



This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

VecHeart: Holistic Four-Chamber Cardiac Anatomy Modeling via Hybrid VecSets

Yihong Chen¹ and Pascal Fua¹

¹CVLAB, EPFL

{yihong.chen, pascal.fua}@epfl.ch

Abstract. Accurate cardiac anatomy modeling requires the model to be able to handle intricate interrelations among structures. In this paper, we propose VecHeart, a unified framework for holistic reconstruction and generation of four-chamber cardiac structures. To overcome the limitations of current feed-forward implicit methods, specifically their restriction to single-object modeling and their neglect of inter-part correlations, we introduce Hybrid Part Transformer, which leverages part-specific learnable queries and interleaved attention to capture complex inter-chamber dependencies. Furthermore, we propose Anatomical Completion Masking and Modality Alignment strategies, enabling the model to infer complete four-chamber structures from partial, sparse, or noisy observations, even when certain anatomical parts are entirely missing. VecHeart also seamlessly extends to 3D+t dynamic mesh sequence generation, demonstrating exceptional versatility. Experiments show that our method achieves state-of-the-art performance, maintaining high-fidelity reconstruction across diverse challenging scenarios. Code is available at <https://github.com/Scalsol/VecHeart>.

Keywords: Cardiac Imaging · Shape Modeling · Generative Modeling

1 Introduction

Accurate cardiac structure modeling is vital for applications like clinical diagnosis and intervention planning [27,15,25]. However, faithfully representing multi-part heart geometry from diverse input configurations remains a significant challenge.

Implicit functions offer significant flexibility and have been used across various fields. But conventional implicit-based shape representations [24,5,40,22] are typically tailored for single-structure modeling, neglecting the correlations among structures. Recent multi-part approaches [29,36,14,18] have begun to address this, but they rely on computationally expensive test-time optimization. Moreover, during test time, they require complete surface information to optimize the latent code. However, the observation is often sparse, and sometimes available only for some structures and not others. For instance, only sparse delineations of the four-chamber structure can be obtained from Long-Axis (LAX) views, while Short-Axis (SAX) MRI stacks cover the left and right ventricles, but not the atrial structures. These constraints severely limit their clinical utility.

To address these limitations, we propose VecHeart, a versatile framework for high-quality four-chamber cardiac structure reconstruction and generation given potentially incomplete data. It keeps the fidelity of the individual parts while maintaining overall consistency across the whole heart.

We start from the VecSet [39] latent representation to represent complex surfaces. Given surface points sampled from a 3D shape, VecSet uses a query vector and an encoder to turn them into a set of latent codes that are fed to a set of attention layers and then decoded into a signed distance function. VecHeart uses a similar encoder and decoder, along with a part-specific query vector. This enables us to introduce three new techniques designed to handle the heart’s complex multi-part anatomy even when the data is missing for some of them:

- **Hybrid Part Transformer.** We sequentially and iteratively apply intra-part attention and inter-part attention to capture the correlations between the parts, as shown in Fig. 1(a).
- **Anatomical Completion Masking.** During training, we randomly mask some chambers and force VecHeart to still output the full anatomy from the incomplete data, as shown by Fig. 1(b). This is key to handling missing data.
- **Modality Alignment.** To allow reconstruction not only from surface samplings but also from potentially sparse CMR slices as depicted by Fig. 1(c), we ensure that codes obtained from such slices are similar to those obtained from denser samplings.

As a result, VecHeart provides superior representation ability and a unified solution for robust cardiac modeling across a wide spectrum of input densities, even when some structures are entirely missing. We further show that with minimal effort, we can use flow matching [17, 7] for 3D+t mesh sequence generation.

2 Method

We aim for a consistent representation of the four cardiac chambers, the Left Ventricle (LV), LV myocardium, Right Ventricle (RV), Left Atrium (LA) and Right Atrium (RA). We seek to obtain this multi-component structure from potentially noisy data from different modalities and sometimes missing altogether.

Among the many existing approaches to modeling complex 3D structures, VecSet [39] is considered one of the most powerful. However, it is designed to model individual objects as opposed to multiple anatomically related objects that must remain consistent with each other. Assigning a separate VecSet model to each part would be a simple solution. However, this would not account for their inter-structure relationships. Instead, we introduce VecHeart, which retains the encoder and decoder of VecSet but uses different latent vectors for each part along with interleaved attention layers to model inter-part relationships.

We briefly review VecSet [39] in Sec. 2.1. In Sec. 2.2 and 2.3, we introduce VecHeart, our proposed framework designed to represent multi-part cardiac structures and handle heterogeneous modalities. Training detail is in 2.4. Finally, we describe how VecHeart can be extended beyond reconstruction to 3D+t shape generation in Sec. 2.5 to showcase the proposed framework’s versatility.

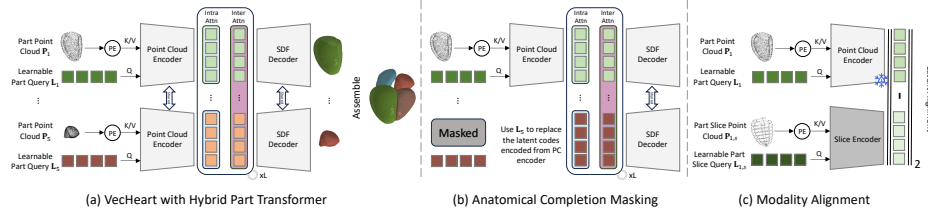


Fig. 1: *VecHeart pipeline*. (a) A shared encoder creates part latent codes from surface point clouds and learnable part queries that are processed by interleaved intra/inter-part attention layers, which are fed to a shared decoder that outputs SDF values for all parts. (b) During training, we randomly mask K chambers while still making the model reconstruct the whole heart. To this end, we replace the encoded latent vectors by corresponding learnable ones. (c) To handle heterogeneous input, we introduce a dual-encoder and an alignment loss.

2.1 Preliminary

VecSet [39] is a transformer-based 3D VAE that encodes surface into compact latent codes. Given a 3D surface S , its pipeline consists of three key steps:

- **Surface Sampling:** Sample a point cloud $\mathbf{P}$ of N points from surface $\mathbf{S}$.
- **Feature Encoding:** $\mathbf{P}$ is encoded into compact latent code $\mathbf{C}$ using a mechanism analogous to Perceiver [11]. This is written as:

$$\mathbf{C} = \text{Enc}(\mathbf{L}, \mathbf{P}) = \text{Attention}(\mathbf{L}, \text{PosEmb}(\mathbf{P})) \in \mathbb{R}^{D \times M}, \quad (1)$$

where $\text{Attention}(Q, K/V)$ is a standard attention block [31]. $\mathbf{L} \in \mathbb{R}^{D \times M}$ is a *learnable query* set with M the set size and D the dimension.

- **Geometry Decoding:** Given query points $\mathbf{Q}_{space}$ sampled from the space, SDF for $\mathbf{Q}_{space}$ are predicted using a combination of cross/self-attention:

$$\text{SDF}^{pred}(\mathbf{Q}_{space}) = \text{Attention}(\text{PosEmb}(\mathbf{Q}_{space}), \text{Attention}^L(\mathbf{C}, \mathbf{C})), \quad (2)$$

where the superscript L denotes applying the attention layers L times.

While achieving superior representation capabilities, VecSet is restricted to modeling single object. This renders it unsuitable for directly representing multi-part structures. Assigning a separate VecSet model to each part seems intuitive. However, this will neglect inter-part correlations, which may degrade the performance when dealing with incomplete or sparse input.

2.2 Structuring Multi-Part Anatomy with VecHeart

To model individual chambers along with myocardium, we introduce the VecHeart architecture depicted by Fig. 1. It expands on VecSet by incorporating a Hybrid Part Transformer (HPT) along with an Anatomical Completion Masking (ACM) strategy. Together, they enable VecHeart to capture inter-part correlations and maintain anatomical integrity, even when the input is incomplete.

Hybrid Part Transformer (HPT). Given a 4-chamber heart-model $\{\mathbf{S}_p\}_{p=1}^5$ where $\mathbf{S}_p$ denotes the surface of the p -th anatomical part, HPT first uses a shared encoder to map each surface into a part-level latent code $\mathbf{C}_p$. To maintain class-wise distinctions while facilitating cross-part information exchange, it relies on *learnable part queries* $\{\mathbf{L}_p\}_{p=1}^5$ to capture part-specific knowledge:

$$\mathbf{C}_p = \text{Enc}(\mathbf{L}_p, \mathbf{P}_p) = \text{Attention}(\mathbf{L}_p, \text{PosEmb}(\mathbf{P}_p)) \in \mathbb{R}^{D \times M}, p = 1, \dots, 5, \quad (3)$$

where $\mathbf{P}_p$ is the point cloud sampled from $\mathbf{S}_p$.

Upon obtaining $\{\mathbf{C}_p\}_{p=1}^5$, we facilitate information exchange across components using interleaved intra/inter-part attentions. Specifically, the intra-part attention is applied independently to each $\mathbf{C}_p$ to capture localized features, ensuring that the local anatomical details are preserved and refined. Global inter-part attention is performed over all latent codes $\{\mathbf{C}_p\}_{p=1}^5$ to model structural dependencies and global interactions. Thus we write $\text{IntraAttn}(\mathbf{C}_p) = \text{Attention}(\mathbf{C}_p, \mathbf{C}_p)$ and $\text{InterAttn}(\mathbf{C}_p) = \text{Attention}(\mathbf{C}_p, \{\mathbf{C}_p\}_{p=1}^5)$.

The SDF for each anatomical part is obtained by querying its part-level latent code through a shared decoder comprising L interleaved attention layers:

$$\begin{aligned} SDF_p^{\text{pred}}(\mathbf{Q}_{\text{space}}) &= \text{Dec}(\mathbf{Q}_{\text{space}}, \mathbf{C}_p), \\ \text{Dec}(\mathbf{Q}_{\text{space}}, \mathbf{C}_p) &= \text{Attention}(\text{PosEmb}(\mathbf{Q}_{\text{space}}), \text{XAttn}^L(\mathbf{C}_p)), \end{aligned} \quad (4)$$

where $\text{XAttn} = \text{InterAttn} \circ \text{IntraAttn}$, and $\circ$ denotes operation composition.

By using part-specific learnable queries and interleaved attention mechanism, HPT maintains class-wise distinction while facilitating cross-part information exchange. In the following, we detail how HPT can be extended to enable complete four-chamber reconstruction from partial observations when integrated with the Anatomical Completion Masking (ACM) strategy.

Anatomical Completion Masking (ACM). Because data can be missing for some chambers, the ability to perform complete reconstructions from incomplete data is highly desirable. Inspired by Masked Autoencoders (MAE) [9], we introduce ACM. Specifically, we randomly mask K parts and substitute their extracted latent codes $\mathbf{C}_p$ with the corresponding part-specific learnable queries $\mathbf{L}_p$ as decoder inputs. Through the interleaved attention, the model is tasked with predicting the SDF for **all** anatomical parts. This encourages the model to infer the missing structures by leveraging both learned anatomical priors and the geometric context provided by other components. We formulate this as:

$$\mathbf{D}_p = (1 - \mathbf{M}_p)\mathbf{C}_p + \mathbf{M}_p\mathbf{L}_p, \quad SDF_p^{\text{pred}}(\mathbf{Q}_{\text{space}}) = \text{Dec}(\mathbf{Q}_{\text{space}}, \mathbf{D}_p), \quad (5)$$

where $\mathbf{M}_p \in \{0, 1\}$ is a binary mask. K is randomly set to 1 or 2 during training.

2.3 Modality Alignment (MA)

The architecture of Fig. 1(b) described above expects point clouds representing the whole chamber as inputs to the encoder. For modalities such as MRI, data is

often only available for a few planar slices and reconstruction from such sparse data is a challenge. We address it using latent representations learned from complete surfaces that serve as priors. To this end, we propose our MA strategy.

Let $\{\mathbf{P}_{p,s}\}_{p=1}^5$ be points from slices corresponding to surfaces $\{\mathbf{S}_p\}_{p=1}^5$. As shown in Fig. 1(c), we introduce an additional slice encoder Enc_s and learnable slice query sets $\{\mathbf{L}_{p,s}\}_{p=1}^5$ that are used to encode the slice points

$$\mathbf{C}_{p,s} = \text{Enc}_s(\mathbf{L}_{p,s}, \mathbf{P}_{p,s}) \in \mathbb{R}^{D \times M}, p = 1, \dots, 5. \quad (6)$$

$\mathbf{C}_{p,s}$ can then be fed to the same frozen decoder as before and the weights of the encoder are learned along the other network weights as described below.

2.4 Training

The training of VecHeart occurs in two stages. First, we train our model using high-quality complete hearts and minimize the composite loss

$$\mathcal{L}_{stageI} = \mathcal{L}_{sdf} + \mathcal{L}_{inter}, \text{ with } \mathcal{L}_{sdf} = \frac{1}{P} \sum_p \mathbb{E}_q [\|s_p^{pred}(q) - s_p^{gt}(q)\|_1], \quad (7)$$

$$\text{and } \mathcal{L}_{inter} = \sum_{q \in \mathcal{C}} \sum_{p \in \mathcal{N}(q)} \tanh(\max(-\sigma \cdot SDF_p^{pred}(q), 0)),$$

where $\mathcal{C}$ is the contact regions among parts, $\mathcal{N}(q)$ are point q 's contacting parts, and σ a scalar. Minimizing $\mathcal{L}_{sdf}$ maximizes agreement with the ground-truth while minimizing $\mathcal{L}_{inter}$ discourages intersections.

Then the model is fine-tuned using SAX and LAX views synthesized as in prior works [19,33,34,35] by incorporating in-plane motion within SAX stacks and spatial misalignments between SAX and LAX views. These slice displacements are modeled as a Gaussian distribution $\mathcal{N}(0, \lambda^2 \mathbf{I})$, where λ is sampled from a uniform distribution $\mathcal{U}(0, 6)$ mm. Additionally, we randomly drop LAX slices at a probability $p = 0.5$, encouraging the model to predict complete chambers from SAX only slices points. We train the encoder of Eq. 6 by minimizing

$$\mathcal{L}_{stageII} = \mathcal{L}_{stageI} + \lambda_{la} \mathcal{L}_{la}, \quad \mathcal{L}_{la} = \frac{1}{P} \sum_{p=1}^P \|\mathbf{C}_{p,s} - \mathbf{C}_p\|_2, \quad (8)$$

where $\mathcal{L}_{la}$ denotes the latent alignment loss. This loss ensures the slice-based encoder inherits rich anatomical priors from the surface-based encoder. Hyperparameters λ_{la}, σ are set to 0.001, 50, respectively.

2.5 Extension to 3D+t Cardiac Shape Generation

We demonstrate the versatility of our framework by extending it to generate 3D+t four-chamber mesh sequences. We adopt an architecture similar to Craftsman [16] to jointly generate latent codes $\{C_p\}$ for all anatomical parts. Inspired by CardiacFlow [19], we incorporate *Periodic Gaussian Kernel (PGK) encoding*, *Learnable Initial Values*, and *Beta Sampling* to enable one-step generative flow. To handle sequences of different lengths, we apply linear interpolation before using PGK encoding. The model is trained using standard v -prediction loss [7].

Table 1: **Heart Shape Reconstruction from Complete/Missing Input.** Gray numbers denote the parts are missing from the input. CD in mm^2 , IoU in %. P-value: statistical tests on average CD.

Method	Myo		LV		RV		LA		RA		p-value (v.s. Ours)
	CD ↓	IoU ↑	CD ↓	IoU ↑	CD ↓	IoU ↑	CD ↓	IoU ↑	CD ↓	IoU ↑	
<i>Complete Input</i>											
ImHeart [36]	2.085	87.19	2.085	93.86	2.809	91.61	2.308	89.53	2.363	91.07	<0.01
PartSDF [29]	0.680	92.60	0.318	96.86	0.457	95.73	0.281	95.73	0.348	95.03	<0.01
SDF4CHD [14]	0.736	90.35	0.393	95.66	0.540	94.01	0.356	93.85	0.448	93.29	<0.01
CardiacFlow [19]	1.609	95.97	1.115	98.09	1.591	97.46	1.396	96.92	1.451	96.84	<0.01
Ours-Sep	0.637	97.78	0.306	99.10	0.397	98.98	0.235	99.06	0.247	99.02	0.0392
Ours	0.617	97.95	0.283	99.16	0.378	99.00	0.215	99.11	0.236	99.05	-
<i>Missing Part</i>											
ImHeart [36]	2.113	86.79	2.164	92.72	2.894	91.26	9.584	81.47	2.432	90.13	<0.01
	2.115	86.32	2.307	92.47	2.853	90.87	2.337	88.99	9.766	85.65	<0.01
PartSDF [29]	0.688	92.46	0.327	96.60	0.455	95.69	13.180	79.38	0.355	94.89	<0.01
	0.686	92.53	0.326	96.68	0.455	95.65	0.287	95.60	11.536	83.83	<0.01
SDF4CHD [14]	0.737	90.13	0.408	95.24	0.532	93.54	4.268	88.80	0.457	92.86	<0.01
	0.739	90.16	0.407	95.23	0.531	93.55	0.364	93.16	5.061	88.88	<0.01
CardiacFlow [19]	1.638	94.42	1.173	97.56	1.723	96.72	5.679	88.70	1.513	95.57	<0.01
	1.665	94.30	1.175	97.50	1.696	96.75	1.447	95.90	5.956	88.22	<0.01
	0.626	97.65	0.290	98.99	0.386	98.82	1.938	92.89	0.246	98.82	-
Ours	0.626	97.65	0.289	99.02	0.387	98.79	0.212	98.94	1.229	94.18	-
	0.640	97.20	0.301	98.81	0.403	98.52	5.979	87.03	4.336	89.48	-

3 Experiments

To train and test VecHeart, we curated a high-quality cardiac mesh dataset comprising 1,060 samples, generated from expert annotations in publicly available datasets, including LAA [38,8], WHS++ [41], and other benchmarks [23,28,13,30]. Each four-chamber mesh is rigidly registered to a WHS atlas [42], with a 6:2:2 train/val/test split. For reconstruction from real slices and 3D+t mesh sequence generation, we use a dataset comprises 835 CMR sequences constructed from three publicly available MRI datasets: *ACDC*, *M&Ms*, and *M&Ms-2* [3,4,20]. As in earlier studies [21,37,32,25], we register and fit a Statistic Shape Model [6,1] to segmentation masks predicted by a publicly available method [2] to generate ground-truth cardiac meshes. We only evaluate LV myocardium, LV and RV as this dataset mainly contain SAX stacks. We use standard L2 Chamfer Distance (CD) and IoU score at resolution 128 to measure 3D reconstruction quality.

In the VecHeart model of Fig. 1, the decoder comprises $L = 3$ interleaved attention layers. For the part-specific learnable queries $\mathbf{L}_p \in \mathbb{R}^{D \times M}$, the set size M is 256 and dimension D is 32. We train the model for 1000/500 epochs for stage I/II using Adam [12]. All experiments are conducted using one V100 GPU.

3.1 Heart Shape Modeling for Complete/Incomplete Data

In this section, we evaluate the shape modeling ability for complete/incomplete data. We use the same setup as conventional approaches [36,14] to reconstruct the test (unknown) samples. We compare against three multi-part auto-decoding implicit methods, PartSDF [29], ImHeart [36], SDF4CHD [14], as well as with CardiacFlow [19], an auto-encoding voxel-based method. For fairness, we train

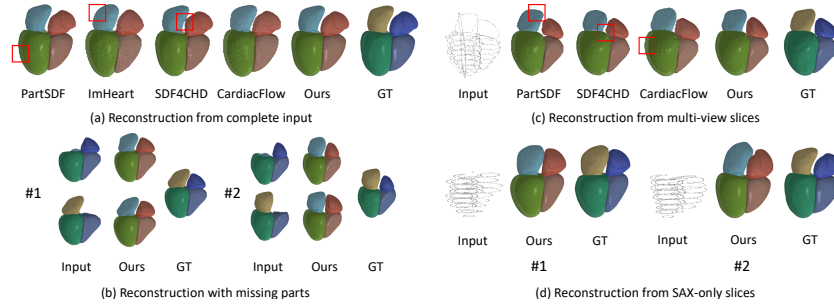


Fig. 2: Reconstruction results from diverse input configurations.

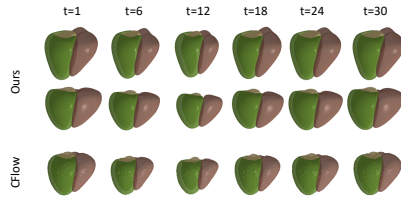


Fig. 3: 3D+t cardiac shape generation results comparison.

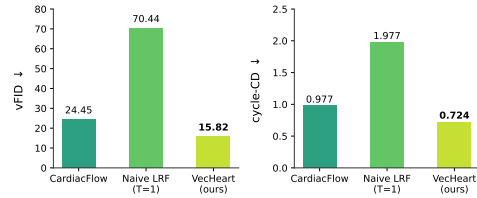


Fig. 4: 3D+t cardiac shape generation evaluated by vFID and cycle-CD.

CardiacFlow using the same ACM training strategy as us. For completeness, we also evaluate *Ours-Sep*, that is, fitting a separate VecSet model to each part.

We present qualitative results in Fig. 2(a)(b) and report quantitative ones in Tab. 1. The gray numbers are CD and IOU for parts which no data was provided and whose shape was therefore inferred from other parts. To obtain these numbers for auto-decoding methods, we compute the loss on available components and minimize it with respect to the latent vectors.

VecHeart outperforms the baselines in both complete and missing data scenarios. ImHeart’s lower accuracy likely stems from its occupancy-based representation. While implicit methods like PartSDF deliver respectable results, they are computationally expensive (e.g., 80s vs. our 0.7s feed-forward inference). While CardiacFlow yields the second-best performance on IoU, it is subject to resolution and discretization constraints. Compared to Ours-Sep, which fail to model inter-part relationships and cannot extend to missing-input scenarios, our full model shows the necessity of modeling inter-part correlation and structural priors. If only HPT is used without the ACM strategy, VecHeart cannot produce any meaningful results. These results validate our HPT and ACM designs.

3.2 Shape Reconstruction from Sparse Slices

We now turn to sparse and noisy CMR slices. CardiacFlow adopts Label Completion U-Net [26,33] as in the original work to recover full volumes from sparse

Table 2: **Shape Reconstruction from Sparse Slices.** Gray numbers denote results for parts for which no input data is supplied. CD in mm^2 , IoU in %.

Method	Myo		LV		RV		LA		RA		p-value (v.s. Ours)
	CD ↓	IoU ↑	CD ↓	IoU ↑	CD ↓	IoU ↑	CD ↓	IoU ↑	CD ↓	IoU ↑	
<i>Synthetic SAX, 2CH, 4CH</i>											
PartSDF [29]	2.434	80.74	2.663	86.35	2.927	86.28	3.034	85.50	8.565	79.97	<0.01
SDF4CHD [14]	1.810	82.39	1.369	91.80	2.058	89.68	2.881	88.85	6.686	82.82	<0.01
CardiacFlow [19]	2.411	86.42	1.728	91.12	2.458	88.43	2.734	87.03	4.997	83.01	<0.01
Ours-SliceOnly	1.430	85.02	1.091	91.10	1.456	89.86	1.481	89.26	2.977	86.32	<0.01
Ours	1.130	88.39	0.732	91.92	1.114	90.25	0.854	90.80	1.571	89.66	-
<i>Synthetic SAX</i>											
PartSDF [29]	4.066	73.37	4.452	83.07	3.956	84.46	37.970	65.36	47.784	61.34	<0.01
SDF4CHD [14]	2.676	79.05	2.445	89.83	2.624	88.62	19.822	69.56	18.568	68.64	<0.01
CardiacFlow [19]	2.591	83.58	2.606	88.35	2.709	86.13	14.207	71.33	11.898	72.95	<0.01
Ours-SliceOnly	2.172	82.02	1.834	88.02	2.427	84.12	14.625	73.73	11.310	74.99	0.0141
Ours	1.816	84.31	1.476	89.26	2.013	87.34	9.224	76.40	7.909	78.72	-
<i>Real SAX (Zero-shot)</i>											
PartSDF [29]	4.721	74.24	4.212	77.50	6.482	70.42	-	-	-	-	<0.01
SDF4CHD [14]	3.653	77.25	3.281	80.64	5.233	73.94	-	-	-	-	<0.01
CardiacFlow [19]	3.168	79.17	2.714	81.05	4.137	76.34	-	-	-	-	<0.01
Ours	2.471	79.87	2.296	81.44	3.621	78.10	-	-	-	-	-
Ours-Finetune	2.108	83.59	1.757	84.57	3.152	82.19	-	-	-	-	-

slices. We experimented with both synthetic and real data. In the synthetic case, following CardiacFlow, we synthesized 10 different multi-view 2D slices at random motion level λ using the displacement strategy of Sec. 2.4 for each test sample. For real experiments, we use the SAX contours of the CMR dataset.

As shown in Fig. 2(c)(d), VecHeart produces plausible four-chamber reconstructions from sparse slices, underscoring the practical utility and robustness of our approach. We report quantitative results for synthetic data in the top rows of Tab. 2 and for real data in the bottom. As before, the gray numbers denote results for structures which is absent from the input. Our model consistently outperforms the baselines in all configurations. We also compare against *Ours-SliceOnly*, which uses the same slice encoder and decoder as the full model but without Stage I training. It works less well, thus confirming the importance of the proposed Modality Alignment strategy.

Note that the results on real-world CMR data were obtained by training on synthetic data and without re-training. They can be further improved by fine-tuning on real data, as shown by the numbers corresponding to *Ours-Finetune*.

3.3 3D+t Cardiac Shape Generation

Now we evaluate VecHeart’s capability to generate 3D+t cardiac sequences. As in CardiacFlow, we employ the Volume Fréchet Inception Distance (vFID) and Cycle-CD metrics to assess the diversity and periodic consistency of the generated sequences [10]. Fig. 3, 4 shows we achieve superior qualitative and quantitative performance, producing diverse and high-fidelity sequences. As a contrast, CardiacFlow suffers from discretization artifacts. The results imply VecHeart’s potential to be integrated with clinical metadata for more controllable and interpretable generation and become a general purpose data augementer.

4 Conclusion

We present VecHeart, a unified framework for holistic reconstruction and generation of four-chamber cardiac anatomy. It delivers high-fidelity reconstruction across a wide spectrum of input densities ranging from dense point clouds to sparse and noisy slices. Furthermore, integrating a flow-matching scheme enables the VecHeart to generate realistic 3D+t mesh sequences.

VecHeart’s flexibility makes it possible to use partially labeled data, giving it the potential for cross-modality data fusion and robust clinical diagnostics in heterogeneous real-world scenarios. Moving forward, we aim to use VecHeart to reconstruct cardiac anatomy from more modalities, including raw MRI and US images acquired over time, within a unified framework.

Acknowledgments. This work was supported by a SNSF grant.

Disclosure of Interests. All authors have no conflict of interest.

References

1. Bai, W., Shi, W., de Marvao, A. et al.: A bi-ventricular cardiac atlas built from 1000+ high resolution mr images of healthy subjects and an analysis of shape and motion. *MIA* (2015) 6
2. Bai, W., Sinclair, M., Tarroni, G. et al.: Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of cardiovascular magnetic resonance* (2018) 6
3. Bernard, O., Lalande, A., Zotti, C. et al.: Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? *TMI* (2018) 6
4. Campello, V.M., Gkontra, P., Izquierdo, C. et al.: Multi-centre, multi-vendor and multi-disease cardiac segmentation: the m&ms challenge. *TMI* (2021) 6
5. Deng, Y., Yang, J., Tong, X.: Deformed implicit field: Modeling 3d shapes with learned dense correspondence. In: *CVPR* (2021) 1
6. Duan, J., Bello, G., Schlemper, J. et al.: Automatic 3d bi-ventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach. *TMI* (2019) 6
7. Esser, P., Kulal, S., Blattmann, A. et al.: Scaling rectified flow transformers for high-resolution image synthesis. In: *ICML* (2024) 2, 5
8. Hansen, B., Pedersen, J., Kofoed, K.F. et al.: A public cardiac ct dataset featuring the left atrial appendage. In: *STACOM* (2025) 6
9. He, K., Chen, X., Xie, S. et al.: Masked Autoencoders Are Scalable Vision Learners. In: *CVPR*. pp. 16000–16009 (2022) 4
10. Heusel, M., Ramsauer, H., Unterthiner, T. et al.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. *NIPS* (2017) 8
11. Jaegle, A., Gimeno, F., Brock, A. et al.: Perceiver: General perception with iterative attention. In: *ICML* (2021) 3
12. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization. In: *arXiv Preprint* (2014) 6

13. Kirişli, H., Schaap, M., Metz, C. et al.: Standardized evaluation framework for evaluating coronary artery stenosis detection, stenosis quantification and lumen segmentation algorithms in computed tomography angiography. *MIA* (2013) [6](#)
14. Kong, F., Stocker, S., Choi, P.S. et al.: Sdf4chd: Generative modeling of cardiac anatomies with congenital heart defects. *MIA* (2024) [1](#), [6](#), [8](#)
15. Li, L., Camps, J., Rodriguez, B. et al.: Solving the inverse problem of electrocardiography for cardiac digital twins: A survey. *IEEE Reviews in Biomedical Engineering* (2024) [1](#)
16. Li, W., Liu, J., Yan, H. et al.: Craftsman3d: High-fidelity mesh generation with 3d native generation and interactive geometry refiner. *CVPR* (2025) [5](#)
17. Lipman, Y., Chen, R.T., Ben-Hamu, H. et al.: Flow Matching for Generative Modeling. In: *arXiv Preprint* (2022) [2](#)
18. Liu, Y., Planche, B., Zheng, M. et al.: Implicit modeling of non-rigid objects with cross-category signals. In: *AAAI* (2024) [1](#)
19. Ma, Q., Meng, Q., Qiao, M. et al.: Cardiacflow: 3d+ t four-chamber cardiac shape completion and generation via flow matching. In: *MICCAI* (2025) [5](#), [6](#), [8](#)
20. Martín-Isla, C., Campello, V.M., Izquierdo, C. et al.: Deep learning segmentation of the right ventricle in cardiac mri: the m&ms challenge. *JBHI* (2023) [6](#)
21. Meng, Q., Bai, W., O'Regan, D.P. et al.: Deepmesh: Mesh-Based Cardiac Motion Tracking Using Deep Learning. *TMI* (2023) [6](#)
22. Mescheder, L., Oechsle, M., Niemeyer, M. et al.: Occupancy Networks: Learning 3D Reconstruction in Function Space. In: *CVPR*. pp. 4460–4470 (2019) [1](#)
23. Metz, C., Schaap, M., Weustink, A. et al.: Coronary centerline extraction from ct coronary angiography images using a minimum cost path approach. *Medical physics* (2009) [6](#)
24. Park, J.J., Florence, P., Straub, J. et al.: Deep sdf: Learning Continuous Signed Distance Functions for Shape Representation. In: *CVPR* (2019) [1](#)
25. Qiao, M., McGurk, K.A., Wang, S. et al.: A personalized time-resolved 3d mesh generative model for unveiling normal heart dynamics. *Nature Machine Intelligence* pp. 1–12 (2025) [1](#), [6](#)
26. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *MICCAI*. pp. 234–241 (2015) [7](#)
27. Sanz, J., Fayad, Z.A.: Imaging of atherosclerotic cardiovascular disease. *Nature* **451**(7181), 953–957 (2008) [1](#)
28. Schaap, M., Metz, C.T., van Walsum, T. et al.: Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. *MIA* (2009) [6](#)
29. Talabot, N., Clerc, O., Demirtas, A. et al.: PartSDF: Part-Based Implicit Neural Representation for Composite 3D Shape Parametrization and Optimization. *TMLR* (2026) [1](#), [6](#), [8](#)
30. Tobon-Gomez, C., Geers, A.J., Peters, J. et al.: Benchmark for algorithms segmenting the left atrium from 3d ct and mri datasets. *TMI* (2015) [6](#)
31. Vaswani, A.: Attention is all you need. In: *NIPS* (2017) [3](#)
32. Xiao, J., Zheng, W., Wang, W. et al.: Slice2mesh: 3d surface reconstruction from sparse slices of images for the left ventricle. *TMI* (2024) [6](#)
33. Xu, H., Muffoletto, M., Niederer, S.A. et al.: Whole heart 3d shape reconstruction from sparse views: leveraging cardiac computed tomography for cardiovascular magnetic resonance. In: *FIMH* (2023) [5](#), [7](#)
34. Xu, H., Williams, S.E., Williams, M.C. et al.: Deep learning estimation of three-dimensional left atrial shape from two-chamber and four-chamber cardiac long axis views. *European Heart Journal-Cardiovascular Imaging* (2023) [5](#)

35. Xu, Y., Xu, H., Sinclair, M. et al.: Improved 3d whole heart geometry from sparse cmr slices. In: STACOM (2024) 5
36. Yang, J., Sedykh, E., Adhinarta, J.K. et al.: Generating anatomically accurate heart structures via neural implicit fields. In: MICCAI (2024) 1, 6
37. Yuan, X., Liu, C., Wang, Y.: 4d myocardium reconstruction with decoupled motion and shape model. In: ICCV (2023) 6
38. Zeng, A., Wu, C., Lin, G. et al.: Imagecas: A large-scale dataset and benchmark for coronary artery segmentation based on computed tomography angiography images. CMIG (2023) 6
39. Zhang, B., Tang, J., Nießner, M. et al.: 3DShape2VecSet: A 3D Shape Representation for Neural Fields and Generative Diffusion Models. In: SIGGRAPH (2023) 2, 3
40. Zheng, Z., Yu, T., Dai, Q. et al.: Deep implicit templates for 3d shape representation. In: CVPR (2021) 1
41. Zhuang, X., Li, L., Payer, C. et al.: Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge. MIA (2019) 6
42. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of mri. MIA (2016) 6