CineMesh4D: Personalized 4D Whole Heart Reconstruction from Sparse Cine MRI

Xiaoyue Liu[1], Xiaohan Yuan[1,2], Mark Y Chan[3,4], Ching-Hui Sia[3,4], and Lei Li[1(✉)]

[1] Department of Biomedical Engineering, National University of Singapore, Singapore
[2] School of Automation, Southeast University, Nanjing, China
[3] Department of Medicine, National University of Singapore, Singapore
[4] Department of Cardiology, National University Heart Centre Singapore, Singapore
lei.li@nus.edu.sg

Abstract. Accurate 3D+t whole-heart mesh reconstruction from cine MRI is a clinically crucial yet technically challenging task. The difficulty of this task arises from two coupled factors: inherently sparse sampling of 3D cardiac anatomy by 2D image slices and the tight coupling between cardiac shape and motion. Current cardiac image-to-mesh approaches typically reconstruct only a subset of cardiac chambers or a single phase of the cardiac cycle. In this work, we propose CineMesh4D, a novel end-to-end 4D (3D+t) pipeline that directly reconstructs patient-specific whole-heart mesh from multi-view 2D cine MRI via cross-domain mapping. Specifically, we introduce a differentiable rendering loss that enables supervision of 3D+t whole-heart mesh from multi-view sparse contours of cine MRI. Furthermore, we develop a dual-context temporal block that fuses global and local cardiac temporal information to capture high-dimensional sequential patterns. In quantitative and qualitative evaluations, CineMesh4D outperforms existing approaches in terms of reconstruction quality and motion consistency, providing a practical pathway for personalized real-time cardiac assessment. The code will be publicly released once the manuscript is accepted.

Keywords: 4D Whole Heart · Differentiable Rendering · Dual-Context Temporal Block · cine MRI · Image-to-Mesh Mapping · Digital Twin

1 Introduction

Cardiovascular diseases remain a leading cause of global mortality, underscoring the need for accurate assessment of cardiac structure and function to guide clinical decisions [18]. Cardiac digital twin, a patient-specific computational model of the heart, holds promise for personalized diagnosis and therapy [22]. A critical step in building such a twin is anatomical twinning: the reconstruction of a complete 3D+t whole-heart anatomy from clinical images. However, standard 2D cine MRI with large slice spacing provides only incomplete volumetric coverage of the heart. This limits patient-specific anatomical context and can compromise

global functional quantification. While 3D MRI or CT offers complete coverage, their extended scan time and requirement for multiple breath hold hinder routine clinical adoption [7]. Reconstructing patient-specific 3D+t whole-heart anatomy directly from multi-view 2D cine MRI therefore represents a key step towards cardiac digital twins [10].

Recently, deep learning-based approaches have emerged as powerful paradigm for 3D cardiac geometry reconstruction [11, 16]. However, many methods are limited to ventricular geometry, reducing their applicability to whole-heart functional twinning and multi-chamber volume dynamics. For instance, Meng et al. [13] proposed a prior-based deformation framework for left ventricular myocardium geometry, whereas MR-Net [1] learned an image-mesh mapping that deformed a template mesh for biventricular anatomy. Instead, Kong et al. [9] predicted whole-heart surface meshes from high-resolution 3D CT and MRI data, but they relied on 3D volumetric acquisitions which are time-consuming and resource-intensive. In contrast, Gaggion et al. [4] proposed a mesh reconstruction framework that achieved whole-heart reconstruction from multi-view sparse cine MRI. While the framework achieves promising shape accuracy, it only considers two cardiac frames instead of the full cardiac cycle. This prevents it from learning a joint representation of cardiac shape and motion or leveraging temporal context, which are essential for modeling cardiac dynamics.

To solve this, we develop CineMesh4D, a novel end-to-end pipeline for reconstructing high-fidelity 4D whole-heart meshes directly from sparse multi-view cine MRI. Unlike prior approaches that rely on volumetric imaging or intermediate cardiac segmentation results, CineMesh4D learns a direct mapping from 2D+t image data to temporally coherent 3D meshes. This is achieved through three key contributions. First, we propose a differentiable rendering loss derived from the Beer–Lambert law, which effectively leverages multi-view 2D+t sparse contours as supervision for 3D+t mesh reconstruction. Second, to model reliable cardiac motion, we incorporate a dual-context temporal block that fuses global and local temporal information, producing a temporally coherent latent representation for mesh generation. Third, the entire pipeline is trained end-to-end, jointly optimizing shape reconstruction and temporal consistency via direct image-to-mesh mapping. This integrated framework offers a robust and practical approach for advancing patient-specific cardiac modeling and functional analysis.

2 Methodology

2.1 Domain-Specific Anatomical Feature Extraction

To obtain anatomical information from cine MRI, we employ a pretrained cardiac U-Net to extract the cardiac chamber regions [19]. Given multi-view cine MRI sequences $\{I_t\}_{t=1}^{N}$, we pass the CMR sequence through the pretrained U-Net encoder E_{seg} to obtain anatomical feature embeddings $\mathbf{z}_t^{\text{anatomy}} \in \mathbb{R}^{d_a}$. Specifically, we use a 2D CNN encoder $E_{\text{seg}}^{\text{2D}}$ for long-axis (LAX) views and a 3D CNN encoder $E_{\text{seg}}^{\text{3D}}$ for short-axis (SAX) stacks.

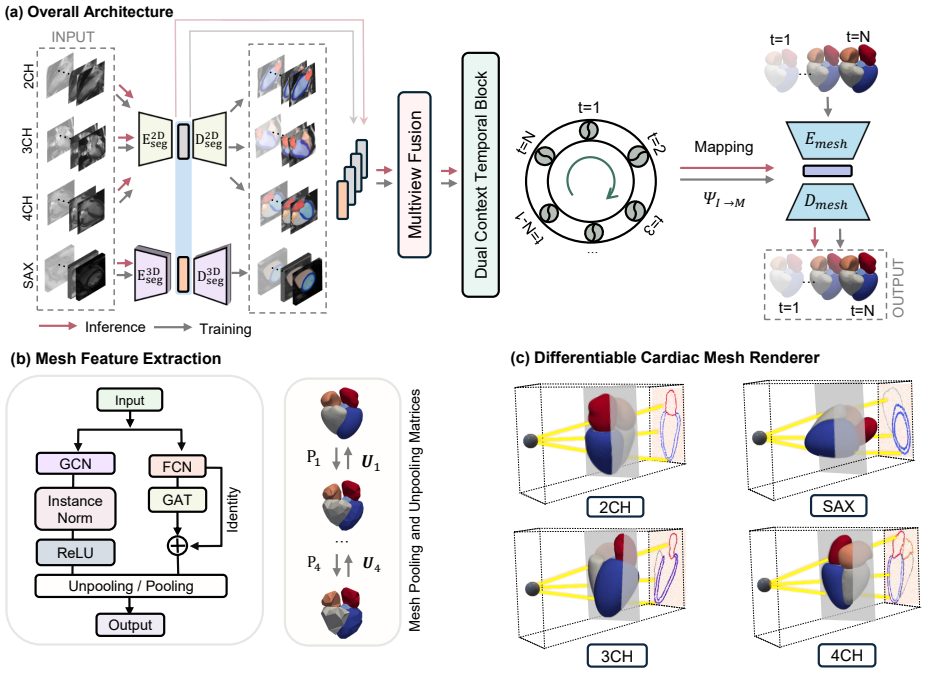

Fig. 1. Framework of the proposed CineMesh4D for direct 4D whole-heart reconstruction from multi-view cine MRI via a novel image-to-mesh mapping architecture.

For the mesh domain, we adopt a variational autoencoder (VAE) to capture cardiac anatomical variability. Each 4D mesh sequence $\{M_t\}_{t=1}^{N}$ is modeled as a sequence of spatial graphs $\{\mathcal{G}_t\}_{t=1}^{N}$, with $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E})$ representing the mesh at time t. Here, $\mathcal{V}_t \in \mathbb{R}^{V\times 3}$ denotes vertex coordinates, where V is the number of mesh nodes, and $\mathcal{E}$ denotes fixed mesh connectivity. As shown in Fig. 1 (b), each mesh feature extraction block applies graph convolutional network (GCN) branch to embed anatomical features. In parallel, we incorporate Exphormer [21] as graph attention pathway and integrate it into our architecture via a residual skip connection. This design complements neighborhood aggregation by enabling interactions between distant regions, strengthening structure-aware representations. Each mesh feature extraction block concludes with a pooling matrix $\mathbf{P}_m$ or an unpooling matrix $\mathbf{U}_m$ for mesh resolution transition [17]. The encoder E_{mesh} maps the input mesh at time step t to the latent Gaussian distribution. The mesh decoder D_{mesh} reconstructs the surface as $\hat{M}_t = D_{\text{mesh}}(\mathbf{z}_t^{\text{mesh}})$. The optimization of the MeshVAE is via:

$$\mathcal{L}_{Mesh}^{\text{recon}} = \frac{1}{N}\sum_{t=1}^{N}\left\|M_t - \hat{M}_t\right\|^2 + \mathcal{L}_{\text{KL}}, \tag{1}$$

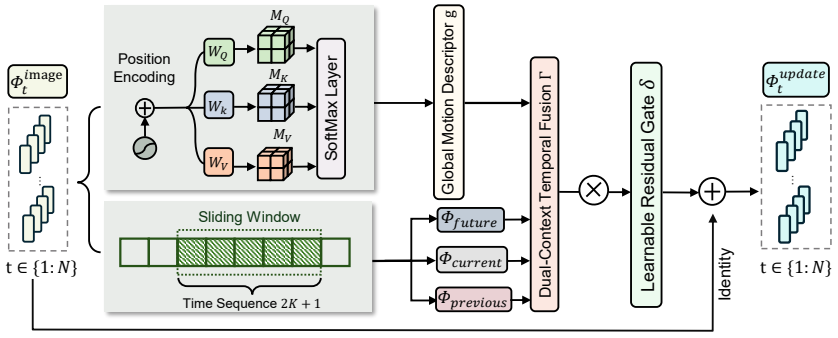

Fig. 2. Dual-context temporal module to learn both global and local cardiac patterns.

where M_t and $\hat{M}_t$ denote the input and reconstructed mesh at time t, and $\mathcal{L}_{\mathrm{KL}}$ is the Kullback-Leibler (KL) divergence term.

2.2 Dual-Context Latent Regularization for Temporal Coherence

Clinically meaningful cardiac function is manifested in both cycle-level temporal patterns (global) and short-term inter-frame consistency (local). We therefore adopt a dual-context temporal design where the global branch summarizes the entire sequence to capture long-range temporal trends, while a local branch enforces short-term consistency among neighboring frames in the latent representation. At each time step t, the anatomical latent embeddings $\mathbf{z}_t^{\mathrm{anatomy}}$ from multiple views are passed through a multi-view fusion module implemented by fully connected layers and linear projection, yielding a unified representation Φ_t^{image}. As illustrated in Fig. 2, the sequential embedding is incorporated into the fused latent sequence via a sinusoidal positional encoding [23]. Temporal self-attention operates on the stacked latent embedding, and the pooled outputs yield a global motion descriptor $\mathbf{g}$ that encodes sequence-wide temporal context. For each frame, neighboring frames within a temporal sliding window $\mathcal{W}_t$ of total size $(2K+1)$ are grouped into three relative temporal categories, namely $\Phi_{\mathrm{previous}} = \{\Phi_{t-j}^{\mathrm{image}}\}_{j=1}^{K}$, $\Phi_{\mathrm{current}} = \Phi_t^{\mathrm{image}}$, and $\Phi_{\mathrm{future}} = \{\Phi_{t+j}^{\mathrm{image}}\}_{j=1}^{K}$, for modeling direction-aware temporal dependencies. We define Γ as a dual-context fusion module that jointly leverages the global context $\mathbf{g}$ and the local neighborhood window $\mathcal{W}_t$ to produce the fused message $\hat{\Phi}_t$. This is implemented by feature-space stacking of sequence-level and sliding-window context, fused by a compact projection head.

$$\hat{\Phi}_t = \Gamma(\mathbf{g}, \mathcal{W}_t),\ \Phi_t^{\mathrm{update}} = \Phi_t^{\mathrm{image}} + \delta \odot \hat{\Phi}_t, \tag{2}$$

Here δ is a learnable residual gating vector that controls temporal fusion. The updated latent Φ_t^{update} is then used as a temporally coherent image representation as the input of D_{mesh}.

2.3 Differentiable Renderer for Contour-Guided Mesh Optimization

Multi-view cine MRI captures cardiac anatomy through sparse view-specific planes, providing complementary yet inherently 2D observations of a 3D heart. To reconstruct a complete 4D mesh under such plane supervision, a differentiable coupling is needed to translate vertex-to-plane distances into continuous per-view contributions, enabling stable contour-guided mesh optimization. We draw inspiration from the Beer-Lambert law [12], where light attenuates exponentially with path length: $\mathsf{I} = \mathsf{I}_0 \exp(-\mu \mathrm{L})$, and μ is the absorption coefficient. This offers a natural analogy: mesh vertices close to an imaging plane should contribute strongly to the projected contour, while distant vertices contribute negligibly.

As shown in Fig. 1 (c), for each case and view $w \in \{2\mathrm{CH}, 3\mathrm{CH}, 4\mathrm{CH}, \mathrm{SAX}\}$, we extract the affine transformation from the fixed image header to recover the view-plane geometry $\mathit{\Pi}^w$ in the world coordinate system. We represent the predicted surface-mesh vertices $\hat{\mathbf{v}}_t^i$ in the same world coordinate system to ensure consistent geometric computation. Here, i indexes the mesh vertices and $\hat{\mathbf{v}}_t^i$ denotes the i-th predicted vertex in the world coordinates. The plane $\mathit{\Pi}^w$ is the view-w imaging plane in world coordinates, represented by a point $\mathbf{c}^w$ on the plane and a unit normal $\mathbf{n}^w$, both in world coordinates. We define the vertex-to-plane normal distance as $R_t^{i,w} = \mathrm{dist}\big(\hat{\mathbf{v}}_t^i, \mathit{\Pi}^w\big) = \big|\mathbf{n}^w \cdot \big(\hat{\mathbf{v}}_t^i - \mathbf{c}^w\big)\big|$. Given the vertex-to-plane distance $R_t^{i,w}$, we denote by $q_t^{i,w}$ the plane-association probability of vertex i to view plane $\mathit{\Pi}^w$, where $\ell_{\mathrm{sigmoid}}(\cdot)$ is a sigmoid-window distance weighting and hyperparameter μ controls the sharpness of the distance-to-probability mapping, as $q_t^{i,w} = 1 - \exp\Big(-\mu\, \ell_{\mathrm{sigmoid}}(R_t^{i,w})\Big)$. Vertex-wise plane-association probabilities $q_t^{i,w}$ are projected onto the corresponding view plane and aggregated to form the probability map Q_t^w on each view, such that vertices closer to the imaging plane contribute more strongly, while vertices farther away are progressively down-weighted. The differentiable rendering (DR) loss can therefore be defined as:

$$\mathcal{L}_{\mathrm{DR}} = \sum_w \mathcal{L}_{\mathrm{B}}(Q_t^w, S_t^w), \tag{3}$$

where $\mathcal{L}_{\mathrm{B}}$ follows the boundary constraint in [8], and S_t^w denotes ground-truth segmentation in the view plane w. We compute $\mathcal{L}_{\mathrm{B}}$ at every time frame t.

2.4 Optimization Objective and Inference

We load pretrained weights for two modality-specific pairs of encoder and decoder, $\{E_{\mathrm{seg}}, D_{\mathrm{seg}}\}$ for the image domain and $\{E_{\mathrm{mesh}}, D_{\mathrm{mesh}}\}$ for the mesh domain. To provide additional flexibility during this process, low-rank adaptation (LoRA) blocks are introduced into the linear layers of the mesh decoder, $\mathrm{LoRA}(D_{\mathrm{mesh}})$, enabling parameter-efficient adaptation under a fixed backbone [6]. Reconstruction loss $\mathcal{L}_{\mathrm{MSE}}$ is defined as the mean squared error between the predicted $\hat{M}$ and the ground-truth mesh M. We apply two regularizers for mesh

optimization: $\mathcal{L}_{\text{edge}}$ penalizes edge-length changes, while $\mathcal{L}_{\text{norm}}$ enforces face-normal consistency [24]. The optimization objective for cross-domain mapping:

$$\mathcal{L}_{\text{Map}} = \lambda_{\text{MSE}}\,\mathcal{L}_{\text{MSE}}(\hat{M},\, M) + \lambda_{\text{DR}}\,\mathcal{L}_{\text{DR}} + \lambda_{\text{edge}}\,\mathcal{L}_{\text{edge}} + \lambda_{\text{norm}}\,\mathcal{L}_{\text{norm}} \quad (4)$$

During inference, the image encoder E_{seg} is frozen and image-to-mesh mapping $\Psi_{I\to M}$ conditions the mesh decoder D_{mesh} to generate subject-specific meshes:

$$\{\hat{M}_t\}_{t=1}^{N} = \left\{\text{LoRA}(D_{\text{mesh}})\left(\Psi_{I\to M}\big(E_{\text{seg}}(I_t)\big)\right)\right\}_{t=1}^{N} \quad (5)$$

3 Experiments and Results

Data Acquisition and Pre-Processing. The dataset consists of 222 subjects scanned with standard multi-view cardiac cine MRI, collected from National University Hospital, and each view comprises 25 frames. All images were center-cropped into a unified size of 150×150 and fed into an automated segmentation pipeline with manual refinement [3]. These refined sparse segmentations were subsequently resampled and converted into dense 3D segmentation via the atlas-deformation [25]. To generate topology-preserved whole heart mesh as reference, we further transformed the 3D segmentation onto a high-resolution template whole-heart surface mesh. Each mesh consists of five anatomical components, namely left ventricle (LV), LV myocardium, right ventricle (RV), left atrium (LA), and right atrium (RA). We randomly split our dataset into 155 training, 10 validation, and 57 test cases.

Implementation. The framework was implemented in PyTorch and trained on one NVIDIA GeForce RTX 4090D GPU. We pretrained MeshVAE for 350 epochs using a learning rate of 1×10^{-4} and trained the image–mesh mapping for 400 epochs using the Adam optimizer with a learning rate of 5×10^{-5}. The temporal sliding window size is set to 5 with $K = 2$. We set $\mu = 8$ in the rasterization loss. Hyperparameters are selected as $\lambda_{\text{MSE}} = 10$, $\lambda_{\text{DR}} = 5$, $\lambda_{\text{edge}} = \lambda_{\text{norm}} = 0.8$.

Evaluation Metrics. Reconstruction accuracy is quantified using vertex-wise Mean Absolute Error (MAE) and Mean Squared Error (MSE) computed between the reference and predicted surface meshes. We report mesh-to-mesh Chamfer Distance (CD), Hausdorff Distance (HD), as well as inference time per mesh, for direct comparison with prior full heart and biventricular pipelines. Mean Contour Distance (MCD) and Boundary F-score (BF) are reported between the reference and predicted contours for each view, where MCD measures the average contour distance and BF quantifies boundary alignment [2, 5]. E_{vol} is reported as the whole-heart volumetric error (mL), averaged over frames, using the absolute volume difference. Mesh jitter (J_m) measures the temporal smoothness of vertex trajectories over time [20, 26].

Table 1. Quantitative evaluation of whole-heart and substructure mesh reconstruction.

	Full Mesh	**LV**	**RV**	**LA**	**RA**
HybridVNet					
MAE (mm) ↓	2.18(0.53)	1.70(0.54)	1.97(0.55)	2.31(0.70)	2.51(0.60)
MSE (mm^2) ↓	8.80(5.31)	5.10(3.67)	7.02(4.72)	9.58(9.24)	11.72(10.14)
J_m ($mm/frame^3$) ↓	2.29(0.23)	1.73(0.16)	2.07(0.15)	2.27(0.25)	2.70(0.28)
CineMesh4D					
MAE (mm) ↓	**1.68(0.31)**	**1.59(0.34)**	**1.64(0.35)**	**1.99(0.48)**	**1.86(0.58)**
MSE (mm^2) ↓	**5.06(1.79)**	**4.45(1.93)**	**4.86(2.15)**	**6.90(3.19)**	**6.26(4.06)**
J_m ($mm/frame^3$) ↓	**0.77(0.17)**	**0.69(0.15)**	**0.78(0.17)**	**0.91(0.21)**	**0.93(0.23)**

Table 2. Summary of 2D contour results of predicted mesh on different views.

	HybridVNet		CineMesh4D (w/o λ_{DR})		**CineMesh4D**	
View	BF (%) ↑	MCD (mm) ↓	BF (%) ↑	MCD (mm) ↓	BF (%) ↑	MCD (mm) ↓
2CH	55.80(12.05)	2.85(4.76)	60.13(9.89)	2.30(0.47)	**65.47(10.41)**	**1.99(0.41)**
3CH	54.62(9.01)	3.60(0.66)	58.27(9.31)	2.32(0.46)	**65.71(10.32)**	**1.97(0.44)**
4CH	57.01(9.14)	3.69(5.10)	59.64(8.68)	2.42(0.46)	**68.24(9.24)**	**1.89(0.38)**
SAX	62.74(13.56)	2.22(0.89)	60.74(11.53)	2.29(0.59)	**62.86(13.74)**	**2.03(0.62)**

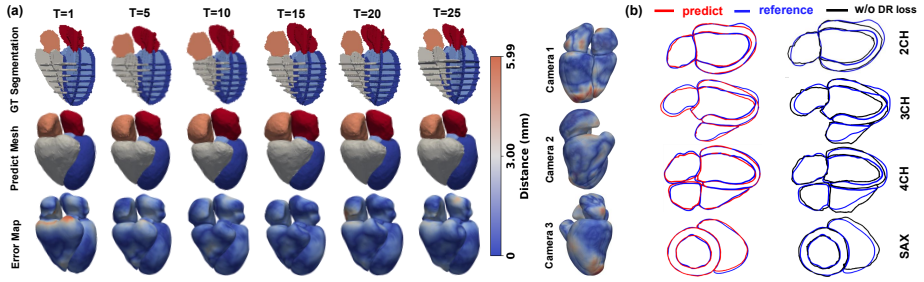

Fig. 3. Visualization of reconstructed whole heart. (a) 3D whole-heart mesh across different phases. (b) 2D contour overlap with and without differentiable rendering (DR).

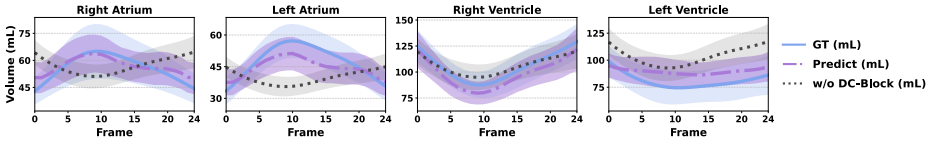

Fig. 4. Illustration of cardiac chamber volume change of all test data. DC: dual-context.

Comparison Study. Table 1 reports overall reconstruction errors and temporal smoothness, where CineMesh4D achieved better performance compared to the baseline across substructures. Ventricular reconstructions were consistently more accurate than atrial ones, as ventricles benefited from dense short-axis coverage, while atria were sparsely sampled through only a few long-axis planes. Table 2 examined boundary fidelity, where BF-score increased by 17.33%, 20.30%,

Table 3. Comparison of cardiac reconstruction results from different methods.

	PointNet++	PU-Net	CPD	MR-Net	HybridVNet	**Ours**
CD (mm) ↓	13.03(2.86)	12.15(2.85)	12.10(6.53)	4.39(1.38)	4.13(1.13)	**3.41(0.62)**
HD (mm) ↓	16.04(3.57)	15.74(3.07)	13.05(7.04)	6.69(1.88)	5.17(1.02)	**5.13(0.80)**
Infer time (s) ↓	< 0.1	< 0.1	37.45	< 0.1	< 0.1	< 0.1

Table 4. Summary of the ablation study results of the proposed method.

	w/o SAX	w/o 2CH	w/o 3CH	w/o 4CH	w/o UNet	w/o VAE	w/o DC	**Ours**
MSE↓	5.64(2.00)	5.59(1.94)	5.87(1.96)	5.99(1.91)	6.43(1.94)	5.67(2.03)	5.32(2.02)	**5.06(1.79)**
MAE↓	1.76(0.31)	1.72(0.29)	1.80(0.30)	1.82(0.28)	1.87(0.28)	1.77(0.32)	1.69(0.29)	**1.68(0.31)**
E_{vol}↓	25.3(10.6)	27.3(12.8)	25.3(11.7)	26.3(9.8)	30.8(14.5)	25.0(10.8)	24.7(13.0)	**17.3(11.2)**

and 19.70% relative to HybridVNet for 2CH, 3CH and 4CH views, respectively. Larger gains in LAX views stemmed from their single-slice nature, with only one 2D plane per view, small 3D deviations translated into pronounced contour misalignment, making direct boundary supervision through differentiable rendering particularly impactful. Table 3 compares surface metrics against PointNet++ [15], PU-Net [27], CPD [14], MR-Net [1], and HybridVNet [4], where our method achieved the lowest CD and HD. Qualitatively, Fig. 3 (a) and a supplementary video visualize reconstruction quality across frames, while Fig. 4 presents chamber volume curves that exhibit expected cardiac motion patterns, confirming physiological consistency of our prediction.

Ablation Study. Table 4 presents the quantitative results of our ablation study. Multi-view feature fusion proved essential, which is expected as SAX provided through-plane resolution while LAX views captured apical and longitudinal anatomy, together enabling complete 3D reconstruction. The exclusion of any single LAX view increased reconstruction error, with the 4CH view being most critical for shape accuracy, as it simultaneously visualized all four chambers. Geometric priors from MeshVAE pretraining benefited downstream optimization by constraining reconstructions to a plausible anatomical manifold. Segmentation-based pretraining outperformed intensity-based self-supervision, as U-Net provided anatomical boundaries during segmentation, yielding cleaner edge-specific features critical for accurate mesh-image alignment. The dual-context block was critical for temporal consistency, particularly for atria where limited short-axis coverage yielded inherently sparser image features. This can be further confirmed by the volume curves presented in Fig. 4. Finally, differentiable rendering supervision via $\mathcal{L}_{\text{DR}}$ enforced contour alignment across imaging planes that vertex-wise reconstruction loss neglected, as presented in Fig. 3 (b). While MSE minimized global vertex-wise distances, it did not penalize misalignment against specific 2D cuts. In contrast, our proposed $\mathcal{L}_{\text{DR}}$ explicitly complement this by rendering and comparing against 2D segmentations.

4 Conclusion

We propose CineMesh4D, a temporally coherent framework for 3D+t whole-heart mesh reconstruction directly from sparse multi-view cine MRI sequences. By introducing an end-to-end 4D image-to-mesh pipeline, a Beer–Lambert inspired differentiable rendering loss, and a dual-context temporal block, CineMesh4D achieves reliable reconstruction accuracy and strong temporal consistency. Our work establishes a practical framework for personalized cardiac function analysis and enables downstream applications. In future work, we will incorporate pathological metadata and clinical signals to enable comprehensive cardiac disease prediction.

Acknowledgments. This work was supported by Singapore National Medical Research Council Open Fund - Young Individual Research Grant (25-1321-A0001), and Singapore Ministry of Education Tier 1 grant (25-1097-P0001).

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

References

1. Chen, X., Ravikumar, N., Xia, Y., Attar, R., Diaz-Pinto, A., Piechnik, S.K., Neubauer, S., Petersen, S.E., Frangi, A.F.: Shape registration with learned deformations for 3d shape reconstruction from sparse and incomplete point clouds. Medical image analysis **74**, 102228 (2021)
2. Cheng, D., Liao, R., Fidler, S., Urtasun, R.: Darnet: Deep active ray network for building segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7431–7439 (2019)
3. Dillon, J.R., Mauger, C., Zhao, D., Deng, Y., Petersen, S.E., McCulloch, A.D., Young, A.A., Nash, M.P.: An open-source end-to-end pipeline for generating 3d+t biventricular meshes from cardiac magnetic resonance imaging. In: International Conference on Functional Imaging and Modeling of the Heart. pp. 372–383. Springer (2025)
4. Gaggion, N., Matheson, B.A., Xia, Y., Bonazzola, R., Ravikumar, N., Taylor, Z.A., Milone, D.H., Frangi, A.F., Ferrante, E.: Multi-view hybrid graph convolutional network for volume-to-mesh reconstruction in cardiovascular mri. Medical Image Analysis **104**, 103630 (2025)
5. Gur, S., Shaharabany, T., Wolf, L.: End to end trainable active contours via differentiable rendering. arXiv preprint arXiv:1912.00367 (2019)
6. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. Iclr **1**(2), 3 (2022)
7. Kawakubo, M., Arai, H., Nagao, M., Yamasaki, Y., Sanui, K., Nishimura, H., Kadokami, T.: Global left ventricular area strain using standard two-dimensional cine magnetic resonance imaging with inter-slice interpolation. Cardiovascular Imaging Asia **2**(4), 187–193 (2018)
8. Kervadec, H., Bouchtiba, J., Desrosiers, C., Granger, E., Dolz, J., Ayed, I.B.: Boundary loss for highly unbalanced segmentation. In: International conference on medical imaging with deep learning. pp. 285–296. PMLR (2019)

9. Kong, F., Wilson, N., Shadden, S.: A deep-learning approach for direct whole-heart mesh reconstruction. Medical image analysis **74**, 102222 (2021)
10. Li, L., Camps, J., Wang, Z.J., Beetz, M., Banerjee, A., Rodriguez, B., Grau, V.: Toward enabling cardiac digital twins of myocardial infarction using deep computational models for inverse inference. IEEE transactions on medical imaging **43**(7), 2466–2478 (2024)
11. Li, L., Ding, W., Huang, L., Zhuang, X., Grau, V.: Multi-modality cardiac image computing: A survey. Medical image analysis **88**, 102869 (2023)
12. Mayerhöfer, T.G., Pahlow, S., Popp, J.: The bouguer-beer-lambert law: Shining light on the obscure. ChemPhysChem **21**(18), 2029–2046 (2020)
13. Meng, Q., Bai, W., O'Regan, D.P., Rueckert, D.: Deepmesh: Mesh-based cardiac motion tracking using deep learning. IEEE transactions on medical imaging **43**(4), 1489–1500 (2023)
14. Myronenko, A., Song, X.: Point set registration: Coherent point drift. IEEE transactions on pattern analysis and machine intelligence **32**(12), 2262–2275 (2010)
15. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems **30** (2017)
16. Qiao, M., Basaran, B.D., Qiu, H., Wang, S., Guo, Y., Wang, Y., Matthews, P.M., Rueckert, D., Bai, W.: Generative modelling of the ageing heart with cross-sectional imaging and clinical data. In: International Workshop on Statistical Atlases and Computational Models of the Heart. pp. 3–12. Springer (2022)
17. Ranjan, A., Bolkart, T., Sanyal, S., Black, M.J.: Generating 3d faces using convolutional mesh autoencoders. In: Proceedings of the European conference on computer vision (ECCV). pp. 704–720 (2018)
18. Rivera, F.A., Ahmad, R., Trejo-Gutierrez, J.: Cardiovascular risk assessment: Practical tips for the internal medicine specialist. European Journal of Internal Medicine p. 106600 (2025)
19. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
20. Shin, S., Kim, J., Halilaj, E., Black, M.J.: Wham: Reconstructing world-grounded humans with accurate 3d motion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2070–2080 (2024)
21. Shirzad, H., Velingker, A., Venkatachalam, B., Sutherland, D.J., Sinop, A.K.: Exphormer: Sparse transformers for graphs. In: International Conference on Machine Learning. pp. 31613–31632. PMLR (2023)
22. Thangaraj, P.M., Benson, S.H., Oikonomou, E.K., Asselbergs, F.W., Khera, R.: Cardiovascular care with digital twin technology in the era of generative artificial intelligence. European Heart Journal **45**(45), 4808–4821 (2024)
23. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
24. Wang, N., Zhang, Y., Li, Z., Fu, Y., Liu, W., Jiang, Y.G.: Pixel2mesh: Generating 3d mesh models from single rgb images. In: Proceedings of the European conference on computer vision (ECCV). pp. 52–67 (2018)
25. Xu, Y., Xu, H., Sinclair, M., Puyol-Antón, E., Niederer, S.A., Chiribiri, A., Williams, S.E., Williams, M.C., Young, A.A.: Improved 3d whole heart geometry from sparse cmr slices. In: International Workshop on Statistical Atlases and Computational Models of the Heart. pp. 43–52. Springer (2024)

26. Yi, X., Zhou, Y., Habermann, M., Shimada, S., Golyanik, V., Theobalt, C., Xu, F.: Physical inertial poser (pip): Physics-aware real-time human motion tracking from sparse inertial sensors. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 13167–13178 (2022)
27. Yu, L., Li, X., Fu, C.W., Cohen-Or, D., Heng, P.A.: Pu-net: Point cloud upsampling network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2790–2799 (2018)