CardiacFlow: 3D+t Four-Chamber Cardiac Shape Completion and Generation via Flow Matching

Qiang Ma[1], Qingjie Meng[2,3], Mengyun Qiao[1,4], Paul M. Matthews[1,5,6], Declan P. O'Regan[7], and Wenjia Bai[1,3,4]

[1] Department of Brain Sciences, Imperial College London, London, UK
[2] School of Computer Science, University of Birmingham, Birmingham, UK
[3] Department of Computing, Imperial College London, London, UK
[4] Data Science Institute, Imperial College London, London, UK
[5] UK Dementia Research Institute, Imperial College London, London, UK
[6] Rosalind Franklin Institute, Harwell Science and Innovation Campus, Didcot, UK
[7] MRC Laboratory of Medical Sciences, Imperial College London, London, UK
q.ma20@imperial.ac.uk

Abstract. Learning 3D+t shape completion and generation from multi-view cardiac magnetic resonance (CMR) images requires a large amount of high-resolution 3D whole-heart segmentations (WHS) to capture shape priors. In this work, we leverage flow matching techniques to learn deep generative flows for augmentation, completion, and generation of 3D+t shapes of four cardiac chambers represented implicitly by segmentations. Firstly, we introduce a latent rectified flow to generate 3D cardiac shapes for data augmentation, learned from a limited number of 3D WHS data. Then, a label completion network is trained on both real and synthetic data to reconstruct 3D+t shapes from sparse multi-view CMR segmentations. Lastly, we propose CardiacFlow, a novel one-step generative flow model for efficient 3D+t four-chamber cardiac shape generation, conditioned on the periodic Gaussian kernel encoding of time frames. The experiments on the WHS datasets demonstrate that flow-based data augmentation reduces geometric errors by 16% in 3D shape completion. The evaluation on the UK Biobank dataset validates that CardiacFlow achieves superior generation quality and periodic consistency compared to existing baselines. The code of CardiacFlow is released publicly at https://github.com/m-qiang/CardiacFlow.

Keywords: Cardiac imaging · Shape modelling · Flow matching.

1 Introduction

As the current gold standard imaging modality for assessing cardiac structure and function, cine cardiac magnetic resonance (CMR) imaging facilitates the quantitative evaluation of cardiac phenotypes, such as chamber volumes and ejection fractions [24]. The reconstruction of 3D+t cardiac shape models [4,6,7,20,35] from CMR images plays an essential role in understanding regional patterns of

cardiac diseases [2] and conducting electrophysiological and biomechanical studies of the heart [23,27,29,32]. However, it is a challenging inverse problem to reconstruct high-resolution four-chamber shapes of the heart from sparse multi-view CMR sequences. The four cardiac chambers, including the left ventricle (LV), right ventricle (RV), left atrium (LA), and right atrium (RA), are delineated by stacks of 2D image slices acquired from multiple view planes. Although there are multiple slices covering the ventricles, the atria are only covered by one or two slices. The challenges also come from potential motion artefacts and slice misalignment caused by respiratory motion during image acquisition.

Previous works demonstrated the feasibility of leveraging shape priors from 3D whole-heart segmentations (WHS) of CT scans [11] to learn 3D+t cardiac shape completion in CMR segmentations [22,36,37,38]. However, this requires a large amount of 3D WHS data, which are not always available due to data regulation issues and the expensive cost of image acquisition and annotation. Existing public WHS datasets [14,21,33,39] only contain dozens of 3D segmentations, which highlight the necessity of data augmentation. Deep generative models [9,12,13] offer a potential solution to enrich the quantity and variety of 3D cardiac shapes. While deep generative models have been utilised to characterise 3D shape distribution [15] or spatio-temporal motion patterns of the heart [25,31], their applications to augmenting, completing and generating 3D+t four-chamber cardiac shapes are still less explored.

Recently, diffusion models have achieved great success in data generation [10,30]. For efficient training and sampling, latent diffusion model [26] operates in the latent space and thus is naturally suitable for 3D medical images. Compared to diffusion models that require considerable training data and thousands of stochastic sampling steps [10,26,34], latest flow matching approaches [5,18,19] are data- and time-efficient by learning a continuous normalising flow between two distributions. In particular, a rectified flow technique [5,17,19] is developed to learn a straight optimal transport path from noise to data, enabling it to capture the target distribution and generate data samples in a few steps.

In this work, we propose a novel solution to learn 3D+t four-chamber cardiac shape completion and generation via flow matching, in which the shapes are represented implicitly by segmentation maps. The main contributions of this work are listed as follows:

- We introduce a latent rectified flow to generate synthetic 3D shapes of four cardiac chambers, learned from a limited number of 3D WHS data via flow matching. Such data augmentation significantly improves the performance of 3D shape completion.
- A label completion network is trained on both real and synthetic 3D shapes, and further applied to reconstruct accurate 3D+t cardiac shapes from multi-view CMR segmentations on the UK Biobank dataset for a large population.
- We propose *CardiacFlow*, an efficient one-step generative flow model for 3D+t four-chamber cardiac shape generation. Incorporating periodic Gaussian kernel encoding of time frames and learnable initial values, CardiacFlow achieves state-of-the-art generation quality and periodic consistency.
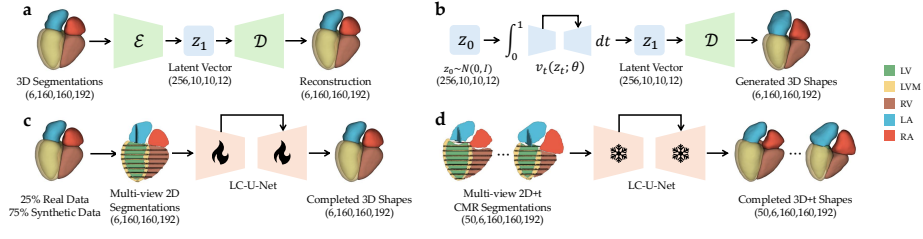
Fig. 1. The procedure of 3D+t cardiac shape completion. (a) An autoencoder is trained to extract latent vectors from 3D shapes. (b) A LRF is learned to generate 3D shapes for data augmentation. (c) A LC-U-Net is trained on both real and synthetic data for 3D shape completion. (d) The trained LC-U-Net is applied to 3D+t shape completion.

2 Methods

2.1 Latent Rectified Flow for Data Augmentation

Learning to complete four-chamber cardiac shapes from multi-view segmentations requires a large amount of training data to capture the shape variation. Given a limited number of 3D WHS data [14,21,33,39], we introduce *flow matching* (FM) [18] to generate 3D cardiac shapes for data augmentation.

Flow Matching. Given a noise distribution $p(x_0) = \mathcal{N}(x_0; 0, I)$ and data distribution $q(x_1)$, FM models a generative process by learning a *continuous normalising flow* from noise $x_0 \sim p(x_0)$ to data $x_1 \sim q(x_1)$ defined by an ODE:

$$dx_t = v_t(x_t; \theta)dt, \ x_0 \sim p(x_0), \ t \in [0, 1], \tag{1}$$

where v_t is a *time-varying vector field* parameterised by a neural network with learnable parameters θ. It defines a probability path p_t with $x_t \sim p_t(x)$. Ideally, the path p_t should flow from noise to data such that $p_0(x) = p(x_0)$ and $p_1(x) = q(x_1)$. To learn the flow model (1), a conditional FM loss function is defined as:

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{t, q(x_1), p_t(x|x_1)} \left[\|v_t(x_t; \theta) - u_t(x_t; x_1)\|^2 \right]. \tag{2}$$

Such a loss trains the flow (1) to match a conditional flow defined by a vector field u_t conditioned on data x_1, where u_t is usually constructed as an optimal transport path $u_t(x_t; x_1) = (x_1 - x_t)/(1 - t)$, *i.e.*, the shortest straight path from noise x_0 to data x_1 [18,19]. In this case, the flow (1) is also called a *rectified flow*. After training, new data samples can be generated by integrating the ODE (1).

Latent Rectified Flow. Instead of learning high-dimensional distributions, FM can be performed efficiently in a latent space [5]. As shown in Fig. 1-a,b, we train an autoencoder $\hat{x} = \mathcal{D}(\mathcal{E}(x))$ with encoder $\mathcal{E}$ and decoder $\mathcal{D}$ to extract latent vectors $z_1 = \mathcal{E}(x)$ from 3D cardiac segmentations $x \sim q(x)$. Then, a *latent rectified flow* (LRF) is defined as $dz_t = v_t(z_t; \theta)dt$, where the initial value $z_0 \sim \mathcal{N}(0, I)$

has the same dimensionality as the latent vector z_1, and v_t is parameterised by a 3D flow U-Net [10,26,28] with *adaptive instance normalisation* layers [12].

To train the LRF, we randomly sample $t \in [0, 1]$, $z_0 \sim \mathcal{N}(0, I)$, and $z_1 = \mathcal{E}(x)$ with $x \sim q(x)$ for each iteration. In order to match the LRF to a straight optimal transport path, z_t is approximated by a linear interpolation $z_t = (1 - t)z_0 + tz_1$ between z_0 and z_1. Hence, the latent FM loss is reformulated as [5,19]:

$$\mathcal{L}_{\text{LFM}}(\theta) = \mathbb{E}_{t,q(x),p_t(z|z_1)} \left[\|v_t((1 - t)z_0 + tz_1; \theta) - (z_1 - z_0)\|^2 \right]. \qquad (3)$$

After training, we sample z_0 and integrate the LRF by forward Euler method with $T{=}100$ steps, in which the flow U-Net v_t is evaluated repeatedly, to generate a latent vector $\hat{z}_1$. The 3D segmentation is reconstructed by $\hat{x} = \mathcal{D}(\hat{z}_1)$ as shown in Fig. 1-b. Note that if the flow trajectory is ideally straight, the forward Euler method can provide exact ODE solution within one step [19]. In spite of learning from a limited number of WHS data [14,21,33,39], the LRF is data efficient and able to generate diverse 3D cardiac shapes for effective data augmentation.

2.2   3D+t Four-Chamber Cardiac Shape Completion

The data augmentation performed by LRF enables us to capture accurate shape distribution for 3D+t shape completion. Given augmented 3D cardiac segmentations, a label completion U-Net (LC-U-Net) [28,36,37] is trained to reconstruct 3D cardiac segmentations from multi-view 2D segmentations (Fig. 1-c). The input multi-view 2D segmentations, including short-axis (SAX) and long-axis (LAX) views, are synthesised following [36,37,38], incorporating in-plane motion between SAX slices and misalignment between SAX and LAX views. The displacements of the slices are simulated as a Gaussian distribution $\mathcal{N}(0, \lambda^2 I)$, where λ is the corruption level following a uniform distribution between 0 and 4 mm. For each training epoch, the training data consist of 25% real data collected from the WHS datasets [14,21,33,39] and 75% synthetic 3D shapes generated by the LRF. After training, as illustrated in Fig. 1-d, we apply the LC-U-Net to complete four-chamber cardiac shapes across $M{=}50$ time frames of multi-view CMR segmentations on the UK Biobank (UKB) database [24], curating a 3D+t four-chamber cardiac shape dataset for a large population.

2.3   CardiacFlow for 3D+t Cardiac Shape Generation

Based on the curated 3D+t cardiac shape dataset, we develop *CardiacFlow*, an efficient one-step generative flow that extends LRF to generate spatio-temporal shapes of four cardiac chambers. Given completed 3D+t segmentations $x \sim q(x)$ from the UKB dataset, we train a 3D autoencoder $\hat{x}_\tau = \mathcal{D}(\mathcal{E}(x_\tau))$ to learn a latent vector $z_{1,\tau} = \mathcal{E}(x_\tau)$ from 3D shape x_τ for time frame $\tau \in \{1, ..., M\}$. As shown in Fig. 2, CardiacFlow defines a LRF from a learnable frame-conditioned initial value $z_{0,\tau}$ to the latent vector $z_{1,\tau}$ for each frame τ, modelled by an ODE:

$$dz_{t,\tau} = v_t\left(z_{t,\tau}; \theta\right) dt, \ \ z_{0,\tau} = f_\Theta\left(\epsilon_x, \mathcal{K}_\sigma(\tau)\right), \ \ t \in [0, 1], \qquad (4)$$
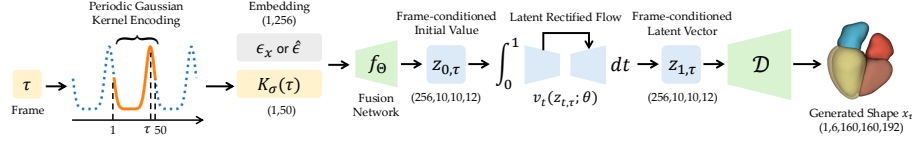
Fig. 2. The architecture of CardiacFlow. Each frame τ is encoded by a PGK $\mathcal{K}_\sigma(\tau)$ and fused with a learnable embedding ϵ_x (train) or a sampled embedding $\hat{\epsilon}$ (test). The fusion network predicts a frame-conditioned initial value $z_{0,\tau}$ for the LRF to generate a latent vector $z_{1,\tau}$, which is decoded to a 3D shape $x_\tau = \mathcal{D}(z_{1,\tau})$.

where v_t is parameterised by a 3D U-Net in the latent space, f_Θ is a fusion network, ϵ_x is a learnable embedding for each training data x, and $\mathcal{K}_\sigma(\tau)$ is a *periodic Gaussian kernel* (PGK) *encoding* of time frame τ.

Periodic Gaussian Kernel Encoding. For cardiac 3D+t shape generation, the heart should have a consistent shape at the start ($\tau=1$) and end ($\tau=M$) of a cardiac cycle. However, existing generative models [25,31] cannot preserve such periodic consistency. Instead of using scalar or one-hot conditioning variable, CardiacFlow employs a PGK $\mathcal{K}_\sigma : \mathbb{R} \to \mathbb{R}^M$ to encode time frame τ, based on the periodicity of heartbeat. The m-th element of the vector $\mathcal{K}_\sigma(\tau)$ is defined by a Gaussian kernel $[\mathcal{K}_\sigma(\tau)]_m \triangleq \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{d(m,\tau)^2}{2\sigma^2})$, where $d(m,\tau) \triangleq |\mathrm{mod}(m - \tau + M/2, M) - M/2|$ is a distance metric, which is symmetric $d(m,\tau) = d(\tau,m)$ and periodic $d(m+M,\tau) = d(m,\tau)$ with period M. As depicted in Fig. 2, the elements of $\mathcal{K}_\sigma(\tau)$ follow a truncated and periodic Gaussian distribution, where $[\mathcal{K}_\sigma(\tau)]_m$ achieves maximum when $m = \tau$. For each frame τ, the PGK encoding allows CardiacFlow to capture the information of neighbouring frames within period M and guarantees the periodic consistency in a cardiac cycle.

Learnable Initial Value. CardiacFlow aims to generate each frame of the 3D+t cardiac shapes in one step. One key factor is to rectify and straighten the flow path, such that one step of numerical integration can provide accurate ODE solution [18,19]. While FM samples random coupling of noise and data for training [18,19], inspired by [15,31], we assign a low-dimensional learnable embedding ϵ_x to each training data $x \sim q(x)$. A fusion network f_Θ with multiple MLP layers is trained, such that the learnable embedding ϵ_x is fused with the frame encoding $\mathcal{K}_\sigma(\tau)$ and upsampled to the same size as the latent vector $z_{1,\tau}$. It provides a learnable frame-conditioned initial value $z_{0,\tau} = f_\Theta(\epsilon_x, \mathcal{K}_\sigma(\tau))$ for the ODE (4), and therefore CardiacFlow can optimise the initial value and flow trajectory simultaneously to learn a straight optimal transport path for one-step generation. For data generation, the learnable embedding ϵ_x is replaced by $\hat{\epsilon} \sim \mathcal{N}(\mu_\epsilon, \Sigma_\epsilon)$, an empirical Gaussian distribution defined by $\{\epsilon_x\}_{x \sim q(x)}$.

Beta Sampling. Another factor for one-step generation is the sampling strategy [5,16,17]. FM typically samples the time steps $t \in [0,1]$ uniformly during training

Algorithm 1 Training	**Algorithm 2** Generation
1: **input:** 3D+t shapes $q(x)$, embeddings ϵ_x	1: $\hat{\epsilon} \sim \mathcal{N}(\mu_\epsilon, \Sigma_\epsilon)$
2: **repeat**	2: **for** $\tau = 1, ..., M$ **do**
3: $x \sim q(x),\ \tau \sim \text{Uniform}(\{1, ..., M\})$	3: $\hat{z}_{0,\tau} = f_\Theta\left(\hat{\epsilon}, \mathcal{K}_\sigma(\tau)\right)$
4: $z_{1,\tau} = \mathcal{E}(x_\tau),\ z_{0,\tau} = f_\Theta\left(\epsilon_x, \mathcal{K}_\sigma(\tau)\right)$	4: $\hat{z}_{1,\tau} = \hat{z}_{0,\tau} + \int_0^1 v_t(\hat{z}_{t,\tau}; \theta)dt$
5: $t \sim \text{Beta}(a, b),\ z_{t,\tau} = (1-t)z_{0,\tau} + tz_{1,\tau}$	5: $\hat{x}_\tau = \mathcal{D}(\hat{z}_{1,\tau})$
6: $\mathcal{L}(\theta, \Theta, \epsilon_x) = \|v_t(z_{t,\tau}; \theta) - (z_{1,\tau} - z_{0,\tau})\|^2$	6: **end for**
7: update θ, Θ, ϵ_x	7: **return** $\hat{x} = \{\hat{x}_\tau\}$

Table 1. The results of 3D shape completion. A LC-U-Net trained without synthetic data is compared to LC-U-Nets with data augmentation using VAE or LRF. The bolded results are significantly better than other methods (t-test, $p < 0.05$).

	Method	LV	LVM	RV	LA	RA
HD95↓ (mm)	LC-U-Net	2.084±0.630	1.995±0.561	2.435±1.007	3.120±1.010	4.385±1.919
	LC-U-Net (VAE)	1.882±0.606	1.870±0.601	2.155±1.161	2.948±0.986	3.999±1.987
	LC-U-Net (LRF)	**1.769±0.521**	**1.693±0.492**	**2.024±0.982**	**2.675±0.884**	**3.581±1.697**
DSC↑	LC-U-Net	0.955±0.017	0.900±0.027	0.944±0.023	0.929±0.021	0.905±0.039
	LC-U-Net (VAE)	0.960±0.014	0.909±0.024	0.952±0.022	0.934±0.021	0.911±0.041
	LC-U-Net (LRF)	**0.963±0.012**	**0.917±0.021**	**0.954±0.020**	**0.939±0.018**	**0.918±0.036**

[18,19], providing very few samples at $t=0$ for one-step generation from initial values [5,16,17]. To alleviate this issue, CardiacFlow samples time t from a *Beta distribution*, $\text{Beta}(a, b)$, defined on $[0, 1]$ with $a=0.1$ and $b=2.0$ [16]. The density function monotonically decreases so that more samples are drawn near $t=0$.

The training and generation procedures of CardiacFlow are summarised in Algorithms 1 and 2. The flow ODE in Algorithm 2 is solved by one step of forward Euler, *i.e.*, $\hat{z}_{1,\tau} = \hat{z}_{0,\tau} + v_0(\hat{z}_{0,\tau}; \theta)$. Note that the Beta sampling used in training also allows few-step generation, which enhances the quality of generated shapes by more integration steps. The 3D shape $\hat{x}_\tau = \mathcal{D}(\hat{z}_{1,\tau})$ at each frame τ is reconstructed by decoding the generated latent vector. The 3D+t cardiac shapes $\hat{x}$ are generated by iterating over all time frames $\tau = 1, ..., M$.

3   Experiments

Dataset. For 3D cardiac shape augmentation and completion, we collect WHS data from multiple public datasets including WHS++ [39] and other challenges [14,21,33], curating 160 3D segmentation maps of 137 CT and 23 MR images after data cleaning. The cardiac four-chamber shapes are represented by segmentation maps without the need of original CT or MR images. All segmentation maps are rigidly aligned to a WHS atlas [40] and clipped to the size of $160 \times 160 \times 192$. The dataset is randomly split by the ratio of $6/1/3$ for training, validation and test.

The 3D+t cardiac shape completion and generation are performed on the segmentation maps of 1,000 UKB CMR scans [24], which are split by the ratio
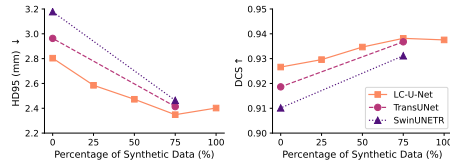
Fig. 3. Shape completion with different ratio of synthetic data and neural networks.
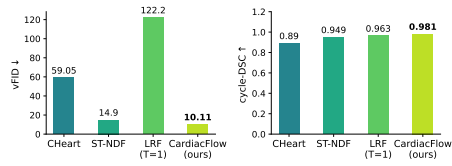
Fig. 4. 3D+t cardiac shape generation results evaluated by vFID and cycle-DSC.

of 6/1/3 as well. The 2D+t segmentations in SAX, two- and four-chamber LAX views are generated using a publicly available method [1] with manual quality control, including the classes of LV, LV myocardium (LVM), RV, LA and RA. All 2D slices are mapped into 3D space and rigidly aligned to the WHS atlas [40]. All experiments are conducted on a NVIDIA 3080 GPU with 12GB memory.

3D Shape Completion. We examine the effects of data augmentation via LRF on 3D cardiac shape completion on the WHS datasets [14,21,33,39]. We train a LRF for 1,000 epochs and it takes 0.266s to generate a 3D segmentation with T=100 steps. A VAE [13] is trained for comparison. We also train a latent diffusion model [26], which however, fails to converge with limited number of training data. Then, we train LC-U-Nets for 600 epochs with data augmentation, which includes 25% real data and 75% synthetic data generated online by VAE [13] or LRF for each epoch, compared to a LC-U-Net trained without synthetic data. For each of 48 3D segmentations in the test set, we simulate 10 different multi-view 2D segmentations at random motion level λ, resulting in 480 test samples for evaluation. The Dice coefficient (DSC) and 95th percentile of Hausdorff distance (HD95) are measured and reported in Table 1. The statistical significance is examined by paired t-test. Compared to the LC-U-Net trained without data augmentation, the LRF enhances the data diversity and reduces the geometric errors significantly by 16.2% in terms of HD95. The higher accuracy in the 3D shape completion task also demonstrates the superior generation quality of the LRF compared to the VAE model.

We further investigate how the ratio of synthetic data and neural network architectures can affect shape completion. We train LC-U-Net, TransUNet [3], and SwinUNETR [8] for 3D cardiac shape completion with synthetic data generated by LRF. Fig. 3 shows that the accuracy of LC-U-Net increases consistently with more synthetic data, until it reaches the best result when 75% synthetic data are used. LC-U-Net performs better than Transformer-based models [3,8] since it is data efficient with inductive bias. Regardless of the network architecture, both HD95 and DSC are improved remarkably by LRF-based data augmentation.

For qualitative results, Fig. 5-a shows a 3D shape completed by LC-U-Net augmented by LRF. Fig. 5-c visualises 3D cardiac shapes generated by LRF. For 3D+t shape completion, we apply LC-U-Net (LRF) to 1,000 UKB data. An example of completed 3D+t cardiac sequence is shown in Fig. 5-b.
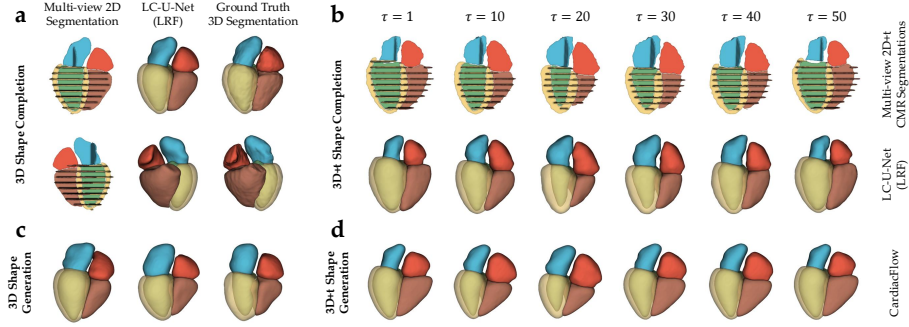
Fig. 5. Qualitative results for (a) 3D shape completion, (b) 3D+t shape completion, (c) 3D shape generation, and (d) 3D+t shape generation.
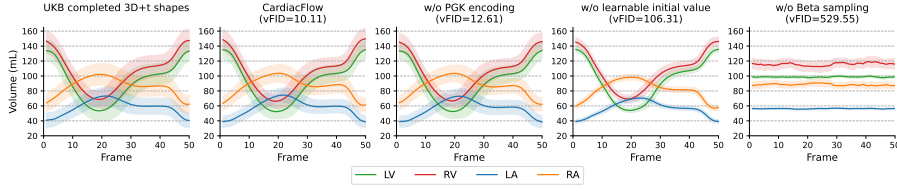

Fig. 6. Four-chamber volumes of synthetic 3D+t cardiac shapes for ablation studies.

3D+t Shape Generation. We train CardiacFlow on the UKB dataset with completed 3D+t cardiac shapes for 100 epochs using $\sigma=1.5$ for PGK encoding. CardiacFlow is compared to CHeart [25], spatio-temporal neural distance field (ST-NDF) [31], and a vanilla LRF with one-step generation ($T=1$). For each baseline method, we randomly generate 1,000 samples for evaluation. The quality of generated 3D+t shapes is measured by the volume Fréchet inception distance (vFID) [9], which is defined as $d_F = \|\mu_1-\mu_2\|^2+\mathrm{tr}(\Sigma_1+\Sigma_2-2(\Sigma_1\Sigma_2)^{\frac{1}{2}})$ between generated 3D+t shapes $\mathcal{N}(\mu_1, \Sigma_1)$ and UKB test set $\mathcal{N}(\mu_2, \Sigma_2)$, where $\mu \in \mathbb{R}^{200}$ is the mean cardiac four-chamber volumes for $M=50$ frames. The periodic consistency is evaluated by cycle-DSC, *i.e.*, the DSC between the start ($\tau=1$) and end ($\tau=50$) frame of a cardiac cycle. The vFID and cycle-DSC are reported in Fig. 4, which shows CardiacFlow achieves superior generation quality and periodic consistency than all baselines. CardiacFlow enables high-quality one-step generation, whereas vanilla LRF fails to capture accurate shape distribution in one step. For runtime, CardiacFlow takes 1.611s to generate an entire 3D+t sequence, which is similar to CHeart (1.467s) and ST-NDF (1.472s). An example of 3D+t shapes generated by CardiacFlow is visualised in Fig. 5-d.

We conduct ablation studies on three components of CardiacFlow: (i) The PGK frame encoding $\mathcal{K}_\sigma(\tau)$ is substituted by a scalar conditioning variable τ. (ii) The learnable embedding ϵ_x is replaced by a Gaussian noise so that the initial value is no longer learnable. (iii) During training, the time t is sampled uniformly instead of Beta sampling. The vFID score and cardiac four-chamber

volumes averaged over 1,000 generated 3D+t shapes are presented in Fig. 6. It indicates that CardiacFlow learns accurate volume distribution from the UKB dataset. The removal of PGK encoding reduces the cycle-DSC from 0.981 to 0.959. Without learnable initial values or beta sampling, CardiacFlow fails to capture the 3D+t shape variation or motion pattern via one-step generation.

4   Conclusion

In this work, we introduce flow matching for 3D+t shape completion and generation of four cardiac chambers. A latent rectified flow is employed to generate 3D cardiac shapes for data augmentation, leading to significant improvement on 3D shape completion. A novel CardiacFlow framework is developed to learn efficient 3D+t cardiac shape generation from the UK Biobank dataset. Instead of learning implicit segmentations, in future work, we will explicitly reconstruct 3D+t meshes for cardiac four chambers. In addition, CardiacFlow could be further extended to generate 3D+t cardiac shapes conditioned on demographic and pathological information, as well as model spatio-temporal distribution for various types of medical images and shapes of organs. Moreover, we plan to validate CardiacFlow clinically on cardiac imaging datasets with diseases to generate plausible shape models of the whole heart for shape analysis, imaging-based biomarker discovery, and digital twin modelling.

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

References

1. Bai, W., Sinclair, M., Tarroni, G., Oktay, O., Rajchl, M., Vaillant, G., et al.: Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. Journal of Cardiovascular Magnetic Resonance **20**(1) (2018)
2. Biffi, C., de Marvao, A., Attard, M.I., et al.: Three-dimensional cardiovascular imaging-genetics: a mass univariate framework. Bioinformatics **34**(1) (2018)
3. Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., et al.: TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers. Medical Image Analysis **97**, 103280 (2024)
4. Deng, Y., et al.: ModusGraph: Automated 3D and 4D mesh model reconstruction from cine CMR with improved accuracy and efficiency. In: MICCAI (2023)
5. Esser, P., Kulal, S., et al.: Scaling rectified flow transformers for high-resolution image synthesis. In: ICML (2024)
6. Fonseca, C.G., et al.: The Cardiac Atlas Project-an imaging database for computational modeling and statistical atlases of the heart. Bioinformatics **27**(16) (2011)
7. Galazis, C., Shepperd, S., Brouwer, E.J., et al.: High-resolution maps of left atrial displacements and strains estimated with 3D cine MRI using online learning neural networks. IEEE Transactions on Medical Imaging (2025)
8. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In: MICCAI BrainLesion Workshop (2021)
9. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs trained by a two time-scale update rule converge to a local Nash equilibrium. NeurIPS **30** (2017)
10. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. NeurIPS (2020)
11. Investigators, S.H.: Coronary CT angiography and 5-year risk of myocardial infarction. New England Journal of Medicine **379**(10) (2018)
12. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: CVPR (2019)
13. Kingma, D.P., Welling, M., et al.: Auto-encoding variational bayes. In: ICLR (2013)
14. Kirişli, H., et al.: Standardized evaluation framework for evaluating coronary artery stenosis detection, stenosis quantification and lumen segmentation algorithms in computed tomography angiography. Medical Image Analysis **17**(8) (2013)
15. Kong, F., Stocker, S., Choi, P.S., et al.: SDF4CHD: Generative modeling of cardiac anatomies with congenital heart defects. Medical Image Analysis **97** (2024)
16. Lee, H., Lee, H., Gye, S., Kim, J.: Beta sampling is all you need: Efficient image generation strategy for diffusion models using stepwise spectral analysis. arXiv:2407.12173 (2024)
17. Lee, S., Lin, Z., Fanti, G.: Improving the training of rectified flows. NeurIPS **37** (2025)
18. Lipman, Y., Chen, R.T., Ben-Hamu, H., et al.: Flow matching for generative modeling. In: ICLR (2022)
19. Liu, X., et al.: Flow straight and fast: Learning to generate and transfer data with rectified flow. In: ICLR (2022)
20. Meng, Q., Bai, W., et al.: DeepMesh: mesh-based cardiac motion tracking using deep learning. IEEE Transactions on Medical Imaging **43**(4) (2023)
21. Metz, C., Schaap, M., Weustink, A., Mollet, N., van Walsum, T., Niessen, W.: Coronary centerline extraction from CT coronary angiography images using a minimum cost path approach. Medical Physics **36**(12) (2009)

22. Muffoletto, M., Xu, H., Burns, R., Suinesiaputra, A., Nasopoulou, A., et al.: Evaluation of deep learning estimation of whole heart anatomy from automated cardiovascular magnetic resonance short-and long-axis analyses in UK Biobank. European Heart Journal - Cardiovascular Imaging **25**(10) (2024)
23. Niederer, S.A., Lumens, J., Trayanova, N.A.: Computational models in cardiology. Nature Reviews Cardiology **16**(2) (2019)
24. Petersen, S.E., Matthews, P.M., et al.: UK Biobank's cardiovascular magnetic resonance protocol. Journal of Cardiovascular Magnetic Resonance **18**(1) (2016)
25. Qiao, M., Wang, S., et al.: CHeart: A conditional spatio-temporal generative model for cardiac anatomy. IEEE Transactions on Medical Imaging **43**(3) (2023)
26. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: CVPR (2022)
27. Roney, C.H., et al.: Universal atrial coordinates applied to visualisation, registration and construction of patient specific meshes. Medical Image Analysis **55** (2019)
28. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: MICCAI (2015)
29. Sermesant, M., Chabiniok, R., Chinchapatnam, P., et al.: Patient-specific electromechanical models of the heart for the prediction of pacing acute effects in CRT: a preliminary clinical validation. Medical Image Analysis **16**(1) (2012)
30. Song, Y., et al.: Score-based generative modeling through stochastic differential equations. In: ICLR (2020)
31. Sørensen, K., et al.: Spatio-temporal neural distance fields for conditional generative modeling of the heart. In: MICCAI (2024)
32. Strocchi, M., Gsell, M.A., Augustin, C.M., et al.: Simulating ventricular systolic motion in a four-chamber heart model with spatially varying robin boundary conditions to model the effect of the pericardium. Journal of Biomechanics **101** (2020)
33. Tobon-Gomez, C., Geers, A.J., Peters, J., Weese, J., et al.: Benchmark for algorithms segmenting the left atrium from 3D CT and MRI datasets. IEEE Transactions on Medical Imaging **34**(7) (2015)
34. Wang, Z., Jiang, Y., Zheng, H., Wang, P., He, P., Wang, Z., Chen, W., Zhou, M., et al.: Patch diffusion: Faster and more data-efficient training of diffusion models. NeurIPS **36** (2023)
35. Xia, Y., Chen, X., Ravikumar, N., Kelly, C., Attar, R., et al.: Automatic 3D+t four-chamber CMR quantification of the UK Biobank: integrating imaging and non-imaging data priors at scale. Medical Image Analysis **80** (2022)
36. Xu, H., Muffoletto, M., Niederer, S.A., Williams, S.E., Williams, M.C., Young, A.A.: Whole heart 3D shape reconstruction from sparse views: leveraging cardiac computed tomography for cardiovascular magnetic resonance. In: FIMH (2023)
37. 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 **24**(5) (2023)
38. Xu, Y., Xu, H., et al.: Improved 3D whole heart geometry from sparse CMR slices. In: Statistical Atlases and Computational Modeling of the Heart Workshop (2024)
39. Zhuang, X., Li, L., et al.: Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge. Medical Image Analysis **58** (2019)
40. Zhuang, X., Shen, J.: Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. Medical Image Analysis **31** (2016)