RePCM: Region-Specific and Phenotype-Adaptive Bi-Ventricular Cardiac Motion Synthesis

Xuan Yang[1], Xiaohan Yuan[1,2], Hao Li[1], Lingyu Chen[3], Yanan Liu[1,4], and Lei Li[1]✉

[1] Department of Biomedical Engineering, National University of Singapore, Singapore
lei.li@nus.edu.sg
[2] School of Automation, Southeast University, Nanjing, China
[3] School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China
[4] School of Information Science and Engineering, Yunnan University, Kunming, China

Abstract. Cardiac motion over a cardiac cycle is crucial for quantifying regional function and is strongly affected by cardiovascular diseases. Since temporally dense cardiac mesh sequences are difficult to obtain in practice, we focus on leveraging the more accessible end-diastolic frame to infer a full-cycle sequence. Due to strong regional and disease-specific differences, traditional methods often oversmooth the data by relying on generative models that are optimized for global patterns. To address this problem, we propose Region-Aware and Phenotype-Adaptive Bi-Ventricular Cardiac Motion Synthesis (RePCM) for single frame Biventricular mesh motion completion. In Stage I, a reconstruction network learns vertex-wise motion descriptors and clustering yields a data-driven functional partition, providing an explicit motion derived region structure. In Stage II, a Region-Specific Injection Module enforces masked, synchronized region exchange within a conditional VAE, preserving localized specific dynamics and restricting cross-region mixing. A Phenotype-Adaptive Mixture-of-Experts prior conditioned on ED shape uses anatomy-guided cues to model latent motion trends and capture inter-disease variability. Experiments on three datasets covering different cardiovascular diseases show consistent gains in geometric and functional metrics and improved preservation of region specific dynamics. Our code is available at https://github.com/yangxuan-nus/RePCM.

Keywords: Cardiac Motion · Conditional Generative Model · Multi-Disease Modeling.

1 Introduction

During the cardiac cycle, the heart undergoes complex three-dimensional deformations, with different chambers and anatomical regions exhibiting distinct

contraction and relaxation patterns [6]. These spatiotemporal motion characteristics are closely associated with a wide spectrum of cardiovascular diseases [18]. Accurate modeling of such dynamics is therefore essential for regional functional assessment, understanding cardiac remodeling, and building individualized computational cardiac models. However, learning physiologically consistent cardiac dynamics remains challenging. Cardiac morphology and motion are intrinsically coupled, while different disease populations exhibit diverse motion patterns [18]. In addition, clinical image sequences often suffer from irregular temporal sampling or missing frames, making it difficult to recover a complete cardiac cycle and maintain cross-phase consistency without densely observed sequences [5].

Early cardiac generative modeling largely relied on template registration and statistical shape models (SSMs), which establish population-level comparisons through consistent parameterization [13]. Although SSMs offer good interpretability, the linear shape space and static modeling assumptions limit their ability to capture complex nonlinear deformations and temporal dynamics. More recently, implicit representations have been explored for complex anatomy [10,19]. SDF4CHD uses signed distance fields to accommodate congenital topological variations [10], and ST-NDF extends neural distance fields to spatiotemporal generation from a single frame [19]. However, implicit methods often lack stable topology and explicit point-wise correspondence, making them less suitable for fine-grained mesh-based dynamic analysis. In contrast, VAE-based frameworks generate explicit, vertex-aligned meshes, enabling reliable regional tracking and functional measurements. CHeart [17] performs conditional spatiotemporal generation using voxel labels and clinical variables, yet spatial and temporal variations are mainly captured in a shared latent space. MeshHeart [16] learns a canonical BiV 3D+t distribution for phenotypic analysis but focuses largely on healthy cohorts. CardioSynth4D [7] improves controllability via shape-motion disentanglement, but still lacks explicit mechanisms for fine-grained regional motion modeling.

To address these limitations, we propose **Re**gion-Aware and **P**henotype-Adaptive Bi-Ventricular **C**ardiac **M**otion Synthesis (**RePCM**), a motion generative framework oriented toward unified-topology bi-ventricular meshes. This method generates full-cycle sequences using a single-phase end-diastolic (ED) mesh as input, simultaneously modeling regional motion constraints and cross-disease dynamic differences during the generation process. Specifically, we incorporate region-aware motion priors to preserve localized dynamics and improve cross-phase consistency, reducing oversmoothing across functionally different regions. In addition, we employ a shape-conditioned, phenotype-adaptive mixture-of-experts (MoE) prior to capture disease-specific motion variability by leveraging the intrinsic coupling between morphology and motion.

2 Method

Our framework aims to learn multi-disease, region-aware cardiac dynamics and performs single-frame cardiac motion completion. Given an end-diastolic (ED)

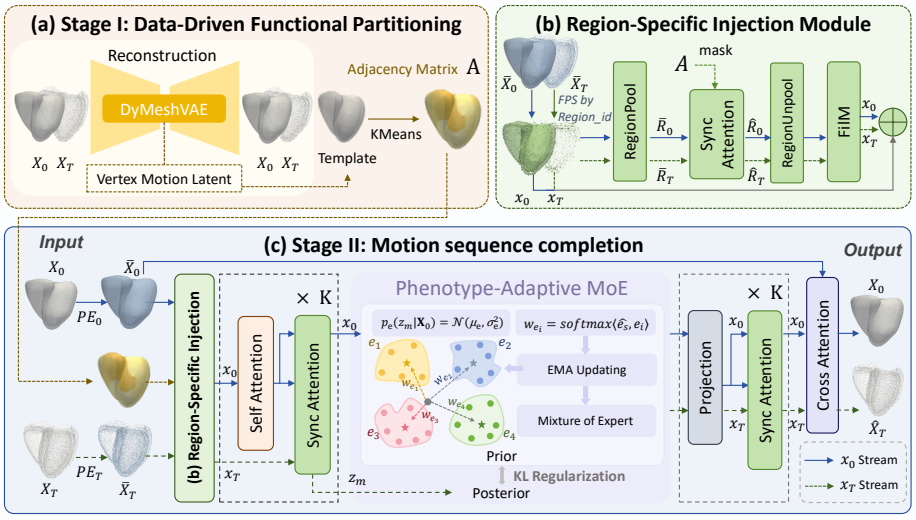

Fig. 1: **Overview of the proposed RePCM framework.** (a) Stage I: Data-Driven Functional Partitioning. (b) Region-Specific Injection Module. (c) Stage II: Cardiac Motion Sequence Completion.

mesh $\mathbf{X}_0 \in \mathbb{R}^{N\times 3}$, our goal is to synthesize a full cardiac cycle $\hat{\mathbf{X}}^{0:T-1} \in \mathbb{R}^{T\times N\times 3}$ by predicting ED-relative trajectories $\hat{\mathbf{X}}_T \in \mathbb{R}^{N\times(T\cdot 3)}$:

$$\hat{\mathbf{X}}^t = \mathbf{X}_0 + \hat{\mathbf{X}}_T^t, \qquad t = 0, \ldots, T-1. \tag{1}$$

As illustrated in Fig. 1, RePCM has two stages: (I) Data-Driven Functional Partitioning: we derive a cohort-level functional partition from reconstruction-based motion descriptors, yielding a vertex-to-region assignment and a region adjacency prior for region-specific motion modeling. (II) Cardiac Motion Sequence Completion: we inject this region prior into a conditional VAE to constrain inter-region communication, and use a Phenotype-Adaptive MoE prior conditioned on $\mathbf{X}_0$ to model phenotype-related motion variability from the ED geometry.

2.1 Stage I: Data-Driven Functional Partitioning

We adopt DyMeshVAE [20] as a motion feature extractor for functional partitioning. Through ED-relative trajectory reconstruction, it learns shape-motion disentangled representations and provides vertex-wise motion descriptors for discovering shared functional regions. The reconstructor encodes ED-relative trajectories $\mathbf{X}_T^t = \mathbf{X}^t - \mathbf{X}_0$ into a vertex-wise motion descriptor $\mathbf{z}^{(v)} \in \mathbb{R}^d$. We then cluster $\{\mathbf{z}^{(v)}\}_{v=1}^N$ using KMeans to obtain a functional partition, assigning each vertex to one of R motion regions. With consistent vertex correspondence across subjects, this vertex-wise clustering yields a shared partition for the cohort. Based on this partition, we further build a binary region adjacency prior

$\mathbf{A} \in \{0,1\}^{R\times R}$ to specify allowable inter-region communication in the completion model.

2.2 Stage II: Cardiac Motion Sequence Completion

Region-Prior Injected Anchor Encoder Given the ED mesh $\mathbf{X}_0$ and relative trajectories $\mathbf{X}_T$, we obtain vertex features by applying distinct positional encodings to their coordinates, $\bar{\mathbf{X}}_0 = PE_0(\mathbf{X}_0)$ and $\bar{\mathbf{X}}_T = PE_T(\mathbf{X}_T)$, following [20]. Farthest point sampling (FPS) then selects K anchor vertices on the input ($K = 512$), and their features are gathered as anchor tokens. Each anchor inherits the region ID of its sampled vertex.

Region-Specific Injection Module To inject region priors into the completion encoder, we introduce a region-aware exchange module. Cardiac motion is highly heterogeneous across anatomical locations [4,12], while global attention tends to over-smooth localized dynamics and mix functionally different parts. Therefore, the learned region prior is injected to explicitly constrain information exchange using the adjacency mask $\mathbf{A}$. Region tokens are obtained by pooling anchor tokens within each region, producing $\bar{R}_0$ and $\bar{R}_T$. Masked SyncAttention is applied at the region level: inter-region routing is computed from x_0, masked by $\mathbf{A}$ via Hadamard product, and reused to update both streams.

$$\begin{aligned}\hat{R}_0 &= \mathrm{Softmax}\left(\frac{\bar{R}_0\,\bar{R}_0^{\top}}{\sqrt{d_k}} \;\odot\; \mathbf{A}\right)\bar{R}_0 + \bar{R}_0,\\ \hat{R}_T &= \mathrm{Softmax}\left(\frac{\bar{R}_0\,\bar{R}_0^{\top}}{\sqrt{d_k}} \;\odot\; \mathbf{A}\right)\bar{R}_T + \bar{R}_T,\end{aligned} \tag{2}$$

where d_k denotes the channel dimension of the projected space and $\odot$ is the Hadamard product. The updated region tokens $\hat{R}_0, \hat{R}_T$ are then broadcast back to anchors according to region IDs, providing region-aware context for anchor tokens x_0 and x_T. Finally, the updated region tokens are injected into anchor features via FiLM [14], which performs region-conditioned feature-wise modulation before the residual update:

$$x_0 \leftarrow x_0 + \mathrm{FiLM}\Big(\mathrm{LN}(x_0);\ \hat{R}_0\Big), \qquad x_T \leftarrow x_T + \mathrm{FiLM}\Big(\mathrm{LN}(x_T);\ \hat{R}_T\Big). \tag{3}$$

This produces region-consistent, adjacency-constrained exchange while preserving stream-specific content, yielding region-aware anchor representations for the trunk. A stack of attention layers further refines the anchor tokens. Each layer applies self-attention on x_0 and a SyncAttention block that reuses routing from x_0 to update both x_0 and x_T, producing the final anchor representations.

Phenotype-Adaptive MoE During training, the encoder infers a motion posterior $q_\phi(\mathbf{z}_m \mid \mathbf{X}_{0:T-1})$ and regularizes it towards the shape-conditioned prior $p_\theta(\mathbf{z}_m \mid \mathbf{X}_0)$. A single disease label may not faithfully reflect motion phenotypes, which are strongly tied to cardiac shape. To capture disease-specific variability,

we parameterize this prior with a shape-conditioned, phenotype-adaptive MoE. Specifically, we compute a global shape embedding $\mathbf{e}_s$ by pooling shape tokens $\mathbf{x}_0$ over anchors. We maintain E prototypes $\{\mathbf{p}_e\}_{e=1}^{E}$ in the same embedding space and obtain soft expert weights by cosine-similarity assignment. Each expert predicts Gaussian prior parameters $(\boldsymbol{\mu}_e, \boldsymbol{\sigma}_e^2) = h_e(\mathbf{x}_0)$, and we combine experts in parameter space to form the final prior:

$$\begin{aligned} w_e &= \mathrm{Softmax}\big(\langle \hat{\mathbf{e}}_s, \hat{\mathbf{p}}_e \rangle\big), \quad e = 1, \dots, E, \\ \boldsymbol{\mu}_p &= \sum_{e=1}^{E} w_e \, \boldsymbol{\mu}_e, \qquad \boldsymbol{\sigma}_p^2 = \sum_{e=1}^{E} w_e \, \boldsymbol{\sigma}_e^2, \end{aligned} \tag{4}$$

where $\hat{\cdot}$ denotes ℓ_2-normalization. This defines $p_\theta(\mathbf{z}_m \mid \mathbf{X}_0) = \mathcal{N}(\boldsymbol{\mu}_p, \mathrm{diag}(\boldsymbol{\sigma}_p^2))$. The prototypes are updated using an exponential moving average (EMA) of assigned embeddings, providing stable adaptive phenotype clustering during training.

Anchor-Guided Trajectory Decoder The sampled latent vector is projected to anchor tokens x_0 and x_T and refined by a stack of SyncAttention layers. We recover dense trajectories by a cross-attention write-back from anchors to vertices, using x_0 as keys and x_T as values:

$$\mathbf{H} = \mathrm{Softmax}\left(\frac{\bar{\mathbf{X}}_0 \, x_0^\top}{\sqrt{d_k}} \right) x_T. \tag{5}$$

The predicted trajectories are obtained by

$$\Delta\hat{\mathbf{X}}_T = \mathrm{Reshape}(\mathrm{Linear}(\mathbf{H})) \in \mathbb{R}^{T \times N \times 3}. \tag{6}$$

2.3 Loss Function

The objective is adapted from the evidence lower bound in [9] and consists of two parts:

$$\mathcal{L} = \mathcal{L}_{\mathrm{rec}} + \beta \, \mathcal{L}_{\mathrm{KL}}. \tag{7}$$

The reconstruction loss $\mathcal{L}_{\mathrm{rec}}$ measures the discrepancy between the predicted mesh sequence and the ground-truth sequence. To encourage accurate estimation of motion patterns, we use an MSE loss over all vertices and time frames:

$$\mathcal{L}_{\mathrm{rec}} = \frac{1}{TN} \sum_{t=0}^{T-1} \sum_{v=1}^{N} \left\| \hat{\mathbf{x}}_t^{(v)} - \mathbf{x}_t^{(v)} \right\|_2^2. \tag{8}$$

The KL loss $\mathcal{L}_{\mathrm{KL}}$ measures the divergence between the approximate posterior and the prior distribution:

$$\mathcal{L}_{\mathrm{KL}} = D_{\mathrm{KL}}(q_\phi(\mathbf{z}_m \mid \mathbf{X}_{0:T-1}) \parallel p_\theta(\mathbf{z}_m \mid \mathbf{X}_0)). \tag{9}$$

3 Experiments and Results

3.1 Datasets and Preprocessing

For validation, we employed three publicly available cardiac cine MRI datasets: ACDC [2], M&Ms [3], and M&Ms-2 [11]. For each dataset, segmentation masks are converted into subject-specific bi-ventricular surface meshes over the full cardiac cycle by fitting an SSM [1] to multi-phase contours, using global alignment, non-rigid refinement, and temporal smoothing to obtain anatomically consistent and reproducible meshes. Our study considers four cardiac phenotypes: normal (NOR), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), and right-ventricular abnormality (RV), covering major variations in ventricular morphology and function with sufficient sample sizes for cross-dataset analysis. The cohort includes 666 subjects from ACDC (120), M&Ms-2 (260), and M&Ms (286), with 194 NOR, 187 DCM, 175 HCM, and 110 RV cases. All meshes are aligned to the template frame via center-of-mass matching and rigid registration applied to all frames, then normalized by a global scaling factor. Sequences are temporally resampled to 25 frames and split by patient into training/validation/test with a 7:1:2 ratio.

3.2 Implementation Details

The proposed framework is implemented in PyTorch. All experiments were conducted on a single NVIDIA RTX 4090 GPU with 24 GB memory. The encoder and decoder are each composed of 8 attention layers. During encoding, 512 anchor tokens are sampled from the input mesh using FPS. The model is optimized using the AdamW optimizer with an initial learning rate of 1×10^{-4} and a batch size of 8. The latent dimension is set to 16. Training was performed for up to 500 epochs with early stopping based on the validation performance. RePCM contains 21.34M parameters and generates a full sequence in 0.026 s per subject, with a peak inference memory of 0.61 GB. Region-aware operations are performed on compact anchor and region tokens, avoiding dense vertex-level communication.

3.3 Evaluation Metrics

We evaluate BiV sequence completion using ASSD, HD95, and vertex-wise RMSE, reported separately for LV and RV. All methods condition on the end-diastolic first-frame mesh X_0 to predict the full cardiac cycle. We compare with a conditional VAE and adapted baselines, including ACTOR [15], Action2Motion [8], CHeart [17], MeshHeart [16], and CardioSynth4D [7], all evaluated under the same unified-topology BiV sequence completion setting.

3.4 Results

Comparison Results As summarized in Table 1, our method achieves the best overall reconstruction accuracy across ASSD, HD95, and vtxRMSE. The largest

Table 1: Comparison on LV/RV surface reconstruction accuracy (mean ± std).

Methods	ASSD (mm)↓		HD95 (mm)↓		vtxRMSE (mm)↓	
	LV	RV	LV	RV	LV	RV
CVAE	2.16 ± 0.66	2.08 ± 0.63	4.34 ± 1.28	4.80 ± 1.75	3.86 ± 1.32	4.25 ± 1.52
ACTOR	2.10 ± 0.63	2.20 ± 0.79	4.28 ± 1.39	4.85 ± 2.30	3.71 ± 1.38	4.16 ± 1.81
Action2Motion	2.28 ± 0.70	2.18 ± 0.77	4.81 ± 1.63	5.07 ± 2.15	4.20 ± 1.51	4.51 ± 1.83
CHeart	2.21 ± 0.60	$\underline{2.04 \pm 0.64}$	4.48 ± 1.27	$\underline{4.73 \pm 1.70}$	3.80 ± 1.23	$\underline{4.14 \pm 1.49}$
MeshHeart	2.09 ± 0.64	2.11 ± 0.62	4.23 ± 1.28	4.78 ± 1.75	3.78 ± 1.32	4.20 ± 1.52
CardioSynth4D	$\underline{2.07 \pm 0.64}$	2.04 ± 0.71	$\underline{4.22 \pm 1.41}$	4.73 ± 1.95	$\underline{3.63 \pm 1.36}$	4.16 ± 1.68
RePCM (Ours)	$\mathbf{1.79 \pm 0.60}$	$\mathbf{2.00 \pm 0.76}$	$\mathbf{3.94 \pm 1.48}$	$\mathbf{4.54 \pm 2.04}$	$\mathbf{3.43 \pm 1.42}$	$\mathbf{4.08 \pm 1.74}$

gain is observed for the LV, where HD95 decreases from 4.22 mm to 3.94 mm compared with CardioSynth4D, suggesting that our model more effectively suppresses localized reconstruction errors and preserves fine-grained LV geometry.

Figure 2 visualizes three representative cases at two time points (t=10, t=18). Across diseases and phases, our predictions exhibit the smallest error regions and better preserve fine-grained ventricular morphology. Notably, RV remains challenging across all methods due to complex anatomy and larger inter-subject variability, but our model reduces both average and near-worst-case RV errors. Consistent with these findings, Fig. 3 (a-b) show that our method follows the reference volume trajectories more closely throughout the cardiac cycle with a narrower uncertainty band, suggesting more temporally consistent dynamics.

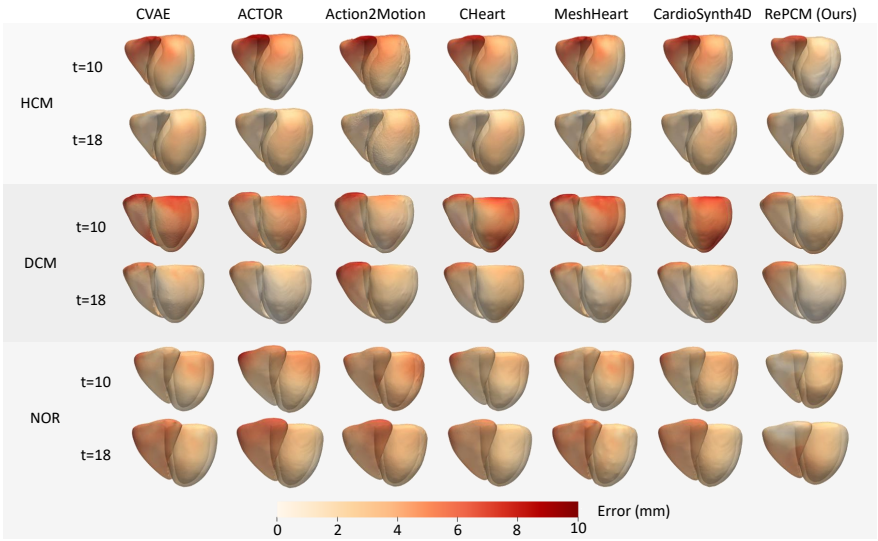

Fig. 2: **Illustration of bi-ventricular motion completion results.** HCM: hypertrophic cardiomyopathy. DCM: dilated cardiomyopathy. NOR: normal.

Table 2: Ablation on the region partition R and expert number E.

Variant	Setting	ASSD (mm) ↓		HD95 (mm) ↓		vtxRMSE (mm) ↓	
		LV	RV	LV	RV	LV	RV
w/o MoE	R=16	1.90 ± 0.69	2.02 ± 0.80	4.05 ± 1.50	4.80 ± 2.22	3.51 ± 1.46	4.17 ± 1.83
	R=24	1.98 ± 0.59	2.06 ± 0.77	4.16 ± 1.46	4.70 ± 2.09	3.56 ± 1.41	4.17 ± 1.82
	R=32	1.97 ± 0.60	2.04 ± 0.72	4.15 ± 1.45	4.91 ± 2.04	3.59 ± 1.43	4.18 ± 1.72
w/ MoE	E=4	1.79 ± 0.60	2.00 ± 0.76	3.94 ± 1.48	4.54 ± 2.04	3.43 ± 1.42	4.08 ± 1.74
	E=6	1.90 ± 0.60	2.07 ± 0.71	4.07 ± 1.41	4.74 ± 1.94	3.55 ± 1.34	4.19 ± 1.63
	E=8	1.86 ± 0.60	2.03 ± 0.74	4.12 ± 1.50	4.80 ± 2.09	3.58 ± 1.43	4.19 ± 1.78

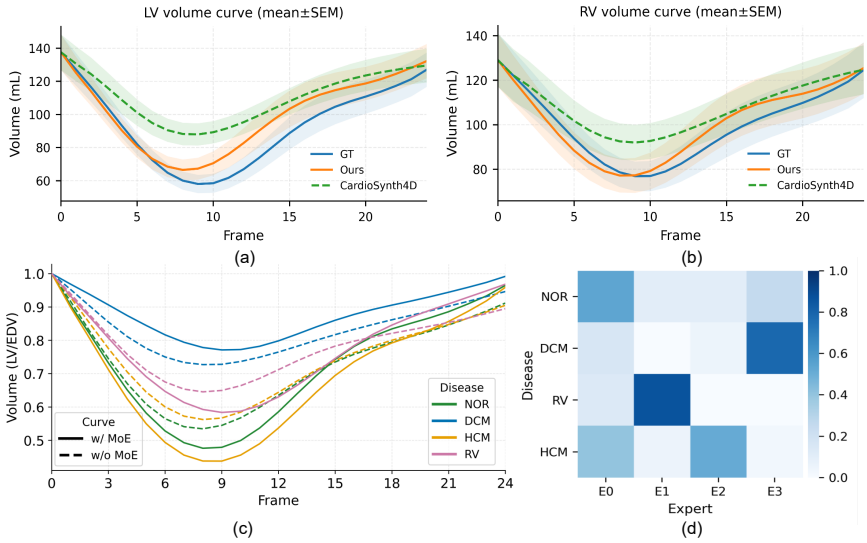

Fig. 3: **Comparison Results:** (a) LV volume curves. (b) RV volume curves. **Ablation Results:** (c) Normalized LV volume/EDV curves by disease. (d) Expert-disease usage matrix showing expert specialization.

Ablation Results Table 2 studies the number of motion regions R and experts E. The best BiV completion performance is achieved with R=16, while larger R slightly degrades performance, suggesting over-partitioning weakens region-specific motion cues and encourages more averaged global dynamics. Adding MoE improves our base framework with the best setting at E=4, whereas larger E reduces accuracy due to fewer samples per expert and less stable routing.

Figure 3 (c) shows disease-wise LV volume trajectories normalized by EDV. Compared with the non-MoE variant, the MoE-enhanced model produces more phenotype-distinct dynamics, suggesting improved capture of phenotype dependent motion patterns. Fig. 3 (d) further visualizes phenotype-aware routing by assigning each test case to the expert with the highest gating weight. The result-

ing usage matrix shows clearer expert preference for DCM and RV, while NOR and HCM exhibit more mixed routing, consistent with their partially overlapping motion characteristics.

4 Conclusion

We propose a multi-disease, region-aware conditional generative framework for single-frame bi-ventricular mesh motion completion. The framework links cardiac shape and motion across disease populations through a two-stage design. First, reconstruction motion descriptors are clustered to obtain cohort-level motion regions and a region adjacency prior. Second, the priors are injected into a conditional VAE through a Region-Specific Injection Module, promoting region-consistent information exchange. A shape-conditioned Phenotype-Adaptive MoE prior captures disease-specific variability. Experiments on three public datasets show improved motion fidelity and better preservation of region-specific dynamics over state-of-the-art methods. Future work will explore learnable region adjacency for subject-specific inter-region coupling and extend the framework to broader phenotypes and whole-heart anatomy.

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. Bai, W., Shi, W., de Marvao, A., Dawes, T.J., O'Regan, D.P., Cook, S.A., Rueckert, D.: 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 **26**(1), 133–145 (2015)
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. Campello, V.M., Gkontra, P., Izquierdo, C., Martin-Isla, C., Sojoudi, A., Full, P.M., Maier-Hein, K., Zhang, Y., He, Z., Ma, J., et al.: Multi-centre, multi-vendor and multi-disease cardiac segmentation: the m&ms challenge. IEEE Transactions on Medical Imaging **40**(12), 3543–3554 (2021)
4. Carruth, E.D., McCulloch, A.D., Omens, J.H.: Transmural gradients of myocardial structure and mechanics: implications for fiber stress and strain in pressure overload. Progress in biophysics and molecular biology **122**(3), 215–226 (2016)
5. Cui, C., Yin, G., Lu, M., Chen, X., Cheng, S., Li, L., Yan, W., Song, Y., Prasad, S., Zhang, Y., et al.: Retrospective electrocardiography-gated real-time cardiac cine mri at 3t: comparison with conventional segmented cine mri. Korean journal of radiology **20**(1), 114–125 (2019)

6. Di Folco, M., Moceri, P., Clarysse, P., Duchateau, N.: Characterizing interactions between cardiac shape and deformation by non-linear manifold learning. Medical image analysis **75**, 102278 (2022)
7. Dou, H., Huang, J., Zakeri, A., Zhou, Z., Mu, T., Duan, J., Frangi, A.F.: 4d cardiosynth: Synthesising dynamic virtual heart populations through spatiotemporal disentanglement. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 3–12. Springer (2025)
8. Guo, C., Zuo, X., Wang, S., Zou, S., Sun, Q., Deng, A., Gong, M., Cheng, L.: Action2motion: Conditioned generation of 3d human motions. In: Proceedings of the 28th ACM international conference on multimedia. pp. 2021–2029 (2020)
9. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
10. Kong, F., Stocker, S., Choi, P.S., Ma, M., Ennis, D.B., Marsden, A.L.: Sdf4chd: Generative modeling of cardiac anatomies with congenital heart defects. Medical image analysis **97**, 103293 (2024)
11. Martín-Isla, C., Campello, V.M., Izquierdo, C., Kushibar, K., Sendra-Balcells, C., Gkontra, P., Sojoudi, A., Fulton, M.J., Arega, T.W., Punithakumar, K., et al.: Deep learning segmentation of the right ventricle in cardiac mri: the m&ms challenge. IEEE Journal of Biomedical and Health Informatics **27**(7), 3302–3313 (2023)
12. Muraru, D., Onciul, S., Peluso, D., Soriani, N., Cucchini, U., Aruta, P., Romeo, G., Cavalli, G., Iliceto, S., Badano, L.P.: Sex-and method-specific reference values for right ventricular strain by 2-dimensional speckle-tracking echocardiography. Circulation: Cardiovascular Imaging **9**(2), e003866 (2016)
13. Ordas, S., Oubel, E., Leta, R., Carreras, F., Frangi, A.F.: A statistical shape model of the heart and its application to model-based segmentation. In: Medical Imaging 2007: Physiology, Function, and Structure from Medical Images. vol. 6511, pp. 490–500. SPIE (2007)
14. Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: Film: Visual reasoning with a general conditioning layer. In: Proceedings of the AAAI conference on artificial intelligence. vol. 32 (2018)
15. Petrovich, M., Black, M.J., Varol, G.: Action-conditioned 3d human motion synthesis with transformer vae. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10985–10995 (2021)
16. 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 **7**(5), 800–811 (2025)
17. Qiao, M., Wang, S., Qiu, H., De Marvao, A., O'Regan, D.P., Rueckert, D., Bai, W.: Cheart: A conditional spatio-temporal generative model for cardiac anatomy. IEEE transactions on medical imaging **43**(3), 1259–1269 (2023)
18. Smiseth, O.A., Rider, O., Cvijic, M., Valkovič, L., Remme, E.W., Voigt, J.U.: Myocardial strain imaging: theory, current practice, and the future. JACC: Cardiovascular Imaging **18**(3), 340–381 (2025)
19. Sørensen, K., Diez, P., Margeta, J., El Youssef, Y., Pham, M., Pedersen, J.J., Kühl, T., De Backer, O., Kofoed, K., Camara, O., et al.: Spatio-temporal neural distance fields for conditional generative modeling of the heart. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 422–432. Springer (2024)
20. Wu, Z., Yu, C., Wang, F., Bai, X.: Animateanymesh: A feed-forward 4d foundation model for text-driven universal mesh animation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13557–13568 (2025)