Mesh4D: A Motion-Aware Multi-View Variational Autoencoder for 3D+t Mesh Reconstruction

Mengyun Qiao[1,2,3*], Jin Zheng[4*], Weitong Zhang[5], Qiang Ma[2], Liu Li[5], Bernhard Kainz[5,6], Declan P. O'Regan[4], Paul M. Matthews[2,8,9], Steven Niederer[7], Wenjia Bai[2,3,5]

[1] Department of Mechanical Engineering, University College London, London, UK
[2] Department of Brain Sciences, Imperial College London, London, UK
[3] Data Science Institute, Imperial College London, London, UK
[4] MRC Laboratory of Medical Sciences, Imperial College London, London, UK
[5] Department of Computing, Imperial College London, London, UK
[6] FAU Erlangen-Nürnberg, Nürnberg, DE
[7] National Heart and Lung Institute, Imperial College London, London, UK
[8] UK Dementia Research Institute, Imperial College London, London, UK
[9] Rosalind Franklin Institute, Harwell Science and Innovation Campus, Didcot, UK
m.qiao21@imperial.ac.uk

Abstract. Reconstructing temporally coherent 3D meshes of the beating heart from multi-view MR images is an important but challenging problem. The challenge is entangled by the complexity in integrating multi-view data, the sparse coverage of a 3D geometry by 2D image slices, and the interplay between geometry and motion. Current approaches often treat mesh reconstruction and motion estimation as two separate problems. Here we propose Mesh4D, a novel motion-aware method that jointly learns cardiac shape and motion, directly from multi-view MR image sequences. The method introduces three key innovations: (1) A cross-attention encoder that fuses multi-view image information, (2) A transformer-based variational autoencoder (VAE) that jointly model the image feature and motion, and (3) A deformation decoder that generates continuous deformation fields and temporally smooth 3D+t cardiac meshes. Incorporating geometric regularisation and motion consistency constraints, Mesh4D can reconstruct high-quality 3D+t meshes (7,698 vertices, 15,384 faces) of the heart ventricles across 50 time frames, within less than 3 seconds. When compared to existing approaches, Mesh4D achieves notable improvements in reconstruction accuracy and motion smoothness, offering an efficient image-to-mesh solution for quantifying shape and motion of the heart and creating digital heart models.

Keywords: 3D+t mesh reconstruction · Shape and motion modelling · Motion-aware VAE · Multi-view fusion

[*] Equal contribution

1 Introduction

Cardiac imaging is essential for assessing the structure and function of the heart, a dynamic organ that undergoes continuous deformation during its rhythmic contraction and relaxation [1,18]. Cine cardiac MR (CMR) is one of the commonly used imaging modalities due to its excellent soft tissue contrast. However, it is a particularly challenging task to reconstruct 3D+t meshes from the cine CMR images, which cover the heart in multiple sparse views with anisotropic resolution. This has become a bottleneck limiting the clinical analyses of cardiac shape and motion, and computational modelling studies for the heart.

Current approaches for 3D+t cardiac mesh reconstruction can be broadly classified into segmentation-based methods and mesh-based methods. Segmentation-based methods extract meshes in two stages, an image segmentation stage, followed by a mesh construction stage [11,25,28]. However, they may struggle to maintain the temporal coherence for meshes between different time frames. Mesh-based methods fit a template mesh to the images and track the deformation of the mesh during cardiac motion [6,7,10,13,15]. Mesh-based methods improve the anatomical plausibility of the meshes, but could not make full use of the boundary information provided by the segmentation. Both segmentation-based and mesh-based methods may integrate statistical shape models (SSMs) to improve robustness by imposing shape priors [9,17,30]. A key limitation of existing methods is that mesh reconstruction and motion estimation are solved in separate steps, leading to suboptimal performance.

Another challenge in cardiac mesh reconstruction is that cardiac MR images have limited through-plane resolution [28], making it difficult to reconstruct high-resolution shape and motion representations. Multi-view imaging has been explored to address this, using images acquired from multiple planes to compensate for the low through-plane resolution [6,13,17,29]. Cross-modality learning, such as integrating cardiac MR and high-resolution CT images, has also been explored to improve shape and motion estimation from anisotropic MR images [7,22]. Existing 3D+t mesh reconstruction methods primarily focus on improving spatial accuracy, and may not always enforce motion consistency across time frames, leaving motion consistency an open challenge.

To overcome these limitations, we introduce `Mesh4D`, a motion-aware multi-view variational autoencoder (VAE) method for reconstructing 3D+t cardiac meshes directly from multi-view cardiac MR image sequences. Unlike traditional methods that treat mesh reconstruction and motion estimation separately, `Mesh4D` jointly learns both cardiac shape and motion from multi-view image sequences. This approach ensures the temporal coherence of the meshes across time. The key innovations include: 1) A cross-attention multi-view encoder, which integrates image features across multiple views; 2) A transformer-based VAE model, which learns long-range dependencies for multi-view features and motion dynamics; 3) A continuous deformation decoder, which reconstructs 3D+t meshes by learning vertex-wise displacements, ensuring anatomical correspondence and temporal coherence. Trained and evaluated on 1,984 multi-view CMR imaging scans, `Mesh4D` significantly improves reconstruction accuracy and

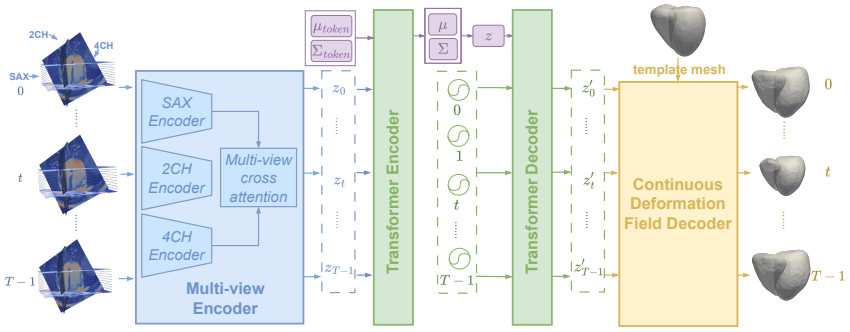

Fig. 1: **Mesh4D framework.** The multi-view encoder learns and integrates features for multi-view cardiac MR image sequences (SAX, 2CH, 4CH). The Transformer-based VAE models temporal dependencies, where the Transformer encoder models the temporal dependencies among the multi-view across time frames, and the Transformer decoder generates a sequence of latent representations. The continuous deformation field decoder learns vertex-wise deformation fields to warp a template mesh, producing anatomically consistent and temporally smooth 3D+t cardiac meshes.

motion smoothness compared to existing methods, offering a unified and efficient framework for high-quality 4D cardiac shape and motion modelling.

2 Methods

This section introduces Mesh4D, a framework for reconstructing 3D+t cardiac meshes directly from multi-view MR image sequences, illustrated in Figure 1. Mesh4D consists of three main components: 1) a multi-view encoder that extracts features from multi-view cardiac MR image sequences, including short-axis (SAX), long-axis two-chamber (2CH), and long-axis four-chamber (4CH) views. The features are integrated using cross-attention to create joint multi-view features. 2) a Transformer-based VAE, where the Transformer encoder captures long-range temporal dependencies among the multi-view features across time frames, and the Transformer decoder generates a sequence of smoothly evolving latent representations $\{z'_t | t = 0, 1, \ldots, T-1\}$, with T denoting the number of time frames. 3) a continuous deformation field decoder that takes the latent representations as input and generates vertex-wise deformation fields to warp a template mesh, creating anatomically consistent and temporally smooth 3D+t cardiac meshes.

2.1 Architecture

Multi-View Encoder In cardiac cine MR, the 3D geometry and motion of the heart is captured by multiple views, including an anisotropic 3D image sequence

from the SAX view and 2D image sequences from the 2CH and 4CH views. To extract spatial features from these multi-view image sequences, three view-specific encoders are developed: a 3D convolutional encoder for the SAX view and two 2D convolutional encoders for the 2CH and 4CH views. These encoders independently process each time frame, producing feature representations for each view.

After feature extraction, 2CH and 4CH features are concatenated into a unified representation to be used as keys and values in the cross-attention module. The SAX features serve as queries, attending to the combined 2CH-4CH features to extract complementary long-axis information. This fusion strategy helps SAX integrate complementary structural information from 2CH and 4CH views to augment limited through-plane resolution of SAX view.

Latent Motion Modelling via Transformer-based VAE To model temporal dependencies in the sequence of multi-view features, the extracted features are processed by a Transformer-based VAE [26,14]. In this formulation, the Transformer encoder serves as the probabilistic encoder of the VAE, mapping the sequence of multi-view features into a latent space. The Transformer decoder then reconstructs temporally smooth latent representations from the VAE latent space.

At each time frame t, the multi-view encoder generates a feature representation, forming a sequence of tokens $Z = \{z_0, z_1, \ldots, z_{T-1}\}$. Two learnable distribution parameter tokens, μ_{token} and Σ_{token}, inspired by the `[class]` token in the vision Transformer [8], are appended at the beginning of the sequence. The Transformer encoder processes the input sequence to estimate the mean (μ) and variance (Σ) of a latent distribution. The latent variable z is sampled from this distribution, using the reparameterisation trick: $z = \mu + \epsilon \cdot \Sigma$. The Transformer decoder takes the sampled latent variable z and sinusoidal temporal embeddings as input. The temporal embeddings are calculated from the time frame index t [8], representing the temporal position in a sequence. The output of the Transformer decoder is a sequence of latent representations $Z' = \{z'_0, z'_1, \ldots, z'_{T-1}\}$.

Continuous Deformation Field Decoder Instead of directly predicting absolute mesh coordinates, `Mesh4D` reconstructs cardiac motion by learning a temporally consistent deformation field that smoothly evolves across time steps [5]. The continuous deformation field decoder takes the sequence of latent representations $\{z'_t\}$ as input, and generates a deformation field of vertex-wise displacements for each time frame. The deformation field is used to warp a publicly available template mesh [3], generating 3D+t cardiac meshes across the time frames.

2.2 Loss Functions

We train the `Mesh4D` model using a combination of loss terms that accounts for both geometric properties and motion consistency.

Boundary Alignment. Following [16], we enforce the alignment between the predicted cardiac mesh V_t at time frame t and the boundary V^* derived from

the ground truth segmentation for all time frames. The boundary alignment loss is defined as a one-sided Chamfer distance:

$$\mathcal{L}_{\text{bound}} = \frac{1}{|V^*|} \sum_{v^* \in V^*} \min_{v_t \in V_t} \|v_t - v^*\|^2, \tag{1}$$

where v_t and v^* denote the predicted mesh vertex coordinate and segmentation boundary coordinate, respectively. This term ensures geometric accuracy by aligning the reconstructed 3D mesh with the segmentation boundary, which are 2D contours.

Template Alignment. We introduce an additional template alignment loss $\mathcal{L}_{\text{temp}}$ to enforce 3D shape consistency by minimising the Chamfer distance between the reconstructed 3D mesh and the 3D template mesh pre-registered to the subject space using non-rigid registration. This term ensures that the predicted mesh maintain a 3D geometry consistent with the template mesh.

Geometric Regularisation. To ensure high-quality mesh geometry and reduce distortions, we apply a loss for edge length alignment $\mathcal{L}_{\text{eg}}$, a loss for normal consistency $\mathcal{L}_{\text{nm}}$ and a loss for Laplacian smoothing $\mathcal{L}_{\text{lap}}$. The loss $\mathcal{L}_{\text{eg}}$ aligns the edge lengths of each face on the predicted mesh with those of the template, preventing excessive stretching or compression. The loss $\mathcal{L}_{\text{nm}}$ reduces the angular disparity between adjacent face normals [27]. The Laplacian smoothing term $\mathcal{L}_{\text{lap}}$ constrains vertex deviations to improve local smoothness [19].

Motion Consistency. Cardiac motion exhibits inherent temporal smoothness due to physiological constraints. The continuity of myocardial tissue and the electromechanical activation process prevents abrupt changes in shape and motion [2,12,24]. To leverage the physiological prior, we introduce a motion consistency loss that enforces temporal consistency of velocities across consecutive time frames. Let $\Delta v_t = v_{t+1} - v_t$ denote the vertex displacement from time t to $t+1$. The vertex velocity is defined as $u_t = \frac{\Delta v_t}{\Delta t}$ and becomes Δv_t if we let the uniform time step to be $\Delta t = 1$. The motion consistency loss is formulated as,

$$\mathcal{L}_{\text{mc}} = \frac{1}{|V_t|} \sum_{t=1}^{T-2} \sum_{v_t \in V_t} \|\Delta v_t - \Delta v_{t-1}\|^2 \tag{2}$$

This term encourages smooth acceleration and penalises abrupt changes in vertex-wise velocities.

Latent Space Regularisation. As in a standard VAE [26,14], the KL divergence loss $\mathcal{L}_{\text{KL}}$ is used to regularise the latent distribution to be close to a normal distribution $\mathcal{N}(0, I)$. The total loss for Mesh4D is defined as:

$$\mathcal{L} = \mathcal{L}_{\text{bound}} + \lambda_{\text{temp}}\mathcal{L}_{\text{temp}} + \lambda_{\text{eg}}\mathcal{L}_{\text{eg}} + \lambda_{\text{nm}}\mathcal{L}_{\text{nm}} + \lambda_{\text{lap}}\mathcal{L}_{\text{lap}} + \lambda_{\text{mc}}\mathcal{L}_{\text{mc}} + \beta\mathcal{L}_{\text{KL}}. \tag{3}$$

3 Experiments and Results

3.1 Dataset and Experimental Setup

This study used a dataset of CMR images from 1,984 subjects in the UK Biobank, split into 1,480 for training, 174 for validation, and 330 for test. For each subject, image sequences were acquired from three standard views: SAX, 2CH, and 4CH, with each sequence consisting of 50 time frames. The template mesh contains three anatomical structures: left ventricular endocardium (LV Endo), left ventricular epicardium (LV Epi), and right ventricular (RV). It consists of 7,698 vertices and 15,384 faces. `Mesh4D` was trained for 200 epochs with a batch size of 1 subject, a learning rate of 1e-3, and a weight decay of 1e-4. The loss function incorporates multiple weighted terms: $\lambda_{temp} = 0.01$, $\lambda_{eg} = 1$, $\lambda_{nm} = 10$, $\lambda_{lap} = 10$, $\lambda_{mc} = 1$, and $\beta = 0.01$. The Adam optimizer was used for training, and early stopping was applied to prevent overfitting. All experiments were conducted on an Nvidia RTX A5000 GPU with 24 GB of GPU memory. For the boundary alignment loss $\mathcal{L}_{\text{bound}}$, the segmentation boundary are 2D contours derived from segmentations SAX, 2CH, and 4CH views at all time frames. The segmentations were generated using a public available model [4] with manual quality control. Contours from different views were aligned into the same world coordinate system using header information from Nifti images.

Evaluation Metrics. The accuracy of the reconstructed 3D+t cardiac meshes is assessed using four metrics: Hausdorff distance (HD), average symmetric surface distance (ASSD), Pearson's correlation coefficient (r), and root mean squared error (RMSE). HD assesses the maximum deviation between the reconstructed and ground truth meshes, whereas ASSD indicates the average bidirectional surface distance. Pearson's r gauges temporal consistency by comparing ventricular volume curves of the reconstructed mesh against those of the ground truth, with values closer to 1 indicating better alignment. RMSE represents the volumetric differences between the reconstructed meshes and ground truth in millilitres (mL), with lower values denoting greater accuracy.

Competing Methods. `Mesh4D` was evaluated against conventional and deep learning approaches for reconstructing 3D+t cardiac meshes. The B-spline free-form deformation method (FFD) [21] performs image registration and deforms a template mesh to the image space for each subject at each time frame. 4DSegment [9] is a two-stage method combining learning-based segmentation and atlas propagation to reconstruct cardiac meshes. MeshHeart [20] performs super-resolution for the images and their segmentations and then aligns the template mesh to the high-resolution image segmentation at ED and ES. Motion tracking is performed between the images of adjacent time frames using Deepali [23] and the resulting deformation field propagates the mesh to all time frames.

http://www.ukbiobank.ac.uk/register-apply

Table 1: Comparison of cardiac mesh reconstruction methods

Methods	HD (mm) ↓	ASSD (mm) ↓	Pearson's r ↑	RMSE: LV (mL) ↓	RMSE: RV (mL) ↓	Time (s)
FFD [21]	$6.517_{\pm 1.666}$	$2.661_{\pm 0.543}$	$0.965_{\pm 0.021}$	$13.752_{\pm 5.041}$	$29.115_{\pm 7.335}$	~ 600
4DSegment [9]	$6.980_{\pm 2.854}$	$2.686_{\pm 0.597}$	$0.983_{\pm 0.013}$	$9.665_{\pm 3.422}$	$19.971_{\pm 7.253}$	~ 1200
MeshHeart [20]	$5.284_{\pm 2.714}$	$2.003_{\pm 0.597}$	$0.984_{\pm 0.012}$	$9.464_{\pm 4.265}$	$9.164_{\pm 5.206}$	445.58
Proposed	$\mathbf{4.350}_{\pm 2.13}$	$\mathbf{1.714}_{\pm 0.548}$	$\mathbf{0.986}_{\pm 0.011}$	$\mathbf{6.979}_{\pm 3.297}$	$\mathbf{8.264}_{\pm 3.981}$	2.89

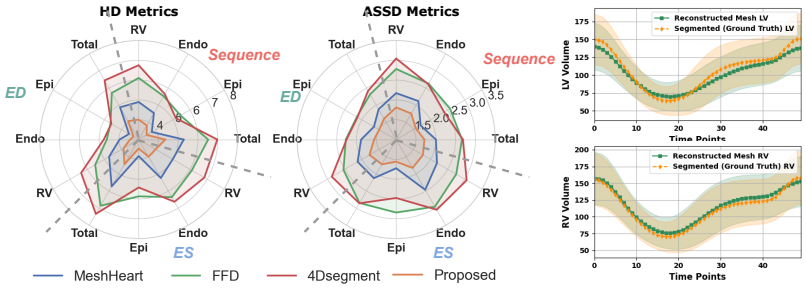

Fig. 2: **Mesh reconstruction performance on the test set.** a) Radar plot summarising HD and ASSD (unit: mm) across cardiac phases (ED, ES, whole sequence) for different structures (LV Endo, LV Epi, RV, total), compared between different methods. b) LV and RV volume curves over time from reconstructed meshes using `Mesh4D` compared to segmentation-derived volume curves. The curves shown are averaged across the test set and the shades denote confidence intervals.

3.2 Results

Comparison with Existing Methods. Table 1 shows that `Mesh4D` outperforms FFD, 4Dsegment, and MeshHeart across all evaluation metrics. The proposed method achieves the lowest HD (4.350 mm) and ASSD (1.714 mm), indicating improved reconstruction accuracy. Pearson's correlation coefficient (r) of 0.986 confirms a strong agreement between the reconstructed LV and RV volumes and segmentation-derived reference volumes. Additionally, `Mesh4D` achieves the lowest RMSE for LV (6.979 mL) and RV (8.264 mL), demonstrating superior volumetric consistency. Beyond performance improvements, `Mesh4D` also substantially reduces computational time. Unlike existing methods that sequentially reconstruct each time frame, `Mesh4D` generates the full 3D+t cardiac mesh sequence simultaneously, completing inference in just 2.89 seconds, whereas other methods require over 400 seconds. Figure 2 further illustrates the performance, where the radar plot (Figure 2a) summarises HD and ASSD across different cardiac phases, and Figure 2b compares LV and RV volume curves over time. The proposed method provides more accurate and stable volume estimations, closely matching the segmentation-derived reference volumes.

Table 2: Ablation studies of `Mesh4D`. The best scores are in bold.

Ablation	HD $(mm)\downarrow$	ASSD $(mm)\downarrow$	Pearson's r $\uparrow$	RMSE: LV $(mL)\downarrow$	RMSE: RV $(mL)\downarrow$
Single view	$5.014_{\pm 1.444}$	$2.039_{\pm 0.543}$	$0.984_{\pm 0.012}$	$9,503_{\pm 3.767}$	$21.529_{\pm 5.684}$
w/out CD	$5.515_{\pm 4.156}$	$1.982_{\pm 0.591}$	$0.984_{\pm 0.012}$	$9.600_{\pm 4.416}$	$8.879_{\pm 5.047}$
w/out $\mathcal{L}_{\text{temp}}$	$4.817_{\pm 1.460}$	$1.983_{\pm 0.537}$	$0.984_{\pm 0.012}$	$9.446_{\pm 3.921}$	$8.337_{\pm 4.047}$
w/out $\mathcal{L}_{\text{mc}}$	$5.015_{\pm 2.110}$	$2.033_{\pm 0.593}$	$0.982_{\pm 0.014}$	$8.512_{\pm 3.106}$	$13.794_{\pm 6.307}$
Proposed	$\mathbf{4.350}_{\pm 2.13}$	$\mathbf{1.714}_{\pm 0.548}$	$\mathbf{0.986}_{\pm 0.011}$	$\mathbf{6.979}_{\pm 3.297}$	$\mathbf{8.264}_{\pm 3.981}$

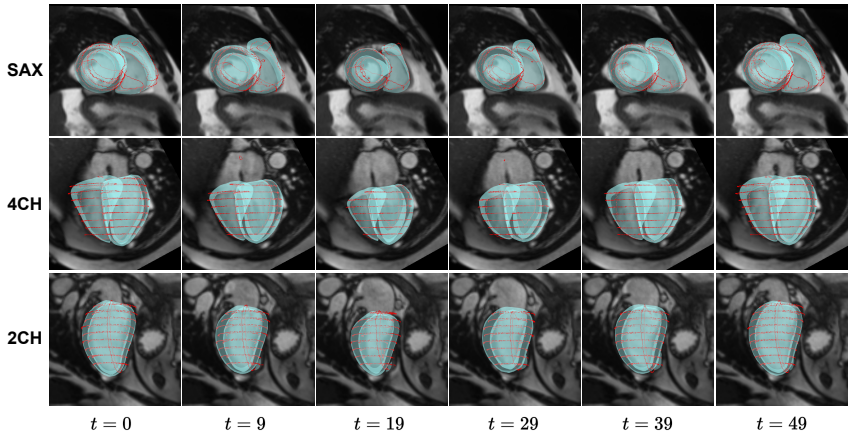

Fig. 3: Visualisation of reconstructed 3D+t meshes(blue), which are closely aligned with the segmentation contours (red) extracted from segmentation map on SAX, 4CH and 2CH view images across time. Our model captures temporal smoothness and anatomical fidelity. Image data reproduced by permission of UK Biobank.

Ablation Study To analyse the contributions of different components in `Mesh4D`, ablation experiments were conducted (Table 2). Removing the continuous deformation field (w/out CD) and replacing it with a linear-layer-based decoder that directly outputs mesh vertices results in an increase in HD (5.515 mm) and ASSD (1.982 mm), demonstrating its role in improving mesh consistency. Excluding alignment loss ($\mathcal{L}_{\text{temp}}$) results in a slight degradation in anatomical accuracy, with HD increasing to 4.817 mm and RMSE increasing for LV and RV. When motion-consistency loss ($\mathcal{L}_{\text{mc}}$) is removed, HD (5.015 mm) and ASSD (2.033 mm) deteriorate, and RMSE for RV significantly increases to 13.794 mL, indicating the necessity of enforcing temporal smoothness for accurate motion tracking. Figure 3 and the supplementary video visualises the reconstructed cardiac meshes overlaid on different views. The meshes closely align with the anatomical boundary of the two ventricles and they move smoothly across time.

4 Conclusion

We propose a novel motion-aware multi-view VAE framework for 3D+t cardiac mesh reconstruction, directly from multi-view image sequences. By integrating multi-view image encoders, Transformer-based latent modelling, and a continuous deformation decoder, `Mesh4D` achieves anatomically consistent and temporally smooth 3D+t mesh reconstruction. Our method provides a computationally efficient tool for 3D+t cardiac shape and motion analysis, and lay the foundation for future work for computational modelling of the heart.

Acknowledgments. This work was supported by the EPSRC DeepGeM Grant (EP/W01842X/1), CVD-Net Programme Grant (EP/Z531297/1) and the BHF New Horizons Grant (NH/F/23/70013). M.Q. is also supported by the Dame Julia Higgins Postdoc Collaborative Research Fund. J.Z. is supported by the Medical Research Council (MC_UP_1605/13). W.Z. is supported by the JADS programme and the UKRI Centre for Doctoral Training in AI for Healthcare (EP/S023283/1). B.K. acknowledges HPC resources provided by NHR@FAU (b143dc and b180dc). NHR@FAU is funded by the state of Bavaria and DFG - 440719683. Additional support was received by the ERC - project MIA-NORMAL 101083647, DFG 513220538, 512819079, and by the state of Bavaria (HTA). D.P.O. is supported by the Medical Research Council (MC_UP_1605/13), the NIHR Imperial College Biomedical Research Centre, and the British Heart Foundation (RG/19/6/34387, RE/24/130023, CH/P/23/80008). P.M.M. acknowledges generous personal support from the Edmond J. Safra Foundation and Lily Safra, an NIHR Senior Investigator Award, Rosalind Franklin Institute, and the UK Dementia Research Institute, which is funded predominantly by the UKRI Medical Research Council. S.N. is supported by the National Institutes of Health (R01-HL152256), the European Research Council (PREDICT-HF 864055), the British Heart Foundation (RG/20/4/34803), the EPSRC (EP/X012603/1 and EP/P01268X/1), the Technology Missions Fund under the EPSRC (EP/X03870X/1), and the Alan Turing Institute. This research was conducted using the UK Biobank Resource under Application Number 18545. Images were reproduced with kind permission of UK Biobank. We thank all UK Biobank participants and staff.

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

References

1. Almutairi, H.M., Boubertakh, R., Miquel, M.E., Petersen, S.E.: Myocardial deformation assessment using cardiovascular magnetic resonance-feature tracking technique. The British Journal of Radiology **90**(1080) (2017)
2. Avazmohammadi, R., , et al.: A contemporary look at biomechanical models of myocardium. Annual Review of Biomedical Engineering **21**(1) (2019)
3. Bai, W., Shi, W., de Marvao, A., Dawes, T.J., O'Regan, D.P., et al.: A bi-ventricular cardiac atlas built from 1000+ high resolution MR images of healthy subjects and an analysis of shape and motion. Medical Image Analysis (2015)
4. Bai, W., Sinclair, M., Tarroni, G., Oktay, O., Rajchl, M., et al.: Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. Journal of Cardiovascular Magnetic Resonance **20**(1), 65 (2018)
5. 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) (2019)
6. Chen, X., Ravikumar, N., Xia, Y., et al.: Shape registration with learned deformations for 3D shape reconstruction from sparse and incomplete point clouds. Medical Image Analysis **74** (2021)
7. Deng, Y., Xu, H., Rodrigo, S., Williams, S.E., et al.: ModusGraph: automated 3D and 4D mesh model reconstruction from cine CMR with improved accuracy and efficiency. In: MICCAI (2023)
8. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. ICLR (2021)
9. Duan, J., Bello, G., Schlemper, J., Bai, W., et al.: Automatic 3d bi-ventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach. IEEE Transactions on Medical Imaging **38** (2019)
10. He, J., Liu, W., Zhao, S., Tian, Y., Wang, S.: Algebraic sphere surface fitting for accurate and efficient mesh reconstruction from cine CMR images. In: MICCAI (2024)
11. He, X., Tan, C., Han, L., et al.: DMCVR: Morphology-guided diffusion model for 3D cardiac volume reconstruction. In: MICCAI (2023)
12. Josephson, M.E., Wellens, H.J.: Implantable defibrillators and sudden cardiac death. Circulation **109**(22) (2004)
13. Joyce, T., Buoso, S., Stoeck, C.T., Kozerke, S.: Rapid inference of personalised left-ventricular meshes by deformation-based differentiable mesh voxelization. Medical Image Analysis **79** (2022)
14. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: ICLR (2014)
15. Kong, F., Shadden, S.C.: Whole heart mesh generation for image-based computational simulations by learning free-from deformations. In: MICCAI (2021)
16. Ma, Q., Li, L., Robinson, E.C., Kainz, B., Rueckert, D.: Weakly supervised learning of cortical surface reconstruction from segmentations. In: MICCAI (2024)
17. Meng, Q., Qin, C., Bai, W., et al.: MulViMotion: Shape-aware 3D myocardial motion tracking from multi-view cardiac MRI. IEEE Transactions on Medical Imaging **41** (2022)
18. Moody, W.E., Taylor, R.J., Edwards, N.C., et al.: Comparison of magnetic resonance feature tracking for systolic and diastolic strain and strain rate calculation with spatial modulation of magnetization imaging analysis. Journal of Magnetic Resonance Imaging **41**(4) (2015)

19. Nealen, A., Igarashi, T., Sorkine, O., Alexa, M.: Laplacian mesh optimization. In: International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia. pp. 381–389 (2006)
20. Qiao, M., McGurk, K.A., Wang, S., Matthews, P.M., O'Regan, D.P., Bai, W.: A personalized time-resolved 3d mesh generative model for unveiling normal heart dynamics. Nature Machine Intelligence pp. 1–12 (2025)
21. 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) (1999)
22. Sander, J., de Vos, B.D., Bruns, S., Planken, N., et al.: Reconstruction and completion of high-resolution 3D cardiac shapes using anisotropic CMRI segmentations and continuous implicit neural representations. Computers in Biology and Medicine **164** (2023)
23. Schuh, A., Qiu, H., HeartFlow Research: deepali: Image, point set, and surface registration in PyTorch (2024)
24. St. John Sutton, M.G., Sharpe, N.: Left ventricular remodeling after myocardial infarction: Pathophysiology and therapy. Circulation **101**(25) (2000)
25. Stolt-Ansó, N., McGinnis, J., Pan, J., Hammernik, K., Rueckert, D.: Nisf: Neural implicit segmentation functions. In: MICCAI (2023)
26. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., et al.: Attention is all you need. In: NeurIPS (2017)
27. Wang, N., Zhang, Y., Li, Z., Fu, Y., Liu, W., Jiang, Y.G.: Pixel2Mesh: Generating 3D mesh models from single RGB images. In: ECCV (2018)
28. Wang, S., Qin, C., Savioli, N., Chen, C., et al.: Joint motion correction and super resolution for cardiac segmentation via latent optimisation. In: MICCAI (2021)
29. Xia, Y., Chen, X., Ravikumar, N., et al.: Automatic 3D+t four-chamber CMR quantification of the UK Biobank: integrating imaging and non-imaging data priors at scale. Medical Image Analysis **80** (2022)
30. Yuan, X., Liu, C., Wang, Y.: 4D myocardium reconstruction with decoupled motion and shape model. In: ICCV (2023)