Slice2Mesh: 3D Surface Reconstruction from Sparse Slices of Images for the Left Ventricle

Jia Xiao, Wen Zheng, Wenji Wang, Qing Xia, Zhennan Yan,
Qian Guo, Xiao Wang, Shaoping Nie, and Shaoting Zhang

Abstract—Cine MRI is a widely used technique to evaluate left ventricular function and motion, as it captures temporal information. However, due to the limited spatial resolution, cine MRI only provides a few sparse scans at regular positions and orientations, which poses challenges for reconstructing dense 3D cardiac structures, which is essential for better understanding the cardiac structure and motion in a dynamic 3D manner. In this study, we propose a novel learning-based 3D cardiac surface reconstruction method, Slice2Mesh, which directly predicts accurate and high-fidelity 3D meshes from sparse slices of cine MRI images under partial supervision of sparse contour points. Slice2Mesh leverages a 2D UNet to extract image features and a graph convolutional network to predict deformations from an initial template to various 3D surfaces, which enables it to produce topology-consistent meshes that can better characterize and analyze cardiac movement. We also introduce As Rigid As Possible energy in the deformation loss to preserve the intrinsic structure of the predefined template and produce realistic left ventricular shapes. We evaluated our method on 150 clinical test samples and achieved an average chamfer distance of 3.621 mm, outperforming traditional methods by approximately 2.5 mm. We also applied our method to produce 4D surface meshes from cine MRI sequences and utilized a simple SVM model on these 4D heart meshes to identify subjects with myocardial infarction, and achieved a classification sensitivity of 91.8% on 99 test subjects, including 49 abnormal patients, which implies great potential of our method for clinical use.

Jia Xiao and Shaoting Zhang are with the SenseTime Research, Beijing 100080, China, and also with the Shanghai Artificial Intelligence Laboratory, Shanghai 200032, China (e-mail: becmamba@gmail.com; zhangshaoting@pjlab.org.cn).

Wenji Wang and Qing Xia are with the SenseTime Research, Beijing 100080, China (e-mail: wangwenji@sensetime.com; xiaqing@sensetime.com).

Zhennan Yan is with the SenseBrain Technology, NJ 08540, USA (e-mail: zhennanyan@sensebrain.site).

Wen Zheng, Qian Guo, Xiao Wang, and Shaoping Nie are with the Center for Coronary Artery Disease, Division of Cardiology, Beijing Anzhen Hospital, Capital Medical University, Beijing 100029, China (e-mail: mosquito99@126.com; gq_guoqian@163.com; spaceeye123@126.com; spnie@ccmu.edu.cn). Xiao Wang is also with the Cardiometabolic Medicine Center, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100037, China.

I. INTRODUCTION

CARDIAC magnetic resonance imaging (MRI) is one of the gold-standard medical imaging modalities for diagnosing cardiac anatomy and function [1]. Most often, cine MRI is a dynamic MRI technique that captures the heart's motion throughout the cardiac cycle. It allows physicians to observe cardiac function and structural changes, such as ventricular contraction and valve movements, essential for diagnosing heart diseases like cardiomyopathy and valvular disorders. However, cine MRI usually images the heart with high soft-tissue contrast in a collection of sparse and intersecting 2D image planes, which are a stack of short-axis (SAX) and a few long-axis (LAX) slices (referred as key slices), resulting in suboptimal image resolution for detailed anatomical assessments, and spatial constraints due to its two-dimensional nature, which can hinder comprehensive representation of the heart's 3D structure and dynamic function [2].

In contrast, 3D mesh reconstruction of the heart from cine MRI holds significant clinical value and advantages. Unlike 3D segmentation masks, which primarily provide volume and surface area metrics, 3D meshes offer detailed geometric and structural insights into cardiac anatomy. This enhanced precision is crucial for diagnosing and monitoring various cardiac conditions such as cardiomyopathies, valve diseases, and myocardial infarctions. These models enhance anatomical understanding and enable quantitative assessments of cardiac volume, wall thickness, and deformation, providing more accurate functional evaluations. Moreover, 3D mesh reconstruction can be integrated with computer-aided design (CAD) and computational fluid dynamics (CFD) for simulating hemodynamics and cardiac mechanics, guiding clinical decisions and surgical planning [3]–[9]. For instance, in assisting device implantation of the left ventricle (LV), 3D mesh models assist in determining optimal placement, reducing surgical risks, and improving outcomes [10].

Generally, traditional cardiac surface reconstruction comprises two steps: cardiac image segmentation and mesh generation from the segmentations. However, cine MRI is limited to capturing the heart's anatomy at a finite number of spatial locations and orientations. As a result, it provides a sparse representation of the actual 3D geometry of the human heart

[4], rendering direct application of traditional isosurfacing algorithms on cardiac segmentation results unfeasible due to the sparsity of the input data. To overcome this limitation, previous research has adopted a mesh adaptation approach. Van et al. [11] introduced a sparse active shape model (SPASM) for the automatic segmentation of cardiac MRI image datasets, featuring multiple planes with arbitrary orientations and significantly undersampled regions. Villard et al. [12] proposed a methodology for reconstructing geometrical surface meshes from sparse, heterogeneous, non-coincidental contours by deformation from a template mesh. Odille et al. [13] employed an implicit B-spline surface reconstruction algorithm for the reconstruction of the LV cavity surface from a sparse set of 2D contours. Attar et al. [14] utilized the SPASM, which consists of a point distribution model (PDM), an intensity appearance model (IAM), and a model matching algorithm, to directly reconstruct 3D segmentation meshes from cine images. Loke et al. [5] applied smooth mappings known as diffeomorphism to adapt the 3D end-diastolic model of the right ventricle (RV) to the 2D segmentation contour points. Hu et al. [15] leveraged a 3D active shape model (3D-ASM) alongside an image intensity model to align initial meshes with ground-truth contours. Bennati et al. [6] morphed a sphere surface to 2D segmentation contours to reconstruct 3D RV meshes based on distance fields. However, these techniques are labor-intensive, with the manual adjustment of optimizer parameters proving to be complex, significantly limiting the feasibility of mesh reconstruction for real-time applications in surgical guidance and navigation [16].

Deep-learning-based methods have also been explored for this task and can be classified into three types: image-interpolation-based, contour-to-mesh-based and image-to-mesh-based methods. Image-interpolation-based methods mainly rely on image interpolation techniques to generate high spatial resolution images and apply sophisticated subsequent segmentation and iso-surface extraction to obtain the final 3D shape. Biffi et al. [17] first segment the 2D SAX images and then reconstruct the 3D high-resolution volume based on a variational autoencoder. Chang et al. [18] propose a latent-space-based generative method to jointly generate 2D SAX segmentation and 3D volume by interpolating latent codes. To transform dense 3D volumes into surface meshes, isosurfacing algorithms such as Marching Cubes [19] are subsequently employed. However, these methods necessitate manual adjustments to segmentation and interpolation outcomes to achieve smoother and more precise 3D meshes, thereby introducing time-consuming artifacts.

Contour-to-mesh-based methods segment the image into contours by a typical image segmentation model and then learn to directly generate the 3D shape from sparse contours or points. Beetz et al. [20] developed Point2mesh-Net, utilizing the Point Convolution Block to directly convert sparse point clouds into meshes. Furthermore, Beetz et al. [21] initially approximate the fitting of the template mesh to the sparse contours, followed by the application of a U-Net based on graph convolution network (GCN) to enhance the mesh's accuracy and smoothness. These techniques typically employ synthetic data from the statistical shape model (SSM) [22]

for training, capturing contours or point clouds. However, the SSM, constrained by extensive manually-defined parameters, proves to be unsuitable for clinical data, characterized by its diverse shapes. Chen et al. [16] embarked on segmenting a stack of SAX images from the clinical dataset of the UK Biobank study (UKBB) [23] to obtain contour points. Subsequently, they utilize a GCN-based network to morph the template mesh into the targeted mesh. It is critical to acknowledge, however, that the ground-truth (GT) meshes utilized for training these methods are generated by models rather than manually annotated, attributable to the complexity of mesh labeling, potentially leading to inaccuracies. Moreover, these methodologies reconstruct meshes solely from contours or points, neglecting the cine image characteristics.

Image-to-mesh-based methods directly reconstruct cardiac meshes from sparse cine images, diverging from the indirect approaches characterized by image-interpolation and contour-to-mesh-based methods. Tothova et al. [24] introduce a probabilistic deep learning strategy for the concurrent 3D surface reconstruction from sparse two-dimensional MR image data, coupled with uncertainty prediction through a Gaussian distribution. Xia et al. [25] employ a point distribution model informed by deep learning to engineer 3D meshes. These techniques utilize reference 3D meshes for supervised training, generated not through manual labeling but by non-rigid registration algorithms. Hence, the fidelity of reference meshes serving as ground truth critically impacts the training's accuracy.

While some works focus on the RV [5], [26], [27] or bi-ventricle reconstruction [15], [16], [28], our work mainly focuses on generating LV meshes. In clinical practice, 3D mesh reconstruction of the LV is prioritized over that of the RV due to the LV's crucial role in the systemic circulation, where it pumps oxygenated blood from the heart to the entire body. The functional state of the LV directly impacts the systemic supply of blood and oxygen. The LV dysfunction, including conditions such as hypertrophy and dilated cardiomyopathy, represents a leading cause of heart failure, significantly affecting patient outcomes. Thus, accurate assessment and continuous monitoring of LV structural and functional changes are essential for the diagnosis and management of heart failure [29], [30]. In contrast, the RV primarily pumps deoxygenated blood to the lungs for gas exchange. Although the RV dysfunction can lead to clinical complications, its evaluation and intervention strategies are relatively simpler. Furthermore, the LV's pivotal role in cardiac surgeries and interventional procedures underscores its precedence in 3D reconstruction efforts. For instance, precise localization of the LV scar tissue is crucial for assessing and treating recovery after myocardial infarction, a concern that is less prevalent with the RV [10], [30].

Given the limitations inherent in existing traditional and deep-learning-based mesh reconstruction techniques, alongside the critical significance of the LV reconstruction, we present a pioneering and pragmatic image-based deep-learning strategy named Slice2Mesh, geared towards the reconstruction of 3D cardiac surfaces, with a particular focus on the LV. The cornerstone contributions of our work are delineated as

follows:

- Our method is image-to-mesh based, allowing the generation of accurate and smooth 3D meshes by directly leveraging sparse SAX and LAX image features without the need for postprocessing.
- Leveraging a template mesh as a foundation for the deformed predicted meshes allows us to integrate prior shape information, compensating for the dearth of spatial information stemming from image sparsity. Moreover, incorporating the As Rigid As Possible (ARAP) energy [31] into our model as a loss function facilitates the optimization of local deformation smoothness, thereby fully utilizing prior shape information.
- We eschew the requirement for 3D meshes as a reference, utilizing solely sparse contour points transformed from 2D segmentation labels to constrain deformation. This yields comparable results to 3D constraints while necessitating less supervision information.
- Our model has been subjected to rigorous training and evaluation using clinical data. Furthermore, we have implemented our methodology in reconstructing 4D (3D+T) cardiac meshes to diagnose myocardial infarction (MI), a prevalent and high-risk cardiovascular condition. Empirical evidence underscores the efficacy and accuracy of our approach.

II. METHODS

In this section, we will introduce the pipeline of our method. Our Slice2Mesh architecture is depicted in Fig. 1. It contains three modules: segmentation module, feature sampling module and mesh deformation module. The segmentation module extracts image features with different resolutions and channels for SAX and LAX views. The feature sampling module is to sample image features to specific vertices of the initial or previously deformed mesh. The deformation module iteratively deforms the initial mesh to the target contour points using the sampled features. Section II-A covers the preprocessing stage, which is the initial step in the subsequent process. Following that, Section II-B, Section II-C and Section II-D provide detailed introductions to the three modules in our model. Finally, Section II-E outlines the loss function used across all three modules.

A. Preprocessing

The preprocessing phase involves correcting the misalignment of image slices, creating a reference point cloud for supervised training, resampling images to the same resolution and cropping them to the same shape, and selecting and processing the template mesh to generate the initial mesh.

1) Misalignment correction: Clinical data has misalignment between slices caused by multiple breath holding and possible body movement during acquisition [32]. Given the contours extracted from segmentation masks, it is possible to optimally align the SAX and LAX slices in 3D space to minimize the misalignments. However, a perfect alignment, where the distances between SAX and LAX contours are zero, is usually not achievable due to motion-related deformations and

inconsistencies in the contouring of the slices [33]. Therefore, we apply an iterative misalignment correction algorithm [34]. In detail, for each slice of both the SAX and LAX slices, we start by locating the intersection points where the slice contour intersects with all other slice planes. Then, we identify the intersection points where the slice plane intersects with all other slice contours. After finding the centroids of both sets of intersection points, we calculate the distance between the target centroid and the source centroid, which gives us the correction value. This correction value is then used to correct the slice. By repeating this process iteratively, we can obtain the final correction matrix for each slice of both the images and masks.

2) Reference point cloud generation: After aligning each image and mask slice, we convert the corrected contour points from image coordinates to world coordinates using the original DICOM information of origin, direction, and spacing. These points can be seen as sparse point clouds and used as a reference for training. It's important to note that the method for correcting misalignments based on contours mentioned above isn't perfect due to the complexity of intra and inter-slice misalignment. This means that the SAX and LAX contours may not be perfectly aligned, leading to chaotic deformation and unsmooth surfaces in these areas. To minimize the impact of the misalignment, we remove the LAX contour points between the top and bottom SAX slices and only keep the LAX contour points in the basal and apical areas to monitor the local deformation where there is no SAX slice image information.

3) Image resampling and crop: As the images and labels have different in-plane resolutions and shapes that can not be inputted into the model, we first resample the images and label to the in-plane resolution of 1 mm×1 mm by linear interpolation while the thickness in the z-orientation remains unchanged, and then center-crop-pad them to the shape of 256×256×N, where N is 16 for SAX images and 1 for LAX images.

4) Mesh initialization: After obtaining sparse reference point clouds to constrain the mesh deformation, we need to generate initial meshes for each sample to deform from. We have two template meshes of the left myocardium, one at the end of diastole (ED) and the other at systole (ES), obtained from a statistical shape model[1] [22]. We use a nonrigid Iterative Closest Point (ICP) algorithm [35] to adjust the ES template to match the ED template, ensuring that both meshes have the same topology. As a result, both the ES and ED template mesh have 13808 vertices and 27612 faces. To make the template mesh more adaptable to various heart shapes in a heart cycle, we create a mesh in the middle phase of ED and ES by linearly interpolating the vertex positions in ED and ES, which acts as the final **template mesh**. It is subsequently flipped in z-orientation and smoothed using Laplacian smoothing [36] in Meshlab [37], in which the smoothing steps are 3, and the 1D boundary smoothing and cotangent weighting are used.

To account for the various locations and shapes of human hearts, we need to roughly align the template mesh before

---

[1] https://wp.doc.ic.ac.uk/wbai/data

training to accurately deform it to match the target contour points. For each sample, we start by uniformly sampling the template surface to create a point cloud with the same number of points as the reference point cloud. Then, we use the rigid ICP algorithm [38] to calculate a rotation matrix and a shift matrix from the sampled point cloud to the reference point cloud in order to make the two point clouds as similar as possible. The transformed template mesh becomes the **initial mesh** for the corresponding sample during the **training** process. In the **inference** stage, the output mask from the segmentation module is transformed to create contour points, which are then used to roughly align the template mesh and generate the initial mesh.

B. Segmenation module

The segmentation module takes resampled and cropped images as input and generates image features with different resolutions of 256×256, 128×128, 64×64 and 32×32 with corresponding channels of 16, 32, 64 and 128. It is based on a 2D UNet [39] architecture with global convolution layers [40] and Resblocks. Incorporating global convolution layers offers significant advantages. A larger kernel size expands the receptive field while employing separate k × 1 × 1 convolutions in different directions yields superior performance compared to using k × k × k convolutions directly, as demonstrated in [40]. To leverage these benefits, we strategically position global convolution layers in a few encoder layers housing higher-resolution features. Additionally, by integrating global convolution layers with multi-scale kernels, we enhance feature diversity to address the potential loss of thin structures in deep intermediate layers, which could otherwise compromise segmentation accuracy. We train three segmentation modules for three views of images and return segmentation masks and image features, respectively.

C. Feature sampling module

The feature sampling module is the critical connection between the segmentation module and the deformation module. It takes the image features from the segmentation module as input and generates grid-sampled features on specific vertices. In this section, we introduce the selection of slice vertices where we sample the image features and the weighted neighborhood feature sampling method, which improves the sampling accuracy.

1) Slice vertices selection: The initial or previously deformed mesh contains complete 3D space information, while only a few slices in 3D space have image features. Therefore, to fill the image features to specific vertices of each slice, we first select the mesh vertices in the slice location. Note that the vertices are not precisely in the slice plane; we first find all faces that the slice plane goes through and then get the nearest vertex in each selected face to the slice plane. Because the mesh with 13808 vertices is dense, the distances away from the plane can be ignored.

2) Weighted neighbourhood feature sampling: The feature vector of each slice is extracted from image features by feature sampling at locations of the slice vertices sampled from the

initial or previously deformed mesh in each iteration. Some works [41] sample features at exact vertex locations referred to as point sampling, while this restricts the sampler's ability to pool information from its neighborhood. While other methods [42] sample features at a fixed neighborhood around the vertex, referred to as neighborhood sampling, they obtain the final feature in the neighborhood by simply calculating the mean value, which restricts the network's ability to learn the weights in the neighborhood. We design a weighted neighborhood feature sampling method to sample the pixel features in the slice.

In detail, we assume that one training sample has N_s slices of SAX and LAX images in total. Each image slice has four image features I_i with different channels and resolutions $C_i \times H_i \times W_i$, where $i = 0, 1, 2, 3$. We first initial the feature vector of all vertices $\mathcal{F} \in \mathbb{R}^{C \times N_v}$ with zeros, where N_v is the total number of vertices and $C = 16 + 32 + 64 + 128 = 240$ is the addition of channels of 4 different resolution image features. To project corresponding image features onto slice vertices, we compute the corresponding image coordinates L_s in the image feature maps from the vertex locations of the initial or previously deformed mesh by transforming the coordinates from world space to image space using the origin, spacing and orientation information of the original dicom data. Then, we utilize bi-linear interpolation to sample the image features into the fixed neighborhood $\mathcal{N}_s$ with 8 neighboring pixels and the computed center pixel itself. We train a neural function to return the weights W_i of 9 pixels, to sum up the 9 features in the neighborhood. The features for each slice $\mathcal{F}_s$ are obtained afterward by concatenating 4 different sampled image features. The procedure can be formulized as,

$$\mathcal{F}_s = \text{CONCAT}(\text{BILINEAR}(I_i, \mathcal{N}_s) \times W_i | i = 0, ..., 3) \quad (1)$$

where $I_i \in \mathbb{R}^{C_i \times H_i \times W_i}$, $\mathcal{N}_s \in \mathbb{R}^{N_s \times 9}$, $\text{BILINEAR}(I_i, \mathcal{N}_s) \in \mathbb{R}^{C_i \times N_s \times 9}$, $W_i \in \mathbb{R}^{9 \times 1}$, $\mathcal{F}_s \in \mathbb{R}^{C \times N_s}$ and $C = \sum_{i=0}^{3} C_i = 240$.

Then, the sampled features of each slice $\mathcal{F}_s$ are filled into corresponding locations of the slice in $\mathcal{F}$. Due to the sparsity of slices, $\mathcal{F}$ can not be fully filled, which means $\mathcal{F}$ has zeros in some slices of vertices. The graph convolutions in the deformation module are used to transfer and collect the features between vertices, which is the interpolation operation in graph space.

D. Deformation module

The deformation module takes the sampled features from the feature sampling module as input and generates the deformed meshes from the initial meshes. In this section, we introduce the graph convolution theory and the iterative deformation mechanism.

1) Graph convolution: A 3D mesh is a collection of vertices, edges and faces, which defines the shape of a 3D object; it can be represented by a graph $\mathcal{M} = (\mathcal{V}, \mathcal{E}, \mathcal{F})$ where $\mathcal{V} = \{v_i\}_{i=1}^{N}$ is the set of N vertices in the mesh, $\mathcal{E} = \{e_i\}_{i=1}^{E}$ is the set of E edges with each connecting two vertices, and $\mathcal{F} = \{f_i\}_{i=1}^{N}$

(a) Architechure

(b) A GCN block
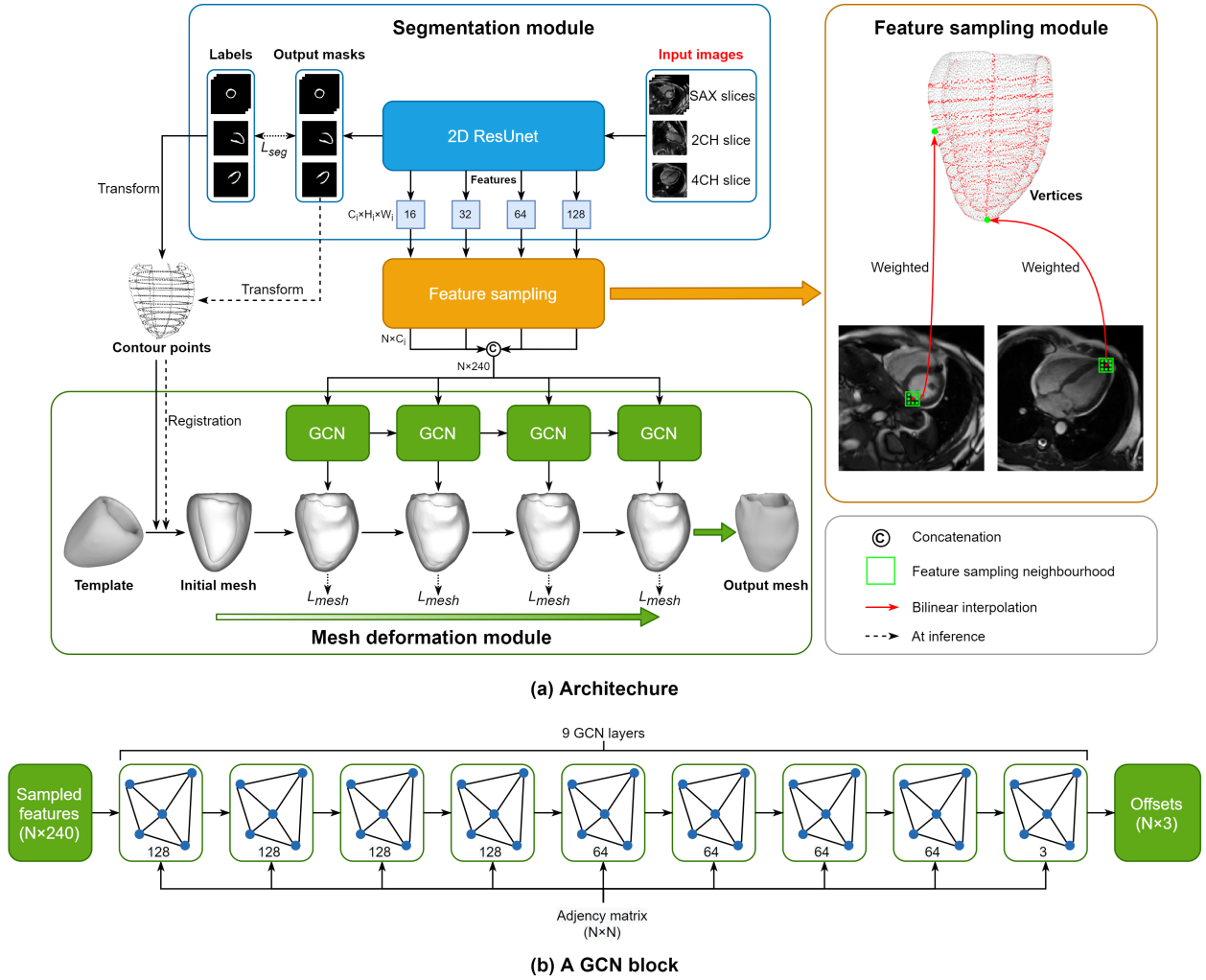
Fig. 1. Diagram of the proposed cardiac surface reconstruction approach, Slice2Mesh. (a) The architecture. It consists of three modules: segmentation module in the blue box, feature sampling module in the brown box and mesh deformation module in the green box. The segmentation module extracts image features with different resolutions and channels. The feature sampling module is to sample image features to specific vertices of the initial or previously deformed mesh. Note that we just use images to show the sampling procedure clearly, and in practice, the sampling operation is applied to image features. The deformation module iteratively deforms the initial mesh to the target contour points using the sampled features in the feature sampling module. The labels are transformed to generate contour points, which are used for coarse registration of the template mesh and supervision for mesh deformation in the **training** phase. In the **inference** stage, the contour points for template mesh registration are transformed from the output masks of the segmentation module (b) A GCN block. It consists of eight graph convolutional layers and one output layer.

are the feature vectors attached on all vertices. A graph-based convolutional layer is defined as

$$\mathcal{F}_{out} = \sigma(\theta_0 \mathcal{F}_{in} + \theta_1 \mathcal{F}_{in} \tilde{L}), \qquad (2)$$

where $\theta_0, \theta_1 \in \mathbb{R}^{c_{out} \times c_{in}}$ are trainable weights, $\mathcal{F}_{in} \in \mathbb{R}^{c_{in} \times N}$, $\mathcal{F}_{out} \in \mathbb{R}^{c_{out} \times N}$ are input and output feature vectors of the graph convolution layer attached on all vertices, and

$$\tilde{L} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} \in \mathbb{R}^{N \times N}, \qquad (3)$$

with adjacency matrix $\tilde{A} = A + I_N$ and degree matrix $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ is the normalized Laplacian matrix which represents the connectivity information between vertices [43]. N is the number of control points. c_{in} and c_{out} are the input and output graph feature channels, respectively.

2) Iterative deformation: The deformation module comprises four deformation blocks (four iterations) with graph convolutional layers that gradually deform the initial meshes toward

reference points based on the concatenated feature vectors $\mathcal{F}$ from the feature sampling module in each iteration. The concatenated feature vectors within each mesh deformation block are processed by four graph residual blocks containing eight graph convolutional layers. We then employ an additional graph convolutional layer to predict offsets as 3D feature vectors and add them with the vertex coordinates of the initial mesh or the mesh from the previous deformation block to obtain the current predicted vertex coordinates.

E. Loss function

We first train three segmentation modules for SAX, 2-chamber LAX and 4-chamber LAX images supervised by reference masks, respectively, by using a hybrid loss function containing cross entropy and dice loss in which the loss balance factor is 0.5. After that, we extract three types of

image features from three segmentation modules and feed them into the deformation module to train it supervised by reference sparse contour points obtained from reference masks. The mesh loss for training the deformation module includes chamfer distance loss, edge length loss, normal loss and Laplacian loss. Besides, we add ARAP loss to make full use of the prior shape information because of the sparsity of image information.

1) Chamfer distance loss: We constrain the deformation using chamfer distance loss $\mathcal{L}_{CD}$ between the predicted and reference contour points in key slices. Chamfer distance loss is applied to capture an overall distance between the predicted points and points of reference. It is defined by,

$$\mathcal{L}_{CD}(\boldsymbol{P}, \boldsymbol{G}) = \frac{1}{|\boldsymbol{P}|} \sum_{\boldsymbol{p} \in P} \min_{\boldsymbol{g} \in G} \|\boldsymbol{p} - \boldsymbol{g}\|_2^2 + \frac{1}{|\boldsymbol{G}|} \sum_{\boldsymbol{g} \in G} \min_{\boldsymbol{p} \in P} \|\boldsymbol{g} - \boldsymbol{p}\|_2^2,$$
(4)

where $\boldsymbol{P}$ and $\boldsymbol{G}$ are the predicted and reference contour points in all key slices. Particularly, the myocardium includes the endocardium and epicardium. To make the deformation more accurate, we calculate the chamfer distance loss separately for endocardium and epicardium points, which means

$$\mathcal{L}_{CD} = \mathcal{L}_{CD\_endo} + \mathcal{L}_{CD\_epi}.$$
(5)

2) Smooth loss: We introduce three regularization terms, edge length loss $\mathcal{L}_{edge}$, normal loss $\mathcal{L}_{normal}$ and Laplacian loss $\mathcal{L}_{lap}$, to improve convergence and smoothness of the output meshes. These losses are implemented by Pytorch3d[2] functions. Besides, we introduce the ARAP loss $\mathcal{L}_{arap}$ to make the local deformation as rigid as possible, which makes full use of the prior shape information.

The ARAP energy, a widely recognized non-rigid deformation energy, was initially introduced and defined in [31] for surface models where the triangular face serves as the fundamental unit. Given a vertex $\boldsymbol{v}_i$ and its neighbor vertices $\mathcal{N}(i)$, a cell $\mathcal{C}_i$ is defined as the unit containing all the triangular faces in the neighborhood. This energy quantifies the deformation between cells before and after deformation, and the overall ARAP energy is computed as the average energy of all cells within the neighborhood.

Given the cell $\mathcal{C}_i$ corresponding to vertex $\boldsymbol{v}_i$, and its deformed version $\mathcal{C}_i'$, we define the approximate rigid transformation between the two cells by observing the edges emanating from the vertex $\boldsymbol{v}_i$. If the deformation $\mathcal{C}_i \to \mathcal{C}_i'$ is rigid, there exists a rotation matrix $\boldsymbol{R}_i$ such that

$$\boldsymbol{e}_{ij}' = \boldsymbol{R}_i \boldsymbol{e}_{ij}, \forall j \in \mathcal{N}(i),$$
(6)

where $\boldsymbol{e}_{ij} = \boldsymbol{v}_i - \boldsymbol{v}_j$ is the edge for $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ in the cell $\mathcal{C}_i$, and similarly for $\boldsymbol{e}_{ij}'$ for the deformed cell $\mathcal{C}_i'$. $\mathcal{N}(i)$ is the neighborhood of vertex $\boldsymbol{v}_i$. When the deformation is not rigid, we can still find the best approximating rotation $\boldsymbol{R}_i$ that fits the Eq. (6) in a weighted least squares sense, i.e., minimizes

$$\mathcal{L}(\mathcal{C}_i, \mathcal{C}_i') = \sum_{j \in \mathcal{N}(i)} \|\boldsymbol{e}_{ij}' - \boldsymbol{R}_i \boldsymbol{e}_{ij}\|_2^2,$$
(7)

where $\mathcal{L}(\mathcal{C}_i, \mathcal{C}_i')$ is the ARAP loss for cell $\mathcal{C}_i$.

[2]https://pytorch3d.org/

To obtain the optimal rotation $\boldsymbol{R}_i$, we first denote

$$\boldsymbol{S}_i = \sum_{j \in \mathcal{N}(i)} \boldsymbol{e}_{ij} \boldsymbol{e}_{ij}'^T = \boldsymbol{E}_i \boldsymbol{E}_i'^T,$$
(8)

where $\boldsymbol{E}_i$ is the $3 \times |\mathcal{N}(i)|$ containing $\boldsymbol{e}_{ij}$s as its columns, and similarly for $\boldsymbol{E}_i'^T$. Then one can derive the optimal $\boldsymbol{R}_i$ from the singular value decomposition of $\boldsymbol{S}_i = \boldsymbol{U}_i \Sigma_i \boldsymbol{V}_i^T$:

$$\boldsymbol{R}_i = \boldsymbol{V}_i \boldsymbol{U}_i^T.$$
(9)

By substituting the optimal $\boldsymbol{R}_i$ into the Eq. (7), we can get the ARAP loss for cell $\mathcal{C}_i$. By averaging the ARAP loss for cells of all vertices, we get the total ARAP loss. ARAP loss tries to make the local deformation of vertices as rigid as possible and not so dramatic away from the vertices before deformation.

Subsequently, the ARAP energy has undergone further refinement and adaptation in various studies, including [44] and [45]. In [44], ARAP energy was integrated into a trainable model for general surface meshes using deep learning methodologies and was extended as a restricted loss function through the decomposition of the Hessian spectrum as an approximation of the ARAP energy. In [45], the original ARAP energy formula for 3D surface model deformation was extended to 3D volumetric mesh deformation and utilized for the reconstruction of aortic valves from 3D CT images. The fundamental unit was transitioned from the triangular face to the hexahedron, and leveraging an equivalent number of neighboring vertices per vertex in the template mesh expedited the computation of the overall ARAP energy through matrix operations.

In contrast to the methodologies presented in [44] and [45], our approach directly leverages the original ARAP energy formula introduced in [5] as the loss function to govern the reconstruction of 3D surface meshes from 2.5D sparse cine images. Our approach refrains from relying on spectral decomposition for approximation and upholds the triangular face as the fundamental unit. Given the variable number of neighboring vertices in our template mesh and the inapplicability of matrix acceleration, we employ a strategy of randomly selecting four vertices in the neighborhood for each vertex during each iteration to expedite matrix calculations.

After getting ARAP loss, the total smooth loss is

$$\mathcal{L}_{smooth} = \lambda_1 L_{normal} + \lambda_2 \mathcal{L}_{lap} + \lambda_3 \mathcal{L}_{edge} + \lambda_4 \mathcal{L}_{arap}. \text{ (10)}$$

The hyper-parameters λ_0, λ_1, λ_2, and λ_3 are empirically determined to be 15.0, 1.0, 20.0, and 5.0, respectively. We write the complete mesh loss to train the deformation module as

$$\mathcal{L}_{mesh} = \sum_{l=1}^{L} (\mathcal{L}_{CD}^l + \mathcal{L}_{smooth}^l),$$
(11)

where L is the number of iterations in the deformation module.

III. Experiments

A. Dataset and implementation

We have obtained two clinical cine MRI datasets. One of the datasets is a private cine MRI dataset from Beijing Anzhen

TABLE I
THE DEMOGRAPHIC INFORMATION OF TWO DATASETS

Type	Dataset	
	Private(MI)	UKBB(healthy)
Subjects number	50	250
Age(years)	56 ± 12	55 ± 9
Male gender(n(%))	42(84%)	161(64.4%)
Weight(kg)	82 ± 7	76 ± 15
Height(cm)	175 ± 8	169 ± 10
Body surface area(m^2)	1.91 ± 0.14	1.83 ± 0.22
Body mass index(kg/m^2)	26.0 ± 2.9	25.1 ± 3.2

Hospital. It consists of 50 abnormal subjects with varying degrees of myocardial infarction, and each subject has 25 phases in a cardiac cycle. The cine images were captured using GE DISCOVERY MR750w scanners. The in-plane resolution ranges from 0.87 mm to 1.72 mm, with a mean value of 1.21 mm. The slice thickness ranges from 8.0 mm to 10.0 mm, with a mean value of 8.24 mm. The in-plane image size ranges from 256 to 320. The number of SAX slices varies from 9 to 15, with a median value of 11, while the number of 2-chamber and 4-chamber LAX slices is both one. The other cine MRI dataset is a subset of the UKBB dataset[3] [23], which consists of 250 healthy subjects of the UK Biobank study, and each subject has 50 phases in a cardiac cycle. The cine images were collected using SIEMENS scanners. The in-plane resolution ranges from 1.83 mm to 2.31 mm, with a mean value of 1.86 mm. The slice thickness is 8.0 mm. The in-plane image size ranges from 138 to 210. The number of SAX slices ranges from 7 to 13, with a median value of 11, while the number of 2-chamber and 4-chamber LAX slices is both one. Additional demographic information for the two datasets can be found in Table I.

We randomly selected 5 phases for each subject in the private dataset while specifically sampling two phases of ED and ES for each subject in the UKBB subset. In total, our dataset consists of 750 samples. We divided the dataset into a training set (600 samples) and a test set (150 samples) at the subject level. Each sample includes a stack of SAX images, one 2-chamber LAX image, one 4-chamber LAX image, and corresponding myocardium segmentation labels, which were manually annotated. The ground-truth sparse contour points used to train the deformation module are transformed from reference segmentation labels.

We use the Adam optimizer in all experiments with a learning rate of 0.001 and a batch size of 4 to train Slice2Mesh for 50 epochs. The network is implemented using Python 3.7.10 and Pytorch 1.10.2, and all experiments are executed on a single NVIDIA GeForce GTX 3090 GPU (24GB).

B. Baseline

For image segmentation, we compare our ResUnet of the segmentation module with Unet [39] and TransUnet [46] for SAX, 2-chamber LAX and 4-chamber LAX images.

For mesh reconstruction, various techniques such as image interpolation followed by segmentation and isosurfacing,

[3]https://biobank.ndph.ox.ac.uk/

deformation from point clouds or contours, point cloud upsampling and mesh reconstruction from a single image address the reconstruction problem in different ways. As we do not have 3D meshes for reference, the deep learning-based methods can not be trained on the same dataset, and we can not compare our method with them directly. So we first choose to compare our method with three traditional nonrigid point set registration techniques, deformation transfer for triangle meshes (DTTM) [47], optimal step nonrigid ICP (NrICP) [35], and nonrigid coherent point drift (NrCPD) [48], where the template mesh is the same as ours, and the hyper-parameters are tuned based on samples from the training and validation set. DTTM and NrICP are implemented by Trimesh[4], and NrCPD is implemented by Probreg[5]. In addition, for better understanding of the performance of our method, we performed an indirect comparison of slice segmentation accuracy and basic physiological indexes at the subject level with three learning-based reconstruction methods: the fully convolutional network (FCN) by Bai et al. [49], the sparse active shape model by Attar et al. [14] and the deep-learning-based point distribution model by Xia et al. [25].

C. Metrics

Dice and IoU (Intersection over Union) metrics are calculated to evaluate the accuracy of three segmentation modules of SAX, 2-chamber LAX and 4-chamber LAX images.

Reconstruction accuracy of the deformation module compared with nonrigid point set registration techniques is measured using 3D chamfer distance (CD), earth mover's distance (EMD) and Hausdorff distance (HD_{3d}). There are no available reference meshes, so these three metrics are calculated between the sparse predicted and reference points sampled in key slices. To fully evaluate the deformation accuracy, we extract the segmentation contours from predicted meshes in SAX planes, transform the contours from world coordinates to image coordinates and calculate the Dice and IoU indexes between the transformed contours and reference contours in specific SAX slices. To evaluate the smoothing degree of the predicted mesh, we calculate Laplacian smoothing loss on the mesh.

Moreover, we conduct the indirect comparison with three learning-based methods using three 2D metrics: Dice, mean contour distance (MCD). [25] and HD (HD_{2d}) between reconstruction contours (obtained from 3D meshes in SAX planes) and reference contours. The MCD is actually the 2D version of CD, with the only difference being an additional division by 2.

The average value and standard deviation are summarised for all evaluation metrics.

D. Accuracy of 3D mesh reconstruction

Table II presents the quantitative segmentation results of the segmentation module in three image views on the test set. Compared with Unet [39] and TransUnet [46], our segmentation model, which incorporates the Unet with global

[4]https://trimsh.org/trimesh.registration.html
[5]https://github.com/neka-nat/probreg

convolution layers and Resblocks, demonstrates improvements in both the Dice and IoU indices across all three image views. Enhanced segmentation outcomes have the potential to yield more precise image features for the subsequent feature sampling module and deformation module.

Fig. 2 shows three samples of our reconstructed 3D cardiac shapes. The shapes are represented as corresponding 2D contours overlaid with the SAX image slices of base, mid and apex positions. The contours predicted by our method are shown in green, while the GT masks are shown in red. It can be seen that the reconstructed contours are close to the reference masks in the base, mid and apex locations, which confirms that our model can produce accurate shape parameters to generate shapes similar to the reference contours.

Table III compares our method with three nonrigid registration baselines in terms of the quantitative 3D mesh reconstruction results on the test set. Our method demonstrates clear advantages in each metric compared with DTTM and NrICP. However, compared with NrCPD, our method is slightly inferior in CD and EMD. This discrepancy arises because the reference contour points are concentrated in sparse slice locations and not distributed throughout the cardiac space. We calculate the corresponding metrics between the reference contour points and the predicted mesh vertices at these slice locations. NrCPD excessively deforms the vertices of the template mesh towards the reference contour points, resulting in insufficient deformation of the vertices at other spatial locations. Consequently, the deformation metrics of CD and EMD are small. In contrast, the smoothness metric of Laplace increases, indicating that the mesh surface generated by NrCPD is very uneven and not ideal. In contrast, our method balances deformation and smoothing to achieve a more even mesh surface. Additionally, our method significantly reduces the inference time for 3D surface reconstruction, making it more efficient for experiments and applications.

Fig. 3 presents the 3D reconstruction results using different methods for four test samples. The first column shows the complete reference contour points to clearly demonstrate the effect. In practice, LAX points between the top and bottom SAX slices are not utilized in the deformation process. The second through fourth columns display the reconstruction results of three nonrigid registration baselines, while the fifth column presents the results of our method. First, we compare the fourth column (NrCPD) with the fifth column (our method). The results of NrCPD confirm our previous observations: NrCPD achieves more accurate deformation at key slices but is insufficient at other positions, resulting in a rough surface unsuitable for static and dynamic heart analysis. Next, we compare the second, third, and fifth columns of DTTM, NrICP and ours. Regarding smoothness, our method is comparable to DTTM and superior to NrICP. Regarding deformation accuracy, DTTM, NrICP, and our method all achieve relatively accurate deformations in the main body of the heart. However, at the basal and apical positions, the short-axis image information is missing, and only a few reference contour points from the long axis supervise the deformation. This limitation hampers the accurate representation of the shape in these regions. Consequently, the reconstruction results

of DTTM and NrICP at the basal and apical positions are suboptimal. In contrast, our method, constrained by ARAP loss, ensures the deformation is not too drastic or sharp, even with limited supervision information at the base and apex. This constraint helps maintain the original shape of the template mesh in these regions, leading to a more accurate and smooth reconstruction.

To thoroughly demonstrate the function of ARAP loss, Fig. 4 shows the results of three test samples with and without ARAP loss. The first row displays the results without ARAP loss, while the second row shows the results with ARAP loss. The comparison indicates that the absence of ARAP loss leads to drastic, sharp, non-rigid, and non-smooth deformation of local areas relative to the template mesh, resulting in apparent folds, protrusions, or concavities. In contrast, using ARAP loss ensures that local deformation remains gentle, rigid, and smooth relative to the template mesh, achieving a better overall effect. In summary, our method effectively balances accuracy and smoothness in deformation, especially achieving better reconstruction results for the heart's base and apex positions.

Table IV provides an indirect comparison of Dice, MCD and HD_{2d} between our method and three learning-based methods on the slice segmentation level. All these three methods conduct segmentation accuracy evaluation between segmentation contours and reference contours on 600 subjects (n=600) of the UKBB dataset. We evaluate our method on 300 subjects (n=300) containing 250 subjects of the UKBB dataset and 50 subjects of the private dataset. LV_{endo} and LV_{epi} represent the endocardium and epicardium of myocardium repectively. In Table IV, the mean Dice of our method for LV_{endo} is 0.92, which is slightly lower than the best Dice of 0.94 of Bai et al. and Attar et al. The mean Dice of our method for LV_{epi} is 0.89, outperforming other methods. Therefore, in terms of Dice, our method is comparable with other learning-based methods, which indicates excellent agreement between manual delineations and automated segmentations. As for MCD, the mean values of MCD for LV_{endo} and LV_{epi} are 0.96 mm and 0.92 mm, which outperform Attar et al. by 0.10 mm and 0.21 mm, respectively. In terms of HD_{2d}, the mean value of HD_{2d} for LV_{epi} is 3.88 mm, which is the best compared with other methods. However, the mean value of HD_{2d} for LV_{endo} is 3.94 mm, which is much higher than Bai et al. and Attar et al. by about 0.8 mm, but the values of HD_{2d} for both LV_{endo} and LV_{epi} of our method are in the same level. The possible reason may be that the endocardium is larger and has more contour points than the epicardium, so the FCN used by Bai et al. and the point distribution model used by Attar et al., which focus on larger reception fields may show better performance in the endocardium segmentation. In contrast, our method models the relationship between the image features and final displacements of vertices in a point-to-point manner, so the distribution of target points does not affect the global deformation much, resulting in the comparable HD_{2d} of endocardium and epicardium.

While our work does not require dense GT meshes for training, we also perform an additional evaluation on a small test set with ground truth dense meshes to demonstrate the effectiveness of Slice2Mesh in the regions between slices.

TABLE II
SEGMENTATION RESULTS OF THE SEGMENTATION MODULE ON THE TEST SET

Model	Dice↑			IoU↑		
	SAX	2CH	4CH	SAX	2CH	4CH
Unet [39]	0.883±0.154	0.881±0.132	0.880±0.126	0.793±0.164	0.790±0.147	0.789±0.114
TransUnet [46]	0.895±0.128	0.890±0.145	0.885±0.153	0.801±0.121	0.798±0.139	0.794±0.145
Ours	**0.912±0.115**	**0.893±0.104**	**0.894±0.127**	**0.840±0.102**	**0.809±0.095**	**0.810±0.116**

TABLE III
RESULTS OF 3D MESH RECONSTRUCTION ON THE TEST SET

Methods	CD(mm)↓	EMD(mm)↓	HD_{3d}(mm)↓	Dice↑	IoU↑	Laplace↓	Infer time(s)↓
DTTM [47]	5.924±3.129	1.341±0.460	12.419±3.254	0.852±0.054	0.746±0.080	0.072±0.007	55.052
NrICP [35]	6.386±3.555	1.422±0.525	12.344±3.514	0.838±0.069	0.728±0.100	0.252±0.019	71.519
NrCPD [48]	**3.573±1.222**	**0.677±0.151**	11.172±2.320	0.869±0.060	0.774±0.091	0.525±0.132	193.993
Ours	3.621±1.551	1.220±0.444	**8.912±2.851**	**0.895±0.025**	**0.810±0.040**	**0.067±0.008**	**1.512**

TABLE IV
SEGMENTATION RESULTS OF 3D MESH RECONSTRUCTION ON ALL SUBJECTS

Method	Dice↑		MCD(mm)↓		HD_{2d}(mm)↓	
	LV_{endo}	LV_{epi}	LV_{endo}	LV_{epi}	LV_{endo}	LV_{epi}
Bai et al. [49](n=600)	0.94 ± 0.04	0.88 ± 0.03	1.04 ± 0.35	1.14 ± 0.40	3.16 ± 0.98	3.92 ± 1.37
Attar et al. [14](n=600)	**0.94 ± 0.04**	0.87 ± 0.03	1.06 ± 0.35	1.13 ± 0.35	**3.15 ± 0.96**	3.90 ± 1.29
Xia et al. [25](n=600)	0.88 ± 0.05	0.78 ± 0.09	1.86 ± 0.87	1.86 ± 0.82	4.74 ± 1.75	4.75 ± 1.76
Ours(n=300)	0.92 ± 0.03	**0.89 ± 0.03**	**0.96 ± 0.26**	**0.92 ± 0.28**	3.94 ± 1.20	**3.88 ± 1.03**

TABLE V
CLINICAL INDEXES ON HEALTHY AND MI DATASETS

Index	UKBB(Healthy)					Private(MI)	
	GT(n=4875)	Bai et al.(n=600)	Attar et al.(n=600)	Xia et al.(n=600)	Ours(n=250)	GT(n=50)	Ours(n=50)
LVEDV(ml)	144 ± 34	144 ± 32	148 ± 35	145 ± 32	142 ± 33	143 ± 32	143 ± 32
LVESV(ml)	59 ± 20	57 ± 18	62 ± 20	59 ± 17	56 ± 16	67 ± 28	71 ± 24
LVEF(%)	59 ± 6	61 ± 7	58 ± 6	59 ± 5	60 ± 8	53 ± 11	50 ± 11
LVM(g)	90 ± 25	89 ± 25	91 ± 24	89 ± 20	87 ± 20	119 ± 20	120 ± 21

The test set consists of 20 subjects with corresponding image sequences. To generate the reference meshes, we first use the rview tool[6] to manually place six initial landmarks on the images. Subsequently, we apply CIMAS[7] (Cardiac Image Multi-Atlas Segmentation pipeline) [50] to segment cardiac MR image sequences at two time frames: ED and ES, using a multi-atlas segmentation method. Given an input image and the six initial landmarks, the CIMAS pipeline performs landmark registration, image registration, and label fusion to produce accurate segmentations and corresponding 3D dense meshes. We treat these output meshes as the ground truth and calculate metrics on dense mesh points across a test set of 40 samples. Table VI presents the CD, EMD, and HD_{3d} results of DTTM, NrICP, NrCPD and out method. Our method outperforms other three registration baselines across all three metrics, which indicates that even in the regions between slices, where there is no direct supervision during training, our method still has advantages in generating surfaces that closely approximate the left ventricular boundary.

E. Accuracy of cardiac function indexes

In this section, we report clinical cardiac function indexes of our method on the UKBB subset (healthy) and private dataset (MI). We processed and quantified volumes across all phases

[6] https://www.doc.ic.ac.uk/ dr/software/download.html
[7] https://github.com/baiwenjia/CIMAS

TABLE VI
RESULTS OF 3D MESH RECONSTRUCTION ON 40 TEST SAMPLES
CALCULATED ON DENSE MESH POINTS

	CD(mm)↓	EMD(mm)↓	HD_{3d}(mm)↓
DTTM [47]	7.577±3.416	1.442±0.503	14.219±3.648
NrICP [35]	7.661±3.573	1.468±0.542	13.859±3.267
NrCPD [48]	5.518±1.541	1.302±0.315	12.532±2.517
Ours	**4.676±1.815**	**1.267±0.465**	**9.811±2.785**

of 250 subjects in the UKBB subset and 50 subjects in the private dataset and reported the following sets of indexes for each subject: the LV end-diastolic volume (LVEDV), the LV end-systolic volume (LVESV), the LV ejection fraction (LVEF), and the LV myocardial mass (LVM). The assessment of these indexes is based on the following volumetric measurements [51].

- LVEDV(ml): the volume of blood in the LV cavity before contraction.
- LVESV(ml): the volume of blood in the LV cavity at the end of contraction.
- LVEF(%): the fraction of blood ejected from a ventricle of the heart with each heartbeat.

$$LVEF = \frac{LVEDV - LVESV}{LVEDV} \quad (12)$$

- LVM(g): the mass of the LV myocardium. The LVM is calculated by multiplying the myocardium volume by the
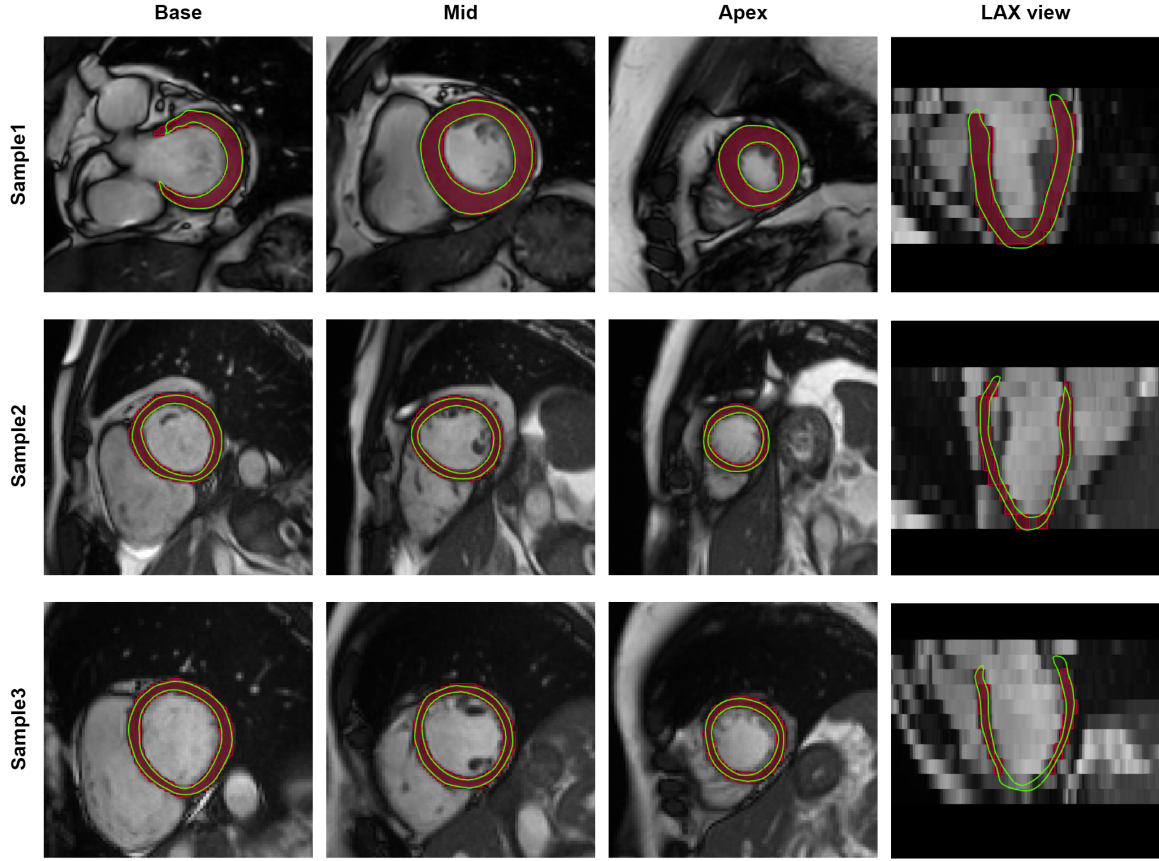
Fig. 2.   Example segmentation results for short-axis slices at the base, mid and apex positions. The GT segmentation masks are red and the segmentation contours by our method are green.

density of the muscle tissue (1.05 g/cm^3).

When calculating the volume, we begin by extracting contours that align with the intersection between our 3D triangular meshes and the SAX image slices. Afterward, we apply Simpson's integration method. This involves approximating the volume of the cardiac 3D structure by summing the areas enclosed within the 2D segmentation contours and then multiplying by the inter-plane spacing.

Table V shows the cardiac function indexes on two datasets. For the UKBB subset (healthy), we compare our results with Bai et al. [49], Attar et al. [14], Xia et al. [25], and the reference values provided by manual segmentation from Petersen et al. [52]. For the private dataset (MI), our results are compared with the reference values provided in the clinical reports from the hospital. Our observations reveal that for the UKBB subset, our clinical indexes exhibit minimal deviations from the ground truth values and demonstrate comparability with the other methodologies despite a smaller subject amount. For the private MI dataset, our proposed method aligns excellently with the ground truth values for LVEDV and LVM indexes. In the case of LVESV and LVEF indexes, the differences are also within acceptable bounds. A comparative analysis of healthy and MI subjects indicates that MI is associated with heightened LVESV and LVM and diminished LVEF, consistent with the findings of Xia et al. [25], which highlight cardiac remodeling and enlargement in MI patients.

F. 3D mesh reconstruction from incomplete contours

We further evaluate the robustness of our method to incomplete data in sparse 2D contours used for 3D mesh reconstruction. We randomly remove 1, 2, or 3 SAX image slices when evaluating the model to see the influences with different SAX slices removed. Additionally, to address the common issue encountered in routine CMR imaging of missing apical/basal information, we remove 2 LAX slices and keep SAX slices to see the results. To show the global effect of incomplete slices, we add the LV volume (LVV) as metrics for each evaluation setting.

The quantitative and qualitative results with different slices of data during inference are shown in Table VII and Fig. 5, respectively. Based on the results from Table VII, we can see that as more SAX slices are removed, the reconstruction metrics (CD, EMD, HD_{3d}, Dice, and IoU) all decrease. This is because removing more image slices leads to fewer image features being used to deform the mesh, resulting in poorer deformation in areas without image information. However, despite the incomplete SAX slices having lower metrics compared to complete SAX slices, we can observe that the LVV does not change significantly with different numbers of SAX slices. This suggests that the overall shape of the heart can still be reconstructed effectively. By comparing columns from 2 to 5 in Fig. 5, we can find that the shapes basically maintain consistency and the surfaces are all smooth, indicating that our
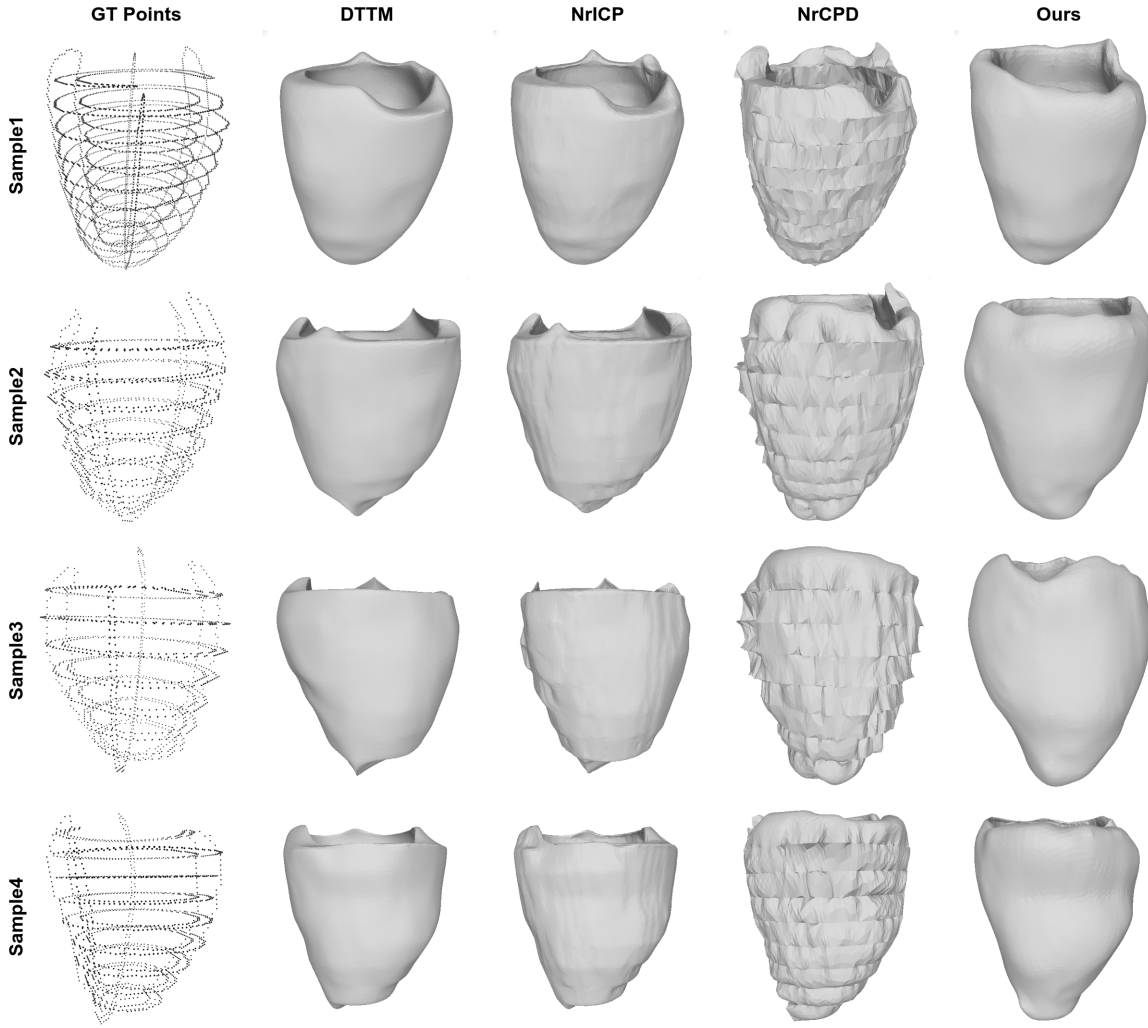
Fig. 3. Qualitative 3D reconstruction results of the myocardium by different methods for 4 test samples. The first column is the reference contour points. The second to fourth columns are the reconstruction results of the baseline methods, and the fifth column is the reconstruction results of our method.
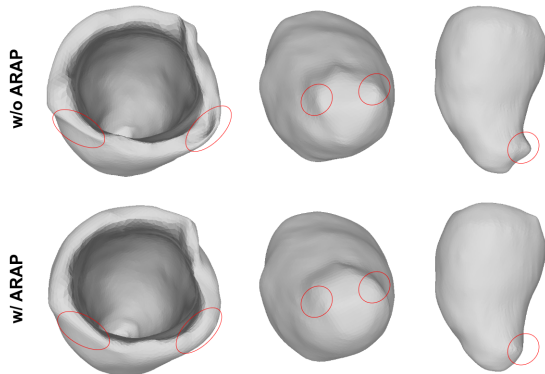

Fig. 4. Different results of three test samples with or without ARAP loss. The first row is the results without ARAP loss, and the second row is the results with ARAP loss.

method can generate accurate reconstructions of cardiac shape even in the presence of missing information. By comparing rows 1 and 5 in Table VII, it can be found that removing 2

LAX slices affects the performance of mesh reconstruction. However, our approach can still achieve a high-quality reconstruction effect in the basal and apical positions, as shown in columns 2 and 6 of Fig. 5, which is because the ARAP loss tries to keep the template shape in the basal and apical area even though there is no image information.

The robustness of our approach to missing slices implies we can reconstruct high-quality cardiac meshes using fewer annotated slices. This provides avenues to reduce scan time in the future.

G. 4D mesh generation and MI classification

Myocardial infarction is a prevalent and high-risk cardiovascular disease. Worldwide, about 15.9 million myocardial infarctions occurred in 2015 [53]. The hearts of healthy individuals exhibit regular and strong systoles and diastoles. In contrast, the hearts of MI patients display abnormal and weak movements. These differences can be better observed by analyzing 4D (3D+T) cardiac meshes with temporal information rather than just considering minimal wall thickness or

TABLE VII
QUANTITATIVE RESULTS WITH INCOMPLETE SLICES OF DATA ON THE TEST SET

| Methods	CD(mm)↓	EMD(mm)↓	HD_{3d}(mm)↓	Dice↑	IoU↑	Laplace↓	LVV(ml)
All slices	**3.621±1.551**	**1.220±0.444**	**8.912±2.851**	**0.895±0.025**	**0.810±0.040**	0.067±0.008	115±20
-1 SAX slice	4.406±1.765	1.307±0.430	9.064±2.737	0.880±0.028	0.797±0.045	0.067±0.008	114±22
-2 SAX slice	5.604±2.210	1.416±0.405	9.396±2.586	0.863±0.035	0.761±0.053	0.066±0.009	114±23
-3 SAX slice	6.818±2.598	1.545±0.430	9.649±2.356	0.845±0.043	0.734±0.064	0.066±0.009	112±23
No LAX slices	5.417±2.455	1.452±0.651	10.184±2.704	0.883±0.031	0.791±0.048	**0.064±0.008**	117±17
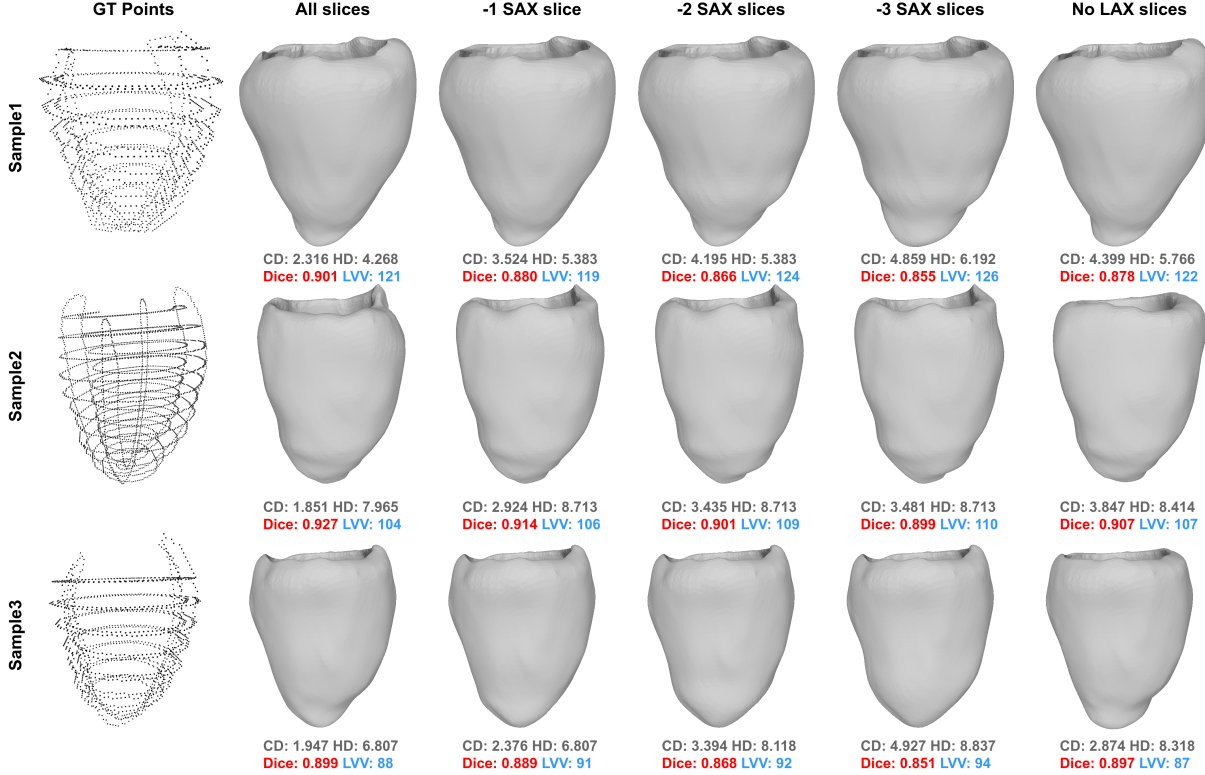
Fig. 5. 3D mesh reconstruction results with incomplete input during inference for three test samples. The first column shows the reference contour points and the second column shows the predicted shapes with all image slices. The third to fifth columns are the reconstruction results with different numbers of SAX slices removed, and the sixth column is the reconstruction results with LAX slices removed. The CD, HD, Dice and LVV are shown on the bottom of each sample. The Dice in red and LVV in blue do not change much as the number of slices decreases, demonstrating the robustness of our method.

deformation in a single cardiac cycle. Therefore, our method can be utilized to generate 4D meshes first and serve as a base for downstream MI classification tasks.

Firstly, we apply our method to 4D cine data consisting of images from multiple frames over the cardiac cycle for each subject to analyze the dynamic cardiac function. Accurate 4D cardiac mesh models can provide an intuitive visualization of cardiac motion and quantitative strain analysis to aid clinical practice, such as disease diagnosis and surgical planning. To this end, we employ the proposed method to reconstruct the 3D surface mesh at each frame of the cardiac cycle. We obtain all frames of all 50 subjects with different degrees of MI from the private dataset as abnormal data (label 1), randomly get 50 normal subjects with all frames in the UKBB dataset as normal data (label 0), and perform cardiac surface mesh reconstruction on the 4D data of these 100 subjects. Fig. 6 shows the cardiac reconstruction results for 9 frames of a 50-frame cardiac cycle for a normal subject.

Then, the 4D cardiac reconstruction results are classified

TABLE VIII
SVM CLASSIFICATION RESULTS ON 99 TEST SUBJECTS WITH
DIFFERENT REDUCED DIMENSIONS

Dimension	Accuracy(%)	Recall(%)	Precision(%)	F1(%)
2	63.6	59.2	64.4	61.7
3	61.6	46.9	65.7	54.7
4	**86.9**	**91.8**	**83.3**	**87.4**
5	79.8	91.8	73.8	81.8
6	84.8	91.8	80.3	85.7

into MI symptoms to distinguish MI patients from normal subjects. Specifically, we calculate the inter-phase displacement of the corresponding mesh vertices along the cardiac cycle, flatten the vectors for each subject, and conduct principal components analysis (PCA) with the linear kernel to reduce the dimensions. Using these 100 subjects of data, we train a linear SVM (Support vector machine) model to classify the symptoms. To evaluate the trained SVM model, we use a test set comprising 49 additional abnormal subjects from Beijing Anzhen Hospital and 50 additional normal subjects from the
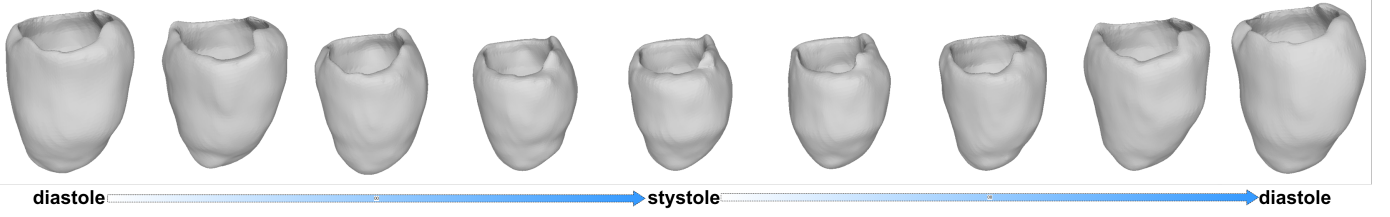
Fig. 6. 4D case study: Reconstruction results of 9 frames in the cardiac cycle of a normal subject, varying according to the process of diastole – systole – diastole.

TABLE IX
THE IMPACT OF THE CD LOSS CALCULATION METHOD AND FEATURE SAMPLING METHOD ON RECONSTRUCTION RESULTS

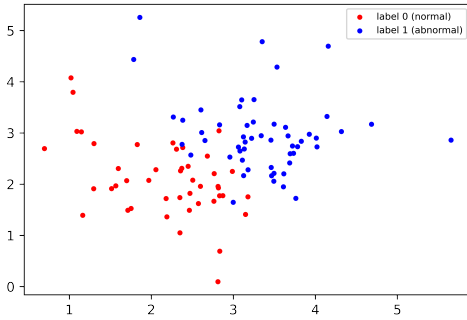| Methods	CD(mm)↓	EMD(mm)↓	HD(mm)↓	Dice↑	IoU↑	Laplace↓
TotalCD	3.716±1.630	1.220±0.396	9.081±3.163	0.892±0.025	0.806±0.041	**0.065±0.008**
PS	3.887±1.652	1.249±0.392	8.934±2.903	0.889±0.027	0.802±0.044	0.067±0.008
MNS	3.832±1.608	1.229±0.388	8.959±2.965	0.888±0.027	0.800±0.043	0.067±0.008
SeparateCD/WNS(Ours)	**3.621±1.551**	**1.220±0.444**	**8.912±2.851**	**0.895±0.025**	**0.810±0.040**	0.067±0.008

Fig. 7. The visualization results of the 4-dimensional data by using t-SNE to reduce the dimension to 2. Red scatters are normal subjects, and blue scatters are abnormal subjects with myocardial infarction.

UKBB dataset. We tried several dimensions to reduce to in PCA and finally got the best results with 4 dimensions. Table VIII shows the classification results of the 99 test subjects and reports the accuracy, recall rate, precision and F1 value at different reduced dimensions. At dimension 4, the four metrics reach the best level, and the recall(sensitivity) rate is 91.8%, which means that the SVM model can classify 45/49 abnormal subjects. Fig. 7 shows the visualization results of the 4-dimension data by using t-SNE (t-distributed Stochastic Neighbor Embedding) [54] to reduce the dimension to 2. Red scatters are normal subjects, and blue scatters are abnormal subjects with myocardial infarction. We can see that the two kinds of scatters can be easily separated by a linear function. This indicates that the trained SVM model can successfully classify the myocardial infarction symptoms of 4D cardiac reconstruction results and verify our method's accuracy and effectiveness for 3D cardiac reconstruction. We also compare our MI classification method based on 4D meshes displacement with a segment-level deep-learning-based baseline [55] and the traditional classification method of minimal wall thickness. The former one extracts local and global motion features from 2D images using a recurrent neural network and an optical flow method and reaches a sensitivity of 89.8%

(44/49) on midcavity segments, which is a little lower than ours, demonstrating our method's superiority. The latter one calculates the minimal wall thickness across all phases for each subject and conducts the classification by logistic regression, reaching the sensitivity of 85.7%(42/49). The gap may be because the displacement of mesh vertices over time, which is a vector, has more latent information to be used than the minimal wall thickness, which is a scalar.

H. Ablation study

An ablation study is performed to analyze the contribution of different components of our method. Table IX shows the influence of the calculation method of CD loss and the feature sampling method on the reconstruction results. TotalCD means we calculate the CD loss of the endocardium and epicardium as a whole, while SeparateCD means we calculate the CD loss of the endocardium and epicardium separately. PS is point sampling, MNS is mean neighborhood sampling, and WNS stands for weighted neighborhood sampling. Firstly, comparing the first and fourth rows, SeparateCD outperforms TotalCD in all metrics except for the slightly worse Laplacian smoothness metric. Then, comparing the second, third, and fourth rows, PS, MNS, and WNS gradually increase in all metrics, indicating that dynamically learning the weights of pixels in the neighborhood can sample the image features more accurately to deform the mesh more precisely.

IV. DISCUSSION AND CONCLUSION

In this study, we present Slice2Mesh, a novel image-based 3D surface reconstruction method for the left ventricle that effectively generates accurate and smooth 3D meshes using sparse image slices. Slice2Mesh surpasses three traditional point cloud registration methods on both accuracy and smoothness of the deformation. The clinical function indexes evaluated on the UKBB test set also underscore the method's precision. One key advantage of Slice2Mesh lies in its efficiency as a supervised method that does not rely on 3D cardiac meshes as references. By requiring only 2D segmentation labels to derive sparse contour points, it reduces

the computational burden and error introduced by generating 3D reference meshes, offering a streamlined alternative to existing deep learning and traditional shape reconstruction techniques. We further demonstrate that our proposed approach can reconstruct precise 3D cardiac meshes even when several SAX and LAX slices of images are missing, showcasing our model's robustness and the effectiveness of the introduced ARAP loss, which plays a critical role in ensuring stable reconstructions even under incomplete data, reinforcing the versatility of the method. Moreover, the method can produce topology-consistent and feature-corresponding 3D shapes by consistently mapping mesh vertices on the templates to different shapes of the same heart. Consequently, this method can display fluent and dynamic changes of the heart and is effective in identifying subjects suffering from myocardial infarction, potentially enabling the efficient construction of a 4D dynamic heart model that captures the motion of a beating heart from time-series cine data. The robustness of Slice2Mesh in reconstructing 3D meshes from sparse data, combined with its potential to integrate with 4D dynamic models, suggests that it could serve as a powerful tool for clinical applications requiring detailed cardiac modeling.

However, some limitations and challenges still remain. Since the method is supervised by contour points derived from segmentation labels, the misalignment between slices during data acquisition may affect the reconstruction quality. Although we apply contour-based misalignment correction, this process is not flawless, and residual misalignments may impact the smoothness of the reconstructed surfaces, particularly in the basal and apical regions. Addressing these limitations in future work could involve more advanced techniques, such as intensity-based or deep learning-based alignment corrections. Moreover, the current pipeline reconstructs 3D meshes individually for each phase using ICP registration with only one initial template, which could result in suboptimal mesh deformation and dynamic motion capture. Using multiple templates to select from may improve the deformation accuracy. Incorporating interframe image or mesh registration would likely improve the continuity and accuracy of dynamic heart motion.

REFERENCES

[1] M. B. Stokes and R. Roberts-Thomson, "The role of cardiac imaging in clinical practice," *Australian prescriber*, vol. 40, no. 4, p. 151, 2017.

[2] X. Yuan, C. Liu, F. Feng, Y. Zhu, and Y. Wang, "Slice-Mask Based 3D Cardiac Shape Reconstruction from CT Volume," in *Computer Vision – ACCV 2022*, L. Wang, J. Gall, T.-J. Chin, I. Sato, and R. Chellappa, Eds. Springer Nature Switzerland, 2023, vol. 13846, pp. 69–85.

[3] M. Vukicevic, B. Mosadegh, J. K. Min, and S. H. Little, "Cardiac 3d printing and its future directions," *JACC: Cardiovascular Imaging*, vol. 10, no. 2, pp. 171–184, 2017.

[4] A. Suinesiaputra, P. Ablin, X. Alba, M. Alessandrini, J. Allen, W. Bai, S. Cimen, P. Claes, B. R. Cowan, J. D'hooge *et al.*, "Statistical shape modeling of the left ventricle: myocardial infarct classification challenge," *IEEE journal of biomedical and health informatics*, vol. 22, no. 2, pp. 503–515, 2017.

[5] Y.-H. Loke, F. Capuano, E. Balaras, and L. J. Olivieri, "Computational Modeling of Right Ventricular Motion and Intracardiac Flow in Repaired Tetralogy of Fallot," *Cardiovascular Engineering and Technology*, vol. 13, no. 1, pp. 41–54, 2022.

[6] L. Bennati, V. Giambruno, F. Renzi, V. Di Nicola, C. Maffeis, G. Puppini, G. B. Luciani, and C. Vergara, "Turbulent blood dynamics in the left heart in the presence of mitral regurgitation: a computational study based on multi-series cine-MRI," *Biomechanics and Modeling in Mechanobiology*, vol. 22, no. 6, pp. 1829–1846, 2023.

[7] D. Collia, L. Zovatto, G. Tonti, and G. Pedrizzetti, "Comparative Analysis of Right Ventricle Fluid Dynamics," *Frontiers in Bioengineering and Biotechnology*, vol. 9, p. 667408, 2021.

[8] I. Fumagalli, P. Vitullo, C. Vergara, M. Fedele, A. F. Corno, S. Ippolito, R. Scrofani, and A. Quarteroni, "Image-Based Computational Hemodynamics Analysis of Systolic Obstruction in Hypertrophic Cardiomyopathy," *Frontiers in Physiology*, vol. 12, p. 787082, 2022.

[9] C. Chnafa, S. Mendez, and F. Nicoud, "Image-Based Simulations Show Important Flow Fluctuations in a Normal Left Ventricle: What Could be the Implications?" *Annals of Biomedical Engineering*, vol. 44, no. 11, pp. 3346–3358, 2016.

[10] H. Xu, M. Muffoletto, S. A. Niederer, S. E. Williams, M. C. Williams, and A. A. Young, "Whole Heart 3D Shape Reconstruction from Sparse Views: Leveraging Cardiac Computed Tomography for Cardiovascular Magnetic Resonance," in *Functional Imaging and Modeling of the Heart*, O. Bernard, P. Clarysse, N. Duchateau, J. Ohayon, and M. Viallon, Eds. Springer Nature Switzerland, 2023, vol. 13958, pp. 255–264.

[11] H. C. Van Assen, M. G. Danilouchkine, A. F. Frangi, S. Ordás, J. J. Westenberg, J. H. Reiber, and B. P. Lelieveldt, "SPASM: A 3D-ASM for segmentation of sparse and arbitrarily oriented cardiac MRI data," *Medical Image Analysis*, vol. 10, no. 2, pp. 286–303, 2006.

[12] B. Villard, V. Grau, and E. Zacur, "Surface mesh reconstruction from cardiac mri contours," *Journal of Imaging*, vol. 4, no. 1, p. 16, 2018.

[13] F. Odille, A. Bustin, S. Liu, B. Chen, P.-A. Vuissoz, J. Felblinger, and L. Bonnemains, "Isotropic 3 D cardiac cine MRI allows efficient sparse segmentation strategies based on 3 D surface reconstruction," *Magnetic Resonance in Medicine*, vol. 79, no. 5, pp. 2665–2675, 2018.

[14] R. Attar, M. Pereañez, A. Gooya, X. Albà, L. Zhang, M. H. de Vila, A. M. Lee, N. Aung, E. Lukaschuk, M. M. Sanghvi *et al.*, "Quantitative cmr population imaging on 20,000 subjects of the uk biobank imaging study: Lv/rv quantification pipeline and its evaluation," *Medical image analysis*, vol. 56, pp. 26–42, 2019.

[15] H. Hu, N. Pan, and A. F. Frangi, "Fully automatic initialization and segmentation of left and right ventricles for large-scale cardiac mri using a deeply supervised network and 3d-asm," *Computer Methods and Programs in Biomedicine*, vol. 240, p. 107679, 2023.

[16] X. Chen, N. Ravikumar, Y. Xia, R. Attar, A. Diaz-Pinto, S. K. Piechnik, S. Neubauer, S. E. Petersen, and A. F. Frangi, "Shape registration with learned deformations for 3d shape reconstruction from sparse and incomplete point clouds," *Medical Image Analysis*, vol. 74, p. 102228, 2021.

[17] C. Biffi, J. J. Cerrolaza, G. Tarroni, A. de Marvao, S. A. Cook, D. P. O'Regan, and D. Rueckert, "3d high-resolution cardiac segmentation reconstruction from 2d views using conditional variational autoencoders," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 1643–1646.

[18] Q. Chang, Z. Yan, M. Zhou, D. Liu, K. Sawalha, M. Ye, Q. Zhangli, M. Kanski, S. Al'Aref, L. Axel *et al.*, "Deeprecon: Joint 2d cardiac segmentation and 3d volume reconstruction via a structure-specific generative method," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part IV*. Springer, 2022, pp. 567–577.

[19] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," *ACM siggraph computer graphics*, vol. 21, no. 4, pp. 163–169, 1987.

[20] M. Beetz, A. Banerjee, and V. Grau, "Point2mesh-net: Combining point cloud and mesh-based deep learning for cardiac shape reconstruction," in *International Workshop on Statistical Atlases and Computational Models of the Heart*. Springer, 2022, pp. 280–290.

[21] M. Beetz, A. Banerjee, and V. Grau, "Reconstructing 3d cardiac anatomies from misaligned multi-view magnetic resonance images with mesh deformation u-nets," in *Geometric Deep Learning in Medical Image Analysis*. PMLR, 2022, pp. 3–14.

[22] W. Bai, W. Shi, A. de Marvao, T. J. Dawes, D. P. O'Regan, S. A. Cook, and D. Rueckert, "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*, vol. 26, no. 1, pp. 133–145, 2015.

[23] S. E. Petersen, P. M. Matthews, F. Bamberg, D. A. Bluemke, J. M. Francis, M. G. Friedrich, P. Leeson, E. Nagel, S. Plein, F. E. Rademakers *et al.*, "Imaging in population science: cardiovascular magnetic resonance in 100,000 participants of uk biobank-rationale, challenges and

approaches," *Journal of Cardiovascular Magnetic Resonance*, vol. 15, no. 1, pp. 1–10, 2013.

[24] K. Tóthová, S. Parisot, M. Lee, E. Puyol-Antón, A. King, M. Pollefeys, and E. Konukoglu, "Probabilistic 3d surface reconstruction from sparse mri information," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23*. Springer, 2020, pp. 813–823.

[25] Y. Xia, X. Chen, N. Ravikumar, C. Kelly, R. Attar, N. Aung, S. Neubauer, S. E. Petersen, and A. F. Frangi, "Automatic 3d+ t four-chamber cmr quantification of the uk biobank: integrating imaging and non-imaging data priors at scale," *Medical Image Analysis*, vol. 80, p. 102498, 2022.

[26] S. R. Mehta, J. W. Eikelboom, M. K. Natarajan, R. Diaz, C. Yi, R. J. Gibbons, and S. Yusuf, "Impact of right ventricular involvement on mortality and morbidity in patients with inferior myocardial infarction," *Journal of the American College of Cardiology*, vol. 37, no. 1, pp. 37–43, 2001.

[27] T. Stiermaier, S. J. Backhaus, J. Matz, A. Koschalka, J. Kowallick, S. de Waha-Thiele, S. Desch, M. Gutberlet, G. Hasenfuß, H. Thiele *et al.*, "Frequency and prognostic impact of right ventricular involvement in acute myocardial infarction," *Heart*, vol. 107, no. 7, pp. 563–570, 2021.

[28] J. Duan, G. Bello, J. Schlemper, W. Bai, T. J. Dawes, C. Biffi, A. de Marvao, G. Doumoud, D. P. O'Regan, and D. Rueckert, "Automatic 3d bi-ventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach," *IEEE transactions on medical imaging*, vol. 38, no. 9, pp. 2151–2164, 2019.

[29] Q. Meng, W. Bai, D. P. O'Regan, and D. Rueckert, "Deepmesh: Mesh-based cardiac motion tracking using deep learning," *IEEE transactions on medical imaging*, 2023.

[30] B. J. Krenning, M. M. Voormolen, and J. R. Roelandt, "Assessment of left ventricular function by three-dimensional echocardiography," *Cardiovascular Ultrasound*, vol. 1, no. 1, p. 12, 2003.

[31] O. Sorkine and M. Alexa, "As-rigid-as-possible surface modeling," in *Symposium on Geometry processing*, vol. 4, 2007, pp. 109–116.

[32] B. Villard, E. Zacur, E. Dall'Armellina, and V. Grau, "Correction of slice misalignment in multi-breath-hold cardiac mri scans," in *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges: 7th International Workshop, STACOM 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers 7*. Springer, 2017, pp. 30–38.

[33] A. Banerjee, J. Camps, E. Zacur, C. M. Andrews, Y. Rudy, R. P. Choudhury, B. Rodriguez, and V. Grau, "A completely automated pipeline for 3d reconstruction of human heart from 2d cine magnetic resonance slices," *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2212, p. 20200257, 2021.

[34] D. N. Metaxas and Z. Yan, "Deformable models, sparsity and learning-based segmentation for cardiac mri based analytics," in *Handbook of Medical Image Computing and Computer Assisted Intervention*. Elsevier, 2020, pp. 273–292.

[35] B. Amberg, S. Romdhani, and T. Vetter, "Optimal step nonrigid icp algorithms for surface registration," in *2007 IEEE conference on computer vision and pattern recognition*. IEEE, 2007, pp. 1–8.

[36] O. Sorkine, "Laplacian Mesh Processing," in *Eurographics 2005 - State of the Art Reports*, Y. Chrysanthou and M. Magnor, Eds. The Eurographics Association, 2005.

[37] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, G. Ranzuglia *et al.*, "Meshlab: an open-source mesh processing tool." in *Eurographics Italian chapter conference*, vol. 2008. Salerno, Italy, 2008, pp. 129–136.

[38] P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in *Sensor fusion IV: control paradigms and data structures*, vol. 1611. Spie, 1992, pp. 586–606.

[39] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[40] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters–improve semantic segmentation by global convolutional network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4353–4361.

[41] N. Wang, Y. Zhang, Z. Li, Y. Fu, W. Liu, and Y.-G. Jiang, "Pixel2mesh: Generating 3d mesh models from single rgb images," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 52–67.

[42] C. Wen, Y. Zhang, Z. Li, and Y. Fu, "Pixel2mesh++: Multi-view 3d mesh generation via deformation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1042–1051.

[43] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[44] Q. Huang, X. Huang, B. Sun, Z. Zhang, J. Jiang, and C. Bajaj, "Arapreg: An as-rigid-as possible regularization loss for learning deformable shape generators," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 5815–5825.

[45] D. H. Pak, M. Liu, T. Kim, L. Liang, R. McKay, W. Sun, and J. S. Duncan, "Distortion energy for deep learning-based volumetric finite element mesh generation for aortic valves," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24*. Springer, 2021, pp. 485–494.

[46] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[47] R. W. Sumner and J. Popović, "Deformation transfer for triangle meshes," *ACM Transactions on graphics (TOG)*, vol. 23, no. 3, pp. 399–405, 2004.

[48] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 12, pp. 2262–2275, 2010.

[49] W. Bai, M. Sinclair, G. Tarroni, O. Oktay, M. Rajchl, G. Vaillant, A. M. Lee, N. Aung, E. Lukaschuk, M. M. Sanghvi *et al.*, "Automated cardiovascular magnetic resonance image analysis with fully convolutional networks," *Journal of cardiovascular magnetic resonance*, vol. 20, no. 1, p. 65, 2018.

[50] W. Bai, W. Shi, D. P. O'regan, T. Tong, H. Wang, S. Jamil-Copley, N. S. Peters, and D. Rueckert, "A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac mr images," *IEEE transactions on medical imaging*, vol. 32, no. 7, pp. 1302–1315, 2013.

[51] A. F. Frangi, W. J. Niessen, and M. A. Viergever, "Three-dimensional modeling for functional analysis of cardiac images, a review," *IEEE transactions on medical imaging*, vol. 20, no. 1, pp. 2–5, 2001.

[52] S. E. Petersen, N. Aung, M. M. Sanghvi, F. Zemrak, K. Fung, J. M. Paiva, J. M. Francis, M. Y. Khanji, E. Lukaschuk, A. M. Lee *et al.*, "Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (cmr) in caucasians from the uk biobank population cohort," *Journal of cardiovascular magnetic resonance*, vol. 19, no. 1, p. 18, 2016.

[53] T. Vos, C. Allen, M. Arora, R. M. Barber, Z. A. Bhutta, A. Brown, A. Carter, D. C. Casey, F. J. Charlson, A. Z. Chen *et al.*, "Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the global burden of disease study 2015," *The lancet*, vol. 388, no. 10053, pp. 1545–1602, 2016.

[54] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.

[55] N. Zhang, G. Yang, Z. Gao, C. Xu, Y. Zhang, R. Shi, J. Keegan, L. Xu, H. Zhang, Z. Fan *et al.*, "Deep learning for diagnosis of chronic myocardial infarction on nonenhanced cardiac cine mri," *Radiology*, vol. 291, no. 3, pp. 606–617, 2019.