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).
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-M
  - yocardium-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 recent 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-
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, 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 regularization 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 \( 4 \times 4 \) matrix \( P \) (details are shown in Sup. Mat.). After that, \( P \) 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 code.
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 \( \bar{s} \) defined in Sec. 3.2.1 and apply the obtained transformation matrix \( P \) to all phases. For a sampling point \( x \in \mathbb{R}^3 \) in the canonical space, its corresponding deformation \( v \) is predicted under the condition of motion code \( c^m_i \in \mathbb{R}^{K^m} \) and phase indicator \( \tau_i \):

\[
f^m : (x, c^m_i, \tau_i) \in \mathbb{R}^{1+3+K^m} \longrightarrow v \in \mathbb{R}^3,
\]

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 \( c^m_i \) is variant for each \( i \)-th phase. With the proposed neural motion model, the corresponding point \( x' \) in ED spatial space can be obtained by \( x' = x + v \). 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:

\[
S = \mathcal{M}(\bar{s} + V^\alpha \alpha),
\]

where \( K^\alpha \) is the dimension of the parameter vector, \( N \) is the number of vertices, \( V^\alpha \in \mathbb{R}^{3N \times K^\alpha} \) is obtained through the singular value decomposition (SVD) of the atlas map \( \bar{s} \) 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 \( \bar{s} \), 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 \( x \in \mathbb{R}^3 \) in the ED spatial space, the corresponding SDF value \( sdf \) can be predicted through an MLP network based on the shape code \( \epsilon^* \in \mathbb{R}^{K^*} \):

\[
f^* : (\epsilon^*, x) \in \mathbb{R}^{K^* + 3} \rightarrow sdf \in \mathbb{R}.
\]

where \( K^* \) is the dimension of the shape code. The implicit shape model learns the shape code \( \epsilon^* \) of a given object, and the weight of network \( f^* \) in the fashion of automatic decoder [31] simultaneously. Note that \( \epsilon^* \) 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:

\[
L_{sdf} = \sum_{i=1}^{S} \sum_{l=1}^{T_N} \sum_{j \in \Omega} |f(x_j) - sdf_j|,
\]

where \( S \) and \( T_N \) represent the number and the length of sequences respectively, \( \Omega \) denotes points in 3D space, \( f = f^*(f^m(\cdot)) \) describes the process of point \( x_j \) 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] \( L_{pw} \) and \( L_{pp} \) to prevent unrealistic large distortion:

\[
L_{pw} = \sum_{i=1}^{S} \sum_{l=1}^{T_N} \sum_{j \in \Omega} h(\|f^m(\epsilon^m, x_j, \tau_j) - x_j\|_2),
\]

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

\[
L_{pp} = \sum_{i=1}^{S} \sum_{l=1}^{T_N} \sum_{j,k \in \Omega, j \neq k} \max(\|\hat{v}_j - \hat{v}_k\|_2 - \epsilon, 0),
\]

where \( \epsilon \) is a parameter controlling the distortion tolerance.

Finally, the total training loss is expressed as:

\[
L = \omega_1 L_{sdf} + \omega_2 L_{pw} + \omega_3 L_{pp} + \omega_4 L_{reg},
\]

where \( L_{reg} = \|\epsilon^m\|_2 + \|\epsilon\|_2 \) is the regularization term of latent codes, and \( \omega \) 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 \( P \) 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 \( \epsilon^m \) and \( \epsilon^* \). Note that the same sequence shares the \( \epsilon^* \), and each phase has its own \( \epsilon^m \).

**Implementation details.** Our neural motion model and ED shape model use a multi-slice perceptron structure. We choose \( K^m = 128 \) and \( K^* = 256 \) 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 categories, each with a slice thickness of 5-10 mm and each sequence covering at least one cardiac cycle, with varying lengths. 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.

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 from the motion code map of the whole sequence from the training set \( \text{Seq} \in \mathbb{R}^{S \times T_N \times K^m} \) \( (S \) 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:

\[
\arg\min_\beta \sum_{i=1}^L \| V_i^m \beta - c_i^m \|_2^2, \tag{8}
\]

where \( V_i^m \in \mathbb{R}^{T_N \times K^m \times K^\beta} \) is obtained through the SVD of \( \text{Seq}, V_i^m \in \mathbb{R}^{K^m \times K^\beta} \) and \( c_i^m \) represent the \( i \)th observation of \( V^m \) and the motion codes respectively.

\[
M = F(\bar{m} + V^m \beta), \tag{9}
\]

where \( \bar{m} \) is the mean motion codes of the sequence, and the operator \( F(\cdot) : \mathbb{R}^{T_N \times K^m} \rightarrow \mathbb{R}^{T_N \times K^m} \). \( M \) 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
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-

ployed 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
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 motion 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 cannot make a good prediction in the systolic pro-

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.
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. DI 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


