

# Learning to Beat: Phenotype-Guided Latent Flow with Regional Motion Priors for Biventricular Motion Synthesis

Xuan Yang<sup>a</sup>, Xiaohan Yuan<sup>a,b</sup>, Hao Li<sup>a</sup>, Lingyu Chen<sup>c</sup>, Yanan Liu<sup>a</sup>, Qingya Li<sup>a</sup>, Lei Li<sup>\*a</sup>

<sup>a</sup>*Department of Biomedical Engineering, National University of Singapore, Singapore*

<sup>b</sup>*School of Automation, Southeast University, Nanjing, China*

<sup>c</sup>*School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China*

## Abstract

Full-cycle biventricular geometry is essential for characterizing cardiac function. However, dense and temporally consistent 3D+t biventricular meshes are not routinely available, whereas end-diastolic (ED) anatomy can often be obtained reliably. We therefore investigate full-cycle biventricular motion synthesis from a single ED mesh. This task is challenging because cardiac deformation is spatially heterogeneous and phenotype dependent, while conventional global generative models often obscure localized motion patterns. In this study, we propose a region-specific and phenotype-adaptive framework that integrates motion-informed functional parcellation with conditional latent flow. A functional partition learned from reconstructed motion organizes the ventricular surface into regions with coherent dynamics and enables topology-aware regional feature exchange. A phenotype-conditioned rectified-flow model subsequently maps the ED anatomy to full-cycle motion latents through fine-grained conditioning and prototype-routed motion adapters. An optional control branch further incorporates available motion descriptors for controllable synthesis. Experiments on ACDC, M&Ms, and M&Ms-2 demonstrate consistent improvements in geometric accuracy and functional fidelity. Under ED-only synthesis, our method achieves biventricular ASSD, HD95, and vRMSE of  $1.49 \pm 0.34$  mm,  $3.77 \pm 1.06$  mm, and  $3.31 \pm 1.03$  mm, respectively, outperforming all competing methods. Complementary functional and robustness evaluations further demonstrate that the synthesized sequences preserve physiologically plausible ventricular dynamics and generalize across cohorts and disease phenotypes. The code will be released publicly upon acceptance of the manuscript for publication.

**Keywords:** Cardiac Motion Synthesis, Multi-Disease Modeling, Rectified Flow, Phenotype Conditioning

## 1. Introduction

Cardiovascular disease is a major cause of death worldwide (Roth et al., 2020). Cardiac abnormalities are reflected not only in changes in anatomy, but also in altered functional and motion patterns throughout the cardiac cycle (Bai et al., 2015; Puyol-Anton et al., 2017). Cardiac magnetic resonance imaging (MRI) can characterize myocardial shape and motion during contraction and relaxation with high spatial and temporal resolution (Wan et al., 2015). However, analyzing such spatiotemporal information only in the voxel or label space makes it difficult to directly represent continuous anatomical surfaces, topological structures, and geometrical changes that are aligned across time. In contrast, time-resolved 3D mesh sequences provide a more natural representation of cardiac surface geometry and temporal correspondence, and offer a useful basis for patient-specific visualization and physics-based computational modeling (Meng et al., 2023; Kong and Shadden, 2022; Marsden and Feinstein, 2015).

Despite these advantages, complete and high-quality 4D (3D+t) cardiac mesh sequences remain difficult to obtain in practice (Yuan et al., 2023; Banerjee et al., 2022), and in real clinical settings only a few key states can usually be acquired reliably. Previous studies (Liu et al., 2026; Kong et al., 2021; Luo et al., 2026; Chen et al., 2026) have shown that even under very limited observations, such as a single frame, sparse slices, or a reference phase, it is still possible to recover patient-specific cardiac dynamics. Among these states, end-diastole (ED) is commonly used as a reference because it provides a relatively stable and complete anatomical configuration and serves as an anchor for subsequent dynamic modeling (Zakeri et al., 2023; Meng et al., 2023; Upendra et al., 2021). These observations motivate using the ED anatomy as an accessible patient-specific condition for synthesizing full-cycle cardiac dynamics.

Generative modeling has become increasingly valuable in cardiac analysis, not only for completing missing dynamics, but also for expanding the available sample space, constructing virtual cohorts, and supporting in silico clinical trials (Niederer et al., 2020; Dou et al., 2023; Sørensen et al., 2024; Dou et al., 2025). Early efforts have largely

URL: lei.li@nus.edu.sg (Lei Li\*)

focused on static cardiac anatomy modeling (Kong et al., 2021; Beetz et al., 2022; Dou et al., 2023; Beetz et al., 2025), including patient-specific mesh reconstruction from cardiac images and conditional generation of anatomical shapes or virtual populations. These methods provide important tools for characterizing inter-subject morphological variability. However, static anatomy alone cannot describe functional changes over the cardiac cycle, nor can it capture the coupled evolution of shape and motion that underlies cardiac function. Recent studies have increasingly shifted toward dynamic cardiac modeling (Qiao et al., 2023; Sørensen et al., 2024; Dou et al., 2025; Qiao et al., 2025), enabling tasks such as motion sequence completion, dynamic shape generation, and joint shape-motion distribution modeling. Despite these advances, many existing dynamic generative models mainly emphasize population-level plausibility and temporal consistency, while explicit modeling of regional motion variation remains limited. This is particularly important because cardiac motion is spatially heterogeneous, with regional wall motion differences reflecting local myocardial function and pathology (Bai et al., 2015; Xue et al., 2018). Without explicitly accounting for such regional characteristics, dynamic generation models may struggle to represent localized abnormalities and fine-grained functional variation.

A second challenge is modeling phenotype-dependent motion variability. Many existing dynamic generation methods are trained primarily on healthy subjects or large normative cohorts (Qiao et al., 2025; Dou et al., 2025), which may limit their ability to represent pathological motion patterns. Although several studies have introduced conditional control for cardiac sequence generation (Qiao et al., 2023; Sørensen et al., 2024; Dou et al., 2025), the conditioning signals are often restricted to general clinical covariates or diagnostic labels, and thus may not fully capture fine-grained, patient-specific functional differences. Clinically meaningful cardiac motion generation therefore requires more informative conditioning mechanisms, particularly phenotype-aware representations that can encode individual disease characteristics and guide the generation of region-specific dynamic patterns.

To address the above challenges, we propose a region-specific and phenotype-adaptive latent generative framework for full-cycle biventricular motion synthesis on unified-topology meshes. The framework uses a patient-specific ED mesh as the anatomical condition and synthesizes full-cycle cardiac dynamics under phenotype guidance, aiming to capture anatomy-dependent and disease-associated motion variation. Specifically, our method first learns a motion-informed functional partition from real cardiac sequences, providing explicit regional priors for modeling local dynamics. The priors are then used to organize a motion variational autoencoder (VAE) for reconstructing full-cycle cardiac sequences on the learned regional structure. Finally, a rectified-flow generator is trained in the learned VAE latent space to enhance phenotype-conditioned mo-

tion synthesis from ED anatomy, with optional motion descriptors used when available for controllable generation. The main contributions of this work are as follows:

- i. We establish a phenotype-adaptive latent motion generation framework for synthesizing full-cycle biventricular motion from a single ED mesh.
- ii. We introduce a motion-informed regional prior and integrate it into a region-structured motion VAE, enabling compact representation of full-cycle cardiac dynamics while preserving regional motion heterogeneity.
- iii. We develop a rectified-flow latent generator with prototype-routed motion adapters, improving phenotype-conditioned motion generation while supporting optional functional control.
- iv. Extensive experiments on multiple public cardiac MRI datasets demonstrate the effectiveness of the proposed method in full-cycle motion synthesis, functional consistency, and cross-disease generalization.

This paper significantly extends a preliminary conference version of the work presented in Yang et al. (2026). Compared with the previous framework, this study redesigns the overall pipeline into a region-specific latent generative framework for full-cycle motion synthesis, strengthening both the structural prior and the latent generation mechanism. Specifically, we replace the fixed binary regional adjacency with a hop-based region-topology prior, improving the modeling of gradual motion propagation across neighboring regions. We also replace the previous VAE-prior-based motion generation with a rectified-flow latent generator in the learned latent space, where fine-grained ED anatomical and phenotype conditioning support more detailed morphology-aware motion synthesis. In addition, an optional functional-control branch enables controllable modulation of the generated dynamics when motion descriptors are available. We further conduct more systematic experiments, including comparisons with competing methods and ablation studies, to evaluate the contribution of these modeling components.

## 2. Related Work

### 2.1. Cardiac Shape and Motion Generative Modeling

Cardiac statistical shape analysis has been used to characterize anatomical variation and its relationship to disease, supporting diagnosis, risk assessment, and treatment planning (Bai et al., 2015; Biglino et al., 2017; Bai et al., 2020). With the development of deep learning, research has gradually shifted toward more expressive generative models. For instance, Dou et al. (2023) proposed a conditional VAE to synthesize virtual anatomical populations of the left ventricle under covariate constraints. Beetz et al. (2021, 2025) used point clouds and VAE to model

three-dimensional cardiac shape variation, and applied it to myocardial infarction-related shape analysis and virtual heart synthesis. Kong et al. (2024) further extended conditional anatomical generation to more complex topologies, enabling controllable synthesis of congenital heart disease anatomy through type-shape disentanglement. Overall, static cardiac mesh generation has evolved from patient-specific anatomical reconstruction to controllable anatomical distribution learning. However, these methods remain limited to static anatomy and cannot directly describe functional changes throughout the cardiac cycle. To solve this, Qiao et al. (2023) proposed a conditional spatiotemporal generative model to jointly model 4D cardiac anatomy and its relationship with non-imaging clinical factors. Qiao et al. (2025) directly learned the shape-motion distribution of 3D+t biventricular meshes and explored cardiac morphology and motion patterns in large-scale population data. Dou et al. (2025) further generated dynamic virtual heart populations through spatiotemporal disentanglement. Nevertheless, existing dynamic models mainly focus on globally coherent motion patterns and are often developed on healthy populations or under global condition signals, leaving disease-related dynamic variation largely underexplored.

## 2.2. Patient-Specific Dynamic Modeling from Sparse Observations

Patient-specific cardiac dynamic modeling has mainly relied on registration, motion tracking, and spatiotemporal deformation estimation from cine imaging (Qiao et al., 2020; Ye et al., 2021; Lu et al., 2023; Qin et al., 2023; López et al., 2023; Yang et al., 2024). With the development of deep learning, these approaches have gradually shifted from traditional optimization-based registration to data-driven dynamic recovery. One group of methods (Qin et al., 2023; Zakeri et al., 2023; Lu et al., 2023), learns temporal deformation representations in the image domain. For example, Qin et al. (2023) achieved left ventricular myocardial motion tracking through a biomechanically constrained latent deformation manifold. DragNet (Zakeri et al., 2023) further recovered temporally consistent full cardiac dynamics from a single-frame cine MRI. However, the outputs of these methods are still essentially image-domain spatiotemporal deformations or generated image sequences, rather than mesh sequences with unified topological constraints.

To obtain more explicit geometric representations, recent studies moved from image-domain displacement fields to mesh-based modeling. DeepMesh (Meng et al., 2023) represented patient-specific cardiac anatomy by deforming a template mesh to the ED geometry and subsequently estimated three-dimensional mesh motion. Laumer et al. (2023) further recovered patient-specific 4D cardiac meshes from 2D echocardiographic videos under a weakly supervised setting. Compared with general image-domain motion fields, mesh sequences can more naturally preserve

surface geometry, fixed topology, and vertex correspondence across time, making them more suitable for subsequent regional prior modeling and local dynamic analysis. Overall, existing patient-specific dynamic modeling studies have shown that recovering full cardiac dynamics from key states, single-frame images, or limited observations is a reasonable and important problem setting. However, their main goal is still tracking, reconstruction, or globally plausible motion recovery, rather than unified generative modeling of disease-related dynamic processes.

## 2.3. Latent and Flow-Based Motion Generation

In complex temporal motion modeling, latent-variable generative models are among the most representative technical routes (Kingma and Welling, 2013; Chung et al., 2015; Sohn et al., 2015). Their core idea is to compress high-dimensional temporal motion into a low-dimensional latent space and then model complex motion distributions within that space. Action2Motion (Guo et al., 2020) used a temporal VAE to model human motion under action conditions. ACTOR (Petrovich et al., 2021) further employed a Transformer to learn action-aware latent motion representations. Furthermore, AnimateAnyMesh (Wu et al., 2025) extended conditional generation in compressed latent spaces to three-dimensional mesh objects with arbitrary topology, enabling unified animation generation for multiple types of mesh objects. This suggests that latent-space modeling is not only suitable for general temporal signals, but can also support motion generation for geometrically more complex and topologically more flexible 3D objects.

Diffusion models (Ho et al., 2020; Dhariwal and Nichol, 2021; Song et al., 2022) achieve high-fidelity generation through iterative denoising and have shown strong performance in modeling complex distributions, but they usually rely on multi-step sampling and large training scales. To improve training and sampling efficiency, recent studies have turned to continuous flow-based modeling. Flow Matching (Lipman et al., 2023) characterizes transport between distributions by learning continuous-time velocity fields along fixed probability paths. Rectified Flow (Liu et al., 2022) further favors learning more direct transport trajectories. Motion Flow Matching (Hu et al., 2023) has applied this idea to human motion generation, completion, and interpolation. However, these methods have mostly been developed for general human motion or generic 3D geometric object scenarios, and they still lack specialized designs for jointly encoding anatomical states, disease-related conditions, and regional priors. Therefore, combining latent-space modeling with flow-based generation, and adapting it to disease-related and regionally heterogeneous cardiac dynamics, remains an important direction for further study.

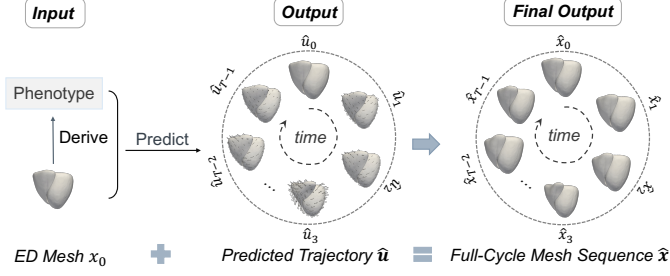


Figure 1: Illustration of the task and its formulation. ED: end-diastolic.

### 3. Methods

#### 3.1. Problem Formulation

Given only the ED biventricular mesh of a subject, our goal is to recover the full dynamic biventricular mesh sequence. Let  $x_t \in \mathbb{R}^{N \times 3}$  denote the biventricular mesh at cardiac frame  $t$ , where  $N$  is the number of vertices and  $t = 0, \dots, T-1$ . The full cardiac mesh sequence is defined as  $\mathbf{x} = \{x_t\}_{t=0}^{T-1} \in \mathbb{R}^{T \times N \times 3}$ , where  $x_0$  corresponds to the ED mesh. All biventricular meshes are represented using a unified template with shared topology and vertex-wise correspondence across subjects and cardiac frames. This allows cardiac motion to be formulated as an ED-referenced displacement trajectory, reducing subject-specific spatial variation and enabling the model to focus on learning cardiac deformation rather than absolute vertex coordinates. Specifically, the displacement at frame  $t$  is defined as  $u_t = x_t - x_0$ . The complete displacement trajectory is therefore given by  $\mathbf{u} = \{u_t\}_{t=0}^{T-1} \in \mathbb{R}^{T \times N \times 3}$ . Under this formulation, cardiac sequence recovery is defined as learning a mapping from the patient-specific ED mesh  $x_0$  to the full displacement trajectory  $\hat{\mathbf{u}}$ , as shown in Fig. 1. The dynamic mesh sequence is then recovered by adding the generated displacement field to the ED mesh:

$$\hat{x}_t = x_0 + \hat{u}_t, \quad t = 0, \dots, T-1. \quad (1)$$

To predict the displacement trajectory  $\hat{\mathbf{u}}$ , we learn a motion generator conditioned on the patient-specific ED mesh  $x_0$  and its ED-derived phenotype descriptors  $c_{ED}^{global}$  and  $c_{ED}^{fine}$ . Following the cardiac MRI-derived phenotype definitions in Bai et al. (2020),  $c_{ED}^{global}$  summarizes global chamber and myocardial morphology, whereas  $c_{ED}^{fine}$  further captures regional ED myocardial wall-thickness information (see Sec. 4.1 for details). When ventricular functional measurements are available, the motion descriptor  $c^{motion}$  can be additionally incorporated as optional inputs.

#### 3.2. Motion-Informed Cardiac Functional Parcellation

To model spatially heterogeneous cardiac deformation, we learn a data-driven functional parcellation directly from training motion sequences, avoiding predefined anatomical divisions that may not reflect regional motion similarity. Defined on the unified biventricular template, the re-

sulting partition provides a fixed, subject-independent regional prior. Specifically, given full cardiac mesh sequences  $\mathbf{x}$ , we first train a dynamic mesh reconstruction network, DyMeshVAE (Wu et al., 2025), to extract vertex-wise motion features  $f(i)$ , as shown in Fig. 2 (a). To obtain a subject-independent descriptor for each template vertex, we average the vertex features across the training set and thus obtain  $\bar{f}(i)$ . We then apply K-means to the template-level descriptors and partition the template vertices into  $N_R$  motion regions:

$$r(i) = \arg \min_{\rho \in \{1, \dots, N_R\}} \|\bar{f}(i) - \mu_\rho\|_2^2, \quad r(i) \in \{1, \dots, N_R\}, \quad (2)$$

where  $i$  is the index of mesh vertex and  $\mu_\rho$  is the centroid of the  $\rho$ -th motion region. The resulting regional prior comprises the fixed vertex-to-region assignments  $\{r(i)\}_{i=1}^N$  and the region adjacency matrix  $Adj$ . Two regions are considered adjacent if at least one mesh edge connects vertices assigned to the two regions.

Cardiac motion involves coordinated deformation across multiple interconnected regions, which may not be fully captured by one-hop connectivity alone. We therefore introduce multi-hop regional attention bias (MRAB) to extend the binary adjacency mask used in Yang et al. (2026) by encoding the multi-hop structure of the learned region graph in the attention logits. Let  $D_{reg}(\rho, \rho')$  denote the shortest-path distance between regions  $\rho$  and  $\rho'$ . Rather than treating all admissible region pairs equally, MRAB converts this distance into an additive attention bias:

$$B_{reg}(\rho, \rho') = \begin{cases} 0, & D_{reg}(\rho, \rho') = 0, \\ (D_{reg}(\rho, \rho') - 1) \log \eta, & 1 \leq D_{reg}(\rho, \rho') \leq L_{hop}, \\ -\infty, & D_{reg}(\rho, \rho') > L_{hop}, \end{cases} \quad (3)$$

where  $0 < \eta < 1$  controls the attenuation across successive hops and  $L_{hop}$  specifies the maximum interaction range. Accordingly, self- and one-hop interactions remain unbiased, whereas every additional hop introduces a progressively stronger negative bias. Region pairs outside the prescribed neighborhood are excluded. MRAB thus provides a distance-aware topological prior for controlled information exchange beyond immediately adjacent regions. The vertex-to-region assignments and  $B_{reg}$  are subsequently used for regional feature aggregation and topology-aware attention, respectively, in the motion VAE described in Sec. 3.3.

#### 3.3. Region-Structured Motion VAE

To learn a compact latent representation for full-cycle cardiac motion, we design a region-structured motion VAE, as illustrated in Fig. 2 (b). The VAE adopts a two-stream design in which the ED stream encodes the static anatomical state  $x_0$ , while the motion stream encodes the ED-relative trajectory  $\mathbf{u}$ . This separation allows the motion stream to model temporal deformation relative to a stable anatomical reference, reducing interference between static



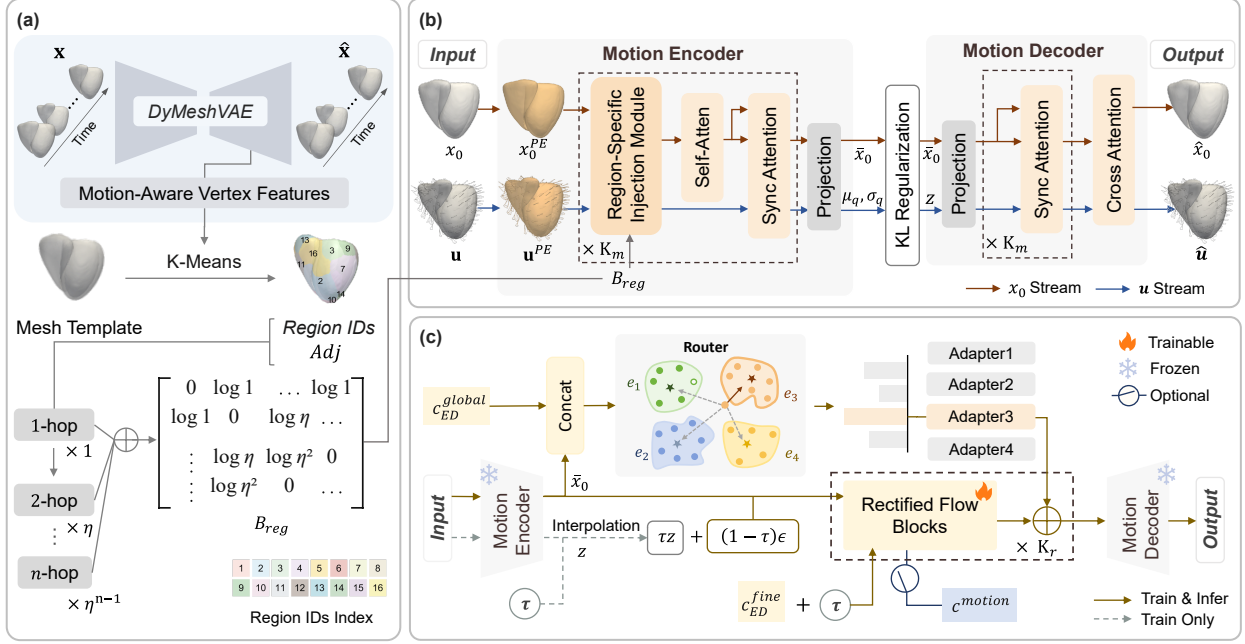


Figure 2: Illustration of the proposed regional motion-informed and phenotype-guided biventricular motion synthesis framework. (a) Motion-driven functional regionalization. (b) Region-topology aware motion variational autoencoder (VAE). (c) Phenotype-guided latent flow for anatomy-to-motion generation. DyMeshVAE: dynamic mesh VAE; PE: positional encoding.

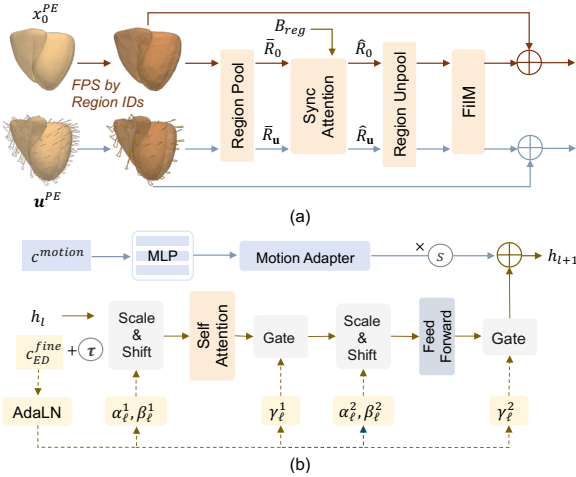


Figure 3: (a) Region-specific injection module. (b) Phenotype-conditioned rectified-flow block with optional functional motion control. FPS: farthest point sampling; MLP: multilayer perceptron.

anatomical coordinates and temporal displacement patterns. Inspired by DyMeshVAE (Wu et al., 2025), we apply distinct positional encoding (PE) to  $\mathbf{x}_0$  and  $\mathbf{u}$ , yielding  $\mathbf{x}_0^{PE}$  and  $\mathbf{u}^{PE}$ , respectively. Region-balanced farthest point sampling (FPS) then selects  $N_a$  anchor vertices, and the same sampling indices are used to gather features from both streams, forming paired ED and motion anchor tokens  $\mathbf{a}_0, \mathbf{a}_{\mathbf{u}} \in \mathbb{R}^{N_a \times d}$ , where  $d$  denotes the token feature dimension. Each token pair inherits the region label of its sampled vertex from the motion-informed partition defined in Sec. 3.2.

To preserve the regional motion heterogeneity captured

by this partition, each encoder block incorporates a region-specific injection module, as shown in Fig. 3 (a). The module mean-pools the anchor tokens within each region to obtain ED and motion region tokens  $\bar{\mathbf{R}}_0, \bar{\mathbf{R}}_{\mathbf{u}} \in \mathbb{R}^{N_R \times d}$ . At the regional level, the two streams retain distinct roles. The ED region tokens determine how information is exchanged between regions, whereas the motion region tokens provide the dynamic information propagated through these interactions. Accordingly,  $B_{reg}$  is added to the attention logits computed from  $\bar{\mathbf{R}}_0$ , and the resulting attention map is applied to both streams:

$$\begin{aligned} \hat{\mathbf{R}}_0 &= \text{Softmax} \left( \frac{\bar{\mathbf{R}}_0 \bar{\mathbf{R}}_0^\top}{\sqrt{d}} + B_{reg} \right) \bar{\mathbf{R}}_0 + \bar{\mathbf{R}}_0, \\ \hat{\mathbf{R}}_{\mathbf{u}} &= \text{Softmax} \left( \frac{\bar{\mathbf{R}}_0 \bar{\mathbf{R}}_0^\top}{\sqrt{d}} + B_{reg} \right) \bar{\mathbf{R}}_{\mathbf{u}} + \bar{\mathbf{R}}_{\mathbf{u}}. \end{aligned} \quad (4)$$

The updated region tokens are then mapped back to their corresponding anchors and injected into both streams through FiLM-based residual modulation (Perez et al., 2017). Stacking  $K_m$  such encoder blocks progressively incorporates topology-constrained regional context into the anchor features while preserving fine-grained spatial representations.

After the final encoder block, the ED anchor tokens are projected into the deterministic anatomical latent  $\bar{\mathbf{x}}_0 \in \mathbb{R}^{N_a \times d_z}$ , while the motion anchor tokens are projected into  $\mu_q$  and  $\sigma_q$ , which parameterize the motion posterior.

$$q_\phi(z | \mathbf{x}_0, \mathbf{u}) = \mathcal{N}(\mu_q, \text{diag}(\sigma_q^2)), \quad (5)$$

where  $z \in \mathbb{R}^{N_a \times d_z}$  denotes a motion latent sampled from the posterior and  $d_z$  is the per-anchor latent feature dimension. The decoder takes the deterministic anatomical

latent  $\bar{x}_0$  and motion latent  $z$  as inputs, projects them back to the ED and motion anchor representations, and refines both streams through SyncAttention. The dense ED features  $x_0^{PE}$  then query the decoded ED anchor representations through cross-attention and aggregate the corresponding motion anchor representations to reconstruct  $\hat{\mathbf{u}} \in \mathbb{R}^{T \times N \times 3}$ . We define the reconstruction term as the mean squared error between the reconstructed and reference dynamic mesh sequences:

$$\mathcal{L}_{rec} = \frac{1}{T} \frac{1}{N} \sum_{t=0}^{T-1} \sum_{i=1}^N \|\hat{x}_t(i) - x_t(i)\|_2^2. \quad (6)$$

The motion VAE is optimized with dynamic mesh reconstruction and KL regularization:

$$\mathcal{L}_{VAE} = \mathcal{L}_{rec} + \lambda_{KL} \mathcal{D}_{KL}[q_\phi(z | x_0, \mathbf{u}) \| \mathcal{N}(0, I)], \quad (7)$$

where  $\lambda_{KL}$  is a balancing parameter. After training, the motion VAE is frozen, providing the latent space and decoder used by the phenotype-guided motion generator described in Sec. 3.4.

### 3.4. Phenotype-Guided Latent Motion Generation

*Latent Rectified-Flow Training.* As described in Sec. 3.3, the frozen motion VAE represents each training sequence using a deterministic anatomical latent  $\bar{x}_0$  and a posterior motion latent  $z$ , from which its decoder reconstructs the displacement trajectory  $\mathbf{u}$ . Its Gaussian prior, however, is independent of subject-specific ED anatomy and phenotype. We therefore train a conditional rectified-flow (RF) model whose velocity field is conditioned on  $\bar{x}_0$ ,  $c_{ED}^{global}$ , and  $c_{ED}^{fine}$ , thereby transporting Gaussian noise into subject-specific motion latents, as illustrated in Fig. 2 (c). Following the standard RF formulation (Liu et al., 2022; Lipman et al., 2023), for each motion latent  $z$  sampled from the frozen VAE posterior, we draw  $\epsilon \sim \mathcal{N}(0, I)$  and  $\tau \sim \mathcal{U}(0, 1)$ , and construct  $z_\tau = (1 - \tau)\epsilon + \tau z$  with target velocity  $z - \epsilon$ . The velocity network comprises  $K_r$  RF blocks, as shown in Fig. 3 (b). The representations of  $z_\tau$  and  $\bar{x}_0$  are summed to initialize the hidden representation. Within each block,  $c_{ED}^{fine}$  and  $\tau$  are incorporated through AdaLN (Peebles and Xie, 2023; Yang et al., 2025), whereas  $\bar{x}_0$  and  $c_{ED}^{global}$  jointly determine the prototype-specific adapter route described below. The RF model is optimized by

$$\mathcal{L}_{RF} = \mathbb{E}_{z, \epsilon, \tau} \left[ \left\| v_\theta \left( z_\tau, \bar{x}_0, c_{ED}^{global}, c_{ED}^{fine}, \tau \right) - (z - \epsilon) \right\|_2^2 \right]. \quad (8)$$

*Prototype-Routed Motion Adapters.* A shared RF backbone models motion patterns common across subjects. To adapt this shared velocity field to heterogeneous ED anatomy and phenotypes without training separate generators, we introduce  $N_e$  lightweight motion adapters, each associated with one routing prototype. For each subject, we form the

Table 1: Composition of the study cohort across datasets and diagnostic groups. NOR: normal; DCM: dilated cardiomyopathy; HCM: hypertrophic cardiomyopathy; RVA: right-ventricular abnormality.

| Dataset | NOR | DCM | HCM | RVA | Total |
|---------|-----|-----|-----|-----|-------|
| ACDC    | 30  | 30  | 30  | 30  | 120   |
| M&Ms    | 89  | 97  | 85  | 15  | 286   |
| M&Ms-2  | 75  | 60  | 60  | 65  | 260   |
| Total   | 194 | 187 | 175 | 110 | 666   |

routing feature  $g_{route}$  by concatenating the mean-pooled anatomical latent  $\text{MeanPool}(\bar{x}_0)$  with the global phenotype descriptor  $c_{ED}^{global}$ . Applying K-means to the training routing features yields  $N_e$  routing prototypes  $\{e_j\}_{j=1}^{N_e}$ . Each subject is assigned to its nearest routing prototype:

$$j^* = \arg \min_{j \in \{1, \dots, N_e\}} \|g_{route} - e_j\|_2^2, \quad (9)$$

where  $e_j$  denotes the  $j$ -th routing prototype. The selected index  $j^*$  activates the corresponding adapter in each RF block, allowing the shared velocity field to specialize according to subject-specific ED anatomy and global phenotype.

*Optional Motion Control.* ED anatomy does not completely determine functional properties such as contraction strength and motion amplitude. When the motion descriptor  $c^{motion}$  is available, it is encoded by a multilayer perceptron (MLP) and injected into each RF block through an additional Motion Adapter. As shown in Fig. 3 (b), the resulting residual is scaled by the control scale  $s$  and added to the hidden representation:

$$h_{\ell+1} \leftarrow h_{\ell+1} + s \mathcal{M}_\ell(\text{MLP}(c^{motion})), \quad (10)$$

where  $h_{\ell+1}$  denotes the output hidden representation of the  $\ell$ -th RF block and  $\mathcal{M}_\ell$  denotes its Motion Adapter. The coefficient  $s$  controls the strength of functional modulation.

*Inference.* Given an ED mesh  $x_0$ , the frozen ED encoder produces the anatomical latent  $\bar{x}_0$ . The routing feature  $g_{route}$  is then constructed from  $\bar{x}_0$  and  $c_{ED}^{global}$ , and Eq. (9) selects the adapter route  $j^*$ , which remains fixed throughout RF integration. Starting from  $z_{\tau=0} = \epsilon$ , where  $\epsilon \sim \mathcal{N}(0, I)$ , the terminal motion latent is obtained by

$$\hat{z} = z_{\tau=1} = \epsilon + \int_0^1 v_\theta^{(j^*)} \left( z_\tau, \bar{x}_0, c_{ED}^{fine}, \tau \right) d\tau, \quad (11)$$

where  $v_\theta^{(j^*)}$  denotes the RF velocity field with the  $j^*$ -th adapter route activated. When  $c^{motion}$  is available, the motion adapter additionally modulates the velocity field throughout integration; otherwise, this branch is disabled. Finally, the terminal latent  $\hat{z} = z_{\tau=1}$ , together with  $\bar{x}_0$ , is passed to the frozen motion decoder to reconstruct the full-cycle ED-relative displacement trajectory  $\hat{\mathbf{u}}$ .

Table 2: Quantitative comparison under the ED-conditioned full-cycle biventricular motion synthesis. ASSD: average symmetric surface distance; HD95: 95th percentile Hausdorff distance; vRMSE: vertex-wise root mean square error; BiV: biventricular; LV: left ventricular; RV: right ventricular. Values are reported as the mean  $\pm$  standard deviation across test subjects. For each subject, each metric was first averaged across 20 independently generated motion sequences.

| Method                           | ASSD (mm) $\downarrow$            |                                   |                                   | HD95 (mm) $\downarrow$            |                                   |                                   | vRMSE (mm) $\downarrow$           |                                   |                                   |
|----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
|                                  | LV                                | RV                                | BiV                               | LV                                | RV                                | BiV                               | LV                                | RV                                | BiV                               |
| CVAE (Sohn et al., 2015)         | 2.16 $\pm$ 0.66                   | 2.08 $\pm$ 0.63                   | 2.15 $\pm$ 0.54                   | 4.34 $\pm$ 1.30                   | 4.78 $\pm$ 1.77                   | 4.71 $\pm$ 1.54                   | 3.86 $\pm$ 1.34                   | 4.25 $\pm$ 1.54                   | 4.11 $\pm$ 1.25                   |
| ACTOR (Petrovich et al., 2021)   | 2.10 $\pm$ 0.64                   | 2.00 $\pm$ 0.80                   | 2.11 $\pm$ 0.60                   | 4.28 $\pm$ 1.41                   | 4.85 $\pm$ 2.33                   | 4.93 $\pm$ 1.98                   | 3.71 $\pm$ 1.39                   | 4.14 $\pm$ 1.83                   | 4.01 $\pm$ 1.52                   |
| Action2Motion (Guo et al., 2020) | 2.28 $\pm$ 0.71                   | 2.18 $\pm$ 0.78                   | 2.22 $\pm$ 0.64                   | 4.81 $\pm$ 1.65                   | 5.07 $\pm$ 2.18                   | 5.16 $\pm$ 1.98                   | 4.20 $\pm$ 1.53                   | 4.51 $\pm$ 1.85                   | 4.21 $\pm$ 1.52                   |
| CHeart (Qiao et al., 2023)       | 2.21 $\pm$ 0.65                   | 2.04 $\pm$ 0.72                   | 2.09 $\pm$ 0.53                   | 4.48 $\pm$ 1.43                   | 4.73 $\pm$ 1.98                   | 4.97 $\pm$ 1.52                   | 3.80 $\pm$ 1.37                   | 4.10 $\pm$ 1.70                   | 3.98 $\pm$ 1.22                   |
| MeshHeart (Qiao et al., 2025)    | 2.09 $\pm$ 0.65                   | 2.01 $\pm$ 0.63                   | 2.08 $\pm$ 0.53                   | 4.23 $\pm$ 1.30                   | 4.64 $\pm$ 1.77                   | 4.68 $\pm$ 1.54                   | 3.78 $\pm$ 1.34                   | 4.16 $\pm$ 1.54                   | 4.13 $\pm$ 1.25                   |
| 4DCardioSynth (Dou et al., 2025) | 2.07 $\pm$ 0.61                   | 2.04 $\pm$ 0.65                   | 2.12 $\pm$ 0.55                   | 4.22 $\pm$ 1.29                   | 4.73 $\pm$ 1.72                   | 4.76 $\pm$ 1.70                   | 3.63 $\pm$ 1.25                   | 4.11 $\pm$ 1.50                   | 3.94 $\pm$ 1.33                   |
| RePCM (Yang et al., 2026)        | 1.79 $\pm$ 0.60                   | 2.00 $\pm$ 0.76                   | 1.69 $\pm$ 0.36                   | 3.94 $\pm$ 1.48                   | 4.54 $\pm$ 2.04                   | 4.27 $\pm$ 1.83                   | 3.43 $\pm$ 1.42                   | 4.08 $\pm$ 1.74                   | 3.61 $\pm$ 1.45                   |
| <b>Ours</b>                      | <b>1.63 <math>\pm</math> 0.48</b> | <b>1.85 <math>\pm</math> 0.50</b> | <b>1.49 <math>\pm</math> 0.34</b> | <b>3.56 <math>\pm</math> 1.09</b> | <b>4.24 <math>\pm</math> 1.23</b> | <b>3.77 <math>\pm</math> 1.06</b> | <b>3.13 <math>\pm</math> 1.09</b> | <b>3.74 <math>\pm</math> 1.12</b> | <b>3.31 <math>\pm</math> 1.03</b> |

## 4. Experiments and Results

### 4.1. Dataset and Pre-processing

We used three public cine cardiac MRI datasets: ACDC (Bernard et al., 2018), M&Ms (Campello et al., 2021), and M&Ms-2 (Martín-Isla et al., 2023). These datasets cover different acquisition protocols and cardiac conditions. Diagnostic labels were harmonized as normal (NOR), dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), and right ventricular abnormality (RVA). The DCM group combined ACDC/ M&Ms DCM with M&Ms-2 dilated-left-ventricle cases. The RVA group combined ACDC/ M&Ms right-ventricular-abnormality cases with M&Ms-2 arrhythmogenic-cardiomyopathy and dilated-right-ventricle cases. These groupings reflect shared LV- or RV-dominant phenotypes and were used only for stratification and subgroup evaluation. Five additional diagnoses were reserved exclusively for diagnostic out-of-distribution (OOD) evaluation. M&Ms hypertensive heart disease (HHD) and ischemic heart disease (IHD) were withheld because of limited sample sizes. M&Ms-2 tetralogy of Fallot (FALL), interatrial communication (CIA), and tricuspid regurgitation (TRI) were not merged into RVA because their defining abnormalities involve anatomical structures or flow phenomena not represented by the modeled biventricular surfaces.

For each subject, the segmentation masks across the cardiac cycle were converted into a unified-topology biventricular surface-mesh sequence. We adopted an SSM-based fitting procedure based on the biventricular atlas of Bai et al. (2015). The template was fitted to the multi-phase contours and refined through global alignment, non-rigid deformation, and temporal smoothing. This process produced anatomically consistent meshes with fixed topology and vertex-wise correspondence across subjects and cardiac frames. All sequences were temporally resampled to  $T = 25$  frames. For model training, all meshes were aligned to the template coordinate system through center-of-mass matching and rigid registration, followed by normalization using a global scaling factor. The same spatial transformation was applied to all frames of each subject to preserve the relative motion trajectory. During evaluation,

the generated meshes were transformed back to their original physical scale before geometric and functional measurements were computed. Following the MRI-derived phenotype definitions in Bai et al. (2020), we extracted ED-derived phenotype descriptors from the unified biventricular meshes. The global descriptor  $c_{ED}^{global} \in \mathbb{R}^6$  comprised the logarithmically transformed LV ED volume, RV ED volume, and LV mass; global myocardial wall thickness; the logarithmic LV-to-RV ED volume ratio; and the logarithmic LV-mass-to-LV-volume ratio. The fine-grained descriptor  $c_{ED}^{fine} \in \mathbb{R}^{22}$  additionally included LV myocardial wall-thickness measurements from the 16 AHA regions. When functional information was available, the optional motion descriptor  $c^{motion} \in \mathbb{R}^6$  comprised the logarithmically transformed LV and RV end-systolic volumes, the logarithmically transformed LV and RV stroke volumes, and the LV and RV ejection fractions. The final in-distribution cohort contained 666 subjects, as summarized in Table 1. The data were divided at the patient level into training, validation, and test sets using a ratio of 7:1:2.

### 4.2. Gold Standard and Evaluation

The SSM-fitted full-cycle mesh sequences were used as the reference standard. All methods were evaluated under the same ED-conditioned synthesis setting defined in Sec. 3.1. Before evaluation, the generated meshes were transformed back to the original physical coordinate system. Geometric agreement between the generated and reference sequences was evaluated using the average symmetric surface distance (ASSD), the 95th-percentile Hausdorff distance (HD95), and the vertex-wise root mean square error (vRMSE). Because all meshes shared fixed vertex correspondence, vRMSE was used to quantify point-wise trajectory errors over the complete cardiac cycle. For each test subject, every method generated 20 stochastic motion samples. Each metric was computed separately for the 20 samples and then averaged within the subject. Functional agreement was assessed using the LV and RV volume trajectories derived from the synthesized sequences. Agreement in LVEF and RVEF was quantified

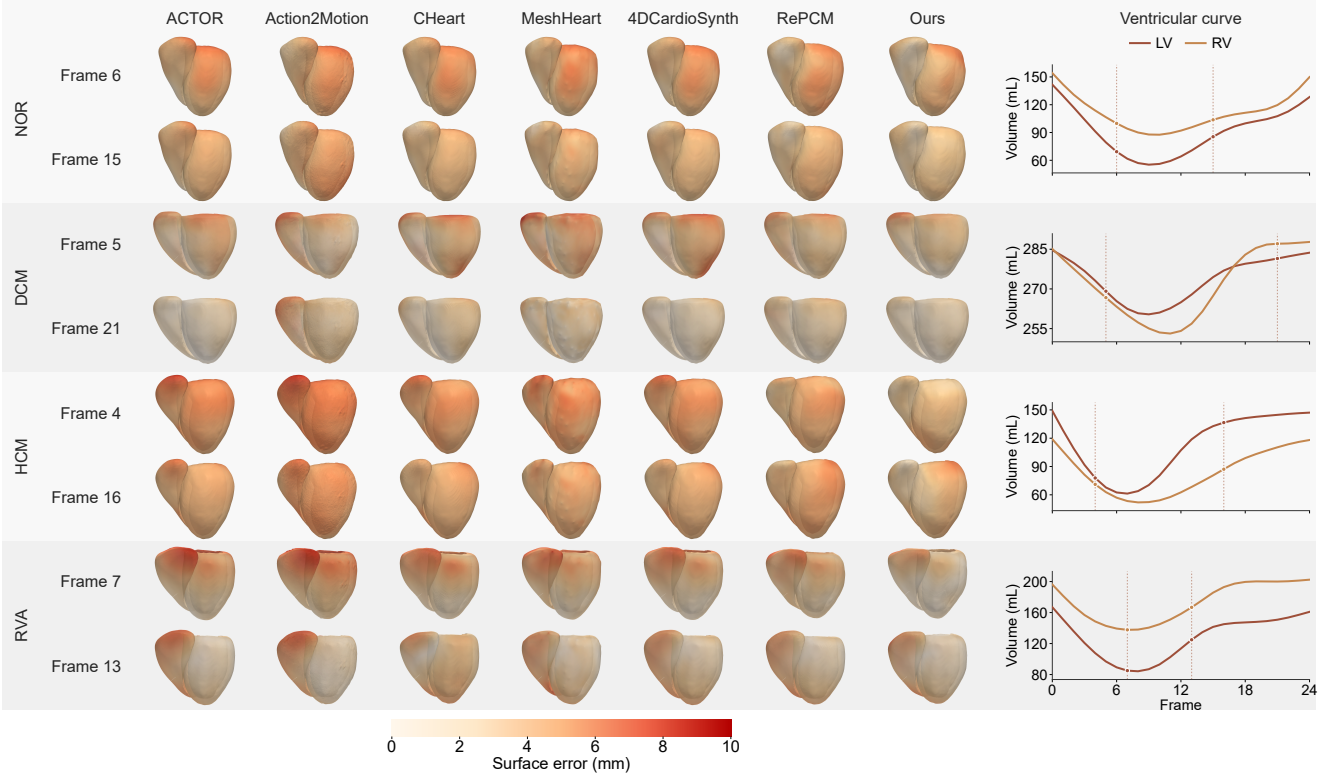


Figure 4: Qualitative comparison of ED-conditioned biventricular motion synthesis. Surface-error maps are shown for two representative cardiac phases from NOR, DCM, HCM, and RVA subjects. Each column corresponds to one competing method, and the color scale denotes the point-wise surface error relative to the reference mesh. The LV and RV volume curves on the right indicate the selected phases using vertical dashed lines.

using Pearson’s correlation coefficient  $r$  and mean absolute error (MAE). Population-level agreement in LVESV, RVESV, LVEF, and RVEF was further evaluated using the KL divergence and Wasserstein distance.

#### 4.3. Implementation

All models were implemented in PyTorch and trained on a single NVIDIA RTX A5500 GPU. The motion-informed functional partition contained  $N_R = 16$  regions, and the motion VAE employed  $N_a = 512$  region-balanced anchor tokens. The token feature dimension and motion latent dimension were set to  $d = 256$  and  $d_z = 32$ , respectively. Both the encoder and decoder contained  $K_m = 8$  attention blocks, with four attention heads in each block. The maximum hop range and attenuation factor of MRAB were set to  $L_{\text{hop}} = 4$  and  $\eta = 0.5$ , respectively. The motion VAE was trained for 500 epochs using Adam with a learning rate of  $1 \times 10^{-4}$ , a batch size of 8, and a KL-divergence weight of  $\lambda_{\text{KL}} = 5 \times 10^{-4}$ . The rectified-flow network contained  $K_r = 4$  blocks with a hidden dimension of 128. The fine-grained descriptor  $c_{ED}^{\text{fine}} \in \mathbb{R}^{22}$  conditioned the RF blocks through AdaLN. The global descriptor  $c_{ED}^{\text{global}} \in \mathbb{R}^6$  was combined with the anatomical latent solely for selecting among  $N_e = 4$  prototype-routed adapters. The RF model was trained for 200 epochs using AdamW with a learning rate of  $1 \times 10^{-5}$ , a batch size of 8, and a weight

decay of  $5 \times 10^{-3}$ . During inference, the rectified-flow ODE was solved using  $N_{\text{Euler}} = 4$  uniform Euler steps over  $\tau \in [0, 1]$ . In the optional functional-control experiments, the six-dimensional motion descriptor  $c^{\text{motion}}$  was used with a control strength of  $s = 1$ . This branch was disabled in the primary ED-conditioned setting.

#### 4.4. Comparison Study

Table 2 reports the quantitative comparison under the ED-conditioned full-cycle motion synthesis setting. We compared the proposed method with representative sequence generation and cardiac motion synthesis baselines, including a conditional VAE (Sohn et al., 2015), ACTOR (Petrovich et al., 2021), Action2Motion (Guo et al., 2020), CHeart (Qiao et al., 2023), MeshHeart (Qiao et al., 2025), 4DCardioSynth (Dou et al., 2025), and RePCM (Yang et al., 2026). For fairness, all baselines used the same pre-processed mesh sequences, ED input, temporal resolution, and patient-level split. One can see that the proposed method achieved the best overall performance across all metrics of different chambers and their combination, indicating more accurate full-cycle motion synthesis than both generic motion-generation baselines and cardiac-specific competing methods. Compared with all external baselines, our method consistently achieved lower biventricular (BiV), left-ventricular (LV), and right-ventricular (RV) surface and vertex-wise

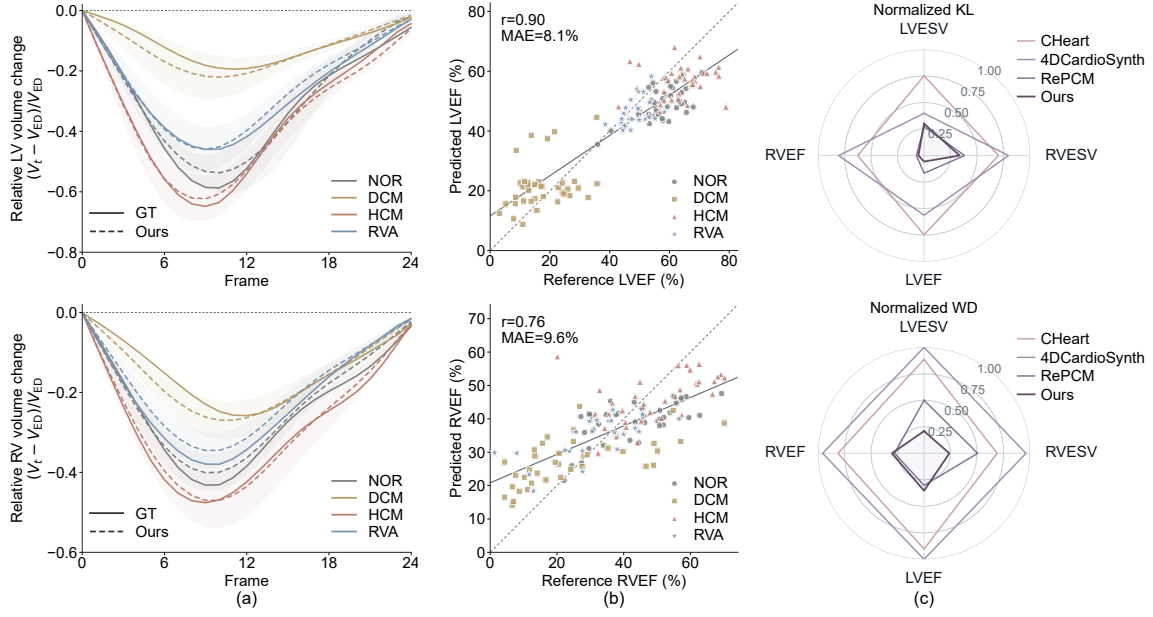


Figure 5: Functional fidelity of the synthesized biventricular sequences. (a) Diagnostic group-stratified relative volume-change trajectories for the LV (top) and RV (bottom). Solid and dashed curves represent the reference and synthesized sequences, respectively, and shaded bands indicate inter-subject variability. (b) Agreement between reference and synthesized left- and right-ventricular ejection fractions (LVEF and RVEF), with the identity line shown for reference. (c) Normalized KL divergence and Wasserstein distance (WD) between the reference and synthesized left- and right-ventricular end-systolic volumes (LVESV and RVESV), LVEF, and RVEF distributions. Lower values indicate better distributional agreement. MAE: mean absolute error.

errors. Compared with the preliminary RePCM framework, our method significantly reduced BiV ASSD from 1.69 to 1.49 mm ( $p < 0.01$ ), BiV HD95 from 4.27 to 3.77 mm ( $p < 0.01$ ), and BiV vRMSE from 3.61 to 3.31 mm ( $p < 0.05$ ), as determined by subject-level paired Wilcoxon signed-rank tests. The RV results are also noteworthy because RePCM showed comparatively limited gains over the external baselines for this chamber. RV motion synthesis is particularly challenging because of the chamber’s thin myocardial wall, complex geometry, and substantial inter-subject variability. Compared with RePCM, our method reduced RV ASSD from 2.00 to 1.85 mm and RV vRMSE from 4.08 to 3.74 mm, indicating more accurate synthesis of challenging RV dynamics.

Figure 4 provides representative surface-error maps for all four diagnostic groups. Competing methods frequently produced spatially extended errors around the basal ventricular surfaces and the RV free wall, particularly near phases of maximal contraction. In contrast, the proposed method yielded smaller and more spatially localized errors across the selected systolic and relaxation phases. The improvement was consistent across NOR, DCM, HCM, and RVA cases, supporting the quantitative findings in Table 2.

Figure 5 presents whether the synthesized sequences preserve clinically relevant ventricular dynamics. The diagnostic group-stratified relative LV and RV volume trajectories closely followed the reference contraction-relaxation patterns, including the reduced LV contraction observed in the DCM group and the stronger relative LV volume reduction in HCM. Across individual subjects, the synthesized

LVEF achieved a correlation of  $r = 0.90$  with an MAE of 8.1%, while RVEF achieved  $r = 0.76$  with an MAE of 9.6%. The proposed method also produced the smallest normalized KL divergence and Wasserstein distance across LVESV, RVESV, LVEF, and RVEF among the compared methods. Thus, the geometric improvements translated into more faithful ventricular volume trajectories and functional-index distributions.

#### 4.5. Ablation Study

##### 4.5.1. Effect of Motion-Informed Parcellation

Table 3 presents the contribution of the motion-informed regional prior and the effect of partition granularity. Compared with the AHA-like partition, the motion-informed parcellation with  $N_R = 8$  and  $N_R = 16$  reduced all BiV errors, confirming the benefit of deriving functional regions directly from cardiac dynamics. The best overall performance was obtained with  $N_R = 16$ , whereas increasing the number of regions to  $N_R = 32$  removed these gains. This indicates that performance does not improve monotonically with finer regionalization, as excessive fragmentation may weaken stable regional aggregation.

Figure 6 provides a more detailed analysis across ventricular surfaces and learned motion regions. As shown in Fig. 6 (a), the  $N_R = 16$  partition achieved the lowest vRMSE across all four ventricular surface groups, while the RV free wall remained the most challenging surface under all partition strategies. Fig. 6 (b) further reports the region-wise vRMSE reduction of the  $N_R = 16$  partition relative to the AHA-like partition, where positive



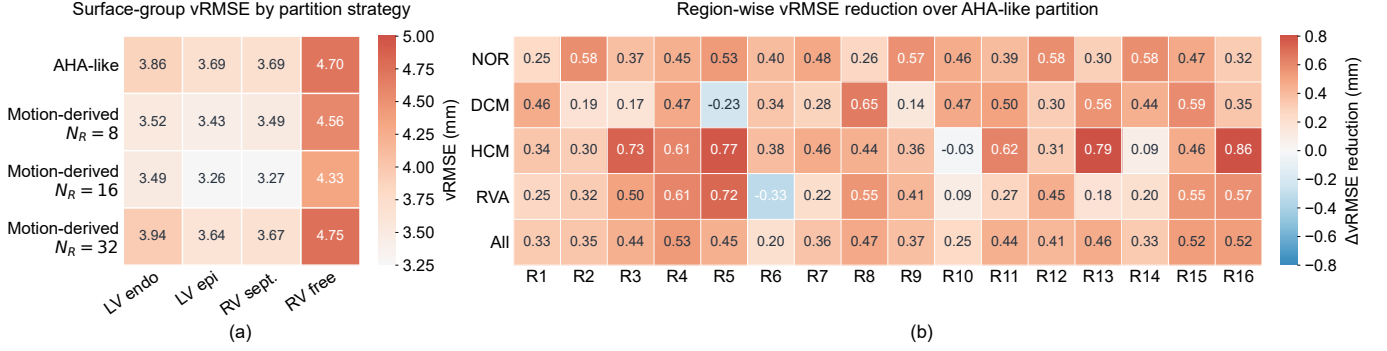


Figure 6: Regional analysis of motion-informed parcellation. (a) Surface-group vRMSE for the AHA-like partition and motion-informed parcellation with different numbers of regions. (b) Region-wise vRMSE reduction obtained by the  $N_R = 16$  motion-informed partition relative to the AHA-like partition, reported for each diagnostic group and the full cohort. Positive values indicate lower errors with motion-informed parcellation.

Table 3: Effect of regional partition strategies on biventricular motion synthesis.

| Partition  | ASSD (mm) ↓                       | HD95 (mm) ↓                       | vRMSE (mm) ↓                      |
|------------|-----------------------------------|-----------------------------------|-----------------------------------|
| AHA-like   | $1.59 \pm 0.44$                   | $4.11 \pm 1.74$                   | $3.62 \pm 1.43$                   |
| $N_R = 8$  | $1.50 \pm 0.37$                   | $3.88 \pm 1.39$                   | $3.40 \pm 1.19$                   |
| $N_R = 16$ | <b><math>1.49 \pm 0.34</math></b> | <b><math>3.77 \pm 1.06</math></b> | <b><math>3.31 \pm 1.03</math></b> |
| $N_R = 32$ | $1.59 \pm 0.38$                   | $4.19 \pm 1.36$                   | $3.62 \pm 1.21$                   |

Table 4: Effect of MRAB on biventricular motion synthesis performance. MRAB: multi-hop regional attention bias.

| Topology setting     | ASSD (mm) ↓                       | HD95 (mm) ↓                       | vRMSE (mm) ↓                      |
|----------------------|-----------------------------------|-----------------------------------|-----------------------------------|
| w/o MRAB             | $1.56 \pm 0.34$                   | $3.92 \pm 1.27$                   | $3.51 \pm 1.08$                   |
| $L_{\text{hop}} = 1$ | $1.55 \pm 0.42$                   | $3.81 \pm 1.71$                   | $3.35 \pm 1.38$                   |
| $L_{\text{hop}} = 2$ | $1.49 \pm 0.38$                   | <b><math>3.75 \pm 1.55</math></b> | $3.31 \pm 1.24$                   |
| $L_{\text{hop}} = 3$ | $1.49 \pm 0.43$                   | $3.86 \pm 1.80$                   | $3.37 \pm 1.44$                   |
| $L_{\text{hop}} = 4$ | <b><math>1.49 \pm 0.34</math></b> | $3.77 \pm 1.06$                   | <b><math>3.31 \pm 1.03</math></b> |
| $L_{\text{hop}} = 5$ | $1.55 \pm 0.40$                   | $3.94 \pm 1.45$                   | $3.41 \pm 1.31$                   |

values indicate lower errors with motion-informed parcellation. At the cohort level, improvements were observed across all 16 learned regions. Similar improvements were found across most diagnostic group-region combinations, with only a few isolated exceptions. These results show that the benefit of motion-informed parcellation was distributed across the ventricular surface rather than being driven by a small subset of regions.

#### 4.5.2. Effect of Multi-Hop Regional Attention Bias

Table 4 summarizes the effect of the MRAB used in the motion VAE. Removing the topology bias increased BiV vRMSE from  $3.31 \pm 1.03$  mm to  $3.51 \pm 1.08$  mm, confirming the contribution of topology-guided regional communication. Introducing MRAB improved performance across most hop ranges, indicating that graded topology-aware communication among the learned regions facilitates coherent motion propagation. The  $L_{\text{hop}} = 4$  configuration achieved comparable mean accuracy to  $L_{\text{hop}} = 2$ , while exhibiting consistently lower inter-subject variability across all three metrics.

Table 5: Ablation of phenotype conditioning, prototype routing, and optional functional control in the rectified-flow generator. Superscripts  $G$  and  $F$  denote the global descriptor  $c_{ED}^{\text{global}}$  and fine-grained descriptor  $c_{ED}^{\text{fine}}$ , respectively. The final row uses the additional motion descriptor  $c_{ED}^{\text{motion}}$  and is not part of the primary ED-only setting.

| $c_{ED}$<br>Condition | Adapter<br>Route | $c_{ED}^{\text{motion}}$<br>Control | ASSD (mm) ↓                       | HD95 (mm) ↓                       | vRMSE (mm) ↓                      |
|-----------------------|------------------|-------------------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
| ×                     | ×                | ×                                   | $1.64 \pm 0.42$                   | $4.17 \pm 1.78$                   | $3.60 \pm 1.43$                   |
| $\checkmark^G$        | ×                | ×                                   | $1.56 \pm 0.37$                   | $4.09 \pm 1.53$                   | $3.47 \pm 1.25$                   |
| $\checkmark^F$        | ×                | ×                                   | $1.51 \pm 0.36$                   | $3.92 \pm 1.12$                   | $3.36 \pm 1.16$                   |
| ×                     | $\checkmark^G$   | ×                                   | $1.52 \pm 0.36$                   | $3.99 \pm 1.47$                   | $3.40 \pm 1.16$                   |
| ×                     | $\checkmark^F$   | ×                                   | $1.54 \pm 0.37$                   | $3.97 \pm 1.44$                   | $3.43 \pm 1.17$                   |
| $\checkmark^G$        | $\checkmark^G$   | ×                                   | $1.51 \pm 0.36$                   | $3.83 \pm 1.52$                   | $3.41 \pm 1.19$                   |
| $\checkmark^G$        | $\checkmark^F$   | ×                                   | $1.56 \pm 0.35$                   | $4.00 \pm 1.23$                   | $3.54 \pm 1.18$                   |
| $\checkmark^F$        | $\checkmark^G$   | ×                                   | <b><math>1.49 \pm 0.34</math></b> | <b><math>3.77 \pm 1.06</math></b> | <b><math>3.31 \pm 1.03</math></b> |
| $\checkmark^F$        | $\checkmark^F$   | ×                                   | $1.54 \pm 0.37$                   | $3.97 \pm 1.44$                   | $3.39 \pm 1.16$                   |
| $\checkmark^F$        | $\checkmark^G$   | $\checkmark$                        | $1.46 \pm 0.39$                   | $3.74 \pm 1.49$                   | $3.21 \pm 1.23$                   |

#### 4.5.3. Effect of Phenotype Conditioning and Prototype Routing

Table 5 evaluates the two descriptor-based conditioning mechanisms used in Sec. 3.4. Relative to the baseline without explicit phenotype conditioning or routing, fine-grained ED conditioning consistently improved all three metrics, while global-descriptor-based routing also yielded clear gains across the three measures. Combining  $c_{ED}^{\text{fine}}$  for latent conditioning with  $c_{ED}^{\text{global}}$  for adapter routing achieved the best ED-only result, with an ASSD of 1.49 mm, HD95 of 3.77 mm, and vRMSE of 3.31 mm. The two mechanisms therefore provide complementary benefits: fine-grained descriptors guide subject-specific latent generation, whereas global descriptors provide a more discriminative signal for routing among the motion adapters.

Figure 7 explains the different roles of the two descriptors. Routing based on  $c_{ED}^{\text{global}}$  assigned DCM and HCM subjects predominantly to distinct adapters, while NOR and RVA subjects were distributed mainly across the remaining routes. In contrast, routing with  $c_{ED}^{\text{fine}}$  produced greater overlap across diagnostic groups and distributed

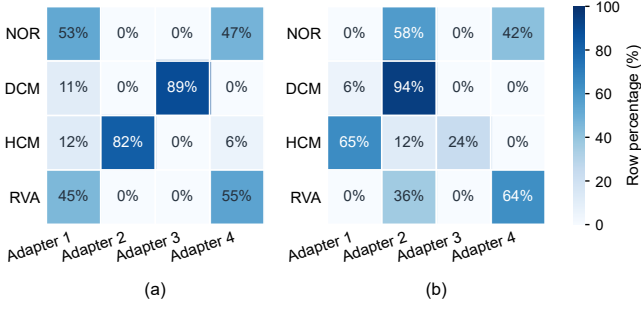


Figure 7: Diagnostic-group distributions across prototype-routed motion adapters. Row-normalized adapter assignments are shown for routing using the anatomical latent together with (a) the global ED descriptor  $c_{ED}^{global}$  or (b) the fine-grained ED descriptor  $c_{ED}^{fine}$ . Each cell reports the percentage of subjects in a diagnostic group assigned to the corresponding adapter.

Table 6: Effect of the number of routing prototypes/ adapters  $N_e$  on biventricular motion synthesis performance.

| $N_e$ | ASSD (mm) ↓                       | HD95 (mm) ↓                       | vRMSE (mm) ↓                      |
|-------|-----------------------------------|-----------------------------------|-----------------------------------|
| 4     | $1.49 \pm 0.34$                   | <b><math>3.77 \pm 1.06</math></b> | <b><math>3.31 \pm 1.03</math></b> |
| 6     | <b><math>1.47 \pm 0.38</math></b> | $3.82 \pm 1.58$                   | $3.36 \pm 1.30$                   |
| 8     | $1.50 \pm 0.40$                   | $3.94 \pm 1.72$                   | $3.41 \pm 1.38$                   |

HCM subjects across several adapters. Fine-grained descriptors are therefore useful as continuous generation conditions, but may contain route-irrelevant local variation that weakens the separation of global motion modes. Accordingly, the final model uses  $c_{ED}^{fine}$  for latent conditioning and  $c_{ED}^{global}$  for prototype routing.

Table 6 evaluates the sensitivity to the number of routing prototypes/ adapters  $N_e$ . Performance was relatively stable across different values of  $N_e$ . Although  $N_e = 6$  achieved a slightly lower ASSD,  $N_e = 4$  produced the best HD95 and vRMSE with lower variance. We used  $N_e = 4$  as the default setting, providing a compact routing module while maintaining stable motion synthesis accuracy.

#### 4.6. In- and Out-of-Distribution Diagnostic-Group Analysis

Figure 8 (a) reports the geometric errors across four in-distribution (ID) diagnostic groups shared by ACDC, M&Ms, and M&Ms-2. Performance remained broadly stable across the four groups, with no marked degradation observed in any single group. DCM and HCM exhibited somewhat higher median ASSD than NOR and RVA, reflecting the greater variability associated with pathological ventricular remodeling and contraction. Nevertheless, their interquartile ranges remained substantially overlapped. Fig. 8 (b) further evaluates five OOD diagnostic groups from M&Ms and M&Ms-2 that were completely excluded from training: HHD, IHD, FALL, CIA, and TRI. Although some OOD groups exhibited greater error variability, particularly CIA, the overall error distributions remained within a comparable range across the two datasets.

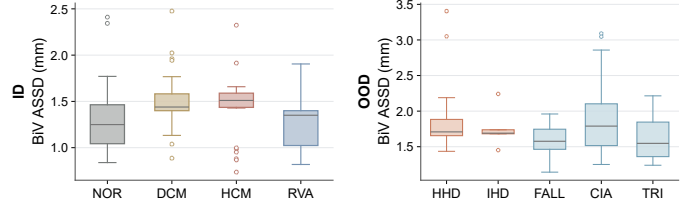


Figure 8: Geometric errors across in-distribution (ID) and out-of-distribution (OOD) diagnostic groups. (a) Four ID groups shared by ACDC, M&Ms, and M&Ms-2. (b) Five OOD groups excluded from training: hypertensive heart disease (HHD) and ischemic heart disease (IHD) from M&Ms, and tetralogy of Fallot (FALL), interatrial communication (CIA), and tricuspid regurgitation (TRI) from M&Ms-2.

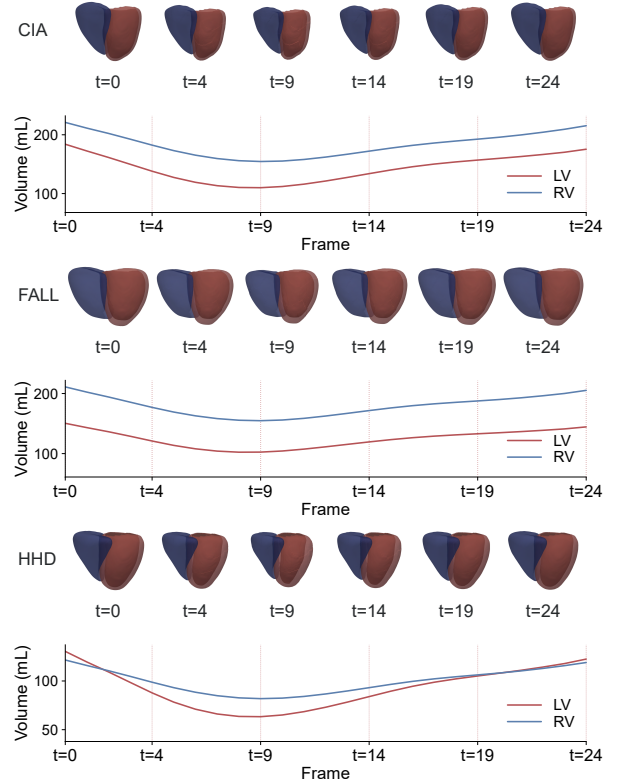


Figure 9: Illustration of full-cycle synthesized mesh sequences for three representative OOD cases and their corresponding LV and RV volume trajectories.

These results suggest that the learned motion representation generalizes beyond the diagnostic groups observed during training rather than relying on a single dominant disease pattern. Representative synthesized sequences are shown in Fig. 9 for OOD cases with CIA, FALL, and HHD. The generated biventricular surfaces evolve smoothly throughout the cardiac cycle and exhibit distinct LV and RV volume trajectories across the three cases. These examples illustrate that the model generates temporally coherent motion with distinct case-specific ventricular dynamics across the three OOD cases.



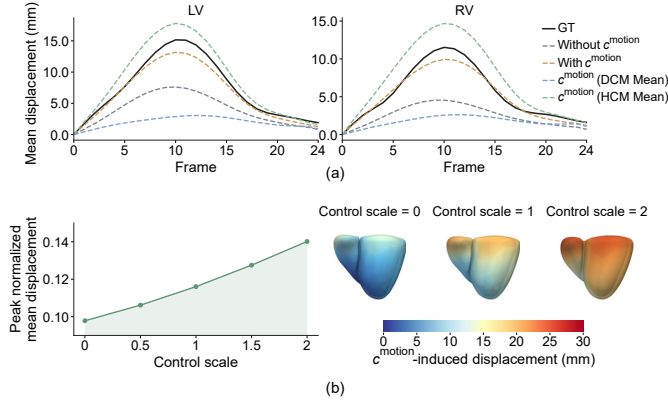


Figure 10: Controllable modulation of synthesized cardiac motion. (a) Mean LV and RV displacement trajectories for the reference, the ED-only synthesis, and predictions obtained using the optional functional-control branch. With  $c^{motion}$  uses the subject-specific motion descriptor of the illustrated NOR case, whereas the DCM-mean and HCM-mean curves use the corresponding group-average descriptors. (b) Response of a representative NOR case to control scale ranging from 0 to 2, including the relationship between the control coefficient and peak normalized mean displacement and the corresponding spatial maps of the  $c^{motion}$ -induced deformation at  $s = 0$ ,  $s = 1$ , and  $s = 2$ .

#### 4.7. Optional Functional Motion Control

The primary model infers motion solely from ED anatomy and ED-derived phenotype descriptors. When additional motion descriptors are available, the optional  $c^{motion}$  branch enables explicit modulation of the generated motion. As shown in the final row of Table 5, incorporating this control reduced BiV ASSD from 1.49 to 1.46 mm and vRMSE from 3.31 to 3.21 mm. Fig. 10 (a) shows that the controlled LV and RV mean-displacement trajectories more closely followed the reference than the uncontrolled synthesis. Conditioning on the mean motion descriptors derived from the DCM and HCM groups produced distinct displacement amplitudes, demonstrating descriptor-dependent modulation of cardiac motion. Fig. 10 (b) further evaluates a representative NOR case under control scale ranging from 0 to 2. The peak normalized mean displacement increased monotonically and approximately linearly with the control scale, while the corresponding spatial maps showed smooth and spatially varying deformation. These results indicate that the optional branch provides interpretable control over motion magnitude without changing the primary ED-only inference setting.

## 5. Discussion and Conclusion

This study presents a region-specific and phenotype-adaptive framework for synthesizing full-cycle biventricular motion from a single ED mesh. Across three public cine MRI datasets and four in-distribution diagnostic groups, the proposed ED-conditioned model consistently outperformed generic motion generators and cardiac-specific synthesis methods in surface- and correspondence-based met-

rics. The improvement over RePCM indicates that replacing global latent generation with structured motion representation and phenotype-conditioned RF generation provides a more effective formulation for this task. The gains were particularly evident for the RV, whose complex geometry and greater inter-subject variability make its motion more difficult to recover. Importantly, the geometric improvements were accompanied by better functional fidelity: the synthesized ventricular volume trajectories preserved diagnostic-group-specific ventricular motion patterns, and the derived LVEF and RVEF showed meaningful agreement with their reference values. The lower distributional discrepancies in ventricular volumes and ejection fractions further suggest that the model captures clinically relevant population-level variation rather than merely minimizing point-wise geometric errors. The OOD evaluation on M&Ms and M&Ms-2 further indicates that the learned representation generalizes beyond the diagnostic groups observed during training, although the increased variability in some OOD groups highlights the remaining challenge of uncommon pathological anatomies.

The ablation studies provide complementary evidence for the principal design choices. Motion-informed functional parcellation consistently outperformed the AHA-like partition, demonstrating that regions learned directly from deformation patterns are better suited to generative motion modeling than predefined anatomical divisions. The strongest performance was obtained with 16 regions, whereas excessive partitioning weakened regional aggregation, indicating that the regional prior should balance local specificity with motion coherence. MRAB further improved synthesis by permitting graded communication between neighboring functional regions while suppressing unrelated interactions. In the latent generator, fine-grained ED phenotype descriptors were more effective for conditioning motion generation, whereas global ED phenotype descriptors produced clearer and more stable prototype routing. This distinction suggests that detailed regional morphology is useful for specifying subject-level dynamics, while global anatomy is better suited to identifying broader motion modes. The optional functional-control branch further modulated motion magnitude in a continuous manner, but its results should be interpreted separately from the primary ED-only setting because it requires additional motion descriptors unavailable from the ED anatomy alone.

Several limitations remain. First, the reference sequences were obtained through SSM fitting of segmentation masks and therefore inherit errors introduced by image segmentation, template fitting, and temporal smoothing; evaluation against independently tracked or manually verified 4D meshes would provide a stronger assessment of motion accuracy. Second, the current study focuses on unified-topology biventricular surfaces and four relatively broad diagnostic groups. Its applicability to atrial motion, myocardial-layer deformation, congenital abnormalities, and more heterogeneous or subtle disease subtypes remains to be investigated. Third, although ED

anatomy provides an accessible patient-specific condition, it cannot uniquely determine functional properties such as contraction strength, electromechanical delay, or regional dysfunction, as also reflected by the remaining errors in RVEF estimation. Future work should therefore incorporate complementary imaging, clinical, or physiological conditions when available, evaluate generalization on independent clinical cohorts and alternative mesh-construction pipelines, and extend the framework toward controllable whole-heart motion generation and virtual population synthesis for downstream computational modeling and in silico studies.

## References

- Bai, W., Shi, W., de Marvao, A., Dawes, T.J., O'Regan, D.P., Cook, S.A., Rueckert, D., 2015. 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, 133–145.
- Bai, W., Suzuki, H., Huang, J., Francis, C., Wang, S., Tarroni, G., Guitton, F., Aung, N., Fung, K., Petersen, S.E., et al., 2020. A population-based phenome-wide association study of cardiac and aortic structure and function. *Nature medicine* 26, 1654–1662.
- Banerjee, A., Zacur, E., Choudhury, R.P., Grau, V., 2022. Automated 3d whole-heart mesh reconstruction from 2d cine mr slices using statistical shape model, in: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE. pp. 1702–1706.
- Beetz, M., Banerjee, A., Grau, V., 2021. Generating subpopulation-specific biventricular anatomy models using conditional point cloud variational autoencoders, in: International Workshop on Statistical Atlases and Computational Models of the Heart, Springer. pp. 75–83.
- Beetz, M., Banerjee, A., Li, L., Camps, J., Rodriguez, B., Grau, V., 2025. 3d cardiac shape analysis with variational point cloud autoencoders for myocardial infarction prediction and virtual heart synthesis. *Computerized Medical Imaging and Graphics* 124, 102587.
- Beetz, M., Corral Acero, J., Banerjee, A., Eitel, I., Zacur, E., Lange, T., Stiermaier, T., Evertz, R., Backhaus, S.J., Thiele, H., et al., 2022. Interpretable cardiac anatomy modeling using variational mesh autoencoders. *Frontiers in cardiovascular medicine* 9, 983868.
- 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., 2018. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE transactions on medical imaging* 37, 2514–2525.
- Biglino, G., Capelli, C., Bruse, J., Bosi, G.M., Taylor, A.M., Schievano, S., 2017. Computational modelling for congenital heart disease: how far are we from clinical translation? *Heart* 103, 98–103.
- 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., 2021. Multi-centre, multi-vendor and multi-disease cardiac segmentation: the m&ms challenge. *IEEE Transactions on Medical Imaging* 40, 3543–3554.
- Chen, Y., Yang, J., Mercadier, D.S., Le, H., Schwitter, J., Fua, P., 2026. End-to-end 4d heart mesh recovery across full-stack and sparse cardiac mri. *arXiv:2509.12090*.
- Chung, J., Kastner, K., Dinh, L., Goel, K., Courville, A.C., Bengio, Y., 2015. A recurrent latent variable model for sequential data. *Advances in neural information processing systems* 28.
- Dhariwal, P., Nichol, A., 2021. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems* 34, 8780–8794.
- Dou, H., Huang, J., Zakeri, A., Zhou, Z., Mu, T., Duan, J., Frangi, A.F., 2025. 4d cardiosynth: Synthesising dynamic virtual heart populations through spatiotemporal disentanglement, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 3–12.
- Dou, H., Ravikumar, N., Frangi, A.F., 2023. A conditional flow variational autoencoder for controllable synthesis of virtual populations of anatomy, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 143–152.
- Guo, C., Zuo, X., Wang, S., Zou, S., Sun, Q., Deng, A., Gong, M., Cheng, L., 2020. Action2motion: Conditioned generation of 3d human motions, in: Proceedings of the 28th ACM international conference on multimedia, pp. 2021–2029.
- Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems* 33, 6840–6851.
- Hu, V.T., Yin, W., Ma, P., Chen, Y., Fernando, B., Asano, Y.M., Gavves, E., Mettes, P., Ommer, B., Snoek, C.G.M., 2023. Motion flow matching for human motion synthesis and editing. *arXiv:2312.08895*.
- Kingma, D.P., Welling, M., 2013. Auto-encoding variational bayes. *arXiv:1312.6114*.
- Kong, F., Shadden, S.C., 2022. Learning whole heart mesh generation from patient images for computational simulations. *IEEE Transactions on Medical Imaging* 42, 533–545.
- Kong, F., Stocker, S., Choi, P.S., Ma, M., Ennis, D.B., Marsden, A.L., 2024. Sdf4chd: Generative modeling of cardiac anatomies with congenital heart defects. *Medical image analysis* 97, 103293.
- Kong, F., Wilson, N., Shadden, S., 2021. A deep-learning approach for direct whole-heart mesh reconstruction. *Medical image analysis* 74, 102222.
- Laumer, F., Amrani, M., Manduchi, L., Beuret, A., Rubi, L., Dubatovka, A., Matter, C.M., Buhmann, J.M., 2023. Weakly supervised inference of personalized heart meshes based on echocardiography videos. *Medical image analysis* 83, 102653.
- Lipman, Y., Chen, R.T.Q., Ben-Hamu, H., Nickel, M., Le, M., 2023. Flow matching for generative modeling. *arXiv:2210.02747*.
- Liu, X., Gong, C., Liu, Q., 2022. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv:2209.03003*.
- Liu, X., Yuan, X., Chan, M.Y., Sia, C.H., Li, L., 2026. Cinemesh4d: Personalized 4d whole heart reconstruction from sparse cine mri. *arXiv:2605.13994*.
- López, P.A., Mella, H., Uribe, S., Hurtado, D.E., Costabal, F.S., 2023. Warppinn: Cine-mr image registration with physics-informed neural networks. *Medical Image Analysis* 89, 102925.
- Lu, J., Jin, R., Wang, M., Song, E., Ma, G., 2023. A bidirectional registration neural network for cardiac motion tracking using cine mri images. *Computers in Biology and Medicine* 160, 107001.
- Luo, Y., Sesia, D., Wang, F., Wu, Y., Ding, W., Hasan, K., Huang, J., Shi, F., Shah, A., Kaura, A., et al., 2026. Explicit differentiable slicing and global deformation for cardiac mesh reconstruction. *Medical image analysis* , 103999.
- Marsden, A.L., Feinstein, J.A., 2015. Computational modeling and engineering in pediatric and congenital heart disease. *Current opinion in pediatrics* 27, 587–596.
- 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., 2023. Deep learning segmentation of the right ventricle in cardiac mri: the m&ms challenge. *IEEE Journal of Biomedical and Health Informatics* 27, 3302–3313.
- Meng, Q., Bai, W., O'Regan, D.P., Rueckert, D., 2023. Deepmesh: mesh-based cardiac motion tracking using deep learning. *IEEE transactions on medical imaging* 43, 1489–1500.
- Niederer, S.A., Aboelkassem, Y., Cantwell, C.D., Corrado, C., Coveney, S., Cherry, E.M., Delhaes, T., Fenton, F.H., Panfilov, A.V., Pathmanathan, P., et al., 2020. Creation and application of virtual patient cohorts of heart models. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 378.
- Peebles, W., Xie, S., 2023. Scalable diffusion models with transformers, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 4195–4205.

- Perez, E., Strub, F., de Vries, H., Dumoulin, V., Courville, A., 2017. Film: Visual reasoning with a general conditioning layer. *arXiv:1709.07871*.
- Petrovich, M., Black, M.J., Varol, G., 2021. Action-conditioned 3d human motion synthesis with transformer vae, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 10985–10995.
- Puyol-Anton, E., Sinclair, M., Gerber, B., Amzulescu, M.S., Langet, H., De Craene, M., Aljabar, P., Piro, P., King, A.P., 2017. A multimodal spatiotemporal cardiac motion atlas from mr and ultrasound data. *Medical image analysis* 40, 96–110.
- Qiao, M., McGurk, K.A., Wang, S., Matthews, P.M., O'Regan, D.P., Bai, W., 2025. A personalized time-resolved 3d mesh generative model for unveiling normal heart dynamics. *Nature Machine Intelligence* 7, 800–811.
- Qiao, M., Wang, S., Qiu, H., De Marvao, A., O'Regan, D.P., Rueckert, D., Bai, W., 2023. Cheart: A conditional spatio-temporal generative model for cardiac anatomy. *IEEE transactions on medical imaging* 43, 1259–1269.
- Qiao, M., Wang, Y., Guo, Y., Huang, L., Xia, L., Tao, Q., 2020. Temporally coherent cardiac motion tracking from cine mri: Traditional registration method and modern cnn method. *Medical Physics* 47, 4189–4198.
- Qin, C., Wang, S., Chen, C., Bai, W., Rueckert, D., 2023. Generative myocardial motion tracking via latent space exploration with biomechanics-informed prior. *Medical Image Analysis* 83, 102682.
- Roth, G.A., Mensah, G.A., Johnson, C.O., Addolorato, G., Ammirati, E., Baddour, L.M., Barengo, N.C., Beaton, A.Z., Benjamin, E.J., Benziger, C.P., et al., 2020. Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the gbd 2019 study. *Journal of the American college of cardiology* 76, 2982–3021.
- Sohn, K., Lee, H., Yan, X., 2015. Learning structured output representation using deep conditional generative models. *Advances in neural information processing systems* 28.
- Song, J., Meng, C., Ermon, S., 2022. Denoising diffusion implicit models. *arXiv:2010.02502*.
- Sørensen, K., Diez, P., Margeta, J., El Youssef, Y., Pham, M., Pedersen, J.J., Kühn, T., De Backer, O., Kofoed, K., Camara, O., et al., 2024. Spatio-temporal neural distance fields for conditional generative modeling of the heart, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 422–432.
- Uppendra, R.R., Hasan, S.K., Simon, R., Wentz, B.J., Shontz, S.M., Sacks, M.S., Linte, C.A., 2021. Motion extraction of the right ventricle from 4d cardiac cine mri using a deep learning-based deformable registration framework, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE. pp. 3795–3799.
- Wan, M., Huang, W., Zhang, J.M., Zhao, X., Tan, R.S., Wan, X., Zhong, L., 2015. Variational reconstruction of left cardiac structure from cmr images. *PloS one* 10, e0145570.
- Wu, Z., Yu, C., Wang, F., Bai, X., 2025. 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.
- Xue, W., Brahm, G., Leung, S., Shmuelovich, O., Li, S., 2018. Cardiac motion scoring with segment-and subject-level non-local modeling, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 437–445.
- Yang, J., Lin, Y., Pu, B., Li, X., 2024. Bidirectional recurrence for cardiac motion tracking with gaussian process latent coding. *Advances in Neural Information Processing Systems* 37, 34800–34823.
- Yang, X., Yuan, X., Li, H., Chen, L., Liu, Y., Li, L., 2026. Repcm: Region-specific and phenotype-adaptive bi-ventricular cardiac motion synthesis. *arXiv:2605.21237*.
- Yang, Z., Teng, J., Zheng, W., Ding, M., Huang, S., Xu, J., Yang, Y., Hong, W., Zhang, X., Feng, G., et al., 2025. Cogvideox: Text-to-video diffusion models with an expert transformer, in: International Conference on Learning Representations, pp. 83048–83077.
- Ye, M., Kanski, M., Yang, D., Chang, Q., Yan, Z., Huang, Q., Axel, L., Metaxas, D., 2021. 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.
- Yuan, X., Liu, C., Wang, Y., 2023. 4d myocardium reconstruction with decoupled motion and shape model, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 21252–21262.
- Zakeri, A., Hokmabadi, A., Bi, N., Wijesinghe, I., Nix, M.G., Petersen, S.E., Frangi, A.F., Taylor, Z.A., Gooya, A., 2023. Dragnet: Learning-based deformable registration for realistic cardiac mr sequence generation from a single frame. *Medical Image Analysis* 83, 102678.