lin, Z., Neerav, K., Christine, T., Ender, K.: Temporal interpolation via motion field prediction. In: Medical Imaging with Deep Learning (2022)
21. Najjar, R.: Redefining radiology: a review of artificial intelligence integration in medical imaging. Diagnostics **13**(17), 2760 (2023)
22. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems **32** (2019)
23. Sartoretti, E., Sartoretti, T., Binkert, C., Najafi, A., Schwenk, Á., Hinnen, M., van Smoorenburg, L., Eichenberger, B., Sartoretti-Schefer, S.: Reduction of procedure times in routine clinical practice with compressed sense magnetic resonance imaging technique. PloS one **14**(4), e0214887 (2019)
24. Wang, W.H., Sung, C.Y., Wang, S.C., Shao, Y.H.J.: Risks of leukemia, intracranial tumours and lymphomas in childhood and early adulthood after pediatric radiation exposure from computed tomography. CMAJ **195**(16), E575–E583 (2023)

25. Wei, T.T., Kuo, C., Tseng, Y.C., Chen, J.J.: Mpvf: 4d medical image inpainting by multi-pyramid voxel flows. IEEE Journal of Biomedical and Health Informatics (2023)
26. Wolterink, J.M., Zwienenberg, J.C., Brune, C.: Implicit neural representations for deformable image registration. In: International Conference on Medical Imaging with Deep Learning. pp. 1349–1359. PMLR (2022)
27. Wu, C.Y., Singhal, N., Krahenbuhl, P.: Video compression through image interpolation. In: Proceedings of the European conference on computer vision (ECCV). pp. 416–431 (2018)
28. Zhang, W., Brady, J.M., Becher, H., Noble, J.A.: Spatio-temporal (2d+ t) non-rigid registration of real-time 3d echocardiography and cardiovascular mr image sequences. Physics in Medicine & Biology **56**(5), 1341 (2011)