

4D Myocardium Reconstruction with Decoupled Motion and Shape Model

Xiaohan Yuan Cong Liu Yangang Wang*

Southeast University, China

Abstract

Estimating the shape and motion state of the myocardium is essential in diagnosing cardiovascular diseases. However, cine magnetic resonance (CMR) imaging is dominated by 2D slices, whose large slice spacing challenges inter-slice shape reconstruction and motion acquisition. To address this problem, we propose a 4D reconstruction method that decouples motion and shape, which can predict the inter-/intra- shape and motion estimation from a given sparse point cloud sequence obtained from limited slices. Our framework comprises a neural motion model and an end-diastolic (ED) shape model. The implicit ED shape model can learn a continuous boundary and encourage the motion model to predict without the supervision of ground truth deformation, and the motion model enables canonical input of the shape model by deforming any point from any phase to the ED phase. Additionally, the constructed ED-space enables pre-training of the shape model, thereby guiding the motion model and addressing the issue of data scarcity. We propose the first 4D myocardial dataset as we know and verify our method on the proposed, public, and cross-modal datasets, showing superior reconstruction performance and enabling various clinical applications.

1. Introduction

Heart attack, also known as myocardial infarction, is the leading cause of death worldwide [1]. Estimating the shape and motion state of the myocardium is an essential step in diagnosing cardiovascular diseases. Cardiac magnetic resonance (CMR) imaging, which can provide successive sequence images including end-diastolic (ED) and end-systolic (ES) phases, is considered the gold standard for cardiac morphological assessment.

*Corresponding author. E-mail: yangangwang@seu.edu.cn. All the authors from Southeast University are affiliated with the Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Nanjing, China. This work was supported in part by the National Natural Science Foundation of China (No. 62076061), the Natural Science Foundation of Jiangsu Province (No. BK20220127).

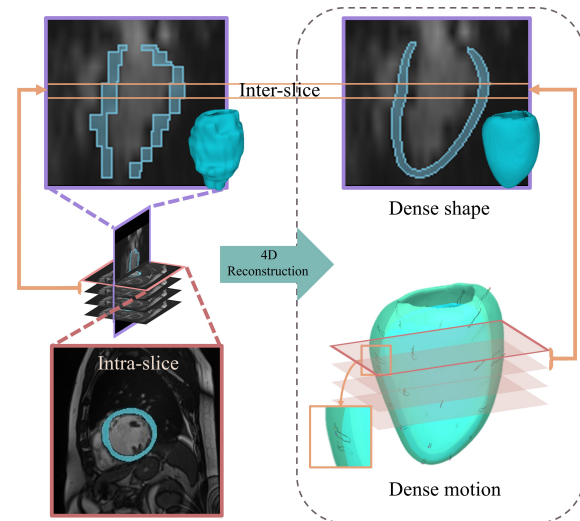


Figure 1. We propose a 4D cardiac reconstruction method predicting the inter-/intra- shape and motion of a cardiac cycle.

However, the standard CMR has a large slice spacing, which leads to the lack of **inter-slice** information as shown in Fig. 1. Meanwhile, the shape and motion observation of the ventricle for clinicians are almost guided by high-resolution two-dimensional slices based on **intra-slice** information, such as short-axis (SAX) or long-axis (LAX) views. The dense and faithful 4D (inter-/intra-) reconstruction of the myocardium would assist medical professionals in intuitively observing and accurately obtaining diverse clinical parameters [3].

The road to inter-slice reconstruction is particularly challenging, as it involves estimating both inter-slice motion and shape. The lack of inter-slice motion data limits the applicability of supervised methods, thus complicating the problem. The large slice thickness of CMR leads to the inaccuracy of direct linear interpolation of pixels between slices, hindering the accurate reconstruction of the inter-slice shape. Few works have considered or fully addressed this problem. Most recent works [43, 34, 42, 46] focus only on motion estimation based on 2D plane registration, ignoring the motion between slices and failing to reflect the myocardium movement truly. Others focus solely on 3D shape reconstruction of a specific phase [4, 21, 19], without con-

sidering motion information. A few methods that attempt simultaneous shape and motion reconstruction [14, 26] have limited resolution and cannot obtain dense point correspondence due to the complexity of cardiac motion.

In this work, we advocate to solve the inter-slice and intra-slice reconstruction as 4D myocardium reconstruction. Our key idea is to **decouple the problem as two steps including motion and shape reconstruction with sparse point clouds as inputs obtained from sequential CMR slices**. Inspired by the implicit neural representation [32, 27, 12, 47], we follow the point sampling strategy for both the motion model and shape model. According to the characteristics of the heart sequence, we build the meaningful **End-Diastolic-space** (ED-space) that consists of a statistical parameter space representing the shape prior and a canonical spatial space suitable for model input. This space enables us to pre-train a reliable implicit shape model, which can learn a continuous boundary and guide the implicit motion model to obtain a dense motion field without the supervision of ground truth deformation.

Specifically, for any query point under a certain phase, the motion model takes its coordinate and phase indicator as input, estimating the deformation of the point to the ED spatial space under the condition of the motion code. The pre-trained shape model estimates the signed distance function (SDF) value of the deformed point under the condition of shape code in the ED spatial space. Finally, the shape reconstruction is completed by extracting the boundary (where SDF value = 0). Our method can be applied to multiple medical image analysis tasks such as point cloud completion, dense motion estimation, and motion interpolation. Furthermore, it is possible to extend our method to other modes of medical imaging, such as CT.

The main contributions of this paper are summarized as follows:

- We propose a new 4D myocardium reconstruction framework via the decoupled motion and shape features.
- We build a ED shape model based on ED-space to alleviate the problem of medical data scarcity.
- We present the first, to our knowledge, 4D myocardial dataset composed of 3D shape sequence and the dataset can be downloaded from <https://github.com/yuan-xiaohan/4D-Myocardium-Reconstruction-with-Decoupled-Motion-and-Shape-Model>.

2. Related work

2.1. Implicit representation

Since the signed distance function (SDF) is a continuous function, implicit functions can represent more detailed and complicated information than mesh-based shape representation and are suitable for generating models [32]. In re-

cent years, combining SDF with deep learning to represent shape has attracted broad interest [15, 12, 47]. Occupancy Flow extends 3D occupancy functions into the 4D domain by assigning a motion vector to every point in space and time [29]. By taking advantage of continuity of representation and delicate reconstruction, spatio-temporal implicit functions have been widely used in dynamic humans and clothing modeling [11, 28, 37, 30, 18]. However, there is few attempt to work on medical imaging at present. The application of implicit function in the medical field still has great potential to be tapped. DISSMs [35] present a new approach that marries the representation power of deep networks with the benefits of SSMs by using an implicit representation. ImplicitAtlas [40] and NDF [36] both learn deep implicit shape templates. The former integrates multiple templates and represents shapes by deforming one of several learned templates. The latter represents the shape as the conditional differential deformation of the template to maintain the shape topology. Only static organs are considered in these works, while cardiac movement needs to consider time-related deformability.

The decoupling of motion and shape codes sets our method apart from other implicit methods and the key benefits include (1) Easier learning and understanding of features related to motion and shape; (2) Better utilization of diverse modal data (CMR and CT share a consistent shape code space; motion knowledge learned from CMR can be interpolated for CT); (3) The decoupled shape codes offer advantages in pre-training the shape model. We define an interpretable canonical space, representing the spatial space at the ED phase, which is distinct from the canonical space used in template-based methods like DIT [47]. This allows us to employ a data augmentation strategy to pre-train the ED shape model, thereby compensating for the data scarcity, as the training data is derived from the existing readily sampled statistical parameter space.

2.2. Medical shape reconstruction

Shape reconstruction methods for the medical image can be divided into parametric-based and deformation-based methods. The former needs to use a large amount of data to build the parameter space. It relies on accurate registration, mainly focuses on iterative optimization, and requires a long reference time [13, 16, 5]. With the development of deep learning, the combination of SSM and convolutional neural networks (CNNs) can achieve better results [4, 48, 2]. However, the lack of medical image data is still the main reason for limiting such methods. The deformation-based method can directly predict the surface mesh of organs from image or volume data [24, 41, 38]. In recent years, the graph convolution neural network also shows the prospect of surface reconstruction [39, 20]. This kind of structured representation needs to build a fixed-

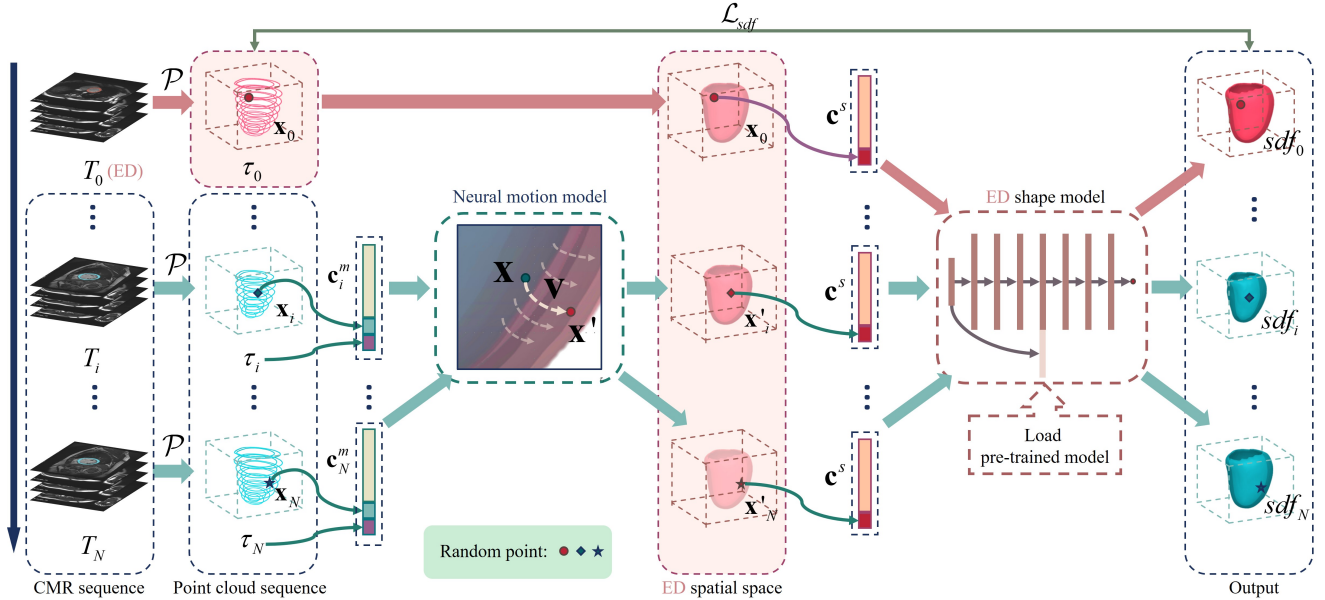


Figure 2. **Method Overview.** The point cloud sequence is first segmented from the input CMR sequence. We register the point cloud at the ED phase to the pre-defined statistical mean shape \bar{s} in the ED spatial space and broadcast the obtained transformation to all phases. For any query point at T_i phase, the neural motion model takes its coordinate and phase indicator as input and predicts the point’s deformation to the ED spatial space under the condition of motion code. The pre-trained implicit ED shape model estimates the SDF value of the deformed point under the condition of shape code. Finally, the shape reconstruction is completed by extracting the boundary.

topology template, which is difficult to obtain dense correspondences and is easily limited by the complexity of the shape. In addition, the reconstruction of complete shapes from sparse point clouds from MR, namely point cloud completion, is also a problem worth studying [7, 10].

2.3. Cardiac motion estimation

Motion estimation is to establish a corresponding deformation field between two phases, and image-based registration combined with deep learning is a more effective method at present. VoxelMorph [6] is a classic unsupervised registration network in medical images, which has served as a foundation for many follow-up studies aimed at its enhancement. Most of the current work is about 2D motion reconstruction [33, 22, 34, 45, 9, 44, 46]. Other methods [14, 25, 26] are carried out on a 3D motion reconstruction. Based on the idea of dense-sparse-dense (DSD), Guo *et al.* [14] extract sparse landmarks from the original image, then calculate the sparse displacement according to the detected sparse landmarks. Finally, they project the sparse motion displacement back to the dense image domain through the motion reconstruction network. However, this motion estimation accuracy will increase with the number of landmarks, and the detailed structural changes cannot be characterized when the number of landmarks is small. In order to solve the problem of cross-plane motion, a 3D motion field is estimated in [26] by fusing the information of LAX and SAX views in latent space, and a shape regular-

ization module is introduced to constrain the consistency of the 3D motion field. Meng *et al.* [25] also model the heart as a 3D geometric mesh and estimate the 3D motion of the cardiac mesh from the multi-view images. However, mesh-based representation requires dense vertices to represent the complex surface details of the heart.

3. Method

The overview of our proposed method is shown in Fig. 2. We decouple the task into a two-step subproblem, including motion reconstruction and ED shape reconstruction, where motion reconstruction focuses on deforming the point at any phase into the ED phase, and ED shape reconstruction can facilitate the reconstruction loss computation. It is noted that the raw input CMR slices might have different directions and positions due to the diversity of data acquisition. Point clouds should be pre-processed and registered. At the beginning, we register the point cloud at the ED phase with the pre-defined statistical mean shape \bar{s} to compute a 4×4 matrix \mathcal{P} (details are shown in [Sup. Mat](#)). After that, \mathcal{P} is applied to the whole sequence to transform all point clouds (shown in Fig. 2) as the sampling input for the neural motion model. For any query point at T_i phase, the motion model (Sec. 3.1) predicts its deformation relative to that point in ED-space (Sec. 3.2.1) under the condition of the motion code. Then, the pre-trained ED shape model (Sec. 3.2.2) estimates the SDF value sdf under the condition of the shape

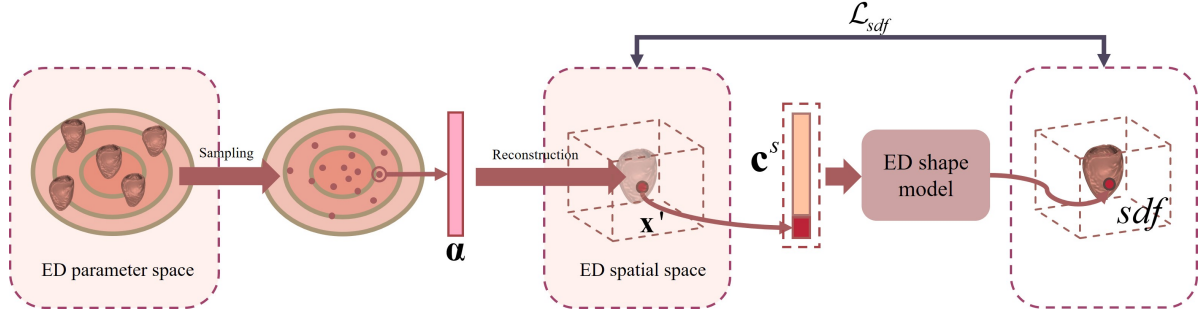


Figure 3. **Pipeline for ED shape model pre-training.** A shape parameter vector is sampled from the ED parameter space to reconstruct a new shape. This reconstructed shape is then normalized into the ED spatial space, and the pre-trained ED shape model is obtained by sampling points within the ED spatial space.

code. Finally, we can easily obtain the myocardium reconstruction mesh via Marching Cubes [23]. Before we train the framework, the constructed ED parameter space is sampled for data augmentation to improve the generalization performance of the shape model.

3.1. Neural motion model

A neural motion model is formulated to predict the deformation of a given point at arbitrary T_i phase to the end-diastole (ED) phase (*i.e.*, $T_0 = 0$). To construct the motion model, all points should be registered and sampled in a canonical space. We align the input ED phase point clouds with a statistical mean shape $\bar{\mathbf{s}}$ defined in Sec. 3.2.1 and apply the obtained transformation matrix \mathcal{P} to all phases. For a sampling point $\mathbf{x} \in \mathbb{R}^3$ in the canonical space, its corresponding deformation \mathbf{v} is predicted under the condition of motion code $\mathbf{c}_i^m \in \mathbb{R}^{K^m}$ and phase indicator τ_i :

$$f^m : (\mathbf{x}, \mathbf{c}_i^m, \tau_i) \in \mathbb{R}^{1+3+K^m} \longrightarrow \mathbf{v} \in \mathbb{R}^3, \quad (1)$$

where K^m is the dimension of the motion code. $\tau_i = T_i/T_N \in [0, 1]$, T_i is the i -th phase of CMR sequence, T_N is the sequence length, and $T_0 = 0$ is for ED phase. It is noted that the motion code \mathbf{c}_i^m is variant for each i -th phase. With the proposed neural motion model, the corresponding point \mathbf{x}' in ED spatial space can be obtained by $\mathbf{x}' = \mathbf{x} + \mathbf{v}$. Our proposed motion model can obtain the deformation from any point at the T_i phase to the ED phase; thus, we can easily establish dense correspondences. To train the motion model in an unsupervised manner, we utilize the reconstruction loss to compare the differences between the input point cloud and the output of the shape model. In order to better guide the motion model, it is crucial to have accurate SDF value predictions. Next, we will incorporate a pre-trained shape model.

3.2. ED shape model

3.2.1 ED-space building

As the first frame of the heart sequence typically represents the pivotal ED phase, we select this frame as the reference. Based on the unique characteristics of the heart sequence, we construct the ED-space that consists of a statistical parameter space representing the shape prior and a canonical spatial space dedicated to the heart as the input space of the shape model (as shown in Fig. 3).

To build the ED parameter space, we introduce the left ventricular ED statistical shape model (SSM) [5], obtained from high-resolution MR images of more than 1,000 healthy subjects, describing individual shape change. For the given shape parameter vector $\alpha \in \mathbb{R}^{K^\alpha}$, we can get the corresponding 3D shape through the following function:

$$\mathbf{S} = \mathcal{M}(\bar{\mathbf{s}} + \mathbf{V}^s \alpha), \quad (2)$$

where K^α is the dimension of the parameter vector, N is the number of vertices, $\mathbf{V}^s \in \mathbb{R}^{3N \times K^\alpha}$ is obtained through the singular value decomposition (SVD) of the atlas map, $\bar{\mathbf{s}}$ is the statistical mean shape, and the operator $\mathcal{M}(\cdot) : \mathbb{R}^{3N} \mapsto \mathbb{R}^{3 \times N}$ maps the parameter vector to the 3D shape. To improve the training of the ED shape model, we generate a multivariate Gaussian distribution centered on real samples for sampling to augment. The newly obtained parameters differ from the original atlas while maintaining anatomical consistency and fidelity. All new shapes obtained from Eq. 2 are naturally registered with the mean shape $\bar{\mathbf{s}}$, and they can be directly normalized into a canonical space called ED spatial space. At the beginning of the whole framework, all our point cloud inputs will be aligned to this space for sampling.

3.2.2 Shape model pre-training

To improve the accuracy of the ED shape model, we use a large number of shapes obtained through the above data

augmentation strategy to pre-train an implicit shape function for the ED-space. As shown in Fig. 3, for a query point $\mathbf{x} \in \mathbb{R}^3$ in the ED spatial space, the corresponding SDF value sdf can be predicted through an MLP network based on the shape code $\mathbf{c}^s \in \mathbb{R}^{K^s}$:

$$f^s : (\mathbf{c}^s, \mathbf{x}) \in \mathbb{R}^{K^s+3} \rightarrow sdf \in \mathbb{R}. \quad (3)$$

where K^s is the dimension of the shape code. The implicit shape model learns the shape code \mathbf{c}^s of a given object, and the weight of network f^s in the fashion of automatic decoder [31] simultaneously. Note that \mathbf{c}^s is shared in the same sequence for all the points. After querying the predicted sdf of each vertex in the spatial grid, the decision boundary of zero isosurfaces is extracted. The final mesh is obtained using Marching Cubes [23]. The obtained pre-trained model will be applied to the joint framework for fine-tuning.

3.3. Loss functions

Given the shapes of a motion sequence, the SDF regression loss is applied:

$$\mathcal{L}_{sdf} = \sum_{i=1}^S \sum_{l=1}^{T_N} \sum_{j \in \Omega} |f(\mathbf{x}_j) - sdf_j|, \quad (4)$$

where S and T_N represent the number and the length of sequences respectively, Ω denotes points in 3D space, $f = f^s(f^m(\cdot))$ describes the process of point \mathbf{x}_j passing through the motion model and the ED shape model in turn, and sdf_j is the ground-truth SDF value. This loss also provides vital guidance for the neural motion model.

In addition, We use the regularization terms [47] \mathcal{L}_{pw} and \mathcal{L}_{pp} to prevent unrealistic large distortion:

$$\mathcal{L}_{pw} = \sum_{i=1}^S \sum_{l=1}^{T_N} \sum_{j \in \Omega} h(\|f^m(\mathbf{c}_l^m, \mathbf{x}_j, \tau_j) - \mathbf{x}_j\|_2), \quad (5)$$

where $h(\cdot)$ is the Huber kernel, and

$$\mathcal{L}_{pp} = \sum_{i=1}^S \sum_{l=1}^{T_N} \sum_{j,k \in \Omega, j \neq k} \max\left(\frac{\|\hat{\mathbf{v}}_j - \hat{\mathbf{v}}_k\|_2}{\|\mathbf{x}_j - \mathbf{x}_k\|_2} - \epsilon, 0\right), \quad (6)$$

where ϵ is a parameter controlling the distortion tolerance.

Finally, the total training loss is expressed as:

$$\mathcal{L} = \omega_1 \mathcal{L}_{sdf} + \omega_2 \mathcal{L}_{pw} + \omega_3 \mathcal{L}_{pp} + \omega_4 \mathcal{L}_{reg}, \quad (7)$$

where $\mathcal{L}_{reg} = \|\mathbf{c}^m\|_2 + \|\mathbf{c}^s\|_2$ is the regularization term of latent codes, and ω are the weights for different loss terms.

3.4. Optimization as inference

After training, we can reconstruct the 4D shape based on the given sparse CMR slice sequence. We manually delineate all the slices of each sequence and apply the segmentation network nn-UNet [17] for training to obtain the contour point clouds of the left myocardium. Here we assume that the slice-level segmentations provided represent the surface of the myocardium, with reference SDF values of 0, and use their supervision to optimize the latent codes. We can use either the short-axis (SAX) or long-axis (LAX) views of CMR for segmentation, but due to the large slice spacing, the resulting point clouds are still sparse at the inter-slice level. Before entering the motion model, we obtain \mathcal{P} by aligning the point cloud at the ED phase to the ED spatial space and applying it to the entire sequence. Thanks to the automatic decoder structure, we fix the network weight here and optimize the latent codes \mathbf{c}^m and \mathbf{c}^s . Note that the same sequence shares the \mathbf{c}^s , and each phase has its own \mathbf{c}^m .

Implementation details. Our neural motion model and ED shape model use a multi-slice perceptron structure. We choose $K^m = 128$ and $K^s = 256$ for the dimensions of codes. In the training set, $T_N = 25$. The learning rate ($1e - 3$) of the motion model is set to be greater than the learning rate ($1e - 4$) of the shape model so that a good shape model can make the motion model learn and deform better. For each loss weight, we set $\omega_1 = 1$, $\omega_2 = 5e - 3$, $\omega_3 = 1e - 4$, $\omega_4 = 1e - 4$.

4. Experiments

Data. We train the model on our proposed CMR dataset, which was acquired at Jiangsu Province Hospital, and composed of 55 healthy subjects. Each subject includes multiple slices (8-10 slices) with a resolution of $1.25 \times 1.25 \times 10mm$. Each slice covers the video sequence of the cardiac cycle (25 phases). The clinical experts manually delineated the left myocardium of all the phases and slices. More information are presented in [Sup. Mat](#).

All the models are trained only on our proposed dataset and evaluate on the proposed dataset, the publicly available ACDC 2017 dataset [8], and the CT dataset. For the proposed internal dataset, we randomly split it into training, testing, and validation sets in a ratio of 60%, 20%, and 20%. The internal validation set is used for optimizing hyperparameters and held out until final evaluation. The ACDC 2017 dataset contains CMRs of 100 patients in five categories, each with a slice thickness of 5-10 mm and each sequence covering at least one cardiac cycle, with varying lengths. The CT dataset was collected in Jiangsu Province Hospital, and each only covers the two critical phases of ED and ES, which will verify the ability of our method to perform the cross-modal interpolation.

Evaluation metrics. Our proposed dataset uses Cham-

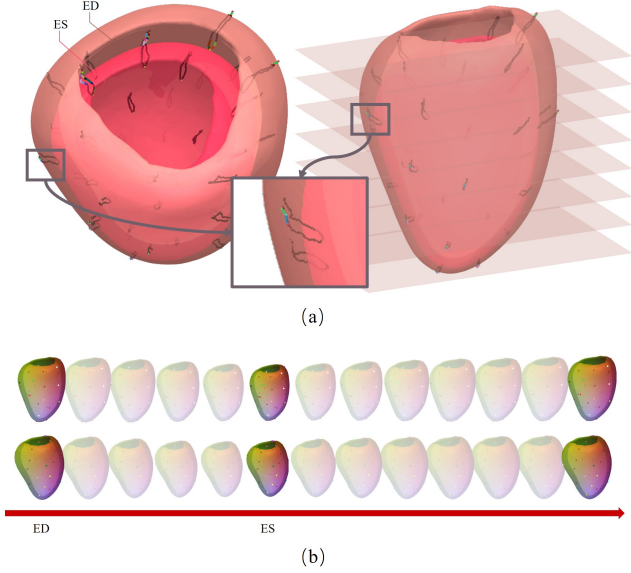


Figure 4. Dense motion estimation. (a) Presentation of dense motion estimation. Some sample points are highlighted. (b) The motion track of some surface points (black curve), one of which is zoomed in. The shapes of the ED phase and ES phase are visualized.

fer Distance (CD) and Earth Mover Distance (EMD) to evaluate the shape similarity between the ground truth and predicted reconstruction. Since the ACDC 2017 and CT datasets have no ground truth shape, we use the Dice score (Dice) and Hausdorff distance (HD) to measure the segmentation results on the intra-slice instead.

4.1. Applications

4.1.1 Dense motion estimation

We select some surface points to visualize the motion track of the cardiac cycle, as shown in Fig. 4 (a). It can be seen that the motion obtained by our method will not be confined to the slice. It will serve the task of 3D motion tracking well and analyze the motion state of key points. The presentation of dense motion estimation is shown in Fig. 4 (b), where the consistent colors across different phases correspond to the same points. We can trace the corresponding point for any given point at any phase. The acquisition of this dense field holds immense potential for analyzing more statistical and clinical indicators.

4.1.2 Motion interpolation

Because our method can decouple the motion and shape, we can estimate the missing shapes by interpolating the motion code from the incomplete sequence. We select L ($L < T_N, T_N$ is the length of the complete sequence) phases as observations. We use PCA to extract features

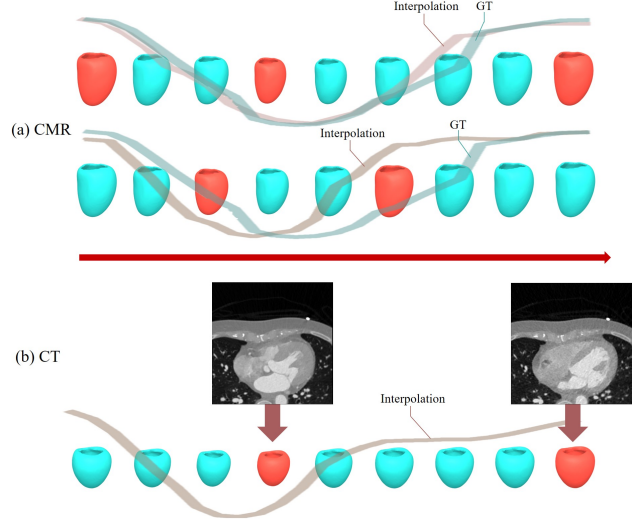


Figure 5. Visualization of motion code interpolation for CMR and CT. Cyan meshes are interpolated phases. The curves represent the corresponding volume change.

from the motion code map of the whole sequence from the training set $\text{Seq} \in \mathbb{R}^{S \times T_N \cdot K^m}$ (S is the number of the sequence), and get the motion parameter vector $\beta \in \mathbb{R}^{K^\beta}$ characterizing the sequence by regressing the observation:

$$\underset{\beta}{\operatorname{argmin}} \sum_{i=1}^L \|\mathbf{V}_i^m \beta - \mathbf{c}_i^m\|_2^2, \quad (8)$$

where $\mathbf{V}^m \in \mathbb{R}^{T_N \cdot K^m \times K^\beta}$ is obtained through the SVD of Seq , $\mathbf{V}_i^m \in \mathbb{R}^{K^m \times K^\beta}$ and \mathbf{c}_i^m represent the i th observation of \mathbf{V}^m and the motion codes respectively.

$$\mathbf{M} = \mathcal{F}(\bar{\mathbf{m}} + \mathbf{V}^m \beta), \quad (9)$$

where $\bar{\mathbf{m}}$ is the mean motion codes of the sequence, and the operator $\mathcal{F}(\cdot) : \mathbb{R}^{T_N \cdot K^m} \mapsto \mathbb{R}^{T_N \times K^m}$. \mathbf{M} contains all motion codes of the entire sequence after completion.

We produce some reasonable results, as shown in Fig. 5 (a). The first column gives keyframes (ED, ES, and the last frame) and interpolates the remaining phases. The second column randomly gives only two frames. The volume curve of the whole sequence is also displayed. It can be seen that the results obtained by our interpolation method can ensure the diastolic-systolic-diastolic of the heart cycle, reflecting the fidelity of motion code learning.

Our application can also be adapted to work with CT images. We aim to convert CT data into a format that our model can handle, and leverage the prior knowledge acquired from analyzing CMR sequences to perform cross-modal motion interpolation. Typically, CT scans have a low temporal resolution and consist of only two keyframes, which is consistent with our assumption. By applying our

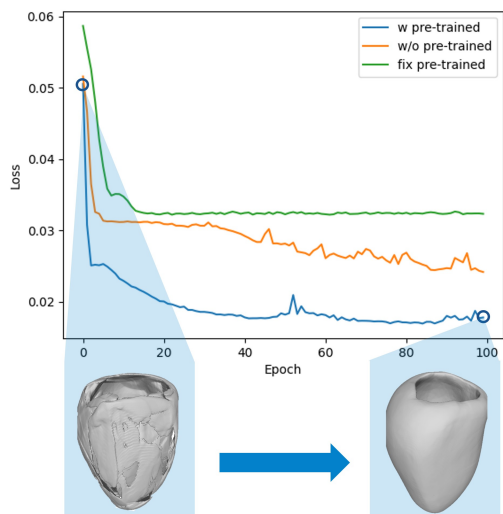


Figure 6. The influence of whether to establish ED-space in advance to pre-train the shape model.

method, we can reconstruct sparse CT data and generate a sequence with a similar temporal resolution to CMR scans, as shown in Fig. 5 (b). This could be particularly useful in clinical settings where doctors need a reliable reference for monitoring motion.

4.2. Ablation study on ED-space building

To demonstrate the effectiveness of pre-training the shape model by building the ED-space in advance, we draw the loss curve as shown in Fig. 6. After constructing the statistical shape model, we sample in the parameter space to obtain many shapes to pre-train an implicit shape function. However, simply fixing the weight of the ED shape model and training only the motion model could result in the network becoming stuck in local optimization. Instead, if the trained shape model is put into the overall framework and trained together with the motion model, it can improve the accuracy of the ED shape model quickly. In our experiment, we set the learning rate of the motion model to be greater than that of the shape model so that a good shape model urges the motion model to learn and deform better. Fig. 6 shows an example where the pre-trained shape model cannot be well estimated initially, but the final inaccuracies will be corrected with the joint training.

4.3. Comparisons

We investigate the representation power of different methods for 4D reconstruction. The quantitative results compared with the above methods are shown in Tab. 1 and Tab. 2.

Comparison with explicit methods. For explicit methods, Voxel2Mesh [39] takes 3D medical volume as input and em-

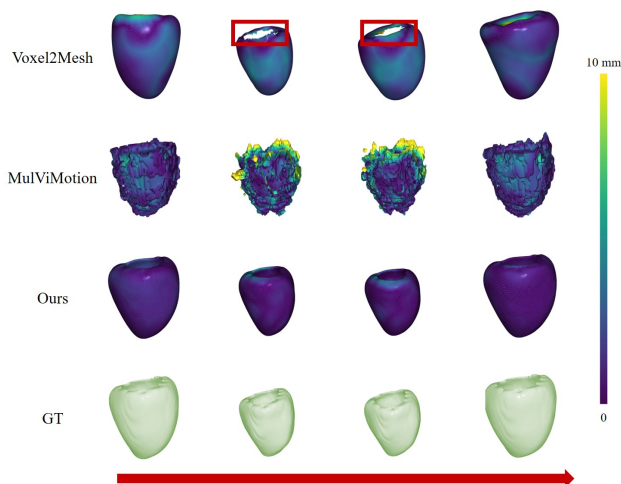


Figure 7. Error distribution of partial 4D reconstruction results of cardiac cycle based on explicit methods.

Table 1. Quantitative results of 4D reconstruction on proposed dataset

Method	Input views	CD↓	EMD↓
Voxel2Mesh [39]	Volume	22.842	7.796
DeepSDF [32]	SAX	3.331	6.447
DIT [47]		7.553	7.645
DIF [12]		2.891	7.010
Ours		2.634	6.173
DeepSDF [32]	SAX+LAX	2.852	6.172
DIT [47]		8.156	6.878
DIF [12]		2.858	6.010
Ours		2.627	3.603
MulViMotion [26]		17.728	7.774

Table 2. Quantitative segmentation results of 4D reconstruction on the ACDC 2017 dataset and CT dataset

Dataset	Method	Dice↑	HD↓
ACDC 2017 dataset	Voxel2Mesh [39]	0.553	3.534
	DeepSDF [32]	0.697	3.257
	DIT [47]	0.702	3.011
	Ours	0.765	2.789
CT dataset	DeepSDF [32]	0.796	3.048
	DIT [47]	0.817	2.646
	Ours	0.842	2.531

plies GCN. MulViMotion [26] is designed for multi-view motion estimation of the heart and predicts the shape simultaneously. As shown in Fig. 7, Voxel2Mesh loses many details and cannot reconstruct complex structures such as the opening and inner wall of the myocardium. MulViMotion has limited resolution and cannot reconstruct high-quality shapes. In contrast, our method utilizes implicit functions to obtain fine shapes and dense motion.

Comparison with implicit methods. To ensure fairness in

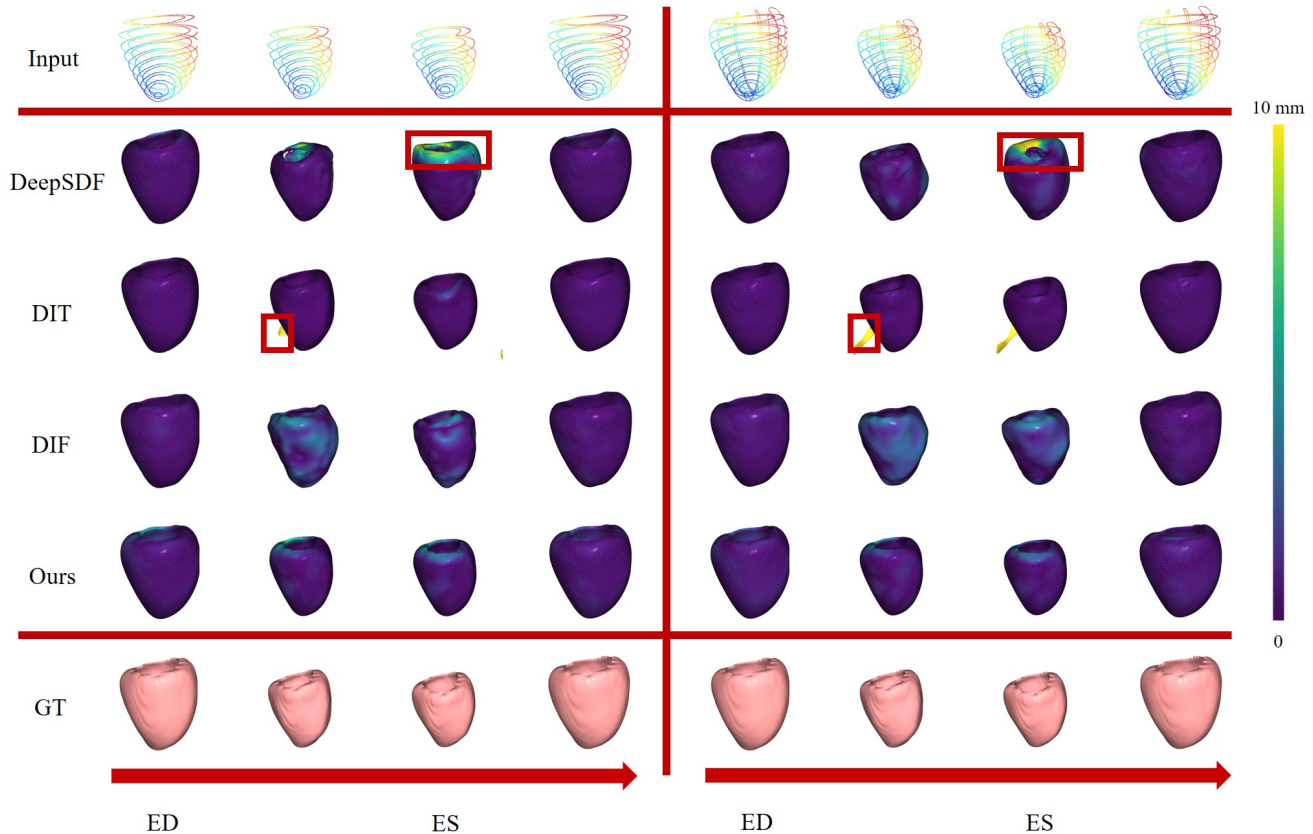


Figure 8. Error distribution of 4D reconstruction results of cardiac cycle based on implicit methods. The left uses the CMR SAX contours as input, and the right uses both the SAX and LAX contours as input. Our results are smoother and more fidelity in time sequence.

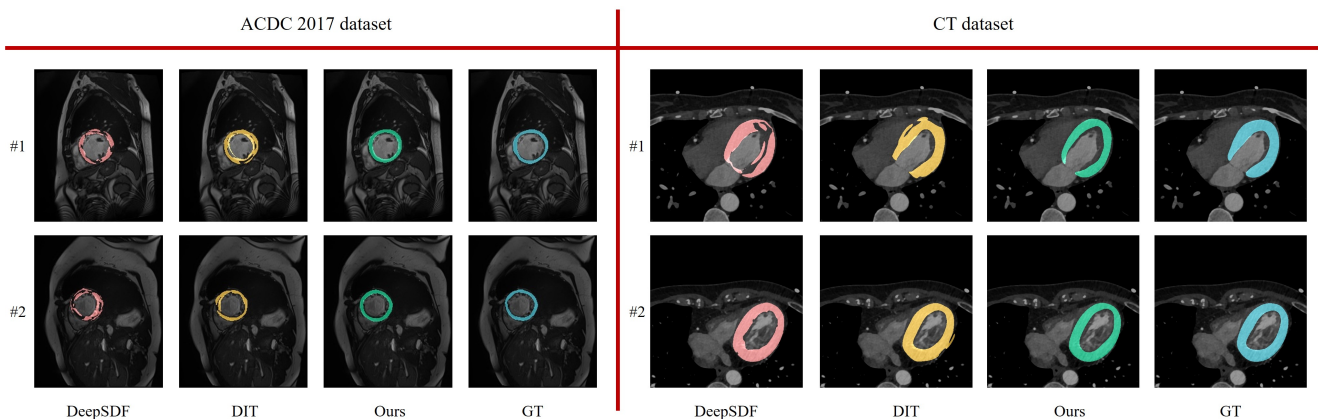


Figure 9. Segmentation results on the ACDC 2017 and CT datasets. We achieve smoother results that are more closely to the ground truth.

the comparative experiments, the pre-processing steps of all implicit methods are consistent with our method, meaning that the input data is the same. DeepSDF [32] is a classic structure without designing a deformation model. Both DIT [47] and DIF [12] are represented by an implicit template field, while their codes are not decoupled. By comparing them, we can prove the significance of decoupling mo-

tion and shape. We evaluate the results of taking only the CMR short-axis (SAX) segmentation contours as input and taking both the short-axis and the long-axis (SAX+LAX) segmentation contours as input on our dataset, and the error visualization is shown in Fig. 8. DeepSDF has good reconstruction results in the time phase close to ED. However, it can not make a good prediction in the systolic pro-

cess with a large difference from ED, which shows that this method is not robust to reconstruct large deformation objects. DIF has the same problem, that is, it is difficult to accurately estimate systolic phases and cannot maintain a smooth topology. DIT can accurately reconstruct the shape of most phases, but there are many unexpected situations, as shown in the red box in Fig. 8, such as some additional parts. Unlike them, due to the separation of motion code, our method adds temporal information to further improve the continuity of motion, making the deformation smooth in time and the overall shape more reasonable.

Fig. 9 displays the segmentation outcomes for the ACDC 2017 dataset and CT dataset. Compared with other implicit methods, the segmentation we estimate on the intra-slice slice is closer to the ground truth. Our overall contour is smoother, and there are no discernible defects. According to Tab. 2, due to limitations imposed by the types of training data, there is still considerable room for improvement in our method when applied to ACDC 2017 dataset with diseases.

5. Conclusion

The diagnosis of cardiovascular diseases depends on the accurate estimation of the shape and motion state of the myocardium, while the sparse information of the CMR slices brings challenges to the task of inter-slice reconstruction. To address this problem, we propose a 4D myocardium reconstruction method, which can predict the complete shape and realize dense motion estimation from a given sparse point cloud sequence. We introduce the first 4D myocardial dataset, and evaluate our method on different datasets. We hope that our dataset and method can bring help and new insights to the field of 4D myocardium reconstruction.

Limitations and future work. Our experiments show that our motion network is capable of learning the motion patterns that correspond to the heart’s movement. However, we have not yet imposed explicit restrictions on the distribution of these motion patterns. While the primary focus of this work is on the 4D reconstruction of healthy hearts, we also aim to validate our method using a dataset containing cardiac diseases (ACDC 2017 dataset). The experimental findings reveal the potential for improvement, particularly in cases with significant morphological disparities, such as dilated cardiomyopathy. Nevertheless, we maintain that our framework, as a general method, can be extended to encompass abnormal hearts given a broader distribution of data. The current progress substantiates the value of applying implicit decoupled representations in the cardiovascular domain, showcasing promising results in myocardial reconstruction. In the future, we plan to explore use of dense inter-slice motion representations in various applications, including the analysis of their correlation with disease incidence.

References

- [1] Who cardiovascular diseases. https://www.who.int/cardiovascular_diseases/about_cvd/en/. Accessed July 29, 2020.
- [2] Jadie Adams, Riddhish Bhalodia, and Shireen Elhabian. Uncertain-deepssm: From images to probabilistic shape models. In *International Workshop on Shape in Medical Imaging*, pages 57–72. Springer, 2020.
- [3] Mihaela Silvia Amzulescu, Mathieu De Craene, Hélène Langet, Agnes Pasquet, David Vancraeynest, Anne-Catherine Pouleur, Jean-Louis Vanoverschelde, and BL Gerber. Myocardial strain imaging: review of general principles, validation, and sources of discrepancies. *European Heart Journal-Cardiovascular Imaging*, 20(6):605–619, 2019.
- [4] Rahman Attar, Marco Pereañez, Christopher Bowles, Stefan K Piechnik, Stefan Neubauer, Steffen E Petersen, and Alejandro F Frangi. 3d cardiac shape prediction with deep neural networks: Simultaneous use of images and patient metadata. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 586–594. Springer, 2019.
- [5] Wenjia Bai, Wenzhe Shi, Antonio de Marvao, Timothy JW Dawes, Declan P O’Regan, Stuart A Cook, and Daniel 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*, 26(1):133–145, 2015.
- [6] Guha Balakrishnan, Amy Zhao, Mert Rory Sabuncu, John V. Guttag, and Adrian V. Dalca. Voxelmorph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, 38:1788–1800, 2019.
- [7] Marcel Beetz, Abhirup Banerjee, and Vicente Grau. Biventricular surface reconstruction from cine mri contours using point completion networks. *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 105–109, 2021.
- [8] Olivier Bernard, Alain Lalonde, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, et al. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE transactions on medical imaging*, 37(11):2514–2525, 2018.
- [9] Pingjun Chen, Xiao Chen, Eric Z Chen, Hanchao Yu, Terrence Chen, and Shanhui Sun. Anatomy-aware cardiac motion estimation. In *International Workshop on Machine Learning in Medical Imaging*, pages 150–159. Springer, 2020.
- [10] Xiang Chen, Nishant Ravikumar, Yan Xia, Rahman Attar, Andrés Diaz-Pinto, Stefan K. Piechnik, Stefan Neubauer, Steffen Erhard Petersen, and Alejandro F. Frangi. Shape registration with learned deformations for 3d shape reconstruction from sparse and incomplete point clouds. *Medical image analysis*, 74:102228, 2021.
- [11] Enric Corona, Albert Pumarola, G. Alenyà, Gerard Pons-Moll, and Francesc Moreno-Noguer. Smplicit: Topology-aware generative model for clothed people. *2021 IEEE/CVF*

- Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11870–11880, 2021.
- [12] Yu Deng, Jiaolong Yang, and Xin Tong. Deformed implicit field: Modeling 3d shapes with learned dense correspondence. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10281–10291, 2021.
- [13] Alejandro F Frangi, Daniel Rueckert, Julia A Schnabel, and Wiro J Niessen. Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modeling. *IEEE transactions on medical imaging*, 21(9):1151–1166, 2002.
- [14] Yuyu Guo, Lei Bi, Dongming Wei, Liyun Chen, Zhengbin Zhu, Dagan Feng, Ruiyan Zhang, Qian Wang, and Jinman Kim. Unsupervised landmark detection based spatiotemporal motion estimation for 4d dynamic medical images. *IEEE transactions on cybernetics*, PP, 2021.
- [15] Zekun Hao, Hadar Averbuch-Elor, Noah Snaveley, and Serge J. Belongie. Dualsdf: Semantic shape manipulation using a two-level representation. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7628–7638, 2020.
- [16] Corné Hoogendoorn, Nicolas Duchateau, Damian Sanchez-Quintana, Tristan Whitmarsh, Federico M Sukno, Mathieu De Craene, Karim Lekadir, and Alejandro F Frangi. A high-resolution atlas and statistical model of the human heart from multislice ct. *IEEE transactions on medical imaging*, 32(1):28–44, 2012.
- [17] Fabian Isensee, Jens Petersen, Simon A. A. Kohl, Paul F. Jäger, and Klaus Maier-Hein. nnu-net: Breaking the spell on successful medical image segmentation. *ArXiv*, abs/1904.08128, 2019.
- [18] Boyan Jiang, Xinlin Ren, Mingsong Dou, Xiangyang Xue, Yanwei Fu, and Yinda Zhang. Lord: Local 4d implicit representation for high-fidelity dynamic human modeling. In *ECCV*, 2022.
- [19] Thomas Joyce, Stefano Buoso, Christian T. Stoeck, and Sebastian Kozerke. Rapid inference of personalised left-ventricular meshes by deformation-based differentiable mesh voxelization. *Medical image analysis*, 79:102445, 2022.
- [20] Fanwei Kong and Shawn C Shadden. Whole heart mesh generation for image-based computational simulations by learning free-from deformations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 550–559. Springer, 2021.
- [21] Fanwei Kong, Nathan M. Wilson, and Shawn C. Shadden. A deep-learning approach for direct whole-heart mesh reconstruction. *Medical image analysis*, 74:102222, 2021.
- [22] Julian Krebs, Tommaso Mansi, Nicholas Ayache, and Hervé Delingette. Probabilistic motion modeling from medical image sequences: application to cardiac cine-mri. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 176–185. Springer, 2019.
- [23] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3d surface construction algorithm. *ACM siggraph computer graphics*, 21(4):163–169, 1987.
- [24] Priyanka Mandikal, KL Navaneet, Mayank Agarwal, and R Venkatesh Babu. 3d-lmnet: Latent embedding matching for accurate and diverse 3d point cloud reconstruction from a single image. *arXiv preprint arXiv:1807.07796*, 2018.
- [25] Qingjie Meng, Wenjia Bai, Tianrui Liu, Declan P O’Regan, and Daniel Rueckert. Mesh-based 3d motion tracking in cardiac mri using deep learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 248–258. Springer, 2022.
- [26] Qingjie Meng, Chen Qin, Wenjia Bai, Tianrui Liu, Antonio de Marvao, Declan P. O’Regan, and Daniel Rueckert. Mulvimotion: Shape-aware 3d myocardial motion tracking from multi-view cardiac mri. *IEEE Transactions on Medical Imaging*, 41:1961–1974, 2022.
- [27] Lars M. Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4455–4465, 2019.
- [28] Marko Mihajlović, Yan Zhang, Michael J. Black, and Siyu Tang. Leap: Learning articulated occupancy of people. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10456–10466, 2021.
- [29] Michael Niemeyer, Lars M. Mescheder, Michael Oechsle, and Andreas Geiger. Occupancy flow: 4d reconstruction by learning particle dynamics. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5378–5388, 2019.
- [30] Pablo Rodríguez Palafox, Aljanz Bovzivc, Justus Thies, Matthias Nießner, and Angela Dai. Npms: Neural parametric models for 3d deformable shapes. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12675–12685, 2021.
- [31] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 165–174, 2019.
- [32] Jeong Joon Park, Peter R. Florence, Julian Straub, Richard A. Newcombe, and S. Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 165–174, 2019.
- [33] Chen Qin, Wenjia Bai, Jo Schlemper, Steffen E Petersen, Stefan K Piechnik, Stefan Neubauer, and Daniel Rueckert. Joint learning of motion estimation and segmentation for cardiac mr image sequences. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 472–480. Springer, 2018.
- [34] Chen Qin, Shuo Wang, Chen Chen, Huaqi Qiu, Wenjia Bai, and Daniel Rueckert. Biomechanics-informed neural networks for myocardial motion tracking in mri. In *MICCAI*, 2020.
- [35] Ashwin Raju, Shun Miao, Chi-Tung Cheng, Le Lu, Mei Han, Jing Xiao, Chien-Hung Liao, Junzhou Huang, and Adam P. Harrison. Deep implicit statistical shape models for 3d medical image delineation. *ArXiv*, abs/2104.02847, 2022.
- [36] Shanlin Sun, Kun Han, Deying Kong, Hao Tang, Xiangyi Yan, and Xiaohui Xie. Topology-preserving shape reconstruction and registration via neural diffeomorphic flow.

- 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20813–20823, 2022.
- [37] Garvita Tiwari, Nikolaos Sarafianos, Tony Tung, and Gerard Pons-Moll. Neural-gif: Neural generalized implicit functions for animating people in clothing. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11688–11698, 2021.
- [38] Yifan Wang, Zichun Zhong, and Jing Hua. Deeporgannet: On-the-fly reconstruction and visualization of 3d / 4d lung models from single-view projections by deep deformation network. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):960–970, 2020.
- [39] Udaranga Wickramasinghe, Edoardo Remelli, Graham Knott, and Pascal Fua. Voxel2mesh: 3d mesh model generation from volumetric data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 299–308. Springer, 2020.
- [40] Jiancheng Yang, Udaranga Wickramasinghe, Bingbing Ni, and P. Fua. Implicitatlas: Learning deformable shape templates in medical imaging. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15840–15850, 2022.
- [41] Meng Ye, Qiaoying Huang, Dong Yang, Pengxiang Wu, Jingru Yi, Leon Axel, and Dimitris Metaxas. Pc-u net: Learning to jointly reconstruct and segment the cardiac walls in 3d from ct data. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 117–126. Springer, 2020.
- [42] Meng Ye, Mikael Kanski, D. Yang, Qi Chang, Zhennan Yan, Qiaoying Huang, Leon Axel, and Dimitris N. Metaxas. Deeptag: An unsupervised deep learning method for motion tracking on cardiac tagging magnetic resonance images. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7257–7267, 2021.
- [43] Hanchao Yu, Xiao Chen, Humphrey Shi, Terrence Chen, Thomas S. Huang, and Shanhui Sun. Motion pyramid networks for accurate and efficient cardiac motion estimation. *ArXiv*, abs/2006.15710, 2020.
- [44] Hanchao Yu, Xiao Chen, Humphrey Shi, Terrence Chen, Thomas S Huang, and Shanhui Sun. Motion pyramid networks for accurate and efficient cardiac motion estimation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 436–446. Springer, 2020.
- [45] Hanchao Yu, Shanhui Sun, Haichao Yu, Xiao Chen, Honghui Shi, Thomas S Huang, and Terrence Chen. Foal: Fast online adaptive learning for cardiac motion estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4313–4323, 2020.
- [46] Xiaoran Zhang, Chenyu You, Shawn S. Ahn, Juntang Zhuang, Lawrence H. Staib, and James S. Duncan. Learning correspondences of cardiac motion from images using biomechanics-informed modeling. *ArXiv*, abs/2209.00726, 2022.
- [47] Zerong Zheng, Tao Yu, Qionghai Dai, and Yebin Liu. Deep implicit templates for 3d shape representation. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1429–1439, 2021.
- [48] Xiao-Yun Zhou, Zhao-Yang Wang, Peichao Li, Jian-Qing Zheng, and Guang-Zhong Yang. One-stage shape instantiation from a single 2d image to 3d point cloud. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 30–38. Springer, 2019.