Dyna3DGR: 4D Cardiac Motion Tracking with Dynamic 3D Gaussian Representation

Xueming Fu[1,2], Pei Wu[1,2], Yingtai Li[1,2], Xin Luo[1,2], Zihang Jiang[1,2,3], Junhao Mei[5], Jian Lu[5], Gao-Jun Teng[5], and S. Kevin Zhou[1,2,3,4] ✉

[1] School of Biomedical Engineering, Division of Life Sciences and Medicine, University of Science and Technology of China (USTC), Hefei Anhui, 230026, China
[2] Center for Medical Imaging, Robotics, Analytic Computing & Learning (MIRACLE), Suzhou Institute for Advance Research, USTC, Suzhou Jiangsu, 215123, China
[3] Jiangsu Provincial Key Laboratory of Multimodal Digital Twin Technology, Suzhou Jiangsu, 215123, China
[4] State Key Laboratory of Precision and Intelligent Chemistry, USTC, Hefei Anhui, 230026, China
[5] Center of Interventional Radiology & Vascular Surgery, Department of Radiology, Medical School, Zhongda Hospital, Southeast University, Nanjing 210009, China

Abstract. Accurate analysis of cardiac motion is crucial for evaluating cardiac function. While dynamic cardiac magnetic resonance imaging (CMR) can capture detailed tissue motion throughout the cardiac cycle, the fine-grained 4D cardiac motion tracking remains challenging due to the homogeneous nature of myocardial tissue and the lack of distinctive features. Existing approaches can be broadly categorized into image-based and representation-based, each with its limitations. Image-based methods, including both traditional and deep learning-based registration approaches, either struggle with topological consistency or rely heavily on extensive training data. Representation-based methods, while promising, often suffer from loss of image-level details. To address these limitations, we propose **Dyna**mic **3D G**aussian **R**epresentation (**Dyna3DGR**), a novel framework that combines explicit 3D Gaussian representation with implicit neural motion field modeling. Our method simultaneously optimizes cardiac structure and motion in a self-supervised manner, eliminating the need for extensive training data or point-to-point correspondences. Through differentiable volumetric rendering, Dyna3DGR efficiently bridges continuous motion representation with image-space alignment while preserving both topological and temporal consistency. Comprehensive evaluations on the ACDC dataset demonstrate that our approach surpasses state-of-the-art deep learning-based diffeomorphic registration methods in tracking accuracy. The code will be available in https://github.com/windrise/Dyna3DGR.

Keywords: Cardiac Motion Tracking · Gaussian Representation.

1   Introduction

Accurate estimation of myocardial motion is essential for evaluating cardiac function and diagnosing myocardial diseases[21]. Dynamic cardiac motion reconstruction provides comprehensive spatiotemporal information throughout cardiac cycle, enabling clinicians to better analyze physiological cardiac dynamics for improved diagnostic accuracy and treatment planning. Tagged cardiac magnetic resonance imaging (t-CMR)[1] serves as the gold standard for assessing myocardial motion using intrinsic markers[26]. However, due to its complex acquisition process, recent research has increasingly focused on estimating motion from untagged CMR images[29,17,13,27]. Technically, cardiac motion estimation approaches can be broadly categorized into two main streams: image-based and representation-based. In the image-based methods, researchers have developed various non-parametric registration methods that rely on mathematical priors and optimization techniques. These include incorporating free-form deformations with B-splines[22], optical flow[4], and biomechanics-informed[19] approaches to achieve accurate correspondence mapping. While these methods have shown promise, they often struggle with preserving topological consistency during deformation. To address this limitation, diffeomorphic registration methods[5,23,2] have introduced topology-preserving constraints. However, these non-parametric registration approaches remain computationally intensive and sensitive to image noise. The advent of deep learning has revolutionized image-based cardiac motion estimation. Data-driven deep registration approaches have demonstrated superior performance in preserving topological consistency and maintaining long-term temporal coherence compared to traditional registration methods [3,11,20,7,15,25]. However, their effectiveness is inherently constrained by the availability of extensive training data, and they often face challenges in generalizing across datasets with different distributions.

While untagged CMR provides clear visualization of cardiac structures that can be precisely segmented, the inherent elasticity and homogeneous nature of myocardial tissue present significant challenges for accurate motion tracking in image space due to the lack of reliable natural landmarks within the tissue. To alleviate this problem, another line of research explores cardiac motion estimation in alternative representation spaces. Guo *et al.* [9] have proposed an unsupervised approach to extract stable landmarks from volumetric images, using optimal transport theory with topological constraints for motion field estimation. However, this approach can be sensitive to noise and may not achieve sufficient precision. Meng *et al.* [16] have adopted fixed-vertex mesh representations with template topology ($\sim$20,000 vertices) as stable identifiers across different subjects and cardiac cycles, and estimate vertex motion from six different view sequences to reconstruct myocardial deformation. While effective, this approach may lose fine-grained image details. Yuan *et al.* [28] have explored using implicit neural representations through signed distance field to model the myocardium, enabling continuous shape representation. However, their approach focuses primarily on global shape deformation modeling, making it challenging to capture fine-grained local motion details. While these representation-based

Table 1. Comparison of different methods.

Methods	Representation	Image Details	No Extra Data
Traditional registration [22,4,19]	Pixel/Voxel	✓	✓
DL-based registration [3,11,20,7,15,25]	Pixel/Voxel	✓	✗
Points Representation[9]	Landmark	✓	✗
Shape Representation[16,28]	Mesh/SDF	✗	✗
Dyna3DGR (ours)	3D Gaussian	✓	✓

approaches show promise in breaking through the performance ceiling of image-based methods, there remains a critical need for a unified framework that can both accurately represent cardiac anatomy and seamlessly bridge the gap between representation space and image space.

To tackle the challenges, we propose **Dyna**mic **3D G**aussian **R**epresentation (**Dyna3DGR**), a novel framework that combines explicit 3D Gaussian representation with implicit neural motion field modeling. Our approach simultaneously optimizes cardiac structure and motion reconstruction in a self-supervised manner, eliminating the need for extensive training data or dense correspondences across cardiac cycles. Through differentiable volumetric rendering, Dyna3DGR efficiently bridges the gap between 3D Gaussian representation and image-space. Our key contributions can be summarized as follows:

1. We propose a self-supervised optimization framework for 4D cardiac motion estimation that simultaneously optimizes cardiac structure and motion estimation, eliminating the dependency on extensive training data that is commonly required by existing image-space methods.
2. Through the unique integration of explicit 3D Gaussian representation and implicit neural deformation field modeling in Dyna3DGR, it effectively fills the gap between representation space and image space. This hybrid design not only preserves topological consistency but also achieves accurate motion tracking without requiring explicit dense correspondence, addressing the shortcomings of existing representation-based approaches.
3. Comprehensive evaluations on the ACDC dataset demonstrate that our approach surpasses state-of-the-art diffeomorphic registration methods in tracking accuracy, validating the effectiveness of our proposed framework.

2   Method

In this section, we present the Dynamic 3D Gaussian Representation (Dyna3DGR) framework in Fig. 1. The framework integrates explicit 3D Gaussians for cardiac volume representation with the implicit neural field for motion field modeling.
Explicit 3D Gaussian Representation: The 3D Gaussians are explicitly used to model cardiac volumetric structures. Following [14,8], each Gaussian G_i has a parameter set $\phi_i = \{xyz_i, \Sigma_i, I_i\}$, where $xyz_i \in \mathbb{R}^3$ denotes the center position of
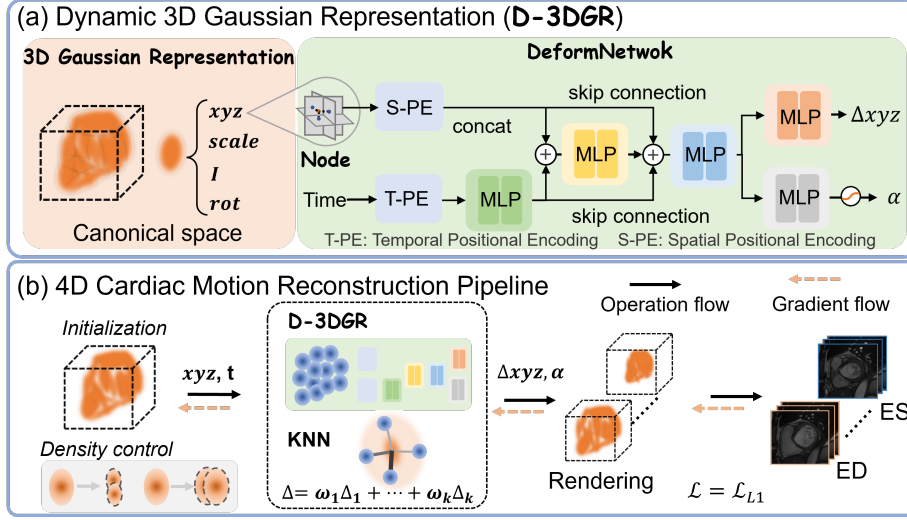
Fig. 1. Overview of Dyna3DGR. (a) The Dyna3DGR consists of an explicit 3D Gaussian representation in canonical space (orange background) for volumetric reconstruction and an implicit motion representation powered by a deformation network (green background) for motion modeling. (b) The end-to-end pipeline that jointly optimizes both volumetric and motion representations.

the Gaussian and $\Sigma \in \mathbb{R}^{3\times3}$ is the covariance matrix. The Σ can be decomposed into two learnable components: a quaternion $\boldsymbol{rot}$ and a scaling $\boldsymbol{scale}$. They can be transformed into the corresponding matrices R and S. The resulting Σ can be expressed as: $\Sigma = RSS^T R^T$. The $I_i \in \mathbb{R}$ determines intensity value of Gaussian center. Each Gaussian's influence to position X is mathematically described by:

$$G_i(X|\phi_i) = I_i \cdot e^{-\frac{1}{2}(X-xyz_i)^T \Sigma_i^{-1}(X-xyz_i)}, \tag{1}$$

where $X \in \mathbb{R}^3$ denotes an position in the 3D space. The exponential term defines the spatial decay of the Gaussian's influence based on the distance from its center. The volumetric value at any point X is computed as a local aggregation of contributions from nearby Gaussians:

$$V(X|\theta_i) = \sum_{i:||X-xyz_i||\leq d_i} G_i(X|\phi_i), \tag{2}$$

where d_i defines the effective radius of influence for each Gaussian.

Implicit Motion Representation: To effectively model cardiac tissue motion and deformation over time, we introduce a control-nodes-based deformation mechanism following [10]. We employ a set of control nodes $C = \{C_i \in \mathbb{R}^3, o_i \in \mathbb{R}^+\}$, where C_i represents learnable position coordinates in the canonical space, o_i is the learnable radius of a radial-basis-function kernel that controls the impact of a control node. For each control node C_i, we learn a Gaussian transformation

parameter $[\Delta xyz_i^t|\alpha_i^t]$ with a neural network. The transformation consists of a translation vector $\Delta xyz \in \mathbb{R}^3$ and a scale factor vector $\alpha \in \mathbb{R}^{+3}$. The deformation network $\mathcal{F}_\theta$ takes both the node position and temporal information as inputs to predict the transformation of these positions:

$$(\Delta xyz, \alpha) = \mathcal{F}_\theta(\gamma(\text{sg}(\boldsymbol{x})), \gamma(t)), \tag{3}$$

where $\text{sg}(\cdot)$ indicates a stop-gradient operation, γ denotes the positional encoding:

$$\gamma(p) = (sin(2^k \pi p), cos(2^k \pi p))_{k=0}^{L-1}. \tag{4}$$

DeformNetwork: The network takes an $N \times 3$ point cloud coordinate matrix and an $N \times 1$ temporal vector as input, outputting an $N \times 3$ translation vector Δxyz and an $N \times 3$ scaling vector α. The framework employs the L1 loss function to guide the optimization process. The loss is computed between the predicted volumetric frames and the ground truth images:

$$\mathcal{L} = \sum \|V(G(xyz + \Delta xyz, \alpha * scale, rot, I)) - V_{gt}\|_1, \tag{5}$$

where V_{gt} denotes the ground truth cardiac voxels.

To compute the dense motion field for each Gaussian in the canonical space, we employ a KNN search to identify the k closest control points. The final transformation for each Gaussian is then derived through Linear Blend Skinning, which interpolates the transformations of the nearby control nodes to determine the position and scale changes of the Gaussian. This approach can be formulated as:

$$[\Delta xyz_i^t|\alpha_i^t] = \sum_{j=1}^{k} w_j \cdot [\Delta xyz_j^t|\alpha_j^t], \tag{6}$$

where w_j represents the blending weight calculated based on the distance between the Gaussian and the nearest k control node. The blending weights are calculated by

$$w_j = \frac{\hat{w}_{ij}}{\sum_{j \in \mathcal{C}_k} \hat{w}_{ij}}, \quad \hat{w}_{ij} = \exp(-\frac{d_{ij}^2}{2o_j^2}), \tag{7}$$

where d_{ij} is the distance between center of Gaussian G_i and the j-th neighboring control node, and o_j is the learned radius parameter of the j-th control point.

3 Experiments

Dataset: Our method is evaluated on the Automated Cardiac Diagnosis Challenge (ACDC) dataset[6], a comprehensive collection of 4D cardiac MRI sequences. The dataset consists 100 clinical cases across five categories: normal subjects (NOR), myocardial infarction with systolic heart failure (MINF), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), and abnormal right ventricle (ARV). Each case includes expert-annotated segmentation
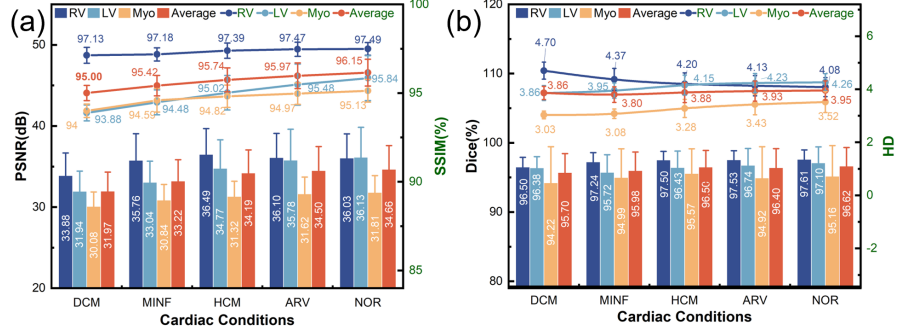
Table 2. Quantitative comparison of different methods. Tracking results evaluated with region-specific (RV,LV,Myo) and aggregate Dice score (%), PSNR (dB) and SSIM (%).(mean ± std, best and second-best results are in **bold** and <u>underline</u>, respectively).

Method	RV ↑	LV ↑	Myo ↑	Avg. ↑	PSNR ↑	SSIM ↑
LDDMM [5]	$73.61_{\pm8.5}$	$65.62_{\pm8.5}$	$56.44_{\pm13}$	$72.39_{\pm18}$	$31.20_{\pm3.8}$	$84.59_{\pm6.0}$
RDMM [23]	$76.43_{\pm7.8}$	$69.50_{\pm9.1}$	$62.19_{\pm14}$	$75.51_{\pm12}$	$31.66_{\pm3.9}$	$84.36_{\pm5.4}$
ANTs (SyN) [2]	$75.30_{\pm7.4}$	$66.92_{\pm8.6}$	$58.03_{\pm11}$	$74.64_{\pm13}$	$30.92_{\pm3.6}$	$84.26_{\pm5.6}$
VoxelMorph [3]	$81.60_{\pm6.5}$	$77.00_{\pm8.6}$	$67.90_{\pm13}$	$79.90_{\pm11}$	$34.68_{\pm3.3}$	$85.01_{\pm5.5}$
DiffuseMorph [11]	$82.10_{\pm6.7}$	$78.30_{\pm8.6}$	$67.80_{\pm15}$	$80.50_{\pm11}$	$34.73_{\pm3.6}$	$84.30_{\pm5.2}$
CorrMLP [15]	$80.33_{\pm6.5}$	$80.07_{\pm7.8}$	$70.51_{\pm14}$	$80.44_{\pm8.6}$	$34.90_{\pm2.9}$	$84.27_{\pm4.5}$
DeepTag [26]	$81.89_{\pm7.0}$	$79.10_{\pm7.5}$	$70.37_{\pm13}$	$80.83_{\pm12}$	$33.64_{\pm3.4}$	$83.09_{\pm4.9}$
Transmatch [7]	$81.22_{\pm7.0}$	$80.34_{\pm6.8}$	$71.21_{\pm12}$	$81.35_{\pm9.8}$	$33.89_{\pm3.3}$	$84.78_{\pm4.9}$
FSDiffReg [20]	$82.70_{\pm6.1}$	$80.90_{\pm7.7}$	$72.40_{\pm12}$	$82.30_{\pm9.6}$	$35.34_{\pm3.5}$	$85.85_{\pm5.2}$
GPTrack [25]	$\underline{82.91}_{\pm5.8}$	$\underline{81.23}_{\pm8.2}$	$\underline{72.86}_{\pm9.0}$	$\underline{82.65}_{\pm10}$	$\mathbf{35.52}_{\pm3.1}$	$\underline{86.19}_{\pm5.0}$
Dyna3DGR (ours)	$\mathbf{97.61}_{\pm1.4}$	$\mathbf{97.10}_{\pm2.3}$	$\mathbf{95.16}_{\pm4.4}$	$\mathbf{96.62}_{\pm2.7}$	$\underline{34.66}_{\pm2.5}$	$\mathbf{97.08}_{\pm0.25}$

masks for three key cardiac structures in the end-diastole (ED) and end-systole (ES): the Left Ventricle (LV), Right Ventricle (RV) and Myocardium (Myo).

Implementation Details: Our 4D cardiac motion estimation framework employs an instance-wise test-time optimization strategy, requiring no pre-training or additional training data. We normalize spatial coordinates to the range [0,1]. The initial positions of 3D Gaussian primitives are obtained through uniform sampling of the ED phase segmentation mask positions, while control nodes are also initialized from these Gaussian positions (with the maximum number equal to the initial number of Gaussians). The framework is implemented in PyTorch and optimized for 20,000 iterations (approximately 11 minutes). The optimization process follows a two-stage strategy: the first 1,000 iterations focus solely on optimizing the canonical space 3D Gaussians to establish stable positions and shapes, followed by joint optimization of 3D Gaussians, control node positions and the deformation network. We employ the Adam optimizer [12] with different initial learning rates for various components: 3D Gaussian positions: 1e-4; Intensity values: 5e-3; Rotation and scale parameters: 1e-4; Control points: 1e-4 (update at iteration 5,000); Deformation network: 1e-6. The learning rates undergo exponential decay from 1e-4 to 1e-7, with optimizer parameters $\beta = (0.9, 0.999)$ and $\epsilon = 1e - 15$. Gaussian densification is performed every 500 iterations starting from the 500th iteration. For the ACDC dataset preprocessing, we follow the state-of-the-art benchmark protocol presented in [25]. All slices are resampled to 1.5×1.5×3.15mm resolution and center-cropped to 128×128×32 dimensions. The image intensities are normalized to [0,1]. All experiments were conducted on a single NVIDIA RTX 3090 GPU.

Evaluation Metrics: We employ multiple metrics to comprehensively evaluate our motion estimation framework. Following [25], the anatomical accuracy is assessed using the Dice score to measure the overlap between predicted cardiac segmentations and ground-truth annotations. Image reconstruction quality

Fig. 2. Quantitative comparison of the estimated ES frame of various cardiac condition.

Table 3. Quantitative comparison of different methods. (mean ± std, best and second-best results are in **bold** and <u>underline</u>, respectively).

Method	HD ↓	$\|\|J\|-1\|$ ↓	$det(J_\phi) \leq 0$ ↓	Params (M) ↓
LDDMM [5]	$6.562_{\pm 2.1}$	$451.8_{\pm 162.3}$	$653.5_{\pm 371.2}$	-
RDMM [23]	$5.728_{\pm 1.5}$	$144.2_{\pm 63.67}$	$266.0_{\pm 165.3}$	-
ANTs (SyN) [2]	$6.242_{\pm 1.6}$	$15.82_{\pm 22.30}$	$57.26_{\pm 37.74}$	-
VoxelMorph [3]	$5.336_{\pm 1.3}$	$0.260_{\pm 0.070}$	$0.079_{\pm 0.058}$	<u>0.327</u>
DiffuseMorph [11]	$3.977_{\pm 1.2}$	$0.237_{\pm 0.068}$	$0.061_{\pm 0.038}$	<u>0.327</u>
CorrMLP [15]	$3.552_{\pm 1.3}$	$0.248_{\pm 0.055}$	$0.059_{\pm 0.022}$	13.36
DeepTag [26]	$3.716_{\pm 1.4}$	$0.185_{\pm 0.067}$	$0.044_{\pm 0.025}$	0.362
Transmatch [7]	$3.361_{\pm 1.1}$	$0.226_{\pm 0.050}$	$0.077_{\pm 0.054}$	70.71
FSDiffReg [20]	<u>$3.283_{\pm 1.2}$</u>	$0.214_{\pm 0.054}$	$0.054_{\pm 0.026}$	1.320
GPTrack [25]	**$3.145_{\pm 1.1}$**	<u>$0.178_{\pm 0.024}$</u>	<u>$0.032_{\pm 0.021}$</u>	1.094
Dyna3DGR (ours)	$3.940_{\pm 0.18}$	**$0.002_{\pm 0.0003}$**	**$0.0_{\pm 0.0}$**	0.604

is evaluated through Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [24] in Table 2. The physical plausibility of predicted Gaussian position displacement field (Δxyz) is evaluated through diffeomorphic properties by examining the Jacobian determinant of the deformation field. Specifically, we count the number of locations where ($det(J_\phi) \leq 0$), which indicates points where the deformation exhibits undesirable properties such as folding or tearing that violate the topology-preserving requirement [26]. Following [18], physiological plausibility is evaluated by the mean absolute deviation between Jacobian determinant of the predicted Gaussian position displacement field (Δxyz) and unity, denoted as $\|J_\phi - 1\|$, to measure the level of volume preservation, where significant deviations suggest violations of cardiac tissue incompressibility. The Haussdorff Distance (HD) results are also provided. Model complexity is reported as the total number of parameters (in millions) for computational efficiency assessment.
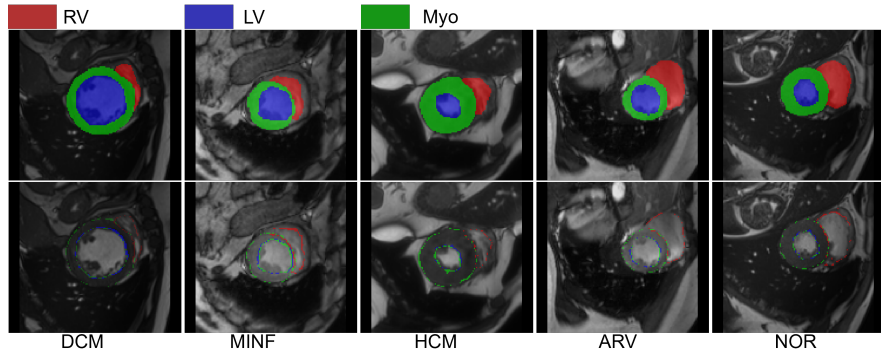
Fig. 3. Visualization of tracking errors (second row) at ES frame and corresponding ground truth (first row) across varying cardiac conditions.

Result: Our experiments demonstrate three key advantages of Dyna3DGR: (1) Superior anatomical accuracy with 17.73% improvement in Dice score and 12.63% increase in SSIM metrics compared to prior arts (paired t-test, p < 0.001); (2) Near-perfect physical plausibility (Jacobian deviation=0.002); (3) Lightweight architecture, fewer than most current methods.

The quantitative results presented in Tables 2 and 3 are evaluated following the same preprocessing protocol as GPTrack for fair comparison. Our method requires no training phase, and the reported metrics represent average performance of all samples. In terms of anatomical accuracy, Dyna3DGR consistently outperforms existing methods, maintaining above 90% accuracy for both normal and pathological cases. Unlike discrete voxel representation, Gaussian ellipsoids influence intensities across a patch of voxels, and the orientation of the ellipsoids naturally encodes structural contour information. Therefore, learning deformations in Gaussian representation compared to learning voxel deformations better preserves contour information during motion, resulting in more accurate tracking and reconstruction of cardiac dynamics.

The results in Fig. 2 show Dyna3DGR's consistent tracking performance across various cardiac conditions, while Fig. 3 provides detailed visualization of tracking errors at the ES frame. The ablation study investigates the impact of varying the number of control nodes on model performance, with results presented in Table 4. The findings demonstrate a positive correlation between the number of control nodes and overall performance across all metrics.

4   Conclusion

This paper presents Dyna3DGR, that uniquely integrates explicit 3D Gaussian representation with implicit neural motion field modeling for cardiac motion estimation. The approach addresses key limitations of existing methods by eliminating the need for training data and bridging the gap between representation space

Table 4. Ablation study on numbers of control node. (mean $\pm$ std, best are in **bold**)

Number	2048	4096	6144	8192	10240
Dice (%) $\uparrow$	$92.51_{\pm2.34}$	$94.60_{\pm1.78}$	$95.27_{\pm1.99}$	$95.70_{\pm2.30}$	$\mathbf{96.62}_{\pm3.36}$
PSNR (dB) $\uparrow$	$31.65_{\pm2.50}$	$32.68_{\pm2.52}$	$33.18_{\pm2.62}$	$33.75_{\pm2.83}$	$\mathbf{34.66}_{\pm2.43}$
SSIM (%) $\uparrow$	$94.78_{\pm0.35}$	$95.99_{\pm0.32}$	$96.40_{\pm0.30}$	$96.72_{\pm0.27}$	$\mathbf{97.08}_{\pm0.35}$
$\|J\|-1 \downarrow$	$0.0041_{\pm0.001}$	$0.0029_{\pm0.001}$	$0.0030_{\pm0.001}$	$0.0028_{\pm0.001}$	$\mathbf{0.0021}_{\pm0.0003}$

and image space. Through self-supervised optimization, Dyna3DGR achieves accurate cardiac motion estimation. The evaluations on the ACDC dataset demonstrate that our method surpasses existing diffeomorphic registration approaches.

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

References

1. Amzulescu, M.S., De Craene, M., Langet, H., Pasquet, A., Vancraeynest, D., Pouleur, A.C., Vanoverschelde, J.L., Gerber, B.: Myocardial strain imaging: review of general principles, validation, and sources of discrepancies. European Heart Journal-Cardiovascular Imaging **20**(6), 605–619 (2019)
2. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical image analysis **12**(1), 26–41 (2008)
3. 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)
4. Becciu, A., van Assen, H., Florack, L., Kozerke, S., Roode, V., ter Haar Romeny, B.M.: A multi-scale feature based optic flow method for 3d cardiac motion estimation. In: Scale Space and Variational Methods in Computer Vision: Second International Conference, SSVM 2009, Voss, Norway, June 1-5, 2009. Proceedings 2. pp. 588–599. Springer (2009)
5. 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**, 139–157 (2005)
6. 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)
7. Chen, Z., Zheng, Y., Gee, J.C.: Transmatch: A transformer-based multilevel dual-stream feature matching network for unsupervised deformable image registration. IEEE transactions on medical imaging **43**(1), 15–27 (2023)

8. Fu, X., Li, Y., Tang, F., Li, J., Zhao, M., Teng, G.J., Zhou, S.K.: 3dgr-car: Coronary artery reconstruction from ultra-sparse 2d x-ray views with a 3d gaussians representation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 14–24. Springer (2024)

9. Guo, Y., Bi, L., Wei, D., Chen, L., Zhu, Z., Feng, D., Zhang, R., Wang, Q., Kim, J.: Unsupervised landmark detection-based spatiotemporal motion estimation for 4-d dynamic medical images. IEEE transactions on cybernetics **53**(6), 3532–3545 (2021)

10. Huang, Y.H., Sun, Y.T., Yang, Z., Lyu, X., Cao, Y.P., Qi, X.: Sc-gs: Sparse-controlled gaussian splatting for editable dynamic scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4220–4230 (2024)

11. Kim, B., Han, I., Ye, J.C.: Diffusemorph: Unsupervised deformable image registration using diffusion model. In: European conference on computer vision. pp. 347–364. Springer (2022)

12. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. In: International Conference on Learning Representations (ICLR) (2015)

13. Krebs, J., Delingette, H., Mailhé, B., Ayache, N., Mansi, T.: Learning a probabilistic model for diffeomorphic registration. IEEE transactions on medical imaging **38**(9), 2165–2176 (2019)

14. Li, Y., Fu, X., Zhao, S., Jin, R., Zhou, S.K.: Sparse-view ct reconstruction with 3d gaussian volumetric representation. arXiv preprint arXiv:2312.15676 (2023)

15. 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)

16. Meng, Q., Bai, W., O'Regan, D.P., Rueckert, D.: Deepmesh: Mesh-based cardiac motion tracking using deep learning. IEEE transactions on medical imaging (2023)

17. Morales, M.A., Izquierdo-Garcia, D., Aganj, I., Kalpathy-Cramer, J., Rosen, B.R., Catana, C.: Implementation and validation of a three-dimensional cardiac motion estimation network. Radiology: Artificial Intelligence **1**(4), e180080 (2019)

18. Qin, C., Wang, S., Chen, C., Bai, W., Rueckert, D.: Generative myocardial motion tracking via latent space exploration with biomechanics-informed prior. Medical Image Analysis **83**, 102682 (2023)

19. Qin, C., Wang, S., Chen, C., Qiu, H., Bai, W., Rueckert, D.: Biomechanics-informed neural networks for myocardial motion tracking in mri. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part III 23. pp. 296–306. Springer (2020)

20. Qin, Y., Li, X.: Fsdiffreg: Feature-wise and score-wise diffusion-guided unsupervised deformable image registration for cardiac images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 655–665. Springer (2023)

21. Reed, G.W., Rossi, J.E., Cannon, C.P.: Acute myocardial infarction. The Lancet **389**(10065), 197–210 (2017)

22. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J.: Non-rigid registration using free-form deformations: application to breast mr images. IEEE transactions on medical imaging **18**(8), 712–721 (1999)

23. Shen, Z., Vialard, F.X., Niethammer, M.: Region-specific diffeomorphic metric mapping. Advances in Neural Information Processing Systems **32** (2019)

24. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)
25. Yang, J., Lin, Y., Pu, B., Li, X.: Bidirectional recurrence for cardiac motion tracking with gaussian process latent coding. Advances in Neural Information Processing Systems **37**, 34800–34823 (2025)
26. Ye, M., Kanski, M., Yang, D., Chang, Q., Yan, Z., Huang, Q., Axel, L., Metaxas, D.: Deeptag: An unsupervised deep learning method for motion tracking on cardiac tagging magnetic resonance images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7261–7271 (2021)
27. Yu, H., Sun, S., Yu, H., Chen, X., Shi, H., Huang, T.S., Chen, T.: Foal: Fast online adaptive learning for cardiac motion estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4313–4323 (2020)
28. Yuan, X., Liu, C., Wang, Y.: 4d myocardium reconstruction with decoupled motion and shape model. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21252–21262 (2023)
29. Zheng, Q., Delingette, H., Ayache, N.: Explainable cardiac pathology classification on cine mri with motion characterization by semi-supervised learning of apparent flow. Medical image analysis **56**, 80–95 (2019)